APPLICATION OF AR VIRTUAL IMPLANTATION TECHNOLOGY BASED ON DEEP LEARNING AND EMOTIONAL TECHNOLOGY IN THE CREATION OF INTERACTIVE PICTURE BOOKS

Sidan Liu*

College of Education, China West Normal University, Nanchong, Sichuan, 637000, China

liusidan11@cwnu.edu.cn

Peng Peng

College of Education, China West Normal University, Nanchong, Sichuan, 637000, China

Lei Cao

College of Education, China West Normal University, Nanchong, Sichuan, 637000, China

Reception: 11/11/2022 **Acceptance**: 16/01/2023 **Publication**: 11/03/2023

Suggested citation:

L., Sidan, P., Peng and C., Lei. (2023). **Application of AR virtual implantation technology based on deep learning and emotional technology in the creation of interactive picture books**. *3C TIC. Cuadernos de desarrollo aplicados a las TIC, 12(1)*, 176-198. https://doi.org/10.17993/3ctic.2023.121.176-198

ABSTRACT

In recent years, the field of deep learning has flourished, not only breaking through many difficult problems that are difficult to be solved by traditional algorithms but also bursting with greater vitality when combined with other fields. For example, product emotional design based on deep learning can integrate users' emotional needs into the actual product design. In this paper, we aim to use deep learning and affective technology in the creation of AR interactive picture books to transform the reading process from static to dynamic, enrich visual stimulation, and increase the fun and interactivity of reading. In this paper, based on the three-level theoretical model of emotion, the emotion labeling results are input to a deep neural network for learning, to establish an emotion-based recognition model for picture book images. The results show that the model can well analyze the emotion of images in AR picture books, and the accuracy of prediction is a big improvement compared with traditional machine recognition algorithms. The application of AR virtual implantation technology in interactive picture books on the market is often just a marketing gimmick while combining deep learning and emotional technology can better create diverse interactive picture books to meet children's emotional reading needs, enhance reading engagement, and stimulate children's creativity.

KEYWORDS

Deep learning; affective technology; AR implantation technology; interactive picture book; three-level theory.

https://doi.org/10.17993/3ctic.2023.121.176-198

PAPER INDEX

https://doi.org/10.17993/3ctic.2023.121.176-198

1. INTRODUCTION

In recent years, the continuous development of artificial intelligence algorithms such as deep learning and the emergence of ever-changing new media have contributed to the growth of various emerging technologies that have gradually changed the way information is relayed [1-3]. At the same time, digital technology, as one of the emerging technologies, has also been developing rapidly, and this technology is especially reflected in the field of picture books, creating a new wave that has a profound impact on the way children read. Augmented reality (AR) is a common and popular over-the-top concept, the principle of which is to use computer technology to transfer virtual information into the real environment to achieve the combination of virtual and reality [4-7], in three-dimensional form, the application of AR technology to interactive picture books can improve children's extraction and recognition of information [8-9], allowing children to feel the overlap of virtual and real scenes to enhance the virtual emotional experience and provide a deeper understanding of knowledge. The application of deep learning and emotional technology will strengthen the emotional analysis in the creation of AR picture books and enrich the emotional education function of interactive picture books.

Research on AR books first started with The Magic Book by Billinghurst et al [10-12]. It is essentially a mixed reality application. In this application, by using a handheld display equipped with a small camera, the experiencer can then experience a realistic virtual world through a paper book. Since then many scholars have studied AR books. Professor Hiromichi Kato and Mark Billinghurst jointly developed the first open-source framework for AR, AR Tool Kit [13-15], through which applications for AR can be easily written to superimpose virtual scenes onto real environments. The literature [16-17] introduces natural feature tracking techniques on AR books. The literature [18-19] focuses on the user interface design and interaction design of AR books. The literature [20-21] focuses on the design and research of interactive 3D books based on AR technology, introducing key technologies and proposing the production process. In the field of AR picture book publishing, Leo Paper Group publishes and designs the augmented reality interactive three-dimensional book *The Search for Wondla* [22], and the German company ArsEdition publishes the augmented reality book *Aliens and UFOs* [23-25]. By installing a special player, it is possible to see the three-dimensional scenes in this children's science fiction story on any camera-equipped computer through the Internet, breaking through the bottleneck of local reading.

In the era of new media, the application of AR interactive picture books is emerging. This paper will analyze the application of augmented reality technology based on deep learning and emotional technology in the creation of interactive picture books with the help of quantitative models based on the concepts of emotional design, deep learning, AR virtual implantation technology and interactive picture books, and explore how deep learning and emotional technology can better serve the application of AR virtual implantation technology in the creation of picture books. This paper establishes a quantitative emotion model based on psychologist Robert Pultchlt's emotion wheel and Donald Arthur Norman's three-level emotional design theory. Then we compare and analyze the algorithmic ideas of several current mainstream deep learning network models, and determine that VGGNet is more suitable for emotion discrimination and prediction of AR picture books. Finally, we establish an AR interactive picture book image sentiment analysis platform through deep learning training in an image sample library.

2. EMOTIONAL AND DEEP LEARNING

This paper uses cutting-edge deep learning techniques to learn product image emotions to promote the development and application of affective design, and affective design and deep learning are the two fundamental cores of this paper. This chapter introduces the concepts and development of these two topics and the lack of research in each of them, to discover the research points of this paper before the specific research is carried out. Among these two topics, affective design is the basic theory of this paper, while deep learning technology is the technical tool of this paper.

2.1. EMOTIONAL DESIGN

People's need for emotion can be found everywhere in life at any time, for example, compared to electronic books, people feel that the paper version is not only more comfortable but there is also a kind of emotional experience that is difficult to imitate with electronic books, perhaps a gentle and solid touch, perhaps a distant memory is touched, perhaps the familiar fragrance of paper brings a sense of freshness. For example, the phone can give me a timely notification of charging when it is out of battery, and give me a warm reminder when the weather is bad, bringing a sense of warmth and surprise. In modern society, machines are becoming more and more common, but people have an innate fear of machines; we feel they are complicated, cold, and even dangerous. Human-to-human communication is the most natural and intimate, so I hope the machine can be more humanized, and realize the humanization of the machine is the main purpose of human-computer interaction research, and the key to humanization is to make the machine have emotion.

Emotional design is the process of making design objects with emotional factors, taking into account the user's physiology and psychology throughout their life cycle, and catering to their innermost emotional needs to induce emotional responses.

2.1.1. SOURCES AND CLASSIFICATION OF EMOTIONS

Emotions were not separated independently long ago and still belonged to the realm of philosophy. It was only in the 19th century that the German psychologist Von Wundt separated the study of human emotions from philosophy and began scientific research. With the development of time and the maturity of theories, the study of human emotion was refined into three major branches, behavioral theory, theory of mind, and cognitive theory.

https://doi.org/10.17993/3ctic.2023.121.176-198

There are various classifications of human emotions, and the earliest days of psychology used a dichotomy to classify emotions into positive and negative emotions. In 1980 psychologist Robert Pultchlt proposed the emotion wheel to describe the association between emotions [26-27]. Robert Pultchlt's psycho-evolutionary theory of emotions is one of the most influential taxonomies of common emotional responses. He believed that there are eight most basic emotional elements as shown in Figure 1, which are anger, fear, sadness, dislike, surprise, curiosity, acceptance, and cheerfulness.

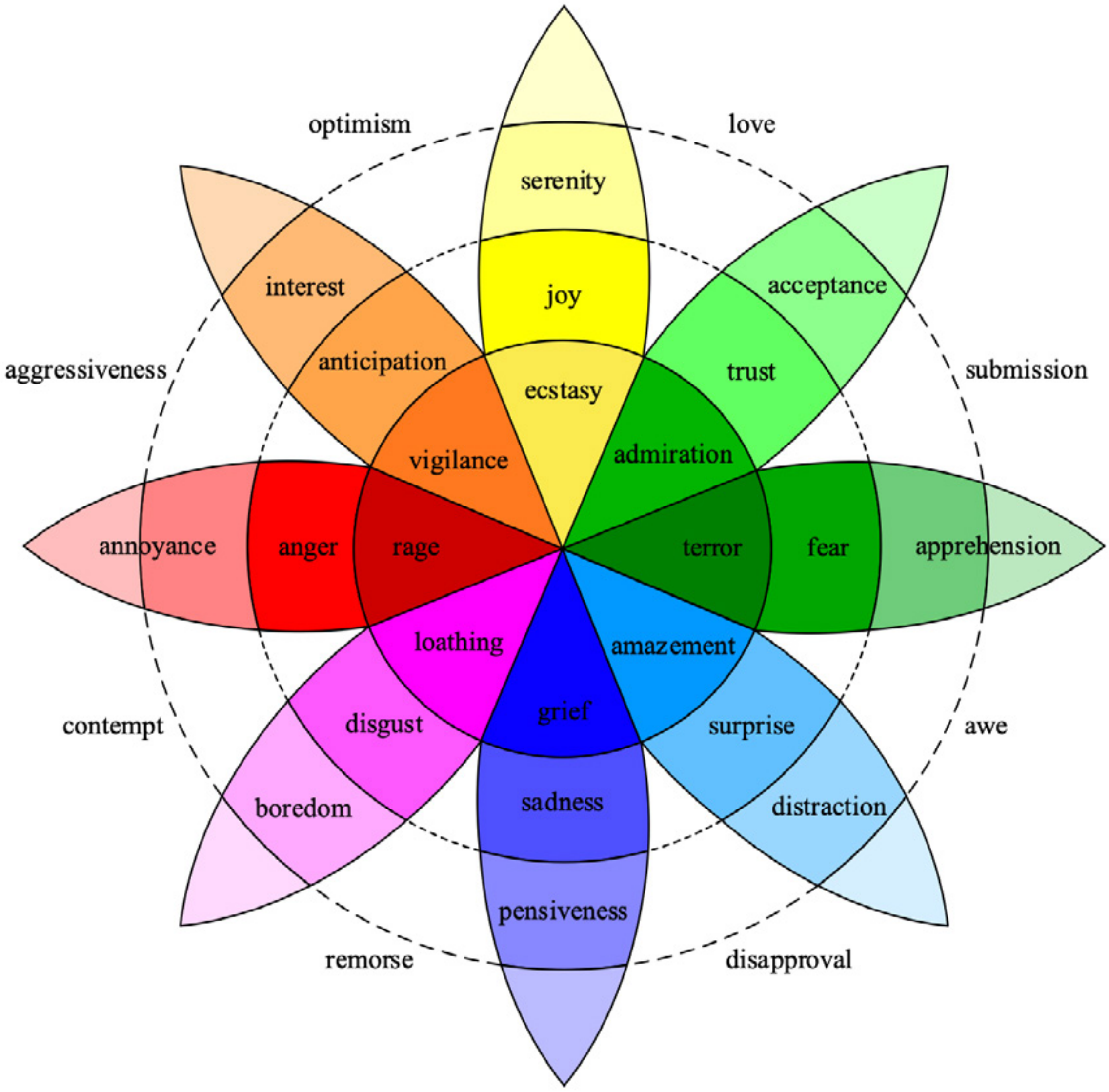


Figure 1. Robert Pultchlt's Emotional Roulette

2.1.2. EMOTIONALITY IN PRODUCT DESIGN

Emotional design is proposed by Donald Arthur Norman, an American cognitive psychologist, and its main idea is to attract users' attention and induce their emotional response through design, to realize the emotional communication and connection between humans and design objects and bring a better experience and deeper impression to users.

https://doi.org/10.17993/3ctic.2023.121.176-198

The purpose of emotional design is to bring positive emotions or complex emotions dominated by positive emotions. Talking about emotion from the perspective of design can even be extended to the category of style. For example, simplicity is style, which can also be said to be an emotion, because it reflects not only the properties of the design itself but also the psychological feelings of the viewer.

Emotions have objective and subjective aspects. The objective aspect refers to the emotional factors embedded in objective things or phenomena, while the subjective aspect refers to the subjective feelings and understanding received by the viewer or experiencer. For design works, the objective aspect is the creator's use of design principles (such as composition principles), combined with psychology, to integrate colors, shapes, and materials to form a complete system to trigger a specific emotional response in the experience, which is the process of injecting emotions into the creator. The subjective aspect of emotion is influenced by the individual differences of the experience, such as the aesthetics, experience, and preferences of different experiences, so each person may have very different emotions towards the same work. It is also influenced by the different states of the same experience, such as the mood at the time. Even so, due to human commonalities, there is a strong convergence in the perception of a particular phenomenon, as evidenced by the research of Dacher Keltner's team.

2.1.3. THREE-LEVEL THEORETICAL MODEL

In Emotional Design, Professor Donald A. Norman divides the emotional design into three levels: instinctive, behavioral, and reflective [28-29]. The instinctive layer of design is concerned with the human sensory experience of the material characteristics of a specific product, which are usually visible or palpable, such as the structure, color, material, and form of the product. The instinctive layer is the most basic part and the first to attract people's attention. The behavioral layer of design is concerned with the interaction process between the product and the user, focusing on the efficiency and enjoyment of the operation, including the functionality, performance and usability of the product. The behavioral layer is about the design of product use so that the function of the product can conform to human behavior to the maximum extent, in this level of design, the shape of the product and design principles are not the most important, the most important is the performance of the product. The reflection layer is essentially also due to the role of the first two levels, but in the three levels is the most difficult to achieve, need to start from the user's common culture and information, resonate with the user, but the impact on the user is far-reaching, which is also the reason why a large number of the current market to give the product a certain quality and story, to create a different and evocative feeling to the user. However, in the actual design process, the three levels of emotional design are not completely independent, but intertwined and difficult to distinguish, as shown in Figure 2 a diagram of the relationship between the three levels, only for different levels of design focus will be different. Instinctive design is the human-computer dialogue of intuitive feeling, behavioral design is the human-computer dialogue in the process of

https://doi.org/10.17993/3ctic.2023.121.176-198

use, and reflective design is a dialogue of deep consciousness activity shown between humans and products.

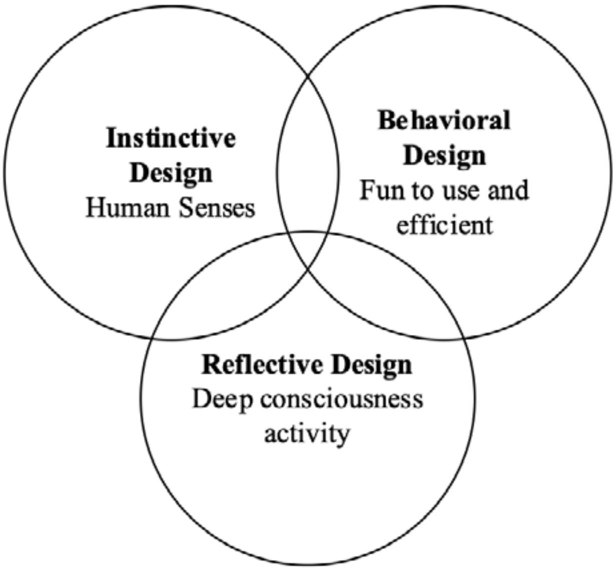


Figure 2. Emotional design three levels of relationship diagram

2.2. DEEP LEARNING

Emotions can be learned. There has been a lot of research using machine learning techniques to perform sentiment analysis for various types of image content. Machine learning is the process of training a model with data, allowing the machine to find patterns and use the knowledge learned to analyze and judge new data. The core part of this process is to design algorithms that allow the computer to learn automatically, so that the whole learning process "comes to life", and to continuously optimize the feature weights through backpropagation to improve the learning effect. With the rise of deep learning, more and more work is being done using deep learning features instead of traditional hand-designed features for analysis. Among them, deep learning is a data-driven feature extraction process. Under the premise of ensuring data quality, the more training samples there are, the deeper features can be obtained and the more ideal the training effect is. Compared with traditional algorithms, the expressiveness of deep learning is significantly more efficient and accurate, and the obtained abstract features are much better in terms of robustness and generalization, and the whole training process is end-to-end, so there is no need for human intervention in the middle.

One of the most important technologies for deep learning is neural networks. A neural network, which mimics the human brain, relies on neurons to transmit and process information, and the entire network is a system that includes a large number of neurons. A neuron is a simple classifier that is used to identify object features.

2.2.1. CONVOLUTIONAL NEURAL NETWORK

https://doi.org/10.17993/3ctic.2023.121.176-198

Convolutional neural network (CNN) [30-31] is arguably the most typical structure of deep learning networks. 1998 LeNet-5, proposed by Lecun et al. is the first real multi-layer network, and its convolutional network structure is shown in Figure 3. It uses convolutional structure to obtain one-dimensional or spatial data, for example, temporal data is a one-dimensional form of data, while pictures are a two-dimensional form of data, and convolutional neural network is especially good at processing picture-like two-dimensional structure data. The process of convolutional data extraction is a special mathematical operation, and convolutional kernels have been used in research related to image content processing for a long time.

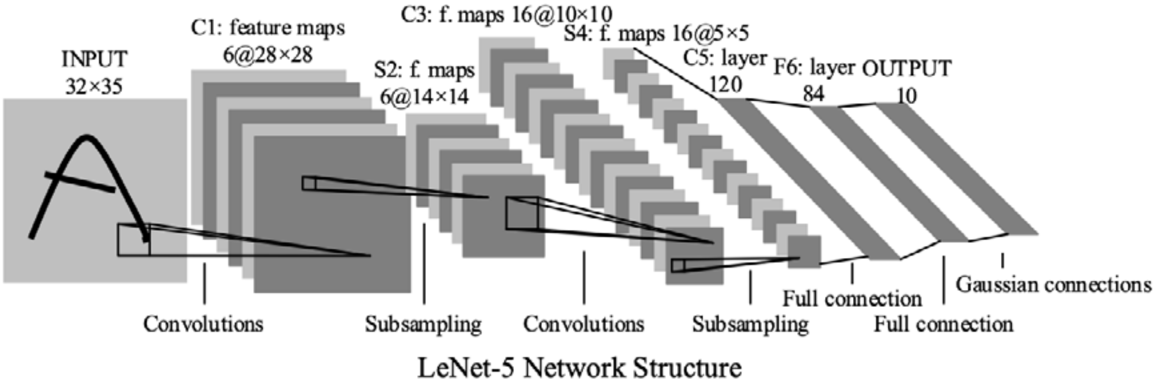


Figure 3. Structure of LeNet-5

The network starts with an input layer, which is typically an image. Next are successive alternating convolutional and pooling layers. The convolutional layer consists of multiple feature maps, i.e., feature maps in Figure 2, and each feature map consists of multiple neurons, which are connected to some local block of the previous layer's feature map by the weights of the convolutional kernel. The convolution kernel is a matrix of weights, for example, it can be 3×3 or 5×5. The extraction of features is performed by the convolution operation of the convolution kernel with the feature maps of the previous layer, and as the number of convolution layers deepens, the feature maps obtained become more abstract. In the same feature map, the weights of the filters are the same, called weight sharing, which has the advantage of reducing the complexity of the model on the one hand and capturing the local features of the input on the other. The locally weighted sum of each convolution kernel after convolution with the previous layer is passed to a nonlinear function (called the activation function) to obtain the output of the convolutional layer. The activation function is an indispensable feature of convolutional neural networks to enhance the expressiveness of the network. The commonly used activation functions are the sigmoid function, tanh function, ReLU function, etc. The calculation formula is as follows.

$$sigmoid(x) = \frac{1}{1 + (1 + e^{-x})} \quad (1)$$

$$tanh(x) = \frac{2}{1 + e^{-2x}} - 1 \quad (2)$$

$$ReLU(x) = max(0,x) \quad (3)$$

https://doi.org/10.17993/3ctic.2023.121.176-198

The learning effect of a convolutional neural network can be defined by the loss function. And the training goal of the convolutional neural network is to minimize the loss of function. The common loss functions are the 0-1 loss function, square loss function, logarithmic loss function, etc. Suppose the input is given as X, the true value is Y, and the predicted value of the convolutional neural network model is $f(x)$, then the squared loss function $L(Y, f(X))$ is

$$L(Y, f(X)) = (Y - f(X))^2 \tag{4}$$

The training process is divided into two processes: forward propagation and backward propagation. First, forward propagation is performed, and the prediction result $f(x)$ is calculated layer by layer according to the given input X. Then the corresponding loss is calculated according to the defined loss function, and then the backpropagation of the gradient is performed according to the loss using the stochastic gradient descent algorithm. The bias of the loss to each parameter in the network is calculated by the chain rule of derivatives, and then the weights are updated, and the update formula is

$$\omega_i \leftarrow \omega_i - \eta \frac{\partial L(Y, f(X))}{\partial \omega_i} \tag{5}$$

where ω_i is one of the weights in the network and η is the learning rate, which is used to control the intensity of the weight update.

2.2.2. DEEP NEURAL NETWORK

A deep neural network (DNN) is a feedforward neural network with multiple hidden layers [32-33], also known as MLP, and its structure is shown in Figure 4. This DNN has a total of *L*+1 layers, where layer 0 is the input layer, layers 1 to *L*-1 are the hidden layers, and layer *L* is the output layer, and adjacent layers are connected by a feedforward weight matrix.

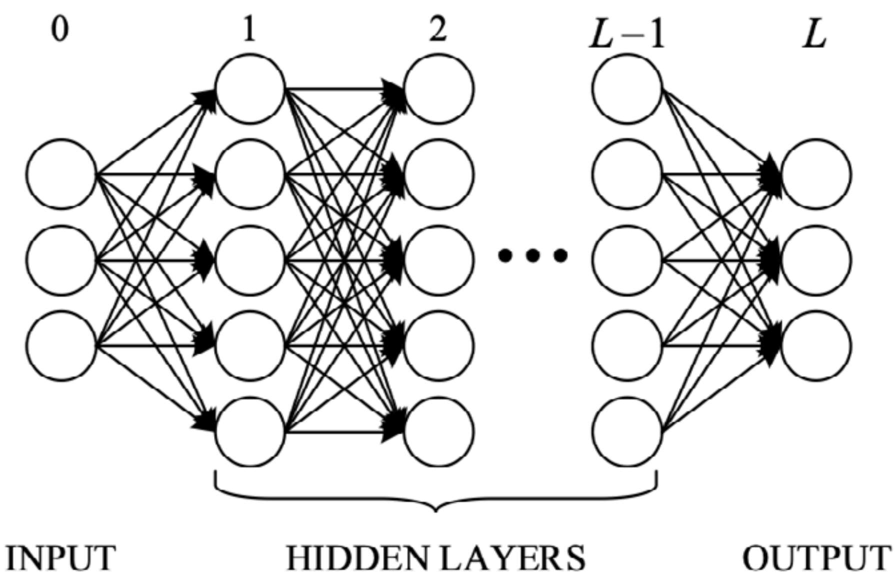


Figure 4. DNN structure schematic

https://doi.org/10.17993/3ctic.2023.121.176-198

Assuming that layer l has n_l neurons, the vector consisting of the inputs of these neurons is $z^{(l)}$ and the vector consisting of the outputs is $h^{(l)}$. Also, we let $u = h^{(l)}$ to distinguish the final output of the DNN from the output of the hidden layer, given a training sample with the feature x, at this point there is $h^{(0)} = z^{(0)} = x$. According to the rules of DNN computation

$$z^{(l)} = W^{(l)} z^{(l-1)} + b^{(l)}, l = 1,2,\cdots,L \quad (6)$$

where $W^{(i)} \in {}^{\circ}\, {}^{n_l \times_{l-1}}$ is the matrix of weights from layer $l-1$ to layer l and $b^{(l)} \in {}^{\circ}\, {}^{n_l}$ is the bias vector of the layer l. Then there is

$$h^{(l)} = f_1\left(z^{(l)}\right) \quad (7)$$

The activation function of the output layer depends on the nature of the problem to be solved by the DNN. Linear activation functions or sigmoid functions are usually used for regression problems, sigmoid functions are usually used for binary classification problems, and for multi-classification problems, the most commonly used is the softmax function, which takes the following form.

$$u = softmax\left(z^{(L)}\right) = \frac{exp\left(z^{(L)}\right)}{\sum_{k=1}^{n_L} exp\left(z^{(L,k)}\right)} \quad (8)$$

$z^{(L,k)}$denotes the krd component of the vector $z^{(L)}$.

Combined with the above process, the features x of the training samples are first sent to the input layer, then propagated through each hidden layer in the direction of the arrow in the figure and finally reach the output layer to obtain the final network output, a process called forward propagation.

DNN can also be trained with a backward propagation algorithm, and the parameters of DNN are

$$\theta = \left\{W^{(l)}, b^{(l)} \mid l = 1,2,\cdots,L\right\} \quad (9)$$

The features of the training sample set used are denoted as

$$\chi = \left\{x_i \in {}^{\circ}\, {}^{n_0} \mid i = 1,2,L\ \ ,S\right\} \quad (10)$$

The corresponding label is noted as

$$R = \left\{r_i \mid i = 1,2,L\ \ ,S\right\} \quad (11)$$

The loss function on the training set is

$$E(\theta, \chi, R) = \frac{1}{S} \sum_{i=l}^{S} E(\theta, x_i, r_i) \quad (12)$$

$E(\theta, x_i, r_i)$is the loss function corresponding to the training sample X_i. The goal of training is to minimize the training set loss function. To obtain satisfactory performance, model optimization is often performed using large-scale data, so the method often used for DNN training is stochastic gradient descent.

Firstly, x is sent to the DNN input layer to complete the forward propagation and $E(\theta, x_i, r_i)$ is calculated based on the output and the label r. Define the error signal of the layer l as

$$\mathbf{e}^{(l)} = \left. \frac{\partial E(\boldsymbol{\theta}, \mathbf{x}, r)}{\partial \mathbf{z}^{(l)}} \right|_{\theta = \ (t)} \tag{13}$$

The error signal $e^{(L)}$ of the output layer can be easily calculated according to the loss function, and for the hidden layer according to the chain rule of derivation

$$\begin{aligned} \mathbf{e}^{(l)} &= \left. \frac{\partial E(\boldsymbol{\theta}, \mathbf{x}, r)}{\partial \mathbf{h}^{(l)}} \right|_{\theta = \ (t)} \text{ e } \frac{\partial \mathbf{h}^{(l)}}{\partial \mathbf{z}^{(l)}} \\ &= \frac{\partial \mathbf{z}^{(l+1)}}{\partial \mathbf{h}^{(l)}} \cdot \left. \frac{\partial E(\ , \ , r)}{\partial \mathbf{z}^{(l+1)}} \right|_{\theta = \ (t)} \text{ e } f_l'\left(\mathbf{z}^{(l)}\right) \\ &= \mathbf{W}^{(l+1)^T} \cdot \mathbf{e}^{(l+1)} \text{ e } f_l'\left(\mathbf{z}^{(l)}\right) \end{aligned} \tag{14}$$

where ⊙ denotes element-by-element multiplication. From this calculation process, it can be seen that the error signal is propagated from layer L to layer 1 through the weight matrix in the opposite direction of forward propagation, hence the name backward propagation. Finally, according to the chain rule, we can get

$$\left. \frac{\partial E(\boldsymbol{\theta}, \mathbf{x}, r)}{\partial \mathbf{W}^{(l)}} \right|_{\theta = \ (t)} = \frac{\partial E(\ , \mathbf{x}, r)}{\partial \mathbf{z}^{(l)}} \cdot \frac{\partial \mathbf{z}^{(l)}}{\partial \mathbf{W}^{(l)}} = \mathbf{e}^{(l)} \cdot \mathbf{h}^{(l)^T} \tag{15}$$

$$\left. \frac{\partial E(\boldsymbol{\theta}, \mathbf{x}, r)}{\partial \mathbf{b}^{(l)}} \right|_{\theta = \ (t)} = \frac{\partial E(\ , \mathbf{x}, r)}{\partial \mathbf{z}^{(l)}} \cdot \frac{\partial \mathbf{z}^{(l)}}{\partial \mathbf{b}^{(l)}} = \mathbf{e}^{(l)} \tag{16}$$

Based on the above derivation process of forward and backward propagation, it can be found that the training of DNN involves a large number of matrix operations, which are well suited to speed up the computation by using graphics processing units (GPUs). The development of GPU technology is credited with driving the deep learning research boom. Since GPUs are suitable for large-scale matrix operations, to give full play to their power, in actual training, we use the mini-batch SGD algorithm, i.e., for each iteration, a small number of samples are randomly taken from the training samples to form a mini-batch, and the gradients corresponding to all the samples in it are calculated at the same time, and this gradient information is used to update the model parameters in this round. the use of SGD can greatly accelerate the training of DNN The use of SGD can greatly accelerate the DNN training speed.

2.2.3. VGGNET NETWORK

In 2014, Simonyan and Zisserman, scholars from Oxford University, proposed the famous VGG family of models (including VGG-11/VGG-13/VGG-16/VGG-19) [34-35] and achieved second place in the classification competition and first place in the localization competition at the ImageNet competition that year.VGGNet, with its good generalization performance, VGGNet has been widely used in the field of computer vision.

https://doi.org/10.17993/3ctic.2023.121.176-198

The network structure of VGGNet is shown in Table 1. VGGNet replaces five convolutional layers with five groups of convolutions, adding the previous part consisting of five convolutional layers superimposed with an activation function, so that each part does not consist of one convolutional layer plus an activation function, but multiple such combinations, with pooling operations between each part[36-37].

Table 1. Architecture of VGGNet

A	A-LRN	B	C	D	E
11 Layers	11 Layers	13 Layers	16 Layers	16 Layers	19 Layers
Input (224×224 RGB) Image					
conv3-64	conv3-64	conv3-64	conv3-64	conv3-64	conv3-64
	LRN	**conv3-64**	conv3-64	conv3-64	conv3-64
Maximum pooling layer					
conv3-128	conv3-128	conv3-128	conv3-128	conv3-128	conv3-128
		conv3-128	conv3-128	conv3-128	conv3-128
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
conv3-256	conv3-256	conv3-256	conv3-256	conv3-256	conv3-256
			conv1-256	**conv3-256**	conv3-256
					conv3-256
Maximum pooling layer					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
			conv1-512	**conv3-512**	conv3-512
					conv3-512
Maximum pooling layer					
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
conv3-512	conv3-512	conv3-512	conv3-512	conv3-512	conv3-512
			conv1-512	**conv3-512**	conv3-512
					conv3-512
Maximum pooling layer					
Fully connected layer-4096					
Fully connected layer-4096					
Fully connected layer-1000					
Softmax layer					

https://doi.org/10.17993/3ctic.2023.121.176-198

VGGNet has the following main advantages.

1. VGGNet uses convolutional kernels with small sensory fields instead of those with large sensory fields, which can reduce the number of network parameters. Since the number of parameters is greatly reduced, one convolutional layer with a large sensory field can be replaced by multiple convolutional layers with a small sensory field, thus increasing the nonlinear expression capability of the network.
2. Starting from VGG-16, VGGNet introduces a convolutional layer with 1×1 kernel size, which enhances the nonlinear expression ability of the network without affecting the size of the feature map.

3. APPLICATION OF AR VIRTUAL IMPLANTATION TECHNOLOGY IN INTERACTIVE PICTURE BOOKS

Augmented Reality (AR) is a branch of VR technology, which is currently used in various fields such as design and life. With the maturity and progress of AR technology, AR technology has been gradually applied to new fields such as education, with a wide range of radiation, such as medical systems, cultural heritage preservation, games and entertainment, children's publishing books, etc. AR technology system extends from a simple desktop to a complex interactive experience and gradually expands to touch screen and portable convenience.

3.1. COMBINED APPLICATION OF AR TECHNOLOGY IN MULTIPLE DIRECTIONS

Currently, in the context of AR technology being widely used in other fields, the technology has also begun to be applied in the book publishing industry. In the process of application, it will also fully combine other forms of technology, such as digital technology, multimedia technology, etc. The use of these technologies provides more possibilities for augmented display technology so that it can be applied to more fields. The static and single content expression is presented in a realistic and three-dimensional way.

3.2. FEATURES OF AR TECHNOLOGY

AR technology is to fuse virtual impact and real impact, and enhances reality to achieve real-time interaction, and through three-dimensional registration to complete the realization of AR products.

1. Virtual reality fusion refers to the enhancement of realistic scenes with the superimposition of virtual objects and realistic environments. Augmented reality technology provides an intermediate transition state for children, this transition state is partly virtual imaginary and partly real, the emergence of this transition state can alleviate children's anxiety due to cognitive uncontrollability after age enhancement and constant exposure to the external world. Potential space, as a kind of

https://doi.org/10.17993/3ctic.2023.121.176-198

intermediary in the transition space, can prompt children to enter reality from the fantasy world, participate in reality and cognize the object.

2. Real-time interaction means that people can interact and operate with the enhanced reality environment machinery through the device. This interaction should meet real-time requirements. Augmented reality technology can convert the static content of science-based children's picture books into a huge network linked by keywords, consisting of graphic images, animation, sound effects and text and other media to form an independent text. Children interact with AR picture books and can get a new reading experience with the help of clicking, touching and hearing, which is incomparable to traditional picture books.
3. Three-dimensional registration (also called three-dimensional alignment) is a one-to-one fusion correspondence between the computer-generated virtual image and the real environment, and maintains accurate positioning and correspondence, while the correspondence between virtual objects and the real environment must be perfectly integrated to facilitate users to use their portable cell phones to use the camera to aim at the QR code for identification, to identify the information materials that match the QR code.

3.3. COMPARISON OF AR INTERACTIVE PICTURE BOOKS AND ORDINARY PICTURE BOOKS

AR interactive picture books and traditional picture books are communication mediums to convey information in the form of images. AR picture books are based on traditional picture books, integrated with augmented reality technology, cell phones scan the picture of the book in three-dimensional presentation. The traditional picture book display form becomes more diversified. It can be said that AR picture books are developed based on traditional picture books, and AR picture books can meet more needs of children. The comparison of the two is shown in Table 2.

https://doi.org/10.17993/3ctic.2023.121.176-198

Table 2. Comparison table of traditional picture books and AR picture books

	Traditional picture books	AR picture books
Vision	Traditional picture books convey information through pictures, the output of knowledge points is more homogeneous.	(1) Cell phones scan the picture of this book, a three-dimensional model and animation display appears on the screen, more vivid image, enhance the emotional experience of children. (2) Augmented reality technology allows the presentation of models' diversification.
Aural	Without parents around to explain, children can only can read with their own understanding.	The content of the science is equipped with audio commentary, subtitles, etc. more memorable for children.
Haptics	Hand-turned book touch only, no interactive operation interaction.	(1) interacting with the model inside the phone screen. through gesture operations can achieve zoom in, zoom out, pan zoom in, zoom out, pan, rotate, etc., which can enhance children's reading interest in reading. (2) AR picture books, by adding game sessions, can to develop children's thinking in a multi-dimensional way.
Other Dimensions	Traditional science picture books do not meet children for digital needs.	AR development supports a wide range of systems to meet the needs of different needs of different models.

3.4. USE OF DEEP LEARNING AND AFFECTIVE TECHNIQUES

AR interactive picture book is a comprehensive picture, voice, click an interactive multimedia product, combined with deep learning product emotional design can make the AR interactive picture book greatly enhance the expressive power, and interactive effect significantly improved.

3.4.1. EMOTION DECODING AND EMOTION LABELING

The most important medium in picture books is images, and how exactly do the underlying features of images generate high-level abstract emotions? It is difficult to explain clearly even for humans, and the machine learning this process is a huge semantic gap, and the process of deep learning is to bridge this gap, but because of the end-to-end nature of deep learning, the machine learning process in terms of which features are extracted is something that no one can say. One cannot understand what rules are used to map the underlying features to the high-level

https://doi.org/10.17993/3ctic.2023.121.176-198

sentiment semantics. This is often referred to as the "black box" of deep learning. The current research in the field of computer science has a lot of neural network feature extraction rules and emotion mapping rules, the object of study is the machine, and the study is "how the machine learns emotion".

Starting from the international award-winning works of high quality and reflecting contemporary design trends, and guided by design emotion theory, we selected product design images with more obvious design emotion characteristics and established a preliminary design gallery with 2048 images covering common product categories. Next, we invited design professionals to extract design emotion vocabulary based on design knowledge and experience and learning of the three-level theory, and to label the images in the design emotion gallery with emotion, identifying the typical emotion label for each design image and the design feature label that generates emotion.

As shown in Table 3, the overall distribution of various sentiments of the picture samples was obtained after the aggregation of all data annotations.

Table 3. Sentiment distribution of image samples

Emotional name	Emotional share	Level of affiliation
Easy	22.70%	2
Comfort	12.15%	2
Simple	11.08%	1
Trust	10.91%	3
Softness	10.36%	1
Happy	8.47%	1
Noble	7.03%	1
Surprise	5.61%	1
Ordered	5.62%	2
Lovely	4.37%	1
Belonging	1.71%	3

In the emotional labeling of product images, the first level of emotion not only has the largest number of words, but also has the highest total percentage, reaching 46.92%, and the second level also has a total percentage of 40.46%, and the highest percentage of ease of use and comfort belongs to the second level, becoming the most prominent emotional words, which also reflects the current trend of emotional design towards the second level. The third level of emotion not only has the least variety, but also accounts for far less than the first two levels, with only 10.91% and 1.71%, and the sense of belonging has become the most unattractive emotion.

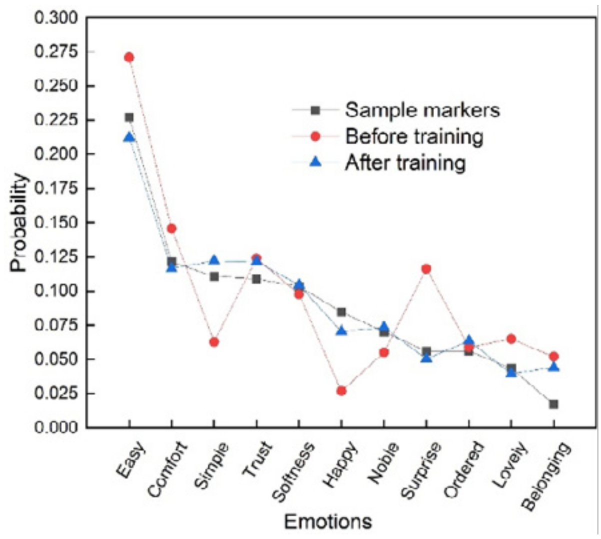
https://doi.org/10.17993/3ctic.2023.121.176-198

3.4.2. EMOTIONAL DEEP LEARNING

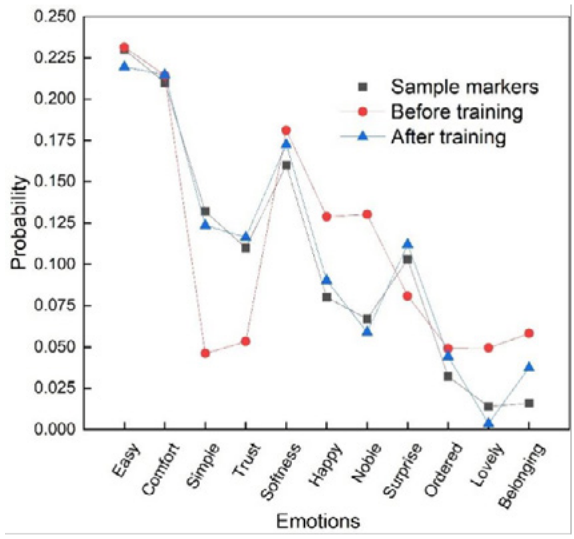
Based on the 2.2.3 VGGNet deep learning network, 1500 of the above 2048 images were used for model training, and the remaining images were used for testing. After the model training is completed, the sentiment with the highest probability for each image is taken as a single label for the test set, the degree of compliance between prediction and actual labeling is compared, and the accuracy is measured in the form of a percentage, which is more intuitive and clear. In this paper, cross-validation is used to calculate the final learning effect. Cross-validation, also known as cyclic evaluation, is done by first dividing the total data set into multiple parts at random, using one part of the data for testing and most of the remaining part for training the model, which can be used to find the prediction accuracy in the test set, and then selecting the second part for testing and the rest for training. This cycle continues until all samples are tested and only tested once, and finally, the average of all predictions is taken. According to this method, we divide more than 2000 images into random equal parts and use one part for testing and the rest for training the model, so that each part is used for prediction once in 10 rounds. The average of the 10 results is taken as the evaluation criterion. The process can also be repeated several times, each time with a different randomly divided sample, to achieve multiple cross-validations for more accurate results.

The predicted results of the sentiment probability distribution of the images are shown in Figure 5.

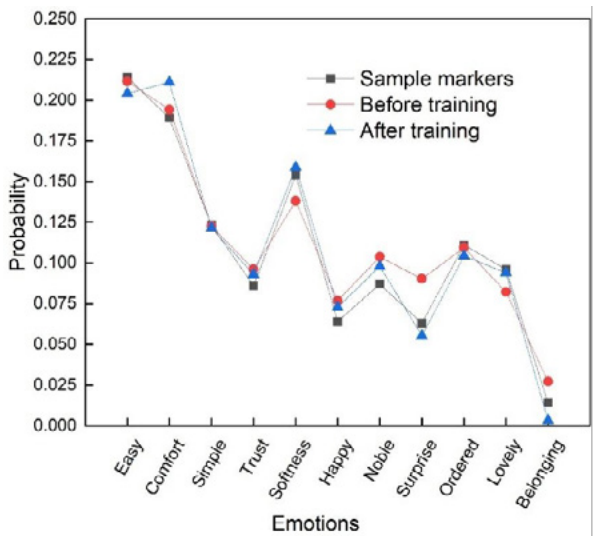
https://doi.org/10.17993/3ctic.2023.121.176-198



(a) Training samples (1-500)



(b) Training samples (501-1000)



(c) Training samples (1001-1500)

Figure 5. Predicted results of the probability distribution of emotions

https://doi.org/10.17993/3ctic.2023.121.176-198

With this calculation method, we obtained a final single-label accuracy of 61.02%, which is a very good learning result compared to the 50% to 60% accuracy of single-label classification of emotions by traditional machine learning methods. The horizontal axis of the figure indicates 11 emotion categories: pleasant, cute, surprise, soft, comfortable, trust, plain, easy to use, noble, sense of order, and sense of belonging. The vertical axis shows the proportion of each emotion in each picture. It can be seen that the prediction results of the learning are all improved compared with those before the learning, and the prediction curves fit better and better as the number of learning samples increases. The evaluation of the learning results in this study includes two parts, the evaluation of the loss of multi-label distribution (KL loss) and the evaluation of the accuracy of single-label classification (cross-validation), and the combination of the two aspects can be a good measure of the learning effect. The evaluation results demonstrate the feasibility and superiority of deep learning in emotion recognition and also show that the emotion-based design combined with deep learning can better serve the AR interactive picture book creation.

4. CONCLUSION

As the book industry becomes more and more depressed, traditional paper picture books are slowly forgotten by children and parents, and many children are even addicted to smartphones, IPads and other mobile devices, rarely reading picture books. The combination of AR technology and traditional industry links can well inherit the culture of traditional paper hand-drawn, giving children's picture books a new connotation, such as AR interactive picture books are more dynamic, enhance the reader and the book presents interactivity, increasing the fun of paper books. At the same time, different from paper picture books, AR interactive picture books can strengthen the interactive communication experience based on deep learning algorithms. By analyzing and summarizing the sources and connotations of emotional design, we focus on the three levels of emotional design, highlighting the second and third levels of emotional needs in the creation of interactive picture books. The creation of AR interactive picture books with such new technology, can bring children an emotionally pleasant experience when reading and subtly educate them.

DATA AVAILABILITY

The data used to support the findings of this study are available from the corresponding author upon request.

CONFLICTS OF INTEREST

The author declares that there is no conflict of interest regarding the publication of this paper.

REFERENCES

https://doi.org/10.17993/3ctic.2023.121.176-198

(1) Standoli, G., Salachoris, G. P., Masciotta, M. G., et al. (2021). **Modal-based FE model updating via genetic algorithms: Exploiting artificial intelligence to build realistic numerical models of historical structures**. *Construction and Building Materials*, 303, Oct. 11.

(2) Maduabuchi, C. (2022). **Thermo-mechanical optimization of thermoelectric generators using deep learning artificial intelligence algorithms fed with verified finite element simulation data.** *Applied Energy*, 315, 118943-.

(3) Gasimova, R. T., & Abbasl, R. N. (2020). **Advancement of the search process for Digital Heritage by utilizing artificial intelligence algorithms**. *Expert Systems with Applications.*

(4) Dias, L., Coelho, A., Rodrigues, A., et al. (2013). **GIS2R — Augmented reality and 360° panoramas framework for geomarketing**. *In Information Systems & Technologies* (pp. 1-6). IEEE.

(5) Moretamartinez, R., Mediavillasantos, L., & Pascau, J. (2021). **Combining Augmented Reality and 3D Printing to Improve Surgical Workflows in Orthopedic Oncology: Smartphone Application and Clinical Evaluation**. *Sensors*, 21(4), 1370.

(6) Laviola, E., Gattullo, M., Manghisi, V. M., et al. (2022). **Minimal AR: visual asset optimization for the authoring of augmented reality work instructions in manufacturing**. *The International Journal of Advanced Manufacturing Technology*, 119(3), 1769-1784.

(7) Zhang, Z., Li, T., & Yoon, S. (2021). **A Feasibility Study of LiDAR-Enhanced Augmented Reality on a Handheld Device for Collision Detection and Patient Positioning**. *International Journal of Radiation Oncology, Biology, Physics*, 111-3S.

(8) Frederico, A. (2016). **The future of the reader or the reader of the future: Children's interactive picturebook apps and multiliteracies.**

(9) Sun, C. (2020). **Interactive Picture Book Read-Alouds to the Rescue: Developing Emerging College EFL Learners' Word Inference Ability**. *Journal of Adolescent & Adult Literacy*, 63.

(10) Kuswaty, M., & Cahyani, I. (2019). **Analysis of musical drama in and the magic book**. *KnE Social Sciences.*

(11) Johnson, A. (2021). **Malory's Magic Book: King Arthur and the Child**, 1862–1980 *by Elly McCausland. The Lion and the Unicorn*, 45(1), 124-127.

(12) Ren, P. (2020). **AR 3D Magic Book: A Healthy Interactive Reading Device Based on AR and Portable Projection**. *In CIPAE 2020: 2020 International Conference on Computers, Information Processing and Advanced Education.*

(13) Min, X. U., Tong, Q., Chen, D. C., et al. (2015). **Design and Production of Augmented Reality Courseware Based on ARTool Kit**. *Modern Computer*.

(14) Khan, D., Ullah, S., & Rabbi, I. (2015). **Classification of Markers in the ARTool Kit Library to Reduce Inter-marker Confusion**. *In International Conference on Frontiers of Information Technology* (pp. 257-262). *IEEE*.

(15) Simon, B. N., Chandrashekar, C. M., & Simon, S. (2011). **Hamilton's turns as a visual tool-kit for designing of single-qubit unitary gates**. *arXiv preprint arXiv,* 1108.1368.

 https://doi.org/10.17993/3ctic.2023.121.176-198

(16) Hermawan, H. D., Yuliana, I., Saputri, A., et al. (2021). **How Does ARBook Promotional Media For MSME Crafts Using Augmented Reality Marker Tracking Works?.**

(17) Chiu, P.-H., Lin, P.-H., Lee, H.-L., et al. (2018). **Interactive Mobile Augmented Reality System for Image and Hand Motion Tracking**. *IEEE Transactions on Vehicular Technology*, 67(6), 5029-5040.

(18) Chrysanthi, A., Papadopoulos, C., & Frankland, T. (2012). **'Tangible pasts': user-centred design of a mixed reality application for cultural heritage**. *Computer Applications & Quantitative Methods in Archaeology*, 2012, 113-120.

(19) Grasset, R., Dünser, A., & Billinghurst, M. (2008). **The design of a mixed-reality book: Is it still a real book?** *In IEEE/ACM International Symposium on Mixed & Augmented Reality* (pp. 73-82). IEEE.

(20) Barvir, R., Vondrakova, A., & Brus, J. (2021). **Efficient Interactive Tactile Maps: A Semi-Automated Workflow Using the TouchIt3D Technology and OpenStreetMap Data**. *ISPRS International Journal of Geo-Information*, 10(5), 316.

(21) Guo, C., Hou, Z. X., Shi, Y. Z., et al. (2017). **A virtual 3d interactive painting method for Chinese calligraphy and painting based on real-time force feedback technology**. *Frontiers of Information Technology and Electronic Engineering*, 18(11), 1799-1810.

(22) Wy, A., Dy, A., Eh, A., et al. (2018). **Acid-activatable oxidative stress-inducing polysaccharide nanoparticles for anticancer therapy**. *Journal of Controlled Release*, 269, 235-244.

(23) Wright, D. (2020). **Encountering UFOs and aliens in the tourism industry**. *Journal of Tourism Futures*, (aop).

(24) Boli, J. (2018). **Small planet in the vastness of space: Globalization and the proliferation of UFOs, aliens, and extraterrestrial threats to humanity**. *Discussion Papers, Research Unit: Global Governance.*

(25) MD Lafayette. (2011). **1520 Things You Don't Know About Ancient Aliens, UFOs, Aliens Technology and U.S. Black Operations** [Ebook].

(26) Shuman, V., Scherer, K., Fontaine, J., et al. (2015). T**he GRID meets the Wheel: Assessing emotional feeling via self-report**. *Oxford University Press.*

(27) Dubreucq, S., Marsicano, G., & Chaouloff, F. (2015). **Emotional consequences of wheel running in mice: which is the appropriate control?** *Hippocampus*, 21(3), 239-242.

(28) Zhang, X. (2016). **Research on Ways to Stimulate Students' Interest in Learning Korean: Based on Dornyei Three-level Theory**. *The Science Education Article Collects.*

(29) Zhao, T., & Zhu, T. (2019). **Exploration of Product Design Emotion Based on Three-Level Theory of Emotional Design**. *Springer, Cham.*

(30) Cui, F., Ning, M., Shen, J., et al. (2022). **Automatic recognition and tracking of highway layer-interface using Faster R-CNN**. *Journal of Applied Geophysics*, 196,

(31) Shin, H. C., Roth, H. R., Gao, M., et al. (2016). **Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset**

Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5), 1285-1298.

(32) Kim, S., Kojima, M., & Toh, K. C. (2016). **A Lagrangian–DNN relaxation: a fast method for computing tight lower bounds for a class of quadratic optimization problems**. *Mathematical Programming*, 156(1-2), 161-187.

(33) Li, J., Zhao, R., Huang, J. T., et al. (2014). **Learning Small-Size DNN with Output-Distribution-Based Criteria**. *In Conference of the International Speech Communication Association. ISCA.*

(34) Wang, L. (2015). **Places205-VGGNet Models for Scene Recognition**. *Computer Science.*

(35) Rao, B. S. (2020). **An Accurate Leukocoria Predictor Based On Deep VGG-Net CNN Technique**. *IET Image Processing*, 14(5).

(36) Singh, A. K., & Sora, M. (2021). **An optimized deep neural network-based financial statement fraud detection in text mining**. *3C Empresa. Investigación y pensamiento crítico*, 10(4), 77-105. https://doi.org/10.17993/3cemp.2021.100448.77-105

(37) Yan Kang, Jinling Song, Mingming Bian, Haipeng Feng, & Salama Mohamed. (2022). **Red tide monitoring method in coastal waters of Hebei Province based on decision tree classification**. *Applied Mathematics and Nonlinear Sciences, 7*(1), 43-60. https://doi.org/10.2478/AMNS.2022.1.00051