Emotion-Aware and Intelligent Internet of Medical Things Toward Emotion Recognition During COVID-19 Pandemic

Tao Zhang, *Member, IEEE*, Minjie Liu, Tian Yuan, and Najla Al-Nabhan, *Member, IEEE*

***Abstract*—The Internet of Medical Things (IoMT) is a brand new technology of combining medical devices and other wireless devices to access to the healthcare management systems. This article has sought the possibilities of aiding the current Corona Virus Disease 2019 (COVID-19) pandemic by implementing machine learning algorithms while offering emotional treatment suggestion to the doctors and patients. The cognitive model with respect to IoMT is best suited to this pandemic as every person is to be connected and monitored through a cognitive network. However, this COVID-19 pandemic still remain some challenges about emotional solicitude for infants and young children, elderly, and mentally ill persons during pandemic. Confronting these challenges, this article proposes an emotion-aware and intelligent IoMT system, which contains information sharing, information supervision, patients tracking, data gathering and analysis, healthcare, etc. Intelligent IoMT devices are connected to collect multimodal data of patients in a surveillance environments. The latest data and inputs from official websites and reports are tested for further investigation and analysis of the emotion analysis. The proposed novel IoMT platform enables remote health monitoring and decision-making about the emotion, therefore greatly contribute convenient and continuous emotion-aware healthcare services during COVID-19 pandemic. Experimental results on some emotion data indicate that the proposed framework achieves significant advantage when compared with the some mainstream models. The proposed cognition-based dynamic technology is an effective solution way for accommodating a big number of devices and this COVID-19 pandemic application. The controversy and future development trend are also discussed.**

***Index Terms*—Cognitive model, Corona Virus Disease 2019 (COVID-19), emotion-aware, healthcare management systems, internet of medical things (IoMT).**

Manuscript received July 1, 2020; revised October 19, 2020; accepted November 12, 2020. Date of publication November 17, 2020; date of current version October 22, 2021. This work was supported in part by the National Science Foundation, China under Grant 61702226; in part by the Natural Science Foundation of Jiangsu Province under Grant BK20170200; in part by the China Postdoctoral Science Foundation under Grant 2019M661722; in part by the Fundamental Research Funds for the Central Universities under Grant JUSRP11854; and in part by the Deanship of Scientific Research at King Saud University through Research Group under Grant RG-1441-331. *(Tao Zhang and Minjie Liu are co-first authors.) (Corresponding author: Tao Zhang.)*

Tao Zhang is with the School of artificial intelligence and computer science, Jiangnan University, Wuxi 214122, China (e-mail: taozhang@jiangnan.edu.cn).

Minjie Liu is with the School of Nursing, Taihu University of Wuxi, Wuxi 214064, China (e-mail: minjieliu@163.com).

Tian Yuan is with the School of Computer Engineering, Nanjing Institute of Technology, Nanjing 210000, China (e-mail: ytian@njit.edu.cn).

Najla Al-Nabhan is with the Department of Computer Science, King Saud University, Riyadh 11682, Saudi Arabia (e-mail: nalnabhan@ksu.edu.sa).

Digital Object Identifier 10.1109/JIOT.2020.3038631

I. Introduction

GLOBAL countries is influenced by the novel coronavirus (SARSCoV-2) that was first found from Wuhan, now named it as Corona Virus Disease 2019 (COVID-19). Up to now, the number of COVID-19 infection cases of the whole world has reached 40 195 729 out of which 1 113 645 deaths are counted and existing 9 323 749 cases are continuously affecting the global countries [1]. In the face of this severe outbreak, global health workers and researchers are now looking for new technologies to screen for and control the spread of the virus. So far, for the sake of reliability and safety, isolating patients is an effective strategy to avoid disease spread, but long-term isolation can also cause psychological problems [2], [3]. In this case, related research with respect to the Internet of Things (IoT) and machine learning (ML) are currently available technology mediators to address major emotional and psychological issues associated with COVID-19.

With the rapid development of IoT devices, wireless technology, and IoT services and applications, the Internet of Medical Things (IoMT) has a new opportunity for development. In the IoMT system, wireless connected devices and smart sensors generate a large amount of signals, including multimodal data. Because of the huge, complex, and multidimensional nature of the wireless big data generated by connected healthcare devices, how to analyze the data effectively becomes a difficulty. Building emotion-aware recognition system with respect to IoT is the key for offering emotional solicitude and improving healthy lives, especially during this COVID-19 epidemic. The above system needs wireless communication frameworks that are aware of big data to support emotionally connected medical big data through better insight into medical signals to offer high quality nursing for infants and young children, the elderly, and the persons with mental problems. Therefore, the IoMT technology has great superiority in supporting these emotions or emotional communication. In this digital age of new epoch, the latest development of the IoT 5G network, artificial intelligence algorithms, big data analysis, cloud computing, Industry 4.0, and edge computing technology can provide effective solutions when confronting the COVID-19 pandemic [4]–[7].

With the development of the IoT and 5G technology, the intelligent healthcare framework with cognition scheme becomes a possibility. Many countries have

developed smart cities that provide their citizens with leading technology. Cutting-edge research hotspots in IoMT is exploiting emotion-aware detection modules in existing IoMT systems [8]–[11]. Suppose in such a monitoring scenario, the elderly and infants are the targets being monitored. This surveillance scene features many smart homes with IoTs. Suddenly, one of them fell to the ground, unable to move. IoTs constantly captures signals, which can be detected by model algorithms on cloud servers. The official workers needs to send paramedics to the accident site. Global positioning system (GPS) is then used to guide the caregiver to the location of the incident; In addition, an accurate human tracking system is designed to accurately locate the customer's fall. There are many IoMT frameworks described in [12]–[16]. The above frameworks also involve some emotion recognition algorithm's design and the network structure design.

At the age of data deluge, a massive amount of data is being produced daily by IoT [17]. While for the object recognition task, the information of the target image can usually be well represented by the manually designed feature description operator. However, the manually designed feature is not only a waste of time, but also requires professional knowledge and experience in related fields [18]. Reviewing information processing mechanism in human vision, visual cortex can not only carry out a visual abstract step by step, and it can also be used in a variety of self-organization, unsupervised or supervised visual cognitive learning approaches to discriminate with invariance of object attribute from complex visual stimulation of the outside world. In recent years, deep learning technology has obtained the amazing achievement, also provides a new way of thought to build simulation human visual information perception model of the optic nerve system; deep learning is used to extract hierarchical abstract information to represent the target image layer by layer. Combining the visual perception system and neuron transmission process, this representation method can build an efficient object recognition system.

In this article, the emotion-aware healthcare framework within IoMT system is proposed through designing a discriminative emotion recognition module. Our contributions are mainly in the following four aspects: 1) designing a novel emotion-aware detection module in the IoMT system for nursing needs; 2) this article constructs a novel local descriptor to describe features from the collected signals to to avoid the feature loss caused by signal compression and other operations; 3) considering the multimodal nature of the data, this article proposes an hierarchical deep cognitive model to classify the emotion types; and 4) with the help Bluetooth and 5G technology, this article constructs a robust tracking model, which will guide the caregivers to find the patients.

The structure of this article is discussed as follows. Recent and related research is discussed in Section II. Section III gives our proposed IoMT-enabled emotion-aware healthcare framework. Section IV presents our experimental results and data analysis. Finally, the controversy and future development trend are also discussed in Section V.

II. Related Work

Emotion recognition is a hot and not a new topic, however, many challenges remain unresolved. Emotion recognition relies heavily on the accuracy of the database, signal acquisition, the environment, the way of the signal acquisition, pattern (voice, images, or video), the types of big data, multiple modal and the available machine power factors, and in this COVID-19 pandemic, how to more effectively integrate it into the IoMT framework, and effectively deal with emotional problems associated with the outbreak, is also a new difficulty. The IoT is a mixed network, including connected physical objects, such as wireless sensors, healthcare devices, smart furniture, home appliances, and smart wearable products. This convenient connectivity enhances the awareness, processing, and interaction capabilities of devices, automatically interact with people and provide quick and convenient service. At present, the COVID-19 epidemic has become a active research topic in medical area. Artificial intelligence and ML technologies may be effective strategy to solve the global crisis. The IoT system, and IoMT system in particular, could address the detection, surveillance, trajectory tracing, and emotion recognition problem during the COVID-19 pandemic.

Samira *et al.* [19] tried to classify and identify emotions in some classic Hollywood movie clips, the multilayer cognitive structure was proposed, the BP algorithm is adopted to optimize and learn network weights, and the relationship projection of the sparse spatial feature is used to reduce model parameters to improve training performance. Some feature processing algorithms are summarized and a robust ML algorithm is used to recognize emotions with respect to EEG signals [20], in terms of the structure, this article takes the local area of the image as the lowest layer of the hierarchical structure, and information is transmitted layer by layer, and each layer obtains the most significant feature of the target through a convolution kernel. Robert *et al.* [21] designed a robust expression recognition system in the framework of IoT, which can be used as an intelligent healthcare system. This model is composed of n convolutional layers and n pooling layers alternately. Convolution operation is carried out through the trainable convolution kernel, and weighted average sum is carried out in the local area of each feature map. In this way, convolutional neural network (CNN) is invariant to considerable spatial translation. The deep belief network structure is adopted to conduct the emotions recognition problem from multimodel input data [22]. They found that the higher order nonlinear relationship is advantageous in describing emotional characteristics and can be classified and recognized effectively.

A large number of speech feature description approaches were proposed in [23] to determine the behavior types of emotion from input data. CNN is an efficient deep learning algorithm. In 2012, Wang *et al.* [24] applied the AlexNet model (a model based on the CNN network structure) in the emotion recognition competition and won the champion, thus setting off a wave of deep learning research. In this model, the local response normalization (LRN) layer is deployed in the CNN-based deep learning model, and the Dropout

algorithm is proposed to solve the overfitting problem. Since then, more scholars have optimized and improved the CNN model from the perspective of structure and algorithm. In 2011, Sun and Pan [25] used an activation function [similar to biological nerve rectified linear unit (ReLU)] to replace the traditional tanh activation function, which further enhanced the expression ability of CNN model. In 2012, Iosifidis *et al.* [26] proposed a DropConnect algorithm, which effectively improved the generalization ability of the neural network at the full connection layer. Song *et al.* [27] proposed a random pooling algorithm, which can prevent overfitting during model training. In 2014, He *et al.* [28] proposed a spatial pyramid pooling (SPP) algorithm to adaptively obtain effective features of object images at different scales, which further approximates the spatial structure of CNN to the actual biological neural network.

Hossain *et al.* [29] constructed a real-time emotion detection module toward big data of IoT. The medical framework of the system was developed based on 5G technology, and the reliability of the system was 83%. Leonardo *et al.* [30] carried out a corresponding MRI experiment on the experimental data group that needed to be tested to find the relationship between emotion and voluntary attention. In this experiment, the main purpose was to investigate the recognition ability of depression. Lisa *et al.* [31] investigated the related factors of facial emotion and corresponding experiments were carried out to verify these factors. Besides, some researchers tried to use ultrasonic radio-frequency signals and the mobile phone to collect the body characteristics [32]–[34], the emotion recognition algorithm will have obvious difference based on their differences in acquisition equipment, and the strength of the signal equipment will also affect the characteristics of the robustness of the collected signals, thus in the later stage, it will affect the feature classification and emotion recognition. This strategy can also be used to enhance the game's level of interaction with the interaction experience.

Of course, there are other emotion recognition systems based on audio-visual [35]–[37]. For example, Melo-frequency Cepstral coefficients and popular formant frequency are described in these papers for audio data, while some metric learning algorithms and time-moving images are used for video data. Finally, some basic classification algorithms (SVM, HMM, extreme learning machine, etc.) are used to conduct recognition task. The problem is that it is too mechanically balanced with the corresponding ML algorithms and cannot give full play to the advantages of ML.

The loss of feature extraction by the convolution operation mainly comes from the increase of the variance of the estimation caused by local neighborhood limitation and the deviation of the average estimation caused by the convolution error of the interlayer [38], [39], but the interaction of positive and negative activators is not considered. Stochastic pooling method fall in between above two methods [40], that is, assign probability value according to pixel value point by point, and solve the probability in the corresponding region through carrying out polynomial distribution location sampling. This is equivalent to the maximum pooling strategy adopted after the pixel points in the action region are adjusted, that is, the maximum pooling criterion is followed on the basis of the mean pooling strategy. The formula of stochastic pooling sampling probability is expressed as: $P_i = a_i / \sum_{k \in R_j} a_k$, where P_i denotes the stochastic pooling sampling probability, a_i denotes the pixel value at position i, R_j denotes jth pooling region. Overlapping pooling technology is also a way to prevent overfitting. The adjacent neurons treated by the pooling layer cells do not overlap in the traditional pooling method, each grid has an interval of the pooling unit of s pixels, each cell is centered on that unit, the range of neurons group is $z \times z$. The size of traditional pooling is selected as $s = z$, overlapping pooling is selected as $s < z$, and resulted dimension conducted by the two kinds of pooling method is the same. It provides a wider range of local invariance features for the CNN model When ensuring a significant reduction in the number of parameters.

The purpose of the normalization layer is to improve the feature accuracy on the basis of maintaining the affine invariance of the input signal. The normalization of local response mainly use the idea of lateral suppression to realize local suppression, especially when activation functions with wide boundary expansibility are activated. $a^i_{x,y}$ denotes the neuron activity at point (x, y), then the expression of the response normalized weight active value is expressed as $b^i_{x,y} = a^i_{x,y} / [k + \alpha \sum_{j=\max(0,1-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2]^\beta$, where n is the sum of the number of adjacent convolution kernels in the same space range, N is the number of convolution nuclei at that layer. Constant values k, n, α, and β are the hyperparameters obtained by debugging on the validation set [41].

Once effective feature representation is obtained, classification is often proceeded by referring some traditional recognition approaches, such as the nearest neighbor scheme [23], artificial network framework [42], and support vector machines [43]. However, the above algorithms rely on the discriminative power of constructed descriptors. Wright *et al.* [44] regulated this kind of classification mechanism at the fist time and applied it on the classical face recognition task. Considering the advantages of shared dictionary in the dictionary learning, many researchers focus on the modifications of representation coefficients [45]. Until recent studies, some more effective dictionary learning models have been proposed in [46]. Although the above algorithms have obtained satisfactory results in some applications, how to construct multicore combination through appropriate rules and learn the optimal kernel weight is a hot research direction in sparse classification at present.

Based on the previous literature research report, effective analysis and identification of emotional type is very difficult in the real world, considering the robustness of the existing healthcare framework is not stable and presents weak generation, so it is necessary to establish one more efficient IoMT-enabled emotion-aware healthcare framework, it will also become an important measure of healthcare service satisfaction degree.

In this article, emotion-aware healthcare framework within the IoMT system is proposed through designing a discriminative emotion recognition module. This article proposes a novel deep learning framework with well constructed modified Weber local descriptor (MWLD) descriptor in order

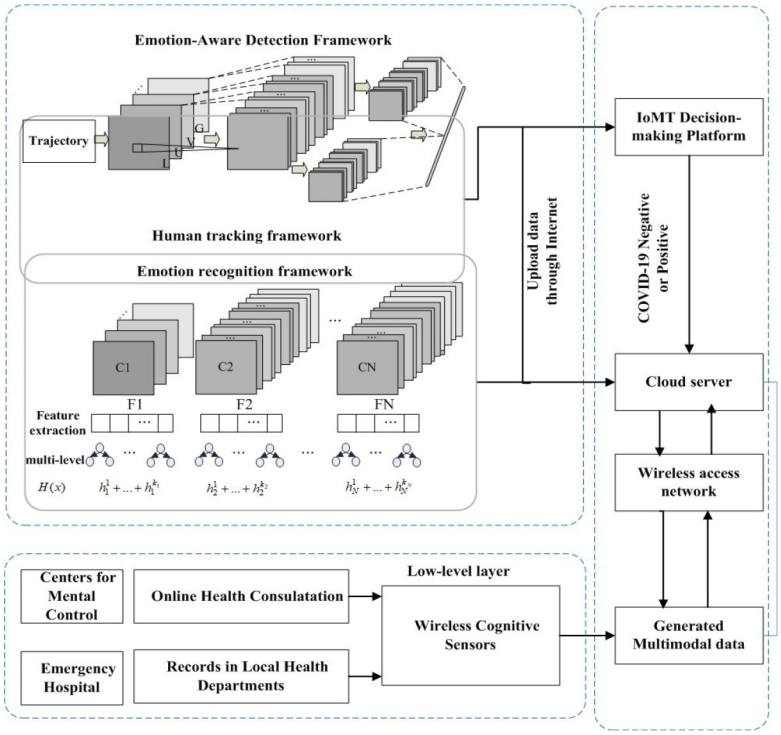


Fig. 1. Proposed IoMT-enabled emotion-aware healthcare framework.

to generate a more effective representation of input data. First, a multichannel MWLD filter was constructed to simulate the response of retinal ganglion cells to wild visual stimuli from the input data. Second, this article introduces the modified nonlinear constraint rule as the activation function. In addition, the spatial pyramid random pool layer is used to obtain the effective description of features so as to avoid the feature loss caused by signal compression and other operations. Subsequently, this article proposes the modified random DropConnect method and classification model to perform recognition task. A classification framework integrating optimized sparse coefficients is then constructed to conduct high-accuracy emotion classification.

III. Proposed IoMT-Enabled Emotion-Aware Healthcare Framework

Fig. 1 presents the proposed IoMT-enabled emotion recognition and human tracking framework. There are four innovative modules in the framework: 1) low-level module; 2) emotion recognition module; 3) human tracking module; and 4) IoMT decision-making module. The low-level module corresponds to the infrastructure structure of IoT, which consists of three submodules: 1) intelligent cognitive devices including some wireless and smart sensors; 2) popular communication devices, including some intelligent radio and 5G networks; and 3) storage and computation devices, including cloud servers and edge computing servers. The low-level module mainly consists of some infrastructures, such as some hospital and health departments, some smart sensors, and some communication devices. As the amount of input data is large, an appropriate optimization technique are required to achieve a flexible and robust framework. Some recent technologies to accomplish network optimization include network function visualization (NFV), software-defined networking, and self-organizing network. Intelligent IoMT devices are connected to collect multimodal data of patients in a surveillance environment. The latest data and inputs from official websites and reports are tested for further investigation and analysis of the emotion analysis. The IoMT decision-making module is mainly in charge of the disposal of resource and effective communication of signals. Because the amount of IoMT data with human emotion is very big, it is needed to establish an appropriate optimization strategy and communication technology so that the hardware equipments can run smoothly in real time with low-power consumption. In addition, the network structure is further optimized through some popular technology, such as NFV solution, network slicing technology, and self-organizing network for variable clustering. The emotion-aware detection module mainly includes a series of feature processing and pattern recognition strategies. The main goal of this module is to select the data to obtain valuable feature and recognize the types of the data based on characteristics and behaviors of targets. The Human tracking layer mainly consist of modules of real-time tracking through some pattern recognition algorithms. The main purpose of this layer is to track the human to get accurate position information and determine the location of accident cases. In the constructed framework, the aim is to identify the types of input emotion data and achieve the accurate location of monitored personnel.

The proposed novel IoMT platform enables remote health monitoring and decision making about the emotion, therefore, immensely contribute convenient and continuous intelligence

healthcare services during the COVID-19 pandemic. The IoMT sensors capture speech and image data, and the data are further utilized to distinguish the emotion types of observers. A robust tracking system is proposed to track the observed that needs help. On the basis of the CNN network structure, the overlapping pooling layer with stochastic strategy and the LRN layer with visual cognitive lateral suppression are combined in this article. According to the three stages of optic nerve information cognition, the convolutional layer, overlapping stochastic pooling layer and local corresponding normalized layer are alternately used to extract signal features. Finally, the fully connected neural network is used to simulate the advanced cognition process of the optic nervous system, and the extracted distributed features are projected to the labeled sample space.

A. Emotion Recognition Framework

This article proposes to construct an innovative emotion recognition framework in the IoMT healthcare management system. The designed module can effectively recognize a kind of emotion of a monitored personnel by inputting two kinds of data: 1) speech and 2) image signals. The main aim of the module is to determine whether the observed presents an abnormal emotion or not. If the well-constructed module shows that the observed presents abnormal emotion, the system will alarm the situation and notify the relevant monitoring personnel, such as official medical doctors, local health departments, hospitals, and other health consultation centers. Then, after obtaining on the monitored data, the IoMT system staff may arrange for ambulances and paramedics to treat depending on the actual situation.

Some intelligent cognitive sensors in the low-level module capture multimodal signals of the monitored personnel. In this case, it may be a smart cellphone, watch, or camera, of course, other sensors that can capture speech and video data are qualified for the role. In the constructed IoMT system, these signals are processed in a cloudlet network where the speech and video signal are divided into some representative frames, which can release more bandwidth and the system process would speed up. The first stage in the constructed cloudlet network is to track the human and crop the human face. Preprocessed voice and video signals are then transmitted to the cloud servers and the require further computation in the decision-making module.

This article uses Weber local law to depict the processing of this visual system. Chen *et al.* [47] constructed a Weber local descriptor (WLD) to achieve the robust description of local features, this kind of WLD descriptor contains two main characteristics, magnitude and orientation, which can be found in [47]. However, the above approaches [47], [48] neglected some of the details of the signal, so the change characteristics of the target cannot be described accurately. To better describe the characteristics of signal, MWLD is constructed.

In our proposed MWLD, the magnitude is defined:

$$\bar{\xi}_m(x_c) = \arctan \alpha < X, I > . \tag{1}$$

Feature Difference in x is defined:

$$\bar{\xi}_{m-x}(x_c) = \arctan \alpha < X, \cos J > . \tag{2}$$

Feature Difference in y is defined:

$$\bar{\xi}_{m-y}(x_c) = \arctan \alpha < X, \sin J > . \tag{3}$$

Thus, we can get the novel orientation as follows:

$$\bar{\xi}_o(x_c) = \arctan\left(\frac{\bar{\xi}_{m-y}(x_c)}{\bar{\xi}_{m-x}(x_c)}\right) \tag{4}$$

where $<,>$ is the inner product operator, $\cos J = (\cos\theta_0, \cos\theta_1, \ldots, \cos\theta_{p-1})^T$, $\sin J = (\cos\theta_0, \cos\theta_1, \ldots, \cos\theta_{p-1})^T$, $I = (1, 1, \ldots, 1)^T$, and $X = ([(x_0 - x_c)/x_c], [(x_1 - x_c)/x_c], \ldots, [(x_{p-1} - x_c)/x_c])^T$. θ_i denotes the angle. Because the angle between x direction and $x_i - x_c$ (i.e., the vector formed by the subtraction of the central pixel's coordinates from the coordinates of the ith neighbor of x_c) is θ_i, for $i = 0, 1, \ldots, p-1$. Then, the projections of X's components to x and y directions could be used to compute the effective contributions of the components for the definitions of MWLD magnitudes in x and y directions [refer to (2) and (3)], respectively. It is worth noting that the *arctangent* function is used to prevent the output from being too large and thus effectively suppressing the side effect of signal noise, and α is a parameter used to balance the difference between different signals.

Compared to traditional WLD, MWLD represents input signals more effectively in detail. To establish the phase relationship between the adjacent cells, in this article, the response characteristics of odd-even symmetric cells were simulated by combining the magnitude of MWLD with orientation of MWLD. Based on above ideas, this article proposes to build the low-level visual feature cognition model, the overall structure is shown in Fig. 2. In this model, a low-level visual feature cognition scheme was constructed to simulate the receptive field visual stimulus response of retinal ganglion cells, and the primary features of the input signal were extracted instead of the convolutional layer of the first layer of traditional CNN. Then, the stochastic pooling layer is adopted to carry out the downsampling operation on the primary feature map, so that the local transformation of our proposed algorithm to the input signal has certain invariance while effectively reducing the data volume. Finally, the lateral inhibition mechanism between nerve cells was introduced, and the response process of neurons to primary features was simulated by LRN layer, and feature amplification and noise screening were performed on the feature map after subsampling. Thus, the low-level visual feature cognition is constructed.

Logistic Sigmoid and Tanh Sigmoid are two different nonlinear Softplus functions that are commonly used in traditional CNN. From the mathematical perspective, the nonlinear Softplus function has a small signal gain in bilateral area, while this signal gain become more and more in the central area, it will lead to a better performance on signal feature space projection. From the perspective of neuroscience, the central area matches with the active state of neuronal and bilateral area matches with of inactive state of neurons, this is equivalent to request activation function with two different states.

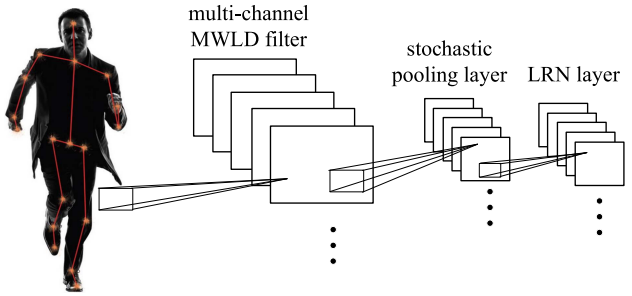


Fig. 2. Low-level visual feature cognition model.

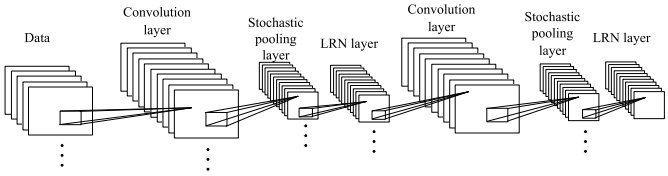


Fig. 3. Constructed intermediate-level visual feature cognition model.

While for CNN, the key features correspond to the central area, the others correspond to the bilateral area. However, from a statistical point of view, almost 50% of the neurons of the traditional Softplus function are activated at the same time, this is far from the truth (5% of the neurons are activated at the same time), which will bring a huge computation problem in the CNN training process [49].

In addition, there is another common activation function, named as ReLUs [41], it can be described as $f(x) = \max(0, x)$. Compared with Softplus activation function, the main changes include three points: unilateral suppression, wide boundary and sparse activation, it should be noted that the sparsity activation of biological nerves is inferred based on the observation and learning of brain energy consumption. Only fewer neurons has strong response when receiving an image of the visual system, most active neurons are weakly or not activated.

Therefore, in this article, smooth characteristic of Softplus function and sparse characteristics of Rectifier function are merged together, a kind of novel rectified softplus units (ReSUs) is proposed, it is defined as $f(x) = \max[0, \ln(e^x + 1)/2]$.

Compared with Softplus activation function, the proposed ReSUs model further enhances the ability of expression of the model through a kind of nonlinear projection operation. Most of the input signal is neglected from the perspective of neurons, only little part of input signals is selectly activated, so it can improve the accuracy of learning and the sparse features were extracted more accurately, and the expression ability of biological activation was better than that of Softplus and ReLUs.

In order to more accurately stimulate cognitive processing mechanism of intermediate visual feature, this article proposes to construct ReSUs activation function to achieve effective expression of nonlinear sparse connection character. The overall structure of intermediate-level visual feature cognition model was shown in Fig. 3.

CNN has the ability to resize target images, that is because that the number of parameters is fixed after the connection weights matrix ω of the fully connected layer is trained. Convolution operation and pooling operation require no limitation for the size of the input image, therefore, SPP approach [28] is used to expand the output feature of the pooling layer to a multilevel feature, and reduce the output feature dimension of the pooling level. In Caffe [50], the CNN framework usually adopts the fixed scale in output layer. Therefore, a new SPP layer is defined in this article, it is assumed that there is $a \times a$ output in the last convolutional layer while training, and SPP has $n \times n$ bins, and the calculation formulas of window (X) and step size (S) are, respectively, defined as $X = \lceil a/n \rceil$ and $S = \lfloor a/n \rfloor$. The SPP layer network structure adopted in this article is the sparse pyramid pooling layer of 3×3, 2×2, 1×1, respectively. The stochastic pooling operation is performed in each layer of the sparse pyramid pooling strategy. Finally, the feature rasterization of different layers is connected as the input of sparse coding, and dictionary learning is conducted to obtain the sparse representation of the input.

In this article, on the basis of the sparse connection spatial structure of neurons in the intermediate visual feature area, the spatial pyramid stochastic pooling layer is added to transform the features into the full-connected layer with the same dimension, so that the AH-CNN structure can receive images with any size. The proposed high-level visual feature cognition model is shown in Fig. 4. DropConnect [26] randomly samples these neurons' connection weight parameters according to certain probability in the network training process, the sampled subnetwork as the current target network is used to iterate temporarily, then DropConnect randomly samples these neurons' connection weight again in the next iteration process, ensure every iteration in different network training, thus, sparse local clusters are constructed to prevent the occurrence of overfitting. In this article, based on the classical DropConnect method, the random item ξ of the discarded neuron connection probability is added, and name it as the stochastic DropConnect method [26], whose output vector formula is described as $r = a(((\xi \cdot M) \cdot *W)v)$, where v denotes the feature vector of $n \times 1$, W denotes the weight matrix of $d \times n$, $(.*)$ represents the multiplication of the corresponding elements of the matrix, M denotes the binary mask matrix of $d \times 1$, nonlinear activation function $a(x)$ satisfies $a(0) = 0$, and probabilistic random variable $\xi \in (0.3, 0.7)$. For example, generated probability random binary mask matrix M, so the number of model combinations obtained is $(1/\xi)^{|M|}$, compared with the Dropout algorithm, the proposed algorithm possesses stronger ability of model combination. The stochastic DropConnect method based on the probability random term absorbs the essence of boosting idea and stochastic simulation idea, hidden nodes connection weights appear in a random probability, the weights updating strategy does not rely on a fixed relationship between the hidden nodes, its learning strategy has the characteristics of sparse constraint tester and can abate the correlation among feature tester, making the network model be more widely applied in all kinds of targets in the scene. Therefore, to simulate the decision of serialized

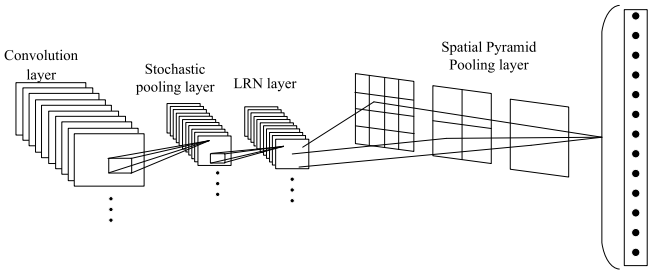


Fig. 4. Constructed high-level visual feature cognition model.

high-level visual features, this article tries to construct the full connection layer and Softmax regression model through the stochastic DropConnect method.

B. Proposed Human Tracking Framework

In sparse representation based classification (SRC) [44], Wright designed a robust classification framework and classified it by balancing reconstruction and redundancy error. In the designed module of IoMT, in order to obtain robust tracking results, this article proposes to modify the SRC model and uses it to design the tracking model.

In the constructed tracking model, the learned dictionary atom is denoted as $D = [D_1, D_2, \ldots, D_K]$, where D_i represents the subdictionary with class label i. Training samples are described as $a_{i,j}, i = 1, 2, \ldots, K, j = 1, 2, \ldots, N$, which represents MWLD descriptor, it can be further described as $A_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,N}] \in \mathbb{R}^{n\times N}, i = 1, 2, \ldots, K$ (n is the feature dimension, N is the sample number). Then, linear predictive classifier $f(a_{i,j}; W) = Wa_{i,j}$ is constructed, the learned dictionary D could be represented as $[d_1, d_2, \ldots, d_k] \in \mathbb{R}^{n\times k}$ ($k > n$ and $k \ll N$), so this article tries to construct the following tracking model:

$$\begin{aligned}\langle D, W, Z\rangle = \arg\min_{D,W,Z} \Bigg\{ &\Bigg(\sum_{i=1}^{K} \|A_i - DZ_i\|_F^2 + \lambda_1\|Z_i\|_1 \\ &+ \lambda_2\|DZ_i - m_i\|_F^2 + \gamma_1\|WZ_i - B\|_F^2\Bigg) \\ &+ \gamma_2\|W\|_F^2\Bigg\} \\ \text{s.t.} \quad &\|d_n\|_2 \le 1 \quad \forall n \end{aligned} \tag{5}$$

where Z_i represents the submatrix over D with respect to A_i, $m_i \in R^{k\times N}$ denotes the operation of sample averaging, $Z_i = [z_1, z_2, \ldots, z_N] \in \mathbb{R}^{k\times N}$. $\|WZ_i - B\|_F^2$ is the classification error term, W indicates the classification regulator and $W \in R^{m\times k}$. $B = [0, 0, \ldots, b_N] \in [0, 0\cdots 1\cdots 0, 0] \in R^{m\times N}$ indicates the corresponding label vector with respect to input signal y_i. λ_1, λ_1, γ_1, and γ_2 denote the scalars adjustment factors of the constrain terms. $\|\cdot\|_F$ is the form of Frobenius norm. Z_i is the sparse coefficients of A_i on dictionary D, therefore, $A_i \approx DZ_i$. It can be noted that Z_i is associated with class i, so Z_i can be well represented by m_i when m_i denotes the mean vector of Z_i of class i. Finally, a conclusion can be drawn that there should exist a case where $\|Z_i - m_i\|_F^2$ is small.

A conclusion can be drawn that (5) is improvement version of the basic class-specific dictionary learning model. Different from the basic SRC model in [44], in our constructed framework, the representation constraint $\phi = \lambda_2\|DZ_i - m_i\|_F^2 + \gamma_1\|WZ_i - B\|_F^2$ and coefficients constraint $\psi = \gamma_2\|W\|_F^2$ are used to enhance the discriminant ability of the model.

Next, the main task will focus on solving the objective function in (5), obviously, it is not convex because it has three variables (D, W, Z). In order to change it into convex optimization problem, this article splits the original objection function (5) into three subfunctions by means of optimizing D, W, and Z, respectively.

Process Of Solving Z: When variables D and W are treated as constant values, the optimization objection in (5) is degraded to $Z = [Z_1, Z_2, \ldots, Z_K]$ convex optimization problem. It is noted that all $Z_j (j \ne i)$ can be treated as constant values when Z_i is solved. Therefore, the convex optimization problem in (5) is rewritten as

$$\begin{aligned}\min_{Z} \Big\{&\|A_i - DZ_i\|_F^2 + \lambda_1\|Z_i\|_1 + \lambda_2\|DZ_i - m_i\|_F^2 \\ &+ \gamma_1\|WZ_i - B\|_F^2\Big\}. \end{aligned} \tag{6}$$

Specially for all Z_i, (6) is further degraded as

$$\begin{aligned}\langle Z_i\rangle = \arg\min_{Z_i} \Big\{&\|a_i - DZ_i\|_2^2 + \lambda_1\|Z_i\|_2^2 + \lambda_2\big\|DZ_i - m_iZ_i^i\big\|_2^2 \\ &+ \gamma_1\|WZ_i - b_i\|_2^2\Big\}. \end{aligned} \tag{7}$$

By computing the above objective function, the solution of the (7) is as follows:

$$\begin{aligned}Z_i = &\left\{D^TD + (\lambda_1 + \lambda_2)I + \gamma_1W^TW\right\}^{-1} \\ &\times \left(D^Ta^i + \lambda_2m_i + \gamma_1W^Tb_i\right). \end{aligned} \tag{8}$$

Finally, the undetermined parameters of (5) could be achieved by solving for all of these objection functions.

Based on the analysis in the above section, this article proposes to design the following model:

$$\hat{\alpha} = \arg\min_{\alpha}\Big\{\|y - D\alpha\|_F^2 + \gamma\|\alpha\|_2\Big\} \tag{9}$$

where γ could be determined in previous training process. $\hat{\alpha} = [\hat{\alpha}^1, \hat{\alpha}^2, \ldots, \hat{\alpha}^K]^T$, $\hat{\alpha}^i$ represents the subvector over D with respect to D_i. In the previous training process, the coefficients constraint term is designed to enhance the discriminant ability of the classifier through determining corresponding adjustment coefficients, therefore suppose sample y belongs to i class, this term $\|y - D_i\hat{\alpha}^i\|_2^2$ could be minimized, and if sample y does not belong to i class, $\|y - D_j\hat{\alpha}^j\|_2^2, j \ne i$ could not be minimized. There is another thing that the coefficient vector $\hat{\alpha}$ appear to differ in different classes. According to the discriminability of residual and coefficient vectors, this article defines the following metrics criterion for measuring the tracking position:

$$l = W\hat{\alpha}. \tag{10}$$

IV. Experiments and Results

A. Data Set

The experimental verification mainly focuses on four challenging data sets: 1) our constructed data set; 2) SEED

data set [51]; 3) eNTERFACE' 05 data set [52]; 4) DEAP datast [53].

Our Constructed Data Set: To measure the designed emotion-aware detection framework in the IoMT system, some audio and video data related to COVID-19 pandemic were collected and created. Fifteen university undergraduates were recruited for this research. The expressions of volunteers were made up of facial and spoken. For each volunteer, the training time is fixed to about three minutes. The smartphone recorded these scenes. Each volunteer is required to perform three expressions in principle: 1) happy state; 2) pain state; and 3) normal state. The total size of collected voice data was about 80 GB, and total size of video data reached to 260 GB.

SEED Data Set: The SEED data set was collected from electrophysiological motor imagery (MI) signals of of network media, the latest medical scanning device was adopted to record the signal of eye movement, the sampling rate is 1000 Hz, and the electrode cap have 62 channels. It consists of many emotions, including positive, negative and neutral types. SMI ETG eye-movement devices were also adopted to represent the character of eye movement. The signals recorded during the first nine movie clips were used as the training data set, and the rest were used as the test samples. This article adopted 32-b and 128-b code method in this article.

eNTERFACE' 05 Data Set: It includes 42 volunteers from 14 different nationalities (81 percent were male and 19 percent were female, 31% wore glasses and 17% had beards). All the experiments were conducted in English. Each volunteer was told to listen to six consecutive short stories, each of which elicited a specific emotion. They then responded to each situation, and two human experts judged whether the response expressed emotion in a specific way. If this is the case, the sample is added to the database.

DEAP Data Set: The DEAP data set consist of a collection of signals, which come from some emotion music videos of network media. The threshold is set to 5, and divided this data set into two categories according to the excitement and titer. It should be noted that the data was preprocessed. 1000 samples are then chosen from each class, there are two types of data to train our proposed model. For the other approaches, about 10 000 samples are used to train.

B. Experiment Settings

Polynomial reduction strategy was adopted to control and adjust parameters, based on eNTERFACE' 05 data set, SEED data set and DEAP database, when the reduced power value was set to 0.5, the learning rate was relatively stable, so the initial network learning rate was set to 0.005, the batch-size is fixed to 20 using polynomial decrease the strategy of vector control. The number of adopted iterations is 40 000 and 25 000, respectively. Results are recorded with acc $\pm$ std as well as ROC curve area (AUC). In addition, while for the sparse model in this article, the selection of superparameters is similar, this article defines $\lambda_1 = 0.005, \lambda_2 = 3, \gamma_1 = 1, \gamma_2 = 0.1$; while for the tracking parameter, the γ is set to 0.01.

TABLE I
COMPARISON OF VARIOUS APPROACHES ON OUR CONSTRUCTED DATA SET

Iteration number	HOG+HVC	SIFT+HVC	WLD+HVC	MWLD+HVC
5000	68.5%	69.9%	70.1%	72.1%
10000	73.9%	74.2%	74.5%	76.6%
15000	79.6%	80.1%	80.2%	81.1%
20000	81.4%	82.3%	82.6%	84.8%
25000	83.2%	84.2%	84.6%	86.2%
30000	85.4%	86.0%	85.9%	87.6%
35000	85.8%	86.5%	86.7%	87.8%

C. Results and Discussion

The proposed model is compared with many classical and recent algorithms: correlated attention network (CAN) [54], motion energy image-based principal component analysis network (MEI-PCANet) [17], appearance and motion deepnet (AMDN) model [55], GoogLeNet in [56] and VGG16 in [57] as well as the model in [44].

Results on Our Constructed Data Set: Table I summarizes the results of various approaches. The results reflect these results using our proposed hierarchical visual cognition (HVC) model paired with WLD, HOG, MWLD and SIFT. As is shown in this table, WLD-based HVG model performs a little better than the SIFT-based HVG model. The proposed MWLD-based HVG model outperforms the above two algorithms, indicating our constructed MoWLD descriptor is very effective for the deep learning framework. The proposed feature descriptor captures the appearance and motion information that are essential for classifying. Moreover, the WLD-based HVG model captures local appearance using an aggregated histogram of the oriented gradient in neighboring regions, so it is tolerant to partial occlusion and deformation. Furthermore, when an interest point is detected, a dominant orientation is calculated and all gradients in the neighborhood are rotated according to the dominant orientation. Therefore, the constructed feature model is rotation invariant. Also, with the increase in the number of iterations, detection rate begins to rise and then remain unchanged. So a conclusion could be drawn that selection of the iteration size is very important.

Table II gives the results of comparison after integrating the constructed classification algorithm into the hierarchical visual cognition (HVG) model. A conclusion could be drawn that the constructed deep learning scheme and classification model performs better than the popular approaches because of the high effectiveness of robust feature description and classification model. It can be concluded that if the level of classical CNN network model is not deep enough, the representation ability is insufficient, and if the level of the network model is too deep, the amount for demanding training data is too large. The AMDN approach is based on feature descriptor of optical flow, however, this feature is easily affected by the image quality and changes, so it performs badly. It can be seen that proposed model surpasses the other models under the same condition. This is because that the constructed detection framework incorporates appropriate deep network layers and effective learning scheme, also the tracking model is based on a distinguishable dictionary learning scheme, thus proposed model beats all the other algorithms.

TABLE II
RECOGNITION RESULTS ON THE OUR CONSTRUCTED DATA SET

Algorithm	acc±std	AUC
CAN [54]	84.77±0.73%	0.8593
MEI-PCANet [17]	85.07±0.59%	0.8623
AMDN [55]	82.27±0.79%	0.8394
GoogLeNet [56]	83.07±0.99%	0.8459
VGG16 [57]	78.8±0.75%	0.7992
SRC [44]	78.6±1.08%	0.7918
Ours	88.1±1.04%	0.8908

TABLE III
CRITERIA OF FEATURE SELECTION

Number	Emotion label	Film clips
1	negative	Wenchuan Earthquake
2	positive	Fly to Moon
3	negative	Lost in Forest
4	positive	Water Margin
5	negative	Atypical Pneumonia
6	neutral	Archeological Area

TABLE IV
RECOGNITION RESULTS ON THE SEED DATA SET

Algorithm	acc±std	AUC
CAN [54]	92.77±1.13%	0.9399
MEI-PCANet [17]	93.07±1.08%	0.9424
AMDN [55]	90.27±1.19%	0.9192
GoogLeNet [56]	92.07±1.09%	0.9359
VGG16 [57]	89.8±1.05%	0.9092
SRC [44]	88.6±1.02%	0.8918
Ours	95.1±1.04%	0.9608

Results on the SEED Data Set: Also, the proposed model is compared with many classical and recent algorithms implemented by this article. Table III gives the criteria of feature selection. The main rules are defined as follows: 1) the length of the description should be moderate; 2) the content of the video is easy to understand; and 3) a single emotion information permitted to be trigged. Each video is seriously treated to generate a coherent feeling.

In Table IV the experimental results with respect to recall rate are given, it can be seen that our designed model performs the best on the SEED data set, compared with other recent approaches. The reason is that this article has developed an effective deep network to describe the feature (see Section III), it is a fact that annotated data is difficult to obtain, this characteristic of the classic CNN network limits the application of the classic CNN network. However, our proposed model in this article shows better performance in the case of insufficient data size. MEI-PCANet approach maintain its priority and GoogLeNet also performs better than the VGG16 model. That is because both MEI-PCANet and GoogLeNet model consider the tiny characteristics of the target, and the multilayer network is used to describe the target, which increases the detection accuracy. Comparing the original CAN model [54], the proposed HVG-based sparse classification scheme has improved the recognition rate effectively, which indicates the efficiency of our proposed hierarchical visual cognition model. Original SRC model achieves the lowest recognition rate.

TABLE V
RECOGNITION RESULTS ON THE ENTERFACE' 05 DATA SET

Algorithm	acc±std	AUC
CAN [54]	86.77±0.77%	0.8792
MEI-PCANet [17]	87.07±0.61%	0.8851
AMDN [55]	83.27±0.79%	0.8498
GoogLeNet [56]	85.07±0.97%	0.8659
VGG16 [57]	79.8±0.75%	0.8092
SRC [44]	78.7±1.08%	0.7969
Ours	89.61±0.17%	0.9178

Also, a curious phenomenon could be found that the CAN model achieves a little better superiority than using GoogLeNet approach. Through the investigation of a large number of experimental data, it can be found that the CAN model introduced more local feature information of target. By observing the performances of comparative approaches on this data set, it is very easy to conclude that the HVG-based sparse classification scheme is superior to other models.

Results on eNTERFACE' 05 Data Set: In this experiment, the dictionary number is set to 21 000 in this data set. Table V gives the results of comparison after integrating the constructed classification algorithm into the robust hierarchical visual cognition (HVG) model. The performance of the VGG16 model is very close to SRC model and they perform not well, that is mainly because the two models are easy to produce overfitting. The AMDN uses deep neural networks to describe the activity feature, however, the network model is not deep enough, the representation ability is insufficient. Also, it can be found that the constructed model outperforms MEI-PCANet, CAN and GoogLeNet models. The proposed hierarchal computational cognition simulation model is based on biological theory and abdominal pathway model, and take the form of framework of CNN that simulates the layered information processing mechanism of human brain, it combined with the recent research progress made in the field of biological vision perfectly, therefore, the model can effectively improve the accuracy of detection and recognition.

Results on DEAP Data Set: The results of the comparison are depicted in Table VI. The performance is evaluated using average accuracy (acc) and standard deviation (std) criterion. This data set gives the detailed description of training samples, which can be found in [58]. Participants use the Sam model of the human body in a discrete nine point scale based on the evaluation of arouse, titer, and dominate degrees. Participants also used an emotional roulette wheel to assess their feelings. As is shown in Table VI, it seems that CAN, MEI-PCANet, and GoogLeNet models play the same performance, it is worth mentioning that all the approaches use the same bits code length.

VGG16 and SRC models have unsatisfactory performance, which is the same acieved on the SEED data set. The performance of the AMDN model is moderate, one reason is that the network model is too deep, and the demand for training data is too great. It can also be seen that our proposed model surpasses the other models under the same condition. This is because that the constructed classification model make full use of the effective deep learning model and distinguishable dictionary learning idea, the proposed model can embody

TABLE VI
RECOGNITION RESULTS ON THE DEAP DATA SET

Algorithm	acc±std	AUC
CAN [54]	87.27±0.79%	0.8862
MEI-PCANet [17]	87.37±0.68%	0.8891
AMDN [55]	84.27±0.81%	0.8505
GoogLeNet [56]	87.07±0.99%	0.8827
VGG16 [57]	80.8±0.79%	0.8141
SRC [44]	78.6±1.07%	0.7947
Ours	89.41±0.17%	0.9151

the highly complex hierarchical characteristics of biological information, which makes the signal and visual recognition inherit strong robustness and fast response time, so the results are very encouraging.

The bandwidth consumption (kb/s) of the designed framework in this article stays around 169 kb/s, which could be considered to be a low consumption in an IoMT system. In addition, to give more statistical analysis, this article performs $t-$ test for the precision obtained by CAN, MEI-PCANet, AMDN model, GoogLeNet, VGG16, SRC, and the model in this article on four data sets under the null hypothesis using a significance level of 0.05. The $P-$value is obtained as 0.00037, 0.00182, 0.00262, and 0.00481, respectively, further demonstrating that the performance of our proposed model is indeed better than other proposed models.

V. CONCLUSION

The IoT establish an emerging paradigm of networked physical systems in which billions of interconnected intelligent objects collect, analyze, and exchange vast amounts of information around the world, and it has played an important role in smart lives, especially during this COVID-19 pandemic. Confronting these challenges, this article proposes an emotion-aware and intelligent IoMT system, which contains discriminative emotion recognition and human detection modules. Intelligent IoMT devices are connected to collect multimodal data of patients in a surveillance environments. The latest data and inputs from official websites and reports are tested for further investigation and analysis of the emotion analysis. The proposed novel IoMT platform enables remote health monitoring and decision making about the emotion, therefore, greatly contribute convenient and continuous emotion-aware healthcare services during the COVID-19 pandemic. The proposed cognition-based dynamic technology is an effective solution way for accommodating a big number of devices and this COVID-19 pandemic application. An indoor tracking module is also exploited.

In addition, in order to achieve the algorithm breakthrough, this article realized the use and evaluation of cognitive models for social media analytics to leverage deeper insights from the vast amount of generated data for IoMT data, this article proposed a robust hierarchical deep cognition model with the discriminative dictionary learning scheme. Moreover, the spatial pyramid random pool layer was used to obtain the effective features of the target so as to avoid the feature loss caused by image compression and other operations. Finally, stochastic DropConnect neural network calculation was proposed to solve common fitting problem. The proposed hierarchical visual cognition (HVC) model have integrated the improved sparse-based classification model very well. The experiment showed that the proposed model presented better accuracy and stability under the same recognition criteria.

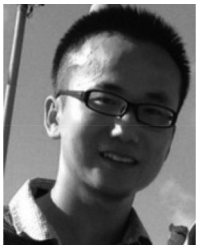
This research can be conducted further in the following aspects.

1) Considering the person being monitored is often not facing the camera, a robust one would require adding more smart devices to capture positive faces when not adding the cost, or developing more robust facial recognition algorithms.
2) Indoor locating technology needs to be further improved, and how to realize real-time tracking should also be considered under the existing hardware conditions.
3) The direct and indirect correlation of the multimodal data needs further study and improvement.
4) How to take advantage of edge computation to save computation and storage load is also a work that needs further study.

REFERENCES

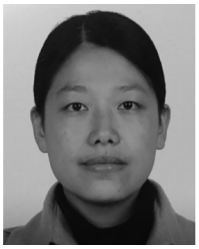
[1] *World Population*. Accessed: Jun. 29, 2020. [Online]. Available: www.worldometers.info
[2] A. Haleem and M. Javaid, "Effects of COVID-19 pandemic in daily life," *Current Med. Res. Pract.*, vol. 10, no. 2, pp. 78–79, 2020.
[3] H. Bai *et al.*, "Performance of radiologists in differentiating COVID-19 from viral pneumonia on chest CT," *Radiology*, vol. 296, Mar. 2020, Art. no. 200823.
[4] X.-B. Pan, "Application of personal-oriented digital technology in preventing transmission of COVID-19, China," *Irish J. Med. Sci.*, vol. 189, pp. 1145–1146, Nov. 2020.
[5] V. Raju, J. Mohd, H. K. Ibrahim, and H. Abid, "Artificial intelligence (AI) applications for COVID-19 pandemic," *Diabetes Metab. Syndrome Clin. Res. Rev.*, vol. 14, pp. 32–43, Jul./Aug. 2020.
[6] J. Mohd, H. Abid, V. Raju, B. Shashi, S. Rajiv, and V. Abhishek, "Industry 4.0 technologies and their applications in fighting COVID-19 pandemic," *Diabetes Metab. Syndrome Clin. Res. Rev.*, vol. 14, pp. 419–422, Jul./Aug. 2020.
[7] E. Waleed and I. Mohamed, "Multiband spectrum sensing and resource allocation for IoT in cognitive 5G networks," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 150–163, Feb. 2018.
[8] S. Emmanouil, M. Angelos, and T. Nikolaos, "A VHO scheme for supporting healthcare services in 5G vehicular cloud computing systems," in *Proc. Wireless Telecommun. Symp. (WTS)*, Phoenix, AZ, USA, 2018, pp. 191–197.
[9] P. S. Ravi, J. Mohd, H. Abid, and S. Rajiv, "Internet of Things (IoT) applications to fight against COVID-19 pandemic," *Diabetes Metab. Syn. Clin. Res. Rev.*, vol. 14, pp. 521–524, Jul./Aug. 2020.
[10] J. X. Li, H. T. Zhao, A. S. Hafid, J. Wei, H. Yin, and B. Ren, "A bio-inspired solution to cluster-based distributed spectrum allocation in high-density cognitive Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9294–9307, Dec. 2019.
[11] L. M. Gladence, H. H. Sivakumar, G. Venkatesan, and S. S. Priya, "Home and office automation system using human activity recognition," in *Proc. Int. Conf. Commun. Signal Process. (ICCSP)*, Chennai, India, 2017, pp. 758–762.
[12] K. L. Terence, S. R. Simon, and D. S. Daniel, "Major requirements for building smart homes in smart cities based on Internet of Things technologies," *Future Gener. Comput. Syst*, vol. 76, pp. 358–369, Nov. 2017.
[13] M. S. Hossain, "Cloud-supported cyber–physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. 11, no. 1, pp. 118–127, Mar. 2017.
[14] M. S. Hossain and G. Muhammad, "Cloud-assisted Industrial Internet of Things (IIoT)—Enabled framework for health monitoring," *Comput. Netw.*, vol. 101, pp. 192–202, Jun. 2016.

[15] Y. B. Miao, Q. Y. Tong, K. K. Raymond Choo, X. M. Liu, R. H. Deng, and H. Li, "Secure online/offline data sharing framework for cloud-assisted industrial Internet of Things," *IEEE Internet Things J.*, vol. 6, no. 5, pp. 8681–8691, Oct. 2019.
[16] L. M. Gladence, V. M. Anu, R. Rathna, and E. Brumancia, "Recommender system for home automation using IoT and artificial intelligence," *J. Ambient Intell. Hum. Comput.*, vol. 72, no. 4, pp. 1–9, Apr. 2020.
[17] A. Amany and A. Saleh, "Human action recognition using short-time motion energy template images and PCANet features," *Neural Comput. Appl.*, vol. 3, no. 10, pp. 983–1009, 2020.
[18] E. L. Jackson, S. G. Spielman, and C. O. Wilke, "Computational prediction of the tolerance to amino-acid deletion in green-fluorescent protein," *PLoS ONE*, vol. 12, no. 4, pp. 164–175, 2017.
[19] E. K. Samira *et al.*, "EmoNets: Multimodal deep learning approaches for emotion recognition in video," *J. Multimodal User Interfaces*, vol. 10, pp. 99–111, Jun. 2016.
[20] J. Robert, P. Angelika, and B. Martin, "Feature extraction and selection for emotion recognition from EEG," *IEEE Trans. Affective Comput.*, vol. 5, no. 3, pp. 327–339, Jul.–Sep. 2014.
[21] J. Robert, P. Angelika, and B. Martin, "Patient state recognition system for healthcare using speech and facial expressions," *J. Med. Syst.*, vol. 40, pp. 212–221, Oct. 2016.
[22] W. Guo and G. Chen, "Human action recognition via multi-task learning base on spatial–temporal feature," *Inf. Sci.*, vol. 320, pp. 418–428, Nov. 2015.
[23] H. Cheng, Z. Su, and N. Xiong, "Energy-efficient node scheduling algorithms for wireless sensor networks using markov random field model," *Inf. Sci.*, vol. 329, pp. 461–477, Feb. 2016.
[24] J. Wang, X.-M. Zhang, Y. Lin, X. Ge, and Q.-L. Han, "Event-triggered dissipative control for networked stochastic systems under non-uniform sampling," *Inf. Sci.*, vol. 447, pp. 216–228, Jun. 2018.
[25] X. Sun and T. Pan, "Static facial expression recognition system using ROI deep neural networks," *Acta Electronica Sinica*, vol. 41, no. 5, pp. 1189–1197, 2017.
[26] A. Iosifidis, A. Tefas, and I. Pitas, "DropELM: Fast neural network regularization with dropout and dropconnect," *Neurocomputing*, vol. 162, pp. 57–66, Aug. 2015.
[27] Z. H. Song *et al.*, "A sparsity-based stochastic pooling mechanism for deep convolutional neural networks," *Neural Netw.*, vol. 105, pp. 340–345, Sep. 2018.
[28] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1904–1916, Sep. 2015.
[29] M. S. Hossain, G. Muhammad, F. A. Mohammed, B. Song, and A. M. Khaled, "Audio-visual emotion recognition using big data towards 5G," *Mobile Netw. Appl.*, vol. 10, pp. 753–762, Jan. 2016.
[30] T. Leonardo, D. Kelly, F. Chloe, J. Sojo, K. Veronica, and F. Thomas, "Functional magnetic resonance imaging correlates of emotion recognition and voluntary attentional regulation in depression: A generalized psycho-physiological interaction study," *J. Affect. Disorders*, vol. 188, pp. 535–544, Jan. 2017.
[31] B. Lisa *et al.*, "Mapping structural covariance networks of facial emotion recognition in early psychosis: A pilot study," *Schizophrenia Res.*, vol. 189, pp. 146–152, Nov. 2017.
[32] M. S. Hossain and G. Muhammad, "An emotion recognition system for mobile applications," *IEEE Access*, vol. 189, pp. 2281–2287, 2017.
[33] M. Zhao, F. Adib, and D. Katabi, "Emotion recognition using wireless signals," in *Proc. 22nd Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2016, pp. 95–108.
[34] S. Wang and W. Guo, "Robust co-clustering via dual local learning and high-order matrix factorization," *Knowl. Based Syst.*, vol. 138, pp. 176–187, Dec. 2017.
[35] F. Luo, W. Guo, Y. Yu, and G. Chen, "A multi-label classification algorithm based on kernel extreme learning machine," *Neurocomputing*, vol. 25, pp. 313–320, Oct. 2017.
[36] S. Jaswini, R. Shyam, A. Vijayendra, and K. M. R. Kumar, "EEG based emotion recognition using wavelets and neural networks classifier," in *Cognitive Science and Artificial Intelligence*. Singapore: Springer, 2018, pp. 122–312.
[37] M. S. Hossain and G. Muhammad, "Audio-visual emotion recognition using multi-directional regression and ridgelet transform," *J. Multimodal User Interfaces*, vol. 10, pp. 325–333, Nov. 2015.
[38] T. Ma, Q. Liu, J. Cao, Y. Tian, A. Al-Dhelaan, and M. Al-Rodhaan, "LGIEM: Global and local node influence based community detection," *Future Gener. Comput. Syst.*, vol. 6, no. 12, pp. 533–546, 2020.
[39] Q. Ye, Z. Li, L. Fu, Z. Zhang, W. Yang, and G. Yang, "Nonpeaked discriminant analysis for data representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 12, pp. 3818–3832, Dec. 2019.
[40] S. Zhang, Y. Xia, and J. Wang, "A complex-valued projection neural network for constrained optimization of real functions in complex variables," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 12, pp. 3227–3238, Dec. 2016.
[41] S. Zhong, T. Chen, and F. He, "Fast Gaussian kernel learning for classification tasks based on specially structured global optimization," *Neural Netw.*, vol. 57, no. 2, pp. 51–62, 2015.
[42] Y. Xia and H. Leung, "Performance analysis of statistical optimal data fusion algorithms," *Inf. Sci.*, vol. 277, pp. 808–824, Sep. 2014.
[43] Y. Yu and Z. Sun, "Sparse coding extreme learning machine for classification," *Neurocomputing*, vol. 261, pp. 50–56, Oct. 2017.
[44] A. Y. Wright, J. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[45] J. Mairal, F. Bach, and J. Ponce, "Task-driven dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 4, pp. 791–804, Apr. 2012.
[46] T. Zhang, W. J. Jia, J. Yang, and X. J. He, "Discriminative dictionary learning with motion weber local descriptor for violence detection," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 3, pp. 696–709, Mar. 2017.
[47] J. Chen *et al.*, "WLD: A robust local image descriptor," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 9, pp. 1705–1720, Sep. 2010.
[48] S. T. Li, D. Y. Gong, and Y. Yuan, "Face recognition using weber local descriptors," *Neurocomputing*, vol. 122, pp. 272–283, Dec. 2013.
[49] Y. Niu, W. Lin, X. Ke, and L. Ke, "Fitting-based optimisation for image visual salient object detection," *IET Comput. Vis.*, vol. 11, no. 2, pp. 161–172, Mar. 2017.
[50] Y. Q. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, 2014, pp. 675–678.
[51] Y. F. Lu, W. L. Zheng, B. B. Li, and B.-L. Lu, "Combining eye movements and EEG to enhance emotion recognition," in *Proc. 24th Int. Conf. Artif. Intell. (IJCAI)*, 2015, pp. 1170–1176.
[52] A. Hosseini *et al.*, "Children activity recognition: Challenges and strategies," in *Proc. IEEE 40th Annu. Int. Conf. Eng. Med. Biol. Soc. (EMBC)*, 2018, pp. 8–14.
[53] Z. Yin, M. Y. Zhao, Y. X. Wang, J. Yang, and J. Zhang, "Recognition of emotions using multimodal physiological signals and an ensemble deep learning model," *Computer Methods and Programs in Biomedicine*, vol. 142, pp. 93–110, Mar. 2017.
[54] J. L. Qiu, X. Y. Li, and K. Hu, "Correlated attention networks for multimodal emotion recognition," in *Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM)*, 2018, pp. 2656–2660.
[55] X. Dan, R. Elisa, Y. Yan, J. K. Song, and S. Nicu, "Learning deep representations of appearance and motion for anomalous event detection," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, 2015, pp. 1–12.
[56] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 2371–2379.
[57] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 2151–2159.
[58] S. Mohammad, P. Maja, and P. Thierry, "Multimodal emotion recognition in response to videos," in *Proc. Int. Conf. Affect. Comput. Intell. Interact.*, Xi'an, China, 2015, pp. 491–497.



Tao Zhang (Member, IEEE) received the bachelor's degree from Henan Polytechnic University, Jiaozuo, China, in 2008, the Ph.D. degree from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China, in 2016.

He was a Visiting Scholar with the University of Technology Sydney, Ultimo, NSW, Australia, in 2015. He is currently an Associate Professor with the Jiangsu Provincial Engineering Laboratory for Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, China. He has led many research projects (e.g., the National Science Foundation and the National Joint Fund). He has authored over 30 quality journal articles and conference papers. His current research interests include medical image processing, medical data analysis, visual surveillance, scene understanding, behavior analysis, object detection, and pattern analysis.



Minjie Liu received the bachelor's degree from Changjiang University, Jingzhou, China, in 2006, and the M.S. degree in physiology from Liaoning Medical University, Jinzhou, China, in 2013.

She is currently a Lecturer with the School of Nursing, Taihu University of Wuxi, Wuxi, China. Her research interests focus on understanding the mechanisms of electromagnetic activities in biological tissue and systems, computational modeling and analysis of organ systems to aid clinical diagnosis of dysfunction in the human body.

Tian Yuan received the master's and Ph.D. degrees from Kyung Hee University, Seoul, South Korea, in 2010 and 2012, respectively.

After that, she joined King Saud University, Riyadh, Saudi Arabia, as an Assistant Professor with the College of Computer and Information Sciences. She is currently an Associate Professor with Nanjing Institute of Technology, Nanjing, China. She has participated more than ten national and industrial projects in Korea and Saudi Arabia, such as National IT industry and National Research Foundation. Besides, she also works as PI and Co-PI in several projects, including National Plan for Science, Technology and Innovation. Her main research interest is information protection and privacy preservation in the area of IoT, location-based service, social networks, cloud computing and healthcare domain.

Dr. Yuan is currently on the editorial boards of several journals, and has been the workshop/session chairs, organization and program committee for several reputable international conferences.

Najla Al-Nabhan (Member, IEEE) received the B.S. degree (Hons.) in computer applications and the master's and Ph.D. degrees in computer science from King Saud University, Riyadh, Saudi Arabia, in 2004, 2008, and 2014, respectively.

She is currently a Professor Assistant and the Head of Academic Unit with the College of Applied Studies and Community Services, King Saud University. She holds a U.S. patent (U.S. 8 606 903) in 2014. Her research interest includes wireless networks (in particular, sensor networks), mobile computing, distributed systems, ubiquitous computing, cloud computing, smart grid, and network security.