IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 26 April 2024, accepted 20 May 2024, date of publication 24 May 2024, date of current version 3 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3404990

RESEARCH ARTICLE

Multi-Task Emotion Recognition Based on Dimensional Model and Category Label

YI HUO [1], (Member, IEEE), AND YUN GE [2]
[1]Department of Educational Information Technology, Beijing Union University, Beijing 100011, China
[2]Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences, Beijing 102488, China

Corresponding author: Yi Huo (sfthuoyi@buu.edu.cn)

This work was supported by China Ministry of Education, Humanities, and Social Science Research Projects under Grant 23YJE880001.

ABSTRACT This study primarily designs a multitask emotion recognition model that combines Valence, Arousal, Dominance (VAD) three-dimensional continuous emotion analysis and discrete emotion classification, providing a more comprehensive and fine-grained emotional measurement tool for intelligent emotional interaction. It utilizes the correlation constraints between the two recognition tasks (category labels are points in the VAD three-dimensional emotion space) to enhance recognition accuracy. First, it provides a method and dataset for multi-dimensional continuous emotion recognition in the VAD three-dimensional space, which can describe emotional states more comprehensively and finely than traditional fixed emotion category labels, especially in dimension D, which is currently less researched. The integration of the Dominance (D) dimension enables a more comprehensive representation of emotional expressions, capturing nuanced variations related to dominance-related behaviors, particularly in contexts where understanding dominance cues is crucial. Second, because fixed category emotion labels represent a point in the VAD three-dimensional space, it utilizes the correlation between them for multi-task joint learning and establishes constraints between emotion categories and the VAD multi-dimensional emotion space. In the experiment, it used the emotional labels available in the existing emotion category dataset FER2013, manually added VAD annotations, and used it for VAD emotional measurement. The prediction results indicate that the average losses for predicting V, A, and D decrease by 0.7%, 6%, and 0.4% respectively, which verifies the effectiveness of the proposed multi-task strategies. The annotated VAD dataset and multi-task emotion recognition codes are available in Github: https://github.com/YeeHoran/Multi-task-Emotion-Recognition.

INDEX TERMS Multi-dimensional emotion recognition, multi-task learning, VAD facial expression recognition dataset.

I. INTRODUCTION

It primarily investigates multi-dimensional facial expression recognition to provide a more comprehensive and accurate emotion measurement tool [1]. Traditional emotion annotation primarily involves emotion category labels, such as seven standard emotion categories: anger, contempt, fear, joy, sadness, surprise, and neutral [2]. Alternatively, a set of emotion category labels is designed for a specific application [3]. However, with the increasing demand for emotion intelligence [4], obtaining more comprehensive, refined, and accurate emotion states has become increasingly urgent. Therefore, this study proposes a multi-dimensional emotion recognition method [5]. Furthermore, it utilizes the relations between the annotation information to conduct multitask learning [6] to further improve recognition performance.

Traditional emotion recognition often utilizes category label datasets [2], making each dataset closely related to the emotional category requirements of its application scenarios, thereby having certain limitations. With the increasing demand for comprehensive emotional representation models in modern emotional interactions, multi-dimensional emotion space models have recently been recognized [3], [7]. AffectNet [8] pioneered the establishment of facial expression

The associate editor coordinating the review of this manuscript and approving it for publication was Maria Chiara Caschera.

recognition datasets annotated with valence-arousal (VA) labels, and found that traditional emotion categories can be observed as points in the VA two-dimensional space. Based on this, VA emotion recognition methods have been developed [9]. Hulliyah et al. [10] also does sentiment analysis based on VA, but utilizes voice intonations, and electroencephalogram (EEG) signals as data source.

However, methods that recognize D-dimensional information are rarely investigated now, which is also the first problem to be addressed in this study. The Valence-Arousal (VA) model, which expresses emotions based on their valence and arousal levels overlooks the role of Dominance(D), a critical dimension in understanding human emotions, particularly in social contexts. By extending the VA model to include the Dominance dimension, it provides a more holistic framework for emotion recognition that captures the multidimensional nature of emotional expressions [11].

Additionally, AffectNet's discovery that categories are points in VA two-dimensional space [8] inspired this research, suggesting that categories are even the points in VAD (Valence, Arousal, Dominance) three-dimensional emotion space, as shown in Figure1. The 7 basic emotion categories are 3D points in VAD 3D space. Based on this hypothesis, it utilizes the correlation between discrete emotion categories and VAD multi-dimensional emotion parameters to establish consistency constraints for training models, thereby improving the accuracy of VAD and category emotion recognition.

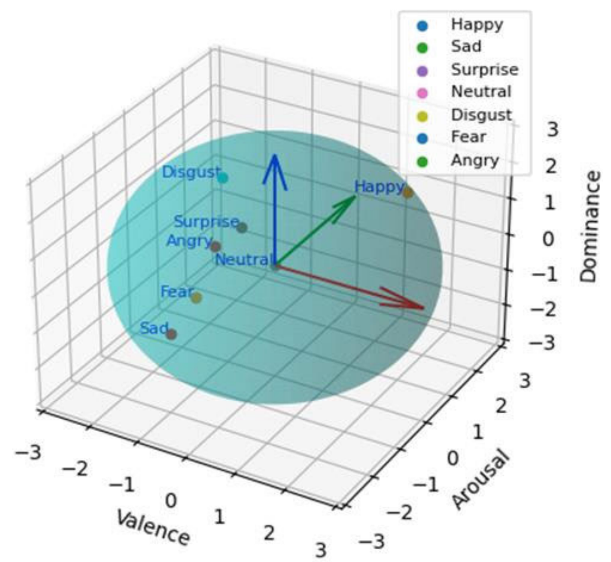

FIGURE 1. 7 emotion categories are the 3D points in VAD space.

Therefore, this study proposes a multi-dimensional multitask emotion recognition method. It first quantifies the FER2013 [12] dataset in VAD three dimensions, especially in the D dimension, and then builds a VAD three-dimensional emotion regression model to predict emotion conditions in VAD three-dimension space. Finally, a joint multi-task training model was established using the correlations between standard emotion categories and VAD three-dimensional continuous emotion values to enhance the computational accuracy of multi-dimensional emotions.

The structure of this paper is as follows: the introduction section introduces the background, problems to be solved, and the proposed solution; the related works section surveys and analyzes relevant content; Section three describes the research plan in detail; the experiment section designs and carries out tests to verify the proposed hypothesis and solutions; the discussion section analyzes the experimental results, summarizes the findings, and discusses its limitations; and finally, the conclusion section ends the entire study by summarizing the whole study, points out possible application scenarios, and identifies future research orientations.

II. RELATED WORK

A. EMOTION PARAMETERS

The idea of ''enabling computers to recognize user emotions'' was initially proposed by Professor Minsky, the father of artificial intelligence at MIT [13]. He points out that emotion is an important factor in achieving computer intelligence. Subsequently, Professor Picard explicitly stated the concept of affective computing in 1997 [14], which refers to the calculation of factors triggered by emotion, related to emotion, and capable of influencing and determining emotional changes. Later, neuroscientists found that emotions play an important role in human cognition and decision making [15]. For example, the work in [16] improved the accuracy of academic risk warnings by introducing VAD emotion parameters. Facial image recognition can also be used to measure emotional states, warn about mental system disorders, and intervene in a timely manner [17].

Currently, there are two main forms of emotion annotation: emotion category labels [18], and dimensional models [19]. Traditional emotion annotation is primarily based on category labels, and there are many classification methods for emotion categories based on different application contexts, such as the seven standard emotions [2] or 20 basic emotions such as boredom, anger, anxiety, surprise, sadness, depression, pride, hope, confusion, and happiness [20].

From another perspective, dimensional emotional parameters have also been proposed. Mehrabian and Russell proposed that emotions should have three dimensions: pleasure-displeasure (valence), arousal-no arousal (arousal), and dominance-submissiveness (dominance). Valence (V) represents the positivity or negativity of the subject's emotional state [21], reflecting the essence of emotion; arousal (A) is the level of physiological activation [22] of the subject's nervous system, related to the degree of activation of bodily energy; and dominance (D) represents the subject's control state [23] over the situation and others, indicating whether the emotional state is subjectively generated by the individual or influenced by the objective environment. Mehrabian's research showed that the use of the VAD's three dimensions [24] can effectively explain human emotions and fully quantify and express human emotions [25]. This study focuses on the second type: the dimensional model.

However, relatively little research has been conducted on the development and use of automated algorithms for measuring emotions using continuous dimensional models (such as valence and arousal). One of the main reasons for this is that creating a large database covering the entire continuous space of valence and arousal is expensive, and there are very limited annotated facial databases in the continuous domain. AffectNet [8] annotated facial emotion images with both V and A dimensions and seven standard categories simultaneously, indicating that category labels represent points in the VA two-dimensional space. There are currently fewer standards for the dominance dimension, but it is an important dimension for describing emotions comprehensively. Therefore, this study simultaneously annotated the VAD three dimensions [26] in a facial emotion recognition dataset to provide data support for a more comprehensive emotional measurement [18].

B. MULTI-DIMENSIONAL JOINT EMOTION RECOGNITION

VAD three-dimensional emotion [27] provides a more comprehensive perception of learners' emotions and technical support for future emotion interaction platforms that require more perspectives and granularities. Existing research on multi-dimensional emotion recognition algorithms either focuses only on the VA two dimensions or uses CNN methods to obtain several emotion values, without considering more perspectives to perceive emotions or generating continuous emotion values. This project considers these aspects and aims to optimize the accuracy of predicting continuous VAD three-dimensional emotion values and learning the continuous emotion recognition of these three dimensions.

Existing research divides the learning of V- and A-dimensional emotion values and the calculation of emotion categories into three independent models for calculating efficiency [28]. However, discrete emotion categories are correlated with the continuous emotion values of the VAD three-dimensional space. By using the correlation between them for VAD continuous emotion regression analysis and emotion classification, sharing part of the calculation models of both, and establishing constraints based on their correlation, the computational efficiency can be improved, and the accuracy of both VAD emotion value analysis and emotion classification can be simultaneously enhanced.

C. MULTI-TASK LEARNING

Most methods in multitask learning (MTL) share a common feature extraction component and then perform different tasks separately. Multi-task learning is a machine learning paradigm in which multiple tasks are learned simultaneously [29], often with the expectation that learning tasks jointly can improve generalization performance compared with learning them independently. There are several main research orientations and approaches for multi-task learning.

- **Shared Representation Learning**: This approach aims to learn a shared representation across multiple tasks [30], [31]. By sharing the representation, the model can leverage the similarities and regularities across tasks to improve performance. Techniques, such as deep neural networks, are often used to learn shared representations.
- **Transfer Learning**: Transfer learning involves leveraging knowledge from one task to improve learning in another related task [32], [33]. In multitask learning, this can involve transferring knowledge from tasks with abundant data to tasks with limited data, or transferring knowledge from related tasks to a target task [30].
- **Task Interaction Modeling**: In cases, tasks may interact with each other in complex ways [34], [35], [36]. Research in this area focuses on modeling these interactions to better understand and exploit the dependencies between tasks.
- **Task Relationship Modeling**: Some approaches focus on explicitly modeling the relationships between tasks [37], [38]. This can include learning task-specific parameters that are influenced by other tasks or incorporating task-relatedness measures into the learning process.

This research falls into the fourth category, which involves modeling the relationships between multiple tasks. However, designing and maintaining a trainable model for multiple tasks is challenging because changes in the training data, loss function, or hyperparameters of one task can affect other tasks. Integrating different architectures into a single model is challenging. Therefore, this study aims to identify the relationships between multiple tasks and incorporate these relationship conditions into the learning process to enhance the capabilities of multiple tasks.

III. METHOD

A. INCORPORATING DOMINANCE(D) AS VAD EMOTION RECOGNITION MODEL

AffectNet [4] indicates that the dimensional model of emotions covers the intensity and different emotion categories in the continuous domain, which can represent facial expression samples in the 2D space of Valence and Arousal. Furthermore, the dimensional model of emotions can finely distinguish between different emotional expressions and encode changes in the intensity of each emotion on a continuous scale.

Building on AffectNet's perspective, this study further incorporated the dominance dimension (D) into the emotion expression model. It proposes a novel emotion recognition framework that extends the VA model into the VAD 3D space by incorporating the Dominance dimension. It collects data with facial expressions, and map them onto the VAD space to capture the full spectrum of emotional expressions.

Therefore, it needs facial expression samples labeled in the 3D space of valence (V), arousal (A), and dominance (D).

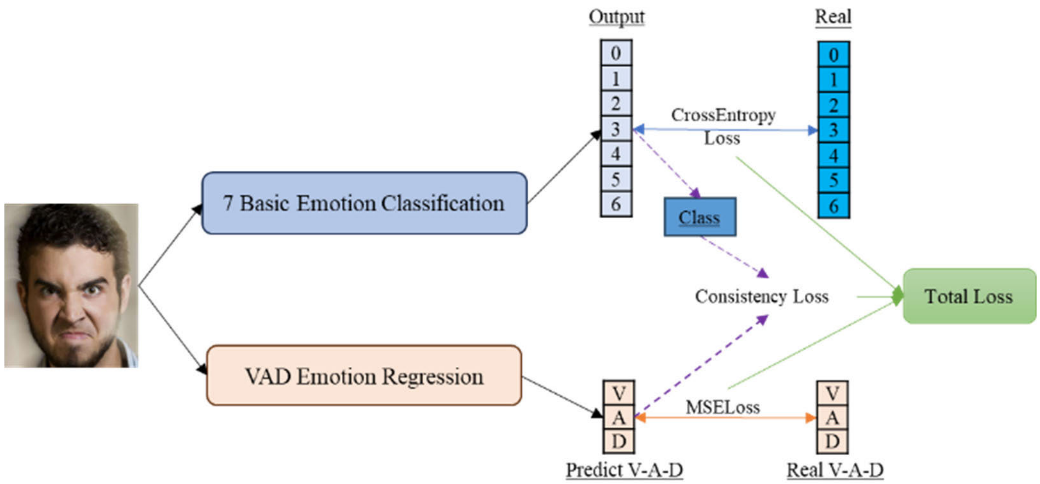

FIGURE 2. **Multi-task joint emotion recognition model.**

However, datasets containing simultaneous VAD (valence-arousal-dominance) facial expression samples are currently scarce. Therefore, this study began by creating a dataset of VAD 3D facial expression samples to provide foundational benchmark for VAD 3D emotion recognition.

B. RELATIONSHIPS BETWEEN VAD-DIMENSIONAL AND CATEGORIAL EMOTION CATEGORIES

Table1 and Figure1 presents the positions of seven standard emotion categories in the Valence-Arousal-Dominance (VAD) 3D space. It can be observed that for Happiness, $V > 0$, $A > 0$, and $D > 0$; for Disgust, $V < 0$, $A > 0$, $D > 0$; for Anger, $V < 0$, $A > 0$, but D may be > 0 (due to personal reasons) or may be < 0 (due to external events); for Fear, $V < 0$, $A < 0$, $D < 0$; for Sadness, $V < 0$, $A < 0$, D may be > 0 (due to personal reasons) or $D < 0$ (due to external influences); for Surprise, A 0, $D < 0$, but V may be > 0 (if pleasantly surprised) or < 0 (if startled). Finally, for Neutral, all the V, A and D were equal to 0. These relationships are illustrated in Figure 1, that is, the seven basic emotion categories are in the VAD three-dimensional space.

TABLE 1. **VAD 3D emotion values for 7 standard emotions.**

	Valence	Arousal	Dominance
Happy	>0	>0	>0
Disgust	<0	>0	>0
Angry	<0	>0	>0 or <0
Fear	<0	<0	<0
Sad	<0	<0	>0 or <0
Surprise	>0 or <0	>0	<0
Neutral	=0	=0	=0

Figure 1 illustrates a comprehensive and refined representation of the valence-arousal-dominance (VAD) three-dimensional emotion model [8]. It has already been added to the FER2013 dataset, thereby making it a dataset that simultaneously possesses annotations for seven standard emotion categories and VAD three-dimensional emotion annotations. This study utilized this newly created dataset for experimentation.

C. THE FRAMEWORK OF MULTI-TASK EMOTION RECOGNITION

In the system of multi-task joint emotion recognition, a model is established by integrating the valence-arousal-dominance (VAD) three-dimensional continuous emotion analysis and seven-category discrete emotion classification, as shown in Figure 2. The two tasks share a dataset of facial images captured in natural scenes that are annotated with VAD values and emotion categories. The input is a facial image, which is fed into two recognition task modules separately: the "7 Basic Emotion Classification" module and the "VAD Emotion Regression" module. Each task generates its own Supervision Loss, representing the loss for the emotion category and regression analysis.

To conduct joint emotion recognition for VAD continuous emotion analysis and discrete emotion classification, it is necessary to express the relationship between these two tasks through Consistency Loss. This Consistency Loss represents the logical or physical relationship between two tasks. As shown in Figure 2, the Consistency Constraint between these two related tasks is utilized to demonstrate the logical relationship between VAD three-dimensional emotion information and emotion category information.

Finally, the overall loss function of this task consists of emotion classification loss, emotion regression loss, and consistency relationship loss. The model is trained to achieve the goal of providing recognition performance for both tasks through the correlation constraint between the two tasks.

D. FORMULATION OF MULTI-TASK JOINT LEARNING

In this project, multitask training will be conducted for emotion classification and valence regression.

Classification and arousal regression, emotion classification and dominance regression, and emotion classification and VAD regression, totaling four models. The formulaic definitions of the four models are as follows:

Assuming that the loss function for emotion classification is $L_{classify}$, Valence loss function for regression is L_{VReg},Arousal loss function for regression L_{Areg}, Dominance loss function for regression L_{Dreg}, and VAD 3D regression loss is L_{VADReg}. Then, the loss functions for each multitask training are defined in the following sections.

1) REGRESS V AND CLASSIFY EMOTION CATEGORY TOGETHER

Given task i, the supervised loss defined for each task are denoted as $L^i_{classify}$ and L^i_{VReg} while the consistency loss for joint training is defined as $L^i_{consist}$. Thus, this project formalizes the joint regression of V and discrete emotion classification as an optimization problem (1):

$$\min_{\theta} \sum_{i=0}^{n} \left[L^i_{classify}(\theta) + L^i_{VReg}(\theta) + \lambda L^i_{consist}(\theta) \right] \tag{1}$$

where Θ is learned by minimizing the total loss consist of $L^i_{classify}(\theta)$, $L^i_{VReg}(\theta)$, and $L^i_{consist}(\theta)$. $L^i_{classify}(\theta)$ is the cross-entropy loss as shown in Figure 2, $L^i_{VReg}(\theta)$ is the MSE loss in the figure, and $L^i_{consist}(\theta)$ is the consistency loss calculated from predicted emotion category and V as illustrated in the figure. These are defined in (2), (3), and (4).

$$L_{classify} = -\sum\nolimits_{i=0}^{n} \left(Real_i \times \log \left(Output_i \right) \right) \tag{2}$$

$$L_{V-reg} = \frac{1}{n} \sum\nolimits_{i=0}^{n} (V_i - \hat{V}_i)^2 \tag{3}$$

$$L_{consist} = \sum\nolimits_{i=0}^{n} Cost_i \left(Class_i, \hat{V}_i \right), \tag{4}$$

where $Real_i$ represents the true emotion category distribution for the i^{th} image, and $Output_i$ represents the emotional category distribution generated by the emotion classification network. L_{V-reg} is defined as $L2$ Loss, representing V and $\hat{V}$ corresponding to the real V value and delivered V value for the i^{th} image. And additionally, $L_{consist}$ is determined by the consistency loss between each of the produced emotion category $Class_i$ and generated $\hat{V}_i$, which is defined in (5).

$$Cost_i \left(Class_i, \hat{V}_i \right) = \begin{cases} 1, & if\ Class_i = 6, \hat{V}_i \neq 0; \\ 1, & if\ Class_i = 3, \hat{V}_i \leq 0; \\ 1, & if\ Class_i = 0, 1, 2, 4, \hat{V}_i \geq 0; \\ 0, & else. \end{cases} \tag{5}$$

where $Class_i$ represents the category obtained from emotion, $\hat{V}$ is the V value obtained through emotion regression. Therefore, $Cost_i$ is the cost loss calculated based on the relationship between $Class_i$ and $\hat{V}_i$. Zero to six correspond to gry, disgust, fear, happy, sad, surprise, and neutral, respectively. (5) are defined in Table 1.

2) REGRESS A AND CLASSIFY EMOTION CATEGORY TOGETHER

Joint A regression and discrete emotion classification are defined as the following optimization problem in (6).

$$\min_{\theta} \sum\nolimits_{i=0}^{n} \left[L^i_{classify}(\theta) + L^i_{AReg}(\theta) + \lambda L^i_{consist}(\theta) \right] \tag{6}$$

where $L^i_{classify}$ is calculated by (2), L^i_{AReg} and $L^i_{consist}$ are defined as (7) and (8), respectively.

$$L_{A-reg} = \frac{1}{n} \sum\nolimits_{i=0}^{n} (A_i - \hat{A}_i)^2 \tag{7}$$

$$L_{consist} = \sum\nolimits_{i=0}^{n} Cost_i \left(Class_i, \hat{A}_i \right) \tag{8}$$

where A_i represents the true A value for the ith image, and $\hat{A}_i$ represents the arousal value calculated through regression analysis for the i^{th} image. The definition of the consistency cost for multitask learning of arousal regression and label classification is given in (9).

$$Cost_i \left(Class_i, \hat{A}_i \right) = \begin{cases} 1, if\ \ Class_i = 6, \hat{A}_i \neq 0; \\ 1, if\ \ Class_i = 4, \hat{A}_i > 0; \\ 1, if\ Class_i = 0, 1, 3, 5, \hat{A}_i \leq 0; \\ 0, else. \end{cases} \tag{9}$$

where $Class_i$ represents the category obtained from emotion classification network as well, A_i is the A value obtained through emotion regression. Therefore, $Cost_i$ is the cost loss calculated from $Class_i$ and $\hat{V}_i$. (9) is defined in Table 1.

3) REGRESS D AND CLASSIFY EMOTION CATEGORY TOGETHER

D regression and emotion classification can be formulated as an optimization problem in (10).

$$\min_{\theta} \sum\nolimits_{i=0}^{n} \left[L^i_{classify}(\theta) + L^i_{DReg}(\theta) + \lambda L^i_{consist}(\theta) \right] \tag{10}$$

where $L^i_{classify}$ is calculated by (2), L^i_{DReg} and $L^i_{consist}$ are defined by (11) and (12), respectively.

$$L_{D-reg} = \frac{1}{n} \sum\nolimits_{i=0}^{n} (D_i - \hat{D}_i)^2 \tag{11}$$

where D_i represents the realistic D value for the i^{th} image and $\hat{D}_i$ is the calculated value. The consistency loss is defined by Equation (12):

$$L_{consist} = \sum\nolimits_{i=0}^{n} Cost_i \left(Class_i, \hat{D}_i \right) \tag{12}$$

where $Class_i$ represents the output emotion class for the i^{th} image, and $\hat{D}_i$ is the calculated D value. The definition of $Cost_i$ is defined by Equation (13).

$$Cost_i \left(Class_i, \hat{D}_i \right) = \begin{cases} 1, & if\ Class_i = 6, \hat{D}_i \neq 0; \\ 1, & if\ Class_i = 4, \hat{D}_i > 0; \\ 1, & if\ Class_i = 0, 1, 3, \hat{D}_i \leq 0; \\ 0, & else. \end{cases} \tag{13}$$

And (13) is defined according to Table 1 as well.

4) REGRESS VAD AND CLASSIFY EMOTION TOGETHER

As the last learning model, the VAD and emotion categories are trained together, as shown in Figure 2, where the optimization problem is defined in (14).

$$\min_{\theta} \sum\nolimits_{i=0}^{n} \left[L^{i}_{classify}(\theta) + L_{VReg}(\theta) + L^{i}_{AReg}(\theta) + L^{i}_{DReg}(\theta) \right. \\ \left. + \lambda L^{i}_{consist}(\theta) \right] \tag{14}$$

In (14), $L^{i}_{classify}$, L_{VReg}, and L^{i}_{DReg} are of the same effect and definition as above, while $L^{i}_{consist}(\theta)$ is defined in (15) and (16).

$$L_{consist} = \sum\nolimits_{i=0}^{n} Cost_i\left(Class_i, \hat{V}_i, \hat{A}_i, \hat{D}_i\right) \tag{15}$$

$$Cost_i\left(Class_i, \hat{V}_i, \hat{A}_i, \hat{D}_i\right) = Cost_i\left(Class_i, \hat{V}_i\right) \\ + Cost_i\left(Class_i, \hat{A}_i\right) \\ + Cost_i\left(Class_i, \hat{D}_i\right) \tag{16}$$

And similarly, $Cost_i\left(Class_i, \hat{V}_i\right)$, $Cost_i\left(Class_i, \hat{A}_i\right)$, and $Cost_i\left(Class_i, \hat{D}_i\right)$ are with the same definitions in above.

IV. EXPERIMENT

A. EXPERIMENTAL ENVIRONMENT SETUP AND DATASET DESCRIPTION

The GPU model of the model training server was 3090ti with 24GB memory, the processor was Intel i9 sixteen-core CPU E5-2678 with 128 GB memory, the operating system was Windows10, and the CUDA version was 11.8. The backbone network was ResNet-18, the batch size was set to 16, the initial learning rate was 0.01, it was decreased by a factor of 10 every 10,000 iterations, and the total number of epochs was 120.

This project uses a publicly available facial expression dataset based on FER2013 with annotations for VAD. It includes both emotion category labels and VAD multi-dimensional emotion parameters.

Under this hardware and software environment, with this model and dataset, the training duration is approximately 2 hours. The duration for using the successfully trained model for sentiment parameter prediction is approximately 1 minute.

B. RESNET-18 STRUCTURE FOR VAD REGRESSION

The model structure of ResNet-18 for regressing V, A, or D is illustrated in Figure 3. The entire model is divided into 4 layers stages with additional beginning and ending operations. Eash layer include 2 blocks. The first blocks in Layer2 to Layer4, i.e., Block3, Block5, and Block7 belong to Type B, while the others belong to Type A, whose detailed frames are elaborated on the right. The short cut is connected from each prior block's output, which is the core promotion method of ResNets.

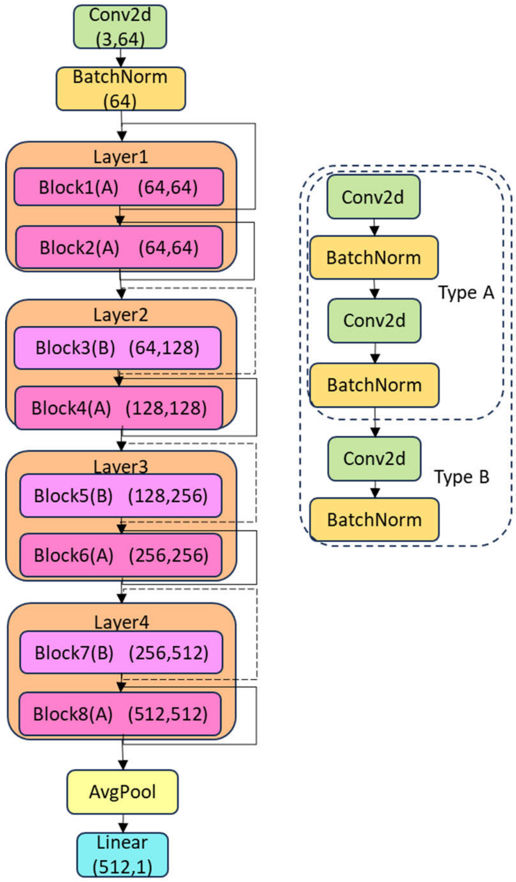

FIGURE 3. Resnet-18 for regressing V, A, or D.

C. PARAMETERS AMOUNT

To present the parameter amount, it lists the concrete calculation processes in Table2, which include operation name, input and output channel number, kernel size, bias information, stride length, and the number of parameters within each operation. For convolution operation, the parameter account in this study is formulated by (17), while the batch-norm operation's is calculated by (18).

$$Conv\_Amount = input\ channel \times output\ channel \\ \times kernel\ size \div \left(stride^2\right) \tag{17}$$

$$BN\_Amount = output\ channel \times 2 \tag{18}$$

The parameter amount for each operation is shown in the rightest column. As it goes deeper, the feature maps become smaller, and the output channels become longer. At last, the features converge to one single neuro that is the final regressed V, A, or D Value. Totally, there are 14,614,081(∼14M) parameters for regressing V, A, or D in this model.

To further improve the prediction accuracy of VAD and emotion categories, it implements orthogonal regularization on a few convolutions which are formatted as bold and italic in Table2. In this way, it extracts more diverse and expressive features than the convolution networks without it.

TABLE 2. Calculating parameter amount for each operation.

Layers and Blocks		operation	input channel	output channel	kernel	bias	stride	amount
beginning		conv1	3	64	3*3	FALSE	(1,1)	1728
		bn1		64				128
Layer 1	1	***conv1***	***64***	***64***	***3*3***	***FALSE***	***(1,1)***	***36,864***
		bn1		64				128
		conv2	64	64	3*3	FALSE	(1,1)	36,864
		bn2		64				128
	2	***conv1***	***64***	***64***	***3*3***	***FALSE***	***(1,1)***	***36,864***
		bn1		64				128
		conv2	64	64	3*3	FALSE	(1,1)	36,864
		bn2		64				128
Layer 2	1	***conv1***	***64***	***128***	***3*3***	***FALSE***	***(2, 2)***	***18,432***
		bn1		128				256
		conv2	128	128			(1,1)	147,456
		bn2		128				256
		(0): Conv	***64***	***128***	***(1,1)***	***FALSE***	***(2, 2)***	***18,432***
		(1): BN		128				256
	2	***conv1***	***128***	***128***	***3*3***	***FALSE***	***(1,1)***	***147,456***
		bn1		128				256
		conv2	128	128	3*3	FALSE	(1,1)	147456
		bn2		128				256
Layer 3	1	***conv1***	***128***	***256***	***3*3***	***FALSE***	***(2, 2)***	***73,728***
		bn1		256				512
		conv2	256	256	3*3		(1,1)	589,824
		bn2		256				512
		(0): Conv	***128***	***256***	***(1,1)***	***FALSE***	***(2, 2)***	***8,192***
		(1): BN		256				512
	2	***conv1***	***256***	***256***	***3*3***	***FALSE***	***(1,1)***	***589,824***
		bn1		256				512
		conv2	256	256	3*3	FALSE	(1,1)	589,824
		bn2		256				512
Layer 4	1	***conv1***	***256***	***512***	***3*3***	***FALSE***	***(2, 2)***	***294,912***
		bn1		512				1024
		conv2	512	512	3*3		(1,1)	7,077,888
		bn2		512				1024
		(0): Conv	***256***	***512***	***(1,1)***	***FALSE***	***(2, 2)***	***32,768***
		(1): BN		512				1024
	2	***conv1***	***512***	***512***	***3*3***	***FALSE***	***(1,1)***	***2,359,296***
		bn1		512				1024
		conv2	512	512	3*3	FALSE	(1,1)	2,359,296
		bn2		512				1024
ending		Avg pool						0
		fc	512	1		TRUE		513
Total								14,614,081

D. EXPERIMENT PROCESS AND METRICS

There are two main objectives: to provide a method for computing continuous VAD multi-dimensional emotion values, and to improve emotion recognition accuracy by utilizing the relationships between emotion categories and VAD parameters.

It performs four multitask learning steps, as described in Section III. For each of them, an ablation experiment was conducted, followed by a comparison of the results between using consistency loss and not using consistency loss. The prediction accuracy for emotion classification is measured by the accuracy, as defined in (19).

$$Accuracy = \frac{Num_{True}}{Num_{Total}} \tag{19}$$

where Num_{True} is the number of correctly detected images and Num_{Total} is the total number of images.

For the prediction accuracies of V, A and D, Average Loss is used as the metric, and they are defined separately from (20) to (22).

$$Loss_{average-V} = \frac{1}{n}\sum\nolimits_{i=0}^{n} V_{RegLoss}^{i} \tag{20}$$

$$Loss_{average-A} = \frac{1}{n}\sum\nolimits_{i=0}^{n} A_{RegLoss}^{i} \tag{21}$$

$$Loss_{average-D} = \frac{1}{n}\sum\nolimits_{i=0}^{n} D_{RegLoss}^{i} \tag{22}$$

where $V_{RegLoss}^{i}$, $A_{RegLoss}^{i}$ and $D_{RegLoss}^{i}$ are the regression losses of V, A, and D for the ith image, respectively. For the last simultaneous prediction of VAD, the accuracy was measured using (23).

$$Loss_{average-VAD} = \frac{1}{n}\sum_{i=0}^{n}(V_{RegLoss} + A_{RegLoss} + D_{RegLoss}) \tag{23}$$

E. EMOTION CLASSIFICATION, V, A, AND D REGRESSION (WITHOUT AND WITH CONSISTENCY LOSS)

1) PREDICTING V AND EMOTION CATEGORY: WITHOUT VS. WITH CONSISTENCY LOSS

The experiment trained and tested V using ablation tests with and without consistency conditions, in which the latter was trained in a multi-task learning framework. The results are shown in Figure 4.

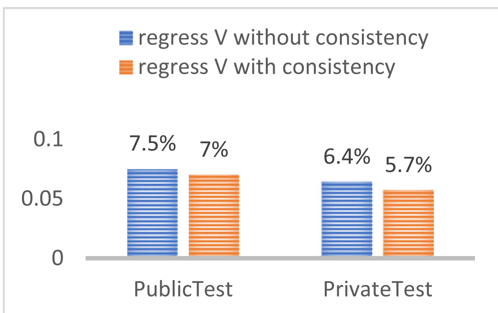

FIGURE 4. Average loss for regressing V.

It can be observed that, in both the public and private tests, the average loss for V without consistency constraints is significantly greater than that without consistency constraints. This validates the hypothesis that using consistency constraints can improve the recognition accuracy of V.

In this experiment, in addition to analyzing the prediction results for V, the results for emotion classification were also obtained, as shown in Figure 5.

The accuracy of the model without consistency is less than that of the model in both public and private tests. This also

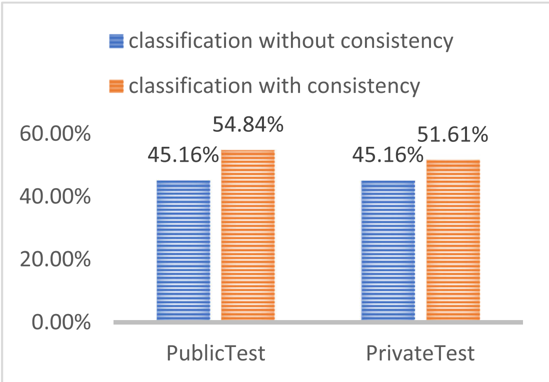

FIGURE 5. **Classification accuracy for emotion classification.**

confirms that it is useful to improve the recognition accuracy in emotion classification using consistency loss.

2) PREDICTING A: WITHOUT VS. WITH CONSISTENCY LOSS

In this experiment, ablation experiments were conducted to compare the average loss in public and private tests with and without consistency constraints, the results of which are shown in Figure 6.

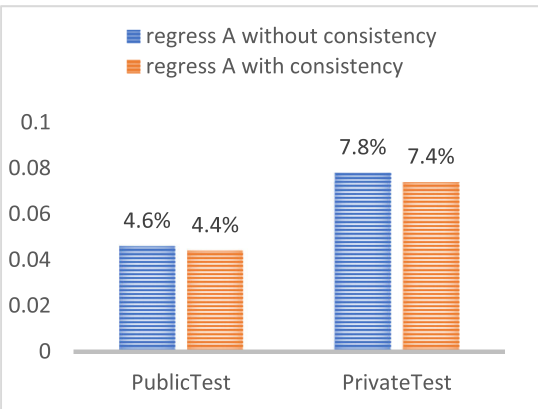

FIGURE 6. **Average loss for regressing A.**

It can be observed that in both the public and private tests, the A regression loss with consistency constraints is lower than that without consistency constraints. This also verifies the hypothesis that improving the recognition accuracy of parameter A could be achieved by enforcing consistency constraints.

3) PREDICTING D: WITHOUT VS. WITH CONSISTENCY LOSS

In this experiment, D was tested by ablation experiments with and without consistency, the results of which are illustrated in Figure. 7. It is obvious that the average loss in both public and private tests, whose learning model has consistency, is less than that without consistency. Thus, this verifies that adding consistency constraints can improve the accuracy of learning D.

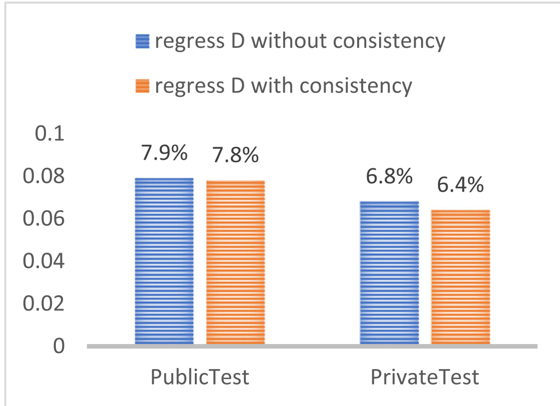

FIGURE 7. **Average loss for regress A.**

It can be observed that, in both the public and private tests, the average loss of D with consistency constraints is less than that without consistency constraints. This confirms the hypothesis that the recognition accuracy in D can also be improved by using consistency constraints.

4) PREDICTING V, A, AND D SIMULTANEOUSLY IN ONE LEARNING MODEL: WITHOUT VERSUS WITH CONSISTENCY LOSS

The results of predicting V, A, and D separately with category joint learning and the results of V, A, and D prediction with four-task joint learning (including consistency constraints between V, A, D, and category, as in equations (16) and (17)) are compared in Table 3.

TABLE 3. **Learning V, A, D respectively with emotion vs. learning V, A, D and emotion simultaneously.**

	Public Test		Private Test	
Reg.V	V&Class	V, A, D&Class	V&C	V, A, D&Class
Ave.Loss	0.07	0.07	0.057	0.057
Reg.A	A&Class	V, A, D&Class	A&C	V, A, D&Class
Ave.Loss	0.044	0.044	0.074	0.074
Reg.D	D&Class	V, A, D&Class	D&C	V, A, D&Class
Ave.Loss	0.078	0.078	0.064	0.064

From Table 3, it is observed that the results generated by the two tasks of V and category, A and category, D, and category are the same as those for the four tasks of V, A, D, and emotion category in both public and private tests.

F. PREDICTING VAD INSTANCES

To verify the generalization ability of the generated VAD multidimensional emotion prediction model, the trained model was tested on three randomly selected real-world facial images. The prediction results are shown in Table 4.

TABLE 4. Prediction results of facial emotion in real-world facial expression images.

			
	Anger	Surprise	Fear
V	-0.3989	-0.3724	-0.4131
A	0.5803	0.6106	0.5878
D	0.2395	0.2169	0.1577

From Table 3, it can be observed that when using the proposed multi-task model for emotion prediction on three randomly selected facial expression images, both the prediction of emotional categories and the prediction of VAD multidimensional emotional parameters are highly accurate. For the image categorized as ''anger'' on the left side, its V value is negative, indicating a negative emotion; the A value is positive, indicating a high arousal state; and the D value is an integer, indicating an autonomous emotional control state. Therefore, the VAD prediction results are consistent with the corresponding ''anger'' emotion. Similarly, for the remaining two images categorized as ''surprise'' and ''fear,'' the emotional prediction results also align with their emotional categories. Hence, this model exhibits good generalization ability on randomly selected test images.

G. USER STUDY FOR PRACTICAL APPLICATIONS

To explore real-world applications of the VAD emotion recognition model, it performs experiments and interviews in both educational and healthcare contexts.

From the aspect of educational application, it conducts a pilot study to assess the effectiveness of the model in predicting students' final grades by incorporating Valence-Arousal-Dominance (VAD) parameters. The results demonstrate a significant increase in prediction accuracy, indicating the potential of this model to support personalized interventions and improve educational outcomes.

For healthcare application, we collaborate with psychologists and healthcare professionals to evaluate the usefulness of our model in healthcare settings. Their feedback highlighted the model's potential in recognizing potential depression among individuals based on their VAD emotional expressions. This early detection enables timely intervention and preventative measures to mitigate the risk of depression onset, demonstrating the clinical significance of our research.

By incorporating these real-world applications and feedback from educators and healthcare professionals, it provides concrete evidence of the practical implications and societal impact of the emotion recognition model. These experiences underscore the versatility and effectiveness of our approach across diverse domains, reinforcing its relevance and applicability in real-world scenarios.

V. DISCUSSION

The study proposes a Valence, Arousal, Dominance (VAD) dimensional emotion recognition baseline that provides more comprehensive emotion states than categorical emotion models and the Valence, Arousal (VA) model, and finds that it is effective to employ consistency regularization in multi-task learning for both VAD emotion parameter regression and categorial emotion classification.

Extending the VAD model with the Dominance dimension offers several key advantages for emotion recognition. It provides a more nuanced representation of emotional expressions by capturing the interplay between valence, arousal, and dominance. This enables to distinguish between subtle variations in emotional states that may convey different levels of dominance-related behaviors. Additionally, integrating dominance cues enhances the interpretability and applicability of emotion recognition systems in real-world scenarios, such as social interactions, workplace dynamics, and human-computer interaction.

In Section IV-C, it conducts ablation experiments on V, A, and D emotion parameter regressions with and without their respective consistency with emotion classes. First, Figure 4 shows the V prediction results for both public and private tests. It is obvious that the average loss with the consistency constraint is lower than that without it by 0.005 and 0.007 for the public and private tests, respectively. This verifies that, for V, the introduction of multitask learning by consistency regularization improves the measurement accuracy. Additionally, Figure 5 shows that in the emotion classification task, the one with the consistency condition was more accurate than that without this condition by 9.68% and 6.45%, respectively. Second, Figure 6 shows the prediction results for A. It is clear that the average loss with consistency is lower than without by 0.002 and 0.004 for the public and private tests, respectively. This verifies that, for A, the use of consistency increases its prediction accuracy. Fig. 7 shows the D prediction results. It is apparent that the average loss with consistency is less than that without it by 0.001 and 0.004 separately for public and private tests, which verifies that for D, using consistency can also enhance precision.

Section IV-D shows whether the performance will be enhanced by training V, A, D and category together. However, Table 2 illustrates that learning V, A, D, and category in a single training model is equivalent to each of them trained with category. This is because the former four in one multi-task model do not introduce any additional conditions compared to the latter schemes. Thus, this implies that combining multiple tasks in one training model without other new conditions cannot improve accuracy but can increase efficiency.

It is worth noting that only the simplest and most straightforward definitions of consistency used has already been effective. For example, if the label is happy, in normal circumstances, V should be more than 0 because happiness itself is a positive and pleasant feeling whose cost should be equal to 0. Therefore, if V is less than 0 at time, it will be contradictory to Happy, and the cost loss should be 1. It is evident that only 0 or 1 is assigned to the cost loss but without further variation in values. In this condition, the results show superior performance in terms of consistency loss. Thus, this implies that the results could be further improved if the loss values are assigned with more varied values that explain the relationships between emotion categories and VAD values in more delicate patterns.

The last but not the least is that in light of the scarcity of studies on VAD 3D emotion recognition from facial expression images, the comparison tests with other baselines are not adequately performed now. But it is necessary for providing it for complete and rigorous research. Therefore, we will keep an eye ongoingly on the development in this area. We will closely monitor the advancements and eagerly await the availability of a new VAD facial expression recognition baseline. Upon its emergence, we will promptly incorporate it into our research. This proactive approach underscores our dedication to contributing to the evolution of understanding and techniques in this nascent field. As we await further developments, we remain poised to integrate new insights into our work, ensuring its relevance and efficacy.

VI. CONCLUSION

In this research, the Valence, Arousal, Dominance (VAD) three-dimensional facial expression recognition method obtains a more complete and detailed interpretation than traditional emotion classification or Valence, Arousal (VA) recognition. It then shows that emotion categories are the points in VAD three-dimensional space and this characteristic is exploited as the relation conditions in a multi-task learning model that combines VAD regression and emotion classification. A comprehensive experiment and results verified its effectiveness. However, it should be noted that combining multiple regression networks with a classification network in one cannot improve the performance compared to regressing them individually with classification without additional new constraints.

For its application, since facial expression is a very genuine and accurate parameter to display feelings, and is also convenient to obtain, this proposed VAD emotion recognition method is valuable and practical in intelligent emotion interactions or early risk warning systems such as failure risk prediction in smart education and depression risk alert in health monitoring software. From the user study for practical applications, this initial validation is encouraging and suggests that this approach holds promise.

However, annotating emotions with VAD dimensions presents inherent difficulties due to the nuanced nature of human emotions. While VAD provides a robust framework, accurately mapping subjective emotional experiences onto these dimensions can be challenging, as emotions are often multifaceted and context-dependent, which may lead to inconsistencies in annotations.

Moreover, in the realm of multitask learning, computational efficiency emerges as a significant hurdle. Balancing multiple tasks while optimizing computational resources requires careful consideration. The trade-off between model complexity and efficiency becomes pronounced, as incorporating additional tasks can strain computational resources and hinder real-time application.

Moving forward, future research directions will firstly include a more extensive and accurate VAD dataset to provide a more reliable benchmark for VAD emotion recognition. Then, another research will be conducted to explore more relations among VAD and category annotations, or to transform the structures of this multi-task network to improve prediction accuracy. Next, it will continue to explore additional datasets and real-world scenarios to further substantiate the generalizability of the model. At last, it will dedicate to apply the model in educational, psychological, healthcare, or other practical scenarios. Possible schemes include but not limited to exploring advanced machine learning techniques such as deep generative models for enhancing emotion annotation accuracy, and developing lightweight architectures for efficient multitask learning.

It emphasizes the importance of empirical studies and interdisciplinary collaborations in advancing these research directions, citing potential collaborations with experts in affective computing, computational neuroscience, and signal processing. As a whole, it aims to position it within the broader context of emotion recognition and multitask learning research.

ACKNOWLEDGMENT

The author Yi Huo thanks all the students who supported the annotation of the dataset with the VAD values. The author Yun Ge thanks H. L. for constructive suggestions on the development problems. All authors thank the editors and reviewers for reviewing this paper.

REFERENCES

[1] L. F. Barrett and C. Westlin, "Chapter 2—Navigating the science of emotion," in *Emotion Measurement*. Cambridge, U.K.: Woodhead, 2021, pp. 39–84.

[2] R. Pekrun, "Progress and open problems in educational emotion research," *Learn. Instruct.*, vol. 15, no. 5, pp. 497–506, Oct. 2005.

[3] Y. Wang, W. Song, W. Tao, A. Liotta, D. Yang, X. Li, S. Gao, Y. Sun, W. Ge, W. Zhang, and W. Zhang, "A systematic review on affective computing: Emotion models, databases, and recent advances," *Inf. Fusion*, vols. 83–84, pp. 19–52, Jul. 2022.

[4] H. Liu, T. Lou, Y. Zhang, Y. Wu, Y. Xiao, C. S. Jensen, and D. Zhang, "EEG-based multimodal emotion recognition: A machine learning perspective," *IEEE Trans. Instrum. Meas.*, vol. 73, pp. 1–29, 2024.

[5] Z. Xiong, X. Weng, and Y. Wei, "Research on the influence of visual factors on emotion regulation interaction," *Frontiers Psychol.*, vol. 12, Jan. 2022, Art. no. 772642.

[6] Y. Zhang and Q. Yang, "A survey on multi-task learning," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.

[7] M. Li, Q. Lu, Y. Long, and L. Gui, "Inferring affective meanings of words from word embedding," *IEEE Trans. Affect. Comput.*, vol. 8, no. 4, pp. 443–456, Oct. 2017.

[8] A. Mollahosseini, B. Hasani, and M. H. Mahoor, "AffectNet: A database for facial expression, valence, and arousal computing in the wild," *IEEE Trans. Affect. Comput.*, vol. 10, no. 1, pp. 18–31, Jan. 2019.

[9] E.-J. Park, M. Kikutani, N. Suzuki, M. Ikemoto, and J.-H. Lee, "Properties of central and peripheral concepts of emotion in Japanese and korean: An examination using a multi-dimensional model," *Frontiers Psychol.*, vol. 13, p. 13, Feb. 2022.

[10] K. Hulliyah, A. Wahab, N. Kamaruddin, S. Erdogan, and Y. Durachman, ''Analysis of Indonesian sentiment text based on affective space model (ASM) using electroencephalogram (EEG) signals,'' in *Proc. Int. Conf. Informat. Comput. (ICIC)*, Mataram, Indonesia, Oct. 2016, pp. 325–328.
[11] Y. Ma, Y. Hao, M. Chen, J. Chen, P. Lu, and A. Kosir, ''Audio-visual emotion fusion (AVEF): A deep efficient weighted approach,'' *Inf. Fusion*, vol. 46, pp. 184–192, Mar. 2019.
[12] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D. H. Lee, and Y. Zhou, ''Challenges in representation learning: A report on ree machine learning contests,'' in *Proc. ICML*, Jul. 2013, pp. 117–124.
[13] M. Minsky, *The Society of Mind*. London, U.K.: Simon & Schuster, 1988.
[14] R. W. Picard, *Affective Computing*. Cambridge, MA, USA: MIT Press, Jul. 2000.
[15] J. A. Marcum, ''The role of emotions in clinical reasoning and decision making,'' *J. Med. Philosophy*, vol. 38, no. 5, pp. 501–519, Oct. 2013.
[16] Y. Huo ''Online Learning risk early warning based on multi-dimensional continuous emotion recognition,'' *Mod. Inf. Technol.*, early access, Jun. 2024.
[17] C. Weitzman and L. Wegner, ''Promoting optimal development: Screening for behavioral and emotional problems,'' *Pediatrics*, vol. 135, no. 2, pp. 384–395, 2015.
[18] M. Dirik, ''Optimized anfis model with hybrid metaheuristic algorithms for facial emotion recognition,'' *Int. J. Fuzzy Syst.*, vol. 25, no. 2, pp. 485–496, Mar. 2023.
[19] J. Tian and Y. She, ''A visual–audio-based emotion recognition system integrating dimensional analysis,'' *IEEE Trans. Computat. Social Syst.*, vol. 10, no. 6, pp. 3273–3282, Dec. 2023.
[20] Z. Shuyuan, ''Study on the academic emotions of college students sed on the control-value theory,'' 2nd Hospital Xiangya, Central South Univ., Tech. Rep., 2013, doi: 10.7666/d.Y2685238.
[21] A.-J. Kyröläinen, J. Luke, G. Libben, and V. Kuperman, ''Valence norms for 3,600 English words collected during the COVID-19 pandemic: Effects of age and the pandemic,'' *Behav. Res. Methods*, vol. 54, no. 5, pp. 2445–2456, Dec. 2021.
[22] H. Han, M. J. Seong, J. Hyeon, E. Joo, and J. Oh, ''Classification and automatic scoring of arousal intensity during sleep stages using machine learning,'' *Sci. Rep.*, vol. 14, no. 1, p. 5983, Mar. 2024.
[23] S. K. Moeller, E. A. E. Lee, and M. D. Robinson, ''You never think about my feelings: Interpersonal dominance as a predictor of emotion decoding accuracy,'' *Emotion*, vol. 11, no. 4, pp. 816–824, 2011.
[24] H. Sourkatti, K. Pettersson, B. van der Sanden, M. Lindholm, J. Plomp, I. Määttänen, P. Henttonen, and J. Närväinen, ''Investigation of different ML approaches in classification of emotions induced by acute stress,'' *Heliyon*, vol. 10, no. 1, Jan. 2024, Art. no. e23611.
[25] A. Mehrabian, *Basic Dimensions for a General Psychological Theory Implications for Personality, Social, Environmental, and developmental Studies*. Cambridge, U.K.: Oelgeschlager, Gunn & Hain, 1980.
[26] X. Guo, S. Li, N. Guo, J. Cao, X. Liu, Q. Ma, R. Gan, and Y. Zhao, ''Disentangled representations learning for multi-target cross-domain recommendation,'' *ACM Trans. Inf. Syst.*, vol. 41, no. 4, pp. 1–27, Oct. 2023.
[27] K. Yang, T. Zhang, H. Alhuzali, and S. Ananiadou, ''Cluster-level contrastive learning for emotion recognition in conversations,'' *IEEE Trans. Affect. Comput.*, vol. 14, no. 4, pp. 3269–3280, Oct./Dec. 2023.
[28] Y. Fang, S. Xiao, M. Zhou, S. Cai, and Z. Zhang, ''Enhanced task attention with adversarial learning for dynamic multi-task CNN,'' *Pattern Recognition*, vol. 128, May 2022, Art. no. 108672, doi: 10.1016/j.patcog.2022.108672.
[29] T. Souč ek, J.-B. Alayrac, A. Miech, I. Laptev, and J. Sivic, ''Multi-task learning of object states and state-modifying actions from web videos,'' *IEEE Trans. Pattern Anal. Mach. Intell*, early access, 2024, doi: 10.1109/TPAMI.2024.3362288.
[30] R. Bachmann, D. Mizrahi, A. Atanov, and A. Zamir, ''MultiMAE: Multi-modal multi-task masked autoencoders,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2022, pp. 348–367.
[31] Z. Chen, Y. Shen, M. Ding, Z. Chen, H. Zhao, E. Learned-Miller, and C. Gan, ''Mod-squad: Designing mixtures of experts as modular multi-task learners,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 11828–11837.
[32] Z. Zhu, K. Lin, A. K. Jain, and J. Zhou, ''Transfer learning in deep reinforcement learning: A survey,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 11, pp. 13344–13362, Nov. 2023.
[33] S. J. Pan and Q. Yang, ''A survey on transfer learning,'' *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010, doi: 10.1109/TKDE.2009.191.
[34] D. Sinodinos and N. Armanfard, ''Attentive task interaction network for multi-task learning,'' in *Proc. ICPR*, Aug. 2022, pp. 2885–2891.
[35] M. Pirvu, A. Marcu, A. Dobrescu, N. Belbachir, and M. Leordeanu, ''Multi-task hype rgraphs for semi-supervised learning using earth observations,'' in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, France, 2023, pp. 3396–3406.
[36] D. Kollias, V. Sharmanska, and S. Zafeiriou, ''Distribution matching for multi-task learning of classification tasks: A large-scale study on faces & beyond,'' in *Proc. AAAI*, 2024, pp. 1–10.
[37] Y. Bing, X. Xueqin, and K. Wanzeng, ''Automatic art analysis of adaptive multi-task learning,'' *Chin. J. Image Graph.*, vol. 27, no. 4, p. 12, 2022.
[38] Y. Huang, E. H. Wang, Z. Liu, L. Pan, H. Li, and X. Liu, ''Modeling task relationships in multivariate soft sensor with balanced mixture-of-experts,'' *IEEE Trans. Ind. Informat.*, vol. 19, no. 5, pp. 6556–6564, May 2023.

YI HUO (Member, IEEE) received the Ph.D. degree in computer science from the Communication University of China, in 2014. She was a Research Assistant with the IBM China Research Laboratory, in 2007. She completed the postdoctoral work with Beijing University of Technology, in 2016. She is currently an Assistant Professor with Beijing Union University. Her main research interests include artificial intelligence, computer vision, smart education technology, and virtual reality.

YUN GE received the Ph.D. degree in computer science from Beijing University of Technology, in 2012. From 2014 to 2017, he was a Postdoctoral Researcher with Beijing University of Technology. He is currently an Associate Professor with the Department of Computer Teaching and Research, University of Chinese Academy of Social Sciences. His current research interests include artificial intelligence, pattern recognition, face recognition, and natural language processing.