Identifying Artistic Expertise Difference in Emotion Recognition in Response to Oil Paintings

Yu-Ting Lan[1], Ze-Chen Li[1], Dan Peng[2], Wei-Long Zheng[1], and Bao-Liang Lu[1,2,†] *Fellow, IEEE*

***Abstract*—Previous studies have been conducted on building emotion recognition frameworks and enhancing their performances using Electroencephalography (EEG) and eye tracking signals. However, the differences between experts in art and non-experts in emotion recognition still remain to be elucidated. In this paper, we systematically evaluate the performance of various computational models for emotion recognition in response to oil paintings and identify the differences between experts in art and non-experts. The experimental results demonstrate that Transformer neural networks achieve the highest accuracies of 65.27% in three-category emotion recognition (negative, neutral, and positive) in response to oil paintings. Although the overall emotion recognition accuracies of the two groups are similar, the mean accuracy of the non-expert group for positive emotion is higher than that of experts, and the expert group has higher recognition accuracy in neutral emotion than the non-expert group. We further investigate the neural patterns of the three emotions in the two groups. The experimental results indicate that neural pattern differences do exist in both emotions and artistic expertise. The parietal and occipital lobes are more activated for positive emotion in the artistic expert group in the alpha, beta, and gamma bands. Our proposed methods provide an understanding of underlying emotion-expertise neurological mechanisms and cognitive processes.**

***Index Terms*—Artistic expertise difference, emotion recognition, oil paintings, EEG, eye tracking.**

I. Introduction

Emotion is an important component of all human activities. In recent decades, extensive efforts have been devoted to the automatic recognition of emotional states from speech, facial expressions, and physiological signals. As many of the physiological signals are bodily responses and hard to be deliberately controlled by users, physiological signals are considered reliable and valid indicators of human emotions evoked by different stimuli. One of the most common physiological signals for measuring emotions is Electroencephalography (EEG). As a physiological signal that directly measures brain activities, EEG has been demonstrated to be a simple, reliable, and easy-to-use solution for recognizing emotions and demonstrating the emotion-related neural patterns [1]. Similar to EEG, eye tracking has been extensively used to detect conscious and unconscious activities [2]. Some complex eye tracking features, such as pupil diameter, gaze behavior, and scan-path, are found to be reliable indicators of personality traits, attention, and emotions.

In addition to designing the emotion recognition framework using EEG and eye tracking signals and enhancing its performance, many previous studies have focused on identifying group differences (such as gender, culture, age, *etc.*) in emotion recognition. For example, Davidson *et al.* [3] revealed significantly greater relative right-hemisphere activation during emotion versus non-emotion trials in females. Males showed no significant task-dependent shifts in asymmetry between conditions. Liu *et al.* [1] reported that, from emotion recognition accuracies, EEG and eye tracking can adapt to Chinese, German, and French cultural diversities and that a cultural in-group advantage phenomenon does exist in emotion recognition with EEG. Langeslag *et al.* [4] found that, in both younger and older groups, recognition accuracy is not affected by emotion and that the response bias is more liberal for unpleasant pictures.

However, few studies have explored the emotion-relevant difference between art experts and non-experts. In fact, some differences between the two groups viewing paintings do exist in terms of aesthetic evaluations. For example, compared with non-experts, experts in the art appreciate original paintings more than altered versions [5] and pay more attention to the composition (such as lines, shapes, colors, *etc.*) [6]. Researchers find that the EEG phase synchronization is higher, particularly in the delta and gamma bands in the right hemisphere and in the posterior brain regions, when experts are asked to imagine a painting after viewing it [7]. However, the emotion-related differences between experts in art and non-experts still need further investigation through quantitative models to understand the underlying emotion-expertise neurological mechanisms and cognitive processes.

In this paper, we systematically evaluate the performance of various computational models for emotion recognition in response to oil paintings and compare the differences between experts in art and non-experts. The experimental results demonstrate that Transformer neural networks achieve the highest accuracies of 65.27% in three-category emotion recognition (negative, neutral, and positive). The mean accuracy for positive emotion of non-experts is higher than that of experts. In contrast, the expert group has higher recognition accuracy in neutral emotion than the non-expert group. We further

This work was supported in part by grants from STI 2030-Major Projects+2022ZD0208500, the National Natural Science Foundation of China (No. 61976135), Shanghai Municipal Science and Technology Major Project (No. 2021SHZDZX), Shanghai Pujiang Program (Grant No. 22PJ1408600), SJTU Global Strategic Partnership Fund, Shanghai Marine Equipment Foresight Technology Research Institute 2022 Fund (No. GC3270001/012), SJTU Global Strategic Partnership Fund (2021 SJTU-HKUST), and GuangCi Professorship Program of RuiJin Hospital Shanghai Jiao Tong University School of Medicine.

[1]Department of Computer Science and Engineering, Shanghai Jiao Tong University. [2]RuiJin-Mihoyo Laboratory, Clinical Neuroscience Center, RuiJin Hospital, Shanghai Jiao Tong University School of Medicine.

†Corresponding author (bllu@sjtu.edu.cn).

2023 11th International IEEE/EMBS Conference on Neural Engineering (NER) | 978-1-6654-6292-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/NER52421.2023.10123777

analyze the classification results and neural patterns in two dimensions: emotions and artistic expertise. The experimental results indicate that neural pattern differences do exist in both emotions and artistic expertise. The parietal and occipital lobes are more activated for positive emotion in the artistic expert group.

II. Emotion Experiment Design

Our emotion recognition and artistic expertise analysis rely on EEG and eye tracking signals recorded simultaneously from participants in response to an oil painting appreciation followed by a rating task. The designed emotion experiment is described in this section.

A. Participants

Twenty-four participants aged from 18 to 39 (9 males and 15 females; age 23.50 $\pm$ 4.3) were recruited from the university for our neurophysiological experiments. Each had normal hearing and self-reported normal or corrected-to-normal vision. Each participant was asked to complete the Assessment of Art Attributes (AAA), which is a classical instrument designed to assess six formal-perceptual and six conceptual-representational attributes [8]. Following the criterion proposed in AAA, 12 of the participants (2 males and 10 females; age 23.83 $\pm$ 5.34) were considered experts in art, and the other 12 participants (7 males and 5 females; age 23.16 $\pm$ 2.97) were considered as non-experts. Each participant was also asked to fill out the Eysenck Personality Questionnaire (EPQ) before the experiment, which could measure the personality of an individual in three independent dimensions: Extroversion/Introversion, Neuroticism/Stability, and Psychoticism/Socialization [9]. Those who turned out to be stable extroverts were selected as the subjects. To avoid the interference of brain laterality effects to our study, all participants enrolled were right-handed.

B. Stimuli

The way in which emotions are evoked is of great importance in emotion experiments. The stimulus materials commonly used in emotion research include sound, movie clips, and static pictures. For our emotion experiments, 114 carefully selected oil paintings from the public artwork dataset WikiArt were used as stimulus materials to evoke three types of emotions (positive, neutral, and negative) [10]. The selected oil paintings were created between the 16^{th} and 19^{th} centuries. The selected paintings covered most major artistic styles (e.g., Baroque, Rococo, Realism, Post-Impressionism) and five genres (i.e., portrait, animal, still life, cityscape, and landscape), thus enriching the diversity of the content of the paintings and better evoking emotions. To eliminate the influence of contrast, clarity, color-saturation and sharpness, all the paintings were uniformly adjusted with Adobe Lightroom Classic. All the paintings were manually screened and rated by at least 7 human assessors into three emotion categories (negative, neutral, and positive).

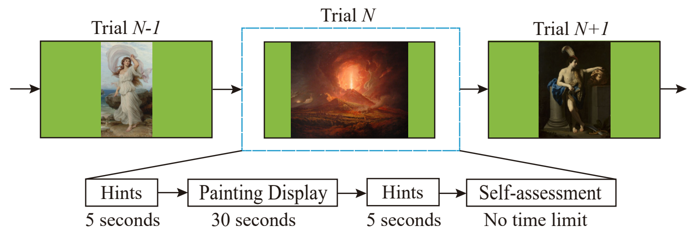


Fig. 1. The protocol used in our emotion experiment.

C. Protocol

The detailed experimental protocol is presented in Figure 1. There was a 5 s hint of start before each oil painting display and a 5 s hint of rating after the oil painting display. Following the rating hint, participants were told to report their emotional responses with both arousal and valence dimensions. In total, there were 60 trials for each experiment. Each trial contained an individual painting selected from a total of 114 paintings. In the whole experiment, EEG data were recorded using an ESI NeuroScan System at a sampling rate of 1000 Hz from a 62-channel active AgCl electrode cap according to the international 10-20 system. Simultaneously, eye tracking signals were recorded at a sampling rate of 120 Hz using a Tobii Pro X3-120 screen-based eye tracker.

III. Method

A. Data Preprocessing

For EEG signals, the raw data are downsampled to 200 Hz to accelerate data processing, reduce redundancy, and increase the amount of information per unit time. A bandpass filter between 1 Hz and 50 Hz is used to eliminate noise caused by environmental factors. Since differential entropy (DE) features performed better in the previous related work [11], we extract the DE features from raw EEG data using short-time Fourier transform (STFT) with 4-s non-overlapping Hanning windows. The DE features are divided into the five frequency bands: δ (1-4 Hz), θ (4-8 Hz), α (8-14 Hz), β (14-31 Hz), and γ (31-50 Hz). As for eye tracking features, the same method used in [12] is applied to extract 23 features, including pupil diameter, fixation, saccade and blink. The detailed features are listed in Table II. Since emotion is indicated as a continuous psycho-physiological state, the temporal dynamics of the emotional state should be taken into account. We apply the linear dynamic system (LDS) approach to smooth our extracted features, which can filter out artifacts that are not associated with emotional states.

B. Classification Models and Implementation Details

We compare the performance of the following models: support vector machine (SVM) with linear kernel, k-nearest neighbors (KNN), logistic regression (LR), and a Transformer neural network [12]. The Transformer model [12] contains L Transformer blocks to extract features and a linear classifier to detect emotions. Each Transformer block consists of two sub-layers: a multi-head self-attention, and a fully connected

TABLE I
THE AVERAGE ACCURACIES (MEAN/STD)% OF DIFFERENT FEATURES AND CLASSIFIERS IN THE EMOTION RECOGNITION TASK.

Model	EEG						Eye Tracking	EEG and Eye Tracking
	Delta	Theta	Alpha	Beta	Gamma	Total		
LR	39.08/9.01	39.31/8.65	37.90/9.67	44.51/7.89	45.59/9.06	45.07/10.53	44.08/7.77	46.49/10.18
KNN	41.10/7.12	43.20/8.59	41.09/8.19	43.91/8.12	46.87/8.79	44.25/8.38	44.58/5.84	45.49/9.40
SVM	46.65/7.71	46.04/7.06	46.72/7.09	50.06/6.62	51.16/7.93	49.20/6.91	49.96/6.04	51.49/7.70
Transformer	56.90/5.52	56.08/5.76	56.15/5.97	59.82/6.81	60.58/7.48	61.56/5.53	60.49/5.25	**65.27/6.26**

TABLE II
THE DETAILS OF FEATURES EXTRACTED FROM EYE TRACKING SIGNALS.

Eye tracking parameters	Extracted features
Pupil diameter (Left and Right)	Mean, standard deviation and DE in four bands: 0-0.2 Hz, 0.2-0.4 Hz, 0.4-0.6 Hz, 0.6-1 Hz
Fixation	Duration mean, duration standard deviation, maximum, and frequency
Saccade	Duration mean, duration standard deviation, latency, and frequency
Blink	Duration mean, duration standard deviation and frequency

feed-forward network. Every sub-layer is started with layer normalization and a residual connection is around each sub-layer to retain the information of the input features and enhance the model stability. The number of multi-head is denoted as H. We feed the input samples with the overlapping time window T to the Transformer model and take the average of the transformed features on each window for emotion recognition. The hyper-parameters of our Transformer model T, L, and H are empirically chosen based on a grid search in [2, 4, 6], [2, 4], and [2, 4, 6], respectively. We choose AdamW as the optimizer. The learning rate is selected from $10^{[-2,-3,-4]}$. The weight decay rate is set to 0.01.

The feature sets we use are: 1) EEG features; 2) eye tracking features; and 3) both features combined. The training and test datasets are divided based on trials, meaning that the samples of a trial are either in the training dataset or in the test dataset. We perform a three-fold cross-validation on the emotion recognition task using the above classification models and take the average accuracy of each subject as the final performance of the models.

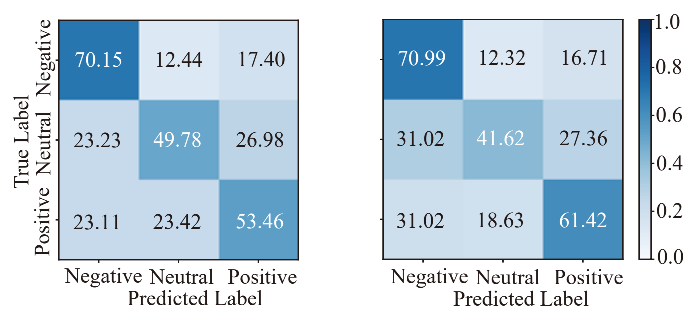
IV. RESULTS

A. Emotion Recognition

To evaluate the recognition performance of EEG, eye tracking and both combined for the three emotions, we perform a three-fold cross-validation for each subject on three feature sets. We compare the performance of several models including LR, KNN, SVM, and Transformer. Table I shows the average recognition accuracies. As shown in Table I, the high frequency bands (beta and gamma bands) outperform the low frequency bands, which is consistent with the previous work [13]. As can be seen, the accuracy of the combination of EEG and eye tracking is significantly higher than single modality. The improvement indicates that the two modalities have strong complementary characteristics for the three emotions. From Table I, we find that Transformer obtains the highest performance of 61.56/5.53, 60.49/52.25, and 65.27/6.26 in percentages on EEG, eye tracking, and the combination of EEG and eye tracking, respectively. The results indicate that the attention mechanism has the ability to capture the emotion-relevant properties and achieves better performance.

B. Expertise Differences in Emotion Recognition and Neural Patterns

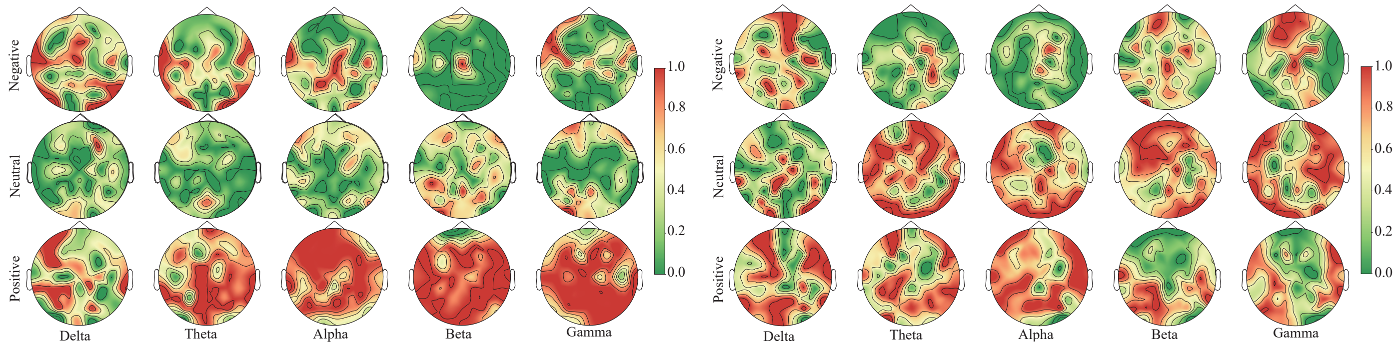
To investigate the expertise differences, the confusion matrices of the expert group and the non-expert group for three-category emotion recognition using Transformer are demonstrated in Figure 2. As illustrated in Figure 2, EEG and eye tracking signals can better distinguish negative emotion compared with other emotions. By comparing Figure 2(a) and Figure 2(b), we can see that the overall emotion recognition accuracies of the two groups are similar. However, the mean accuracy for positive emotion of non-experts is higher than that of experts. In contrast, the expert group has higher recognition accuracy in neutral emotion than the non-expert group.



(a) Expert (b) Non-expert

Fig. 2. Confusion matrices of the expert group, and the non-expert group for three-category emotion recognition using the signals of the combination of EEG and eye tracking. Each row of the confusion matrices represents the target class and each column represents the predicted class. The element (i, j) is the percentage of samples in class i that is classified as class j.

To further explain the artistic expertise differences in recognizing the three emotions, we project the average EEG features to the brain topography and obtain the stable patterns of the three emotions for experts and non-experts, respectively. Figure 3 depicts the average neural patterns for experts and non-experts. By comparing Figure 3(a) and Figure 3(b), we find that the general responses of experts are more active



(a) Expert (b) Non-expert

Fig. 3. The average neural patterns of the three emotions in the expert and the non-export group. The EEG features are normalized to scales between 0 and 1 and averaged within each group, respectively.

for positive emotions, while the responses of non-experts are more active for neutral emotions. In both beta and gamma bands, the prefrontal lobe is less activated for both experts and non-experts, while the temporal lobes on both hemispheres have higher activations, which is consistent with the previous work [1]. However, the parietal and occipital lobes of experts are more activated than those of non-experts in the alpha, beta, and gamma bands for positive emotion. The existing studies have shown that the beta and gamma activities in EEG reflect cognitive processing, such as active thinking and object recognition [14], and the EEG phase synchronization in the gamma bands is higher when experts are asked to imagine a painting after viewing it [7]. When appreciating the oil paintings, the experts may analyze the oil painting with their artistic knowledge and end up with more activations of the parietal and occipital lobes. This result could also explain why positive emotion evocations are less effective for expert groups in Figure 2, as the experts pay more attention to analyzing the oil paintings using their artistic skills, thus causing the neural patterns to be more diverse and reducing the recognition performance.

V. Conclusion

In this paper, we have systematically evaluated the performance of various models for emotion recognition in response to oil paintings and identified the difference between artistic experts and non-experts. The experimental results demonstrate that Transformer neural networks achieve the highest accuracies in three-category emotion recognition (negative, neutral, and positive). The overall emotion recognition accuracies of the two groups are similar. However, the mean accuracy for positive emotion of non-experts is higher than that of experts, and the expert group has higher recognition accuracy in neutral emotion than the non-expert group. We have investigated the neural patterns of two groups. The result shows that the neural patterns are different in both emotions and artistic expertise. The parietal and occipital lobes are more activated for positive emotion in the artistic expert group. Our proposed approaches can provide potential quantitative models of human emotions and artistic expertise to analyze both human mental states and behaviors.

References

[1] W.-L. Zheng, J.-Y. Zhu, and B.-L. Lu, "Identifying stable patterns over time for emotion recognition from EEG," *IEEE Transactions on Affective Computing*, vol. 10, no. 3, pp. 417–429, 2017.

[2] S. Berkovsky, R. Taib, I. Koprinska, E. Wang, Y. Zeng, J. Li, and S. Kleitman, "Detecting personality traits using eye-tracking data," in *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 2019, pp. 1–12.

[3] R. J. Davidson, G. E. Schwartz, E. Pugash, and E. Bromfield, "Sex differences in patterns of EEG asymmetry," *Biological Psychology*, vol. 4, no. 2, pp. 119–137, 1976.

[4] S. J. Langeslag and J. W. Van Strien, "Age differences in the emotional modulation of ERP old/new effects," *International Journal of Psychophysiology*, vol. 70, no. 2, pp. 105–114, 2008.

[5] P. Hekkert and P. C. Van Wieringen, "Beauty in the eye of expert and nonexpert beholders: A study in the appraisal of art," *The American Journal of Psychology*, pp. 389–407, 1996.

[6] C. F. Nodine, P. J. Locher, and E. A. Krupinski, "The role of formal art training on perception and aesthetic judgment of art compositions," *Leonardo*, vol. 26, no. 3, pp. 219–227, 1993.

[7] J. Bhattacharya and H. Petsche, "Shadows of artistry: cortical synchrony during perception and imagery of visual art," *Cognitive Brain Research*, vol. 13, no. 2, pp. 179–186, 2002.

[8] A. Chatterjee, P. Widick, R. Sternschein, W. B. Smith, and B. Bromberger, "The assessment of art attributes," *Empirical Studies of the Arts*, vol. 28, no. 2, pp. 207–222, 2010.

[9] S. B. Eysenck, H. J. Eysenck, and P. Barrett, "A revised version of the psychoticism scale," *Personality and Individual Differences*, vol. 6, no. 1, pp. 21–29, 1985.

[10] S. Luo, Y.-T. Lan, D. Peng, Z.-Y. Li, W.-L. Zheng, and B.-L. Lu, "Multimodal emotion recognition in response to oil paintings," in *44nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society*, 2022, pp. 4167–4170.

[11] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *6th International IEEE/EMBS Conference on Neural Engineering*, 2013, pp. 81–84.

[12] Y. Wang, W.-B. Jiang, R. Li, and B.-L. Lu, "Emotion transformer fusion: Complementary representation properties of EEG and eye movements on recognizing anger and surprise," in *2021 IEEE International Conference on Bioinformatics and Biomedicine*, 2021, pp. 1575–1578.

[13] W.-B. Jiang, L.-M. Zhao, P. Guo, and B.-L. Lu, "Discriminating surprise and anger from EEG and eye movements with a graph network," in *2021 IEEE International Conference on Bioinformatics and Biomedicine*, 2021, pp. 1353–1357.

[14] E. Başar, C. Başar-Eroğlu, S. Karakaş, and M. Schürmann, "Are cognitive processes manifested in event-related gamma, alpha, theta and delta oscillations in the EEG?" *Neuroscience Letters*, vol. 259, no. 3, pp. 165–168, 1999.