AI-Powered Digital Media Platform and Its Applications

Fang Liang[1]

[1]Qianjiang Collegee, Hangzhou Normal University, Hangzhou, 310018, China
360949215@qq.com

ABSTRACT

Previously, the main method to create digital media such as animation was by humans. With the development of artificial intelligence technology, the work of making media content from raw materials is gradually replaced by computers. In this paper, we conceive a digital media production and interaction platform supported by artificial intelligence, based on some existing artificial intelligence and virtual reality display technologies. We also discuss the feasibility of some possible application scenarios for this platform.Based on our literature review and feasibility analysis, we conclude that current technology development is capable of supporting the AI-powered digital media platform that we propose in this paper.

CCS CONCEPTS

CCS· Computing methodologies· Artificial intelligence

KEYWORDS

Artificial Intelligence; Virtual Reality; Biometric Authentication

1 INTRODUCTION

Digital media technology, as a new technology with special forms of artistic expression, can make abstract information become sensible, manageable and interactive. In the process of social practice, digital media technology not only has a very broad application direction, but also has a very prominent role.

Artificial intelligence technology is the advanced form of computer technology development. It fully simulates the operation process of human brain and realizes the overall intellectualization of technology. To some extent, the application of artificial intelligence technology can also effectively reduce people's workload and reduce human errors. In practice, the application of artificial intelligence technology can not only carry out autonomous work in accordance with the program, but also carry out creative work based on its own characteristics. Nowadays, with the comprehensive and rapid development of digital media technology, the wide application of artificial intelligence technology can optimize the usage of digital media technology to a great extent. In particular, artificial intelligence can enable the system to mimic human thinking, which is the innovation ability needed by digital media technology.

Character design is one of the main application of artificial intelligence in digital media technology. Before the emergence of artificial intelligence, designers may need a lot of tedious and repetitive work in the character design process. However, relying on artificial intelligence technology, it can effectively improve the speed and quality of character design and comprehensively optimize the overall effect of character design. In the process of character design, the maximum utilization of information and resources can be realized by relying on artificial intelligence technology. In the process of design, artificial intelligence technology can help designers carry out the design work quickly and in batches. The designers can rely on artificial intelligence technology to generalize the design framework, and promote the quality of the design and performance through the framework of adjustment and optimization, effectively making up for the cumbersome traditional design.

Artificial intelligence is also very useful in game design, especially the online game. When the majority of users play games, in addition to their own role settings, they also include other game roles in the game. The application of artificial intelligence in game programming is mainly reflected in these aspects. Through artificial intelligence technology, it can provide rich battle roles for the vast number of users, which enhances the fun of the game. At the same time, relying on artificial intelligence technology, users will have a different experience every time they play the game, which can enhance their interest in participating. If the game program setting is the same every time, it undoubtedly limits the user's participation in the fun to a great extent. But at the same time, in the process of game programming, the use of artificial intelligence technology is to optimize the game role, but also to ensure the game strategy that can be analyzed, so that users can study the possibility of deep participation in the game through analysis and judgment.

In addition, in the aspect of network security, and many game will involve a lot of network supervision, early are with the method of artificial audit, audit on all users to upload content, engineering of the huge, boring, artificial intelligence technology to help audit department will account for more than for the vast majority of normal content filtering, so that greatly reduce the workload of artificial audit, more help to the popularization and application of the game.

CAIH2020, October 23–25, 2020, Taiyuan, China

ACM ISBN 978-1-4503-8864-1/20/10…$15.00
https://doi.org/10.1145/3433996.3434018

In this paper, based on some existing artificial intelligence technologies and virtual reality display technologies, we conceive a digital media production and interaction platform supported by artificial intelligence, which is composed of three modules. We also propose the future possible application scenarios for this platform, including but not limited to education, companion, healthcare, cultural Heritage and public welfare services.

The rest of this article is organized as follows:

In Section 2, we briefly discussed the basic technologies involved in the Ai-driven digital media platform.

In Section 3, we explained the architecture of the system and the working process of the system.

In Section 4, we propose some possible application scenarios of the platform in the future.

In Section 5, we summarize this article.

2 FOUNDATIONAL TECHNOLOGIES

In this section, we give a short discussion about the foundational technologies that consist of the basic components of our AI-powered digital media platform.

2.1 Virtual Reality

Virtual reality is a term used to describe a three-dimensional computer-generated environment that can be explored and interacted with by humans [1]. The characteristics of multi-perception, immersion and interactivity in virtual reality enable people to enter the generated virtual environment, interact with it and communicate. Through the interaction between the participants and the simulated environment, and with the help of the person's own perception and cognitive ability of the things they are in contact with, it helps to inspire the participants' thinking, so as to obtain all kinds of spatial information contained in the environment in an all-round way, just like being on the scene.

With the development of artificial intelligence, virtual reality technology has been widely used in many fields such as medicine, entertainment, business, competitive sports.

Online fitting room is an example of virtual reality [2]. You do not have to worry about buttons, zippers and little details about specific sizes. Consumers can choose a standard model or create a model similar to your own body shape, skin color, hair style, body features, height, weight, and specific details such as eye color, nose, and lips. Once your registration is confirmed, you can design for your model with all the clothes sold at H&M, including swimsuits, ready-to-wear and knickknacks. When the consumer completes the satisfied combination, the storage enters "your wardrobe", prints out the combination in the wardrobe, and takes it to the shopping mall.

Virtual reality systems are often controlled by more advanced operating systems, such as virtual reality gloves [3]. While early attempts included Nintendo's power gloves, today's virtual reality systems are controlled by platforms such as Razer Hydra, which is widely used in head-mounted displays such as the Oculus Rift. Running on a machine that emits a weak magnetic field, Razer Hydra USES hardened circuits, a digital signal processor and positioning algorithms to turn data collected by the controller into position and direction information, which is then used as a control element in a computer graphics environment. Virtual reality combines several technologies to create a believable virtual world for its users. The most important accessory is the head-mounted display, which USES a technique called stereoscopic observation to allow our eyes to see slightly different images. The Oculus Rift has a single LCD screen with a 24-bit color depth and a 1,000-hertz adjacency reality tracker that reduces image lag and allows for more realistic image quality when the head-mounted display moves. In addition, the lens at the front of each eye provides depth of field information, so it gives a complete picture of the wearer's vision. The head-mounted display also includes tracking software that allows the user to move his head and the virtual world to move with it.

2.2 Biometric Authentication

The traditional identity authentication method is based on identification objects such as certificates or keys, or knowledge such as passwords. There are obvious problems such as the loss or forgery of personal possessions, and the biological characteristics inherent in our body provide the most convenient way for this, which is more reliable and more convenient than traditional identity authentication.

Biometric Authentication is based on the unique features of each person's body, such as fingerprints, irises, etc [4]. Synaptic Natural ID fingerprint sensor solution, for example, encodes fingerprint template data with aes256-bit encryption, which is critical to the rapid deployment of biometric authentication on smartphones.

Fingerprint identification ensures the security of personal information [5]. On May 30, 2012, the newly revised "law on resident identity cards" clearly stipulated that the registration items of resident identity cards include fingerprint information in China. The Ministry of Public Security has formulated the task objectives, working measures and implementation plans for the preparation of fingerprint information for the registration of resident identity cards by the end of 2012, and public security departments and bureaus around the country have set up leading groups and offices to formulate preliminary work plans. In January 2013, China start to register fingerprint information on id cards and suspend the use of first-generation cards.

Brainwave authentication is a novel method of authentication that researchers have come up with [6]. It is a new identification technology that scans the brain, analyzes brainwaves, judges a user's emotional and mental health, and decides whether he or she should be allowed to use certain systems or applications. According to Digital Trends, researchers have come up with a new biometric method that analyzes a person's brain waves,

determines their mood and mental health, and then decides whether to let him or her use a device or app. In a paper, Violeta Tulceanu, a scientist at Alexandru Ioan Cuza University in Romania, describes the neural activity that she and her colleagues associate with different emotions, and whether that neural activity -- a barometer of health psychology -- can be used for identity verification [7].

The contact method of fingerprint identification is more direct. But this is also a disadvantage because the contact image pollution would affect the reliability. Finger contamination can also cause legal disputes, as has happened in Hong Kong. In addition, the device cannot be made smaller than the thumb. Extracting single finger image affects reliability, while extracting multiple fingerprint images affects practicability. It is easier to make false fingerprints and fingerprints on the contact surface are easy to be stolen.

Compared with fingerprint recognition, it can be seen that iris recognition has omni-directional advantages and will become a mainstream product in the market segment of human body biological characteristics recognition [8]. The iris is located between the sclera and the pupil. It contains the richest texture information and is one of the most unique structures in the human body.Iris recognition or iris scanning is the process of taking high-contrast photos of human iris using visible light and near-infrared light. It is also a better recognition way to be combined with virtual reality when the touch of fingers is not available. The high uniqueness, stability and unchangeable characteristics of the iris are the material basis that the iris can be used for identity identification.

2.3 Artificial Intelligence

Represented by the latest techniques of deep learning, artificial intelligence can be used for many tasks and bring dramatic improvements [9-14]. In this part, we discuss the role of artificial intelligence in photograph repair, video classification, and lip-reading recognition, which are fundamental in our system.

2.3.1 Photograph Repair [15]

Before the popularization of artificial intelligence technology, repairing pictures is a more professional job and is often done by the picture-editing software, e.g., Photoshop. Professional photographers have higher requirements for image processing. They can refine a picture for at least 20 minutes or even a day or two.

However, with the rapid development of artificial intelligence, AI is also increasingly being used in image processing software, meanwhile the optimization of photos is becoming more and more intelligent. For example, most smartphones now have AI, and in the process of taking pictures and videos, they directly "optimize" portraits, plants, blue sky, pets, food, etc., in line with the public's aesthetic standards. This makes it easier and easier for amateur photographers to process photos.

Old black and white photos can be turned into colorful with artificial intelligence too. Some old photos in the past are scanned or reproduced into digital photos after being taken with a film camera. Because of digital technology in older days, the digital photos had insufficient pixels. With the baptism of time, the old photos become blurred and unclear, which is a pity. Nowadays, we can not only recover the details of old photos, but also increase the resolution, making the photos clearer. Artificial intelligence is used to restore the original appearance of the previous characters in the photos to the people.

2.3.2 Video Classification [16]

Video technology is inseparable from artificial intelligence, and has become an important means to achieve artificial intelligence. In the early days, most video classifications were based on features designed by hand and typical machine learning methods. With the development of deep learning, methods based on artificial features Fading out people's vision, the trend of video classification is developing towards the general multi-label video classification.

In recent years, deep learning has achieved the highest level of effect in video classification. video classification algorithms involve many technologies, including video frame feature extraction, video frame feature integration, multi-modal video information extraction (video picture, voice, object motion, scene, etc) and other aspects.

ACM MM is the world's top machine vision conference. LSVC, as an important part of ACM MM, is called Large-Scale Video Classification Challenge, which mainly tests the ability of participating teams in large-scale video classification algorithms. The competition data set contains approximately 160,000 videos from Youtube for a total of 8,000 hours. The participating teams need to identify the 500 types of content in the video, including social events (such as football games), objects (such as pandas), scenes (such as beaches), and actions (such as making cakes). In 2017, ACM MM LSVC announced the best results of the year, and the Alibaba iDST team won the championship with an average accuracy of 87.41%.

The Alibaba iDST team used Inception-Resnet-v2 and Squeeze-and-Excitation Networks to extract video frame features, and used NetVLAD to integrate the extracted video frame features. After combining the fusion of multi-modal information, the average accuracy of the single model on the validation set reached 84.85%, and the fusion of multiple models reached 87.41%.

At present, many LSVC's award-winning video analysis algorithms have been applied to multiple businesses including Youku, Tudou, UC, Xianyu, etc., which effectively improves the user's experience in video search, recommendation, and editing.

2.3.3 Lip-reading Recognition [17]

The lip recognition system uses machine vision technology to continuously recognize a face from the image and extract the

continuous lip-change features. The continuously-changing features are then input into the lip-recognition model to identify the corresponding speech type. Pronunciation is used to calculate the most likely expression sentence. Lip recognition technology can be applied to fields such as daily life, social welfare, and ruling of sports events. It can help deaf-mute people and the elderly communicate, and can also achieve "lip unlock" for language violence in sports events, which is convenient for referees. In addition, lip recognition technology can also play an auxiliary role in speech interaction and image recognition. It is more suitable for places with excessive noise or difficulty in making sounds, which can not only guarantee the accuracy and stability of recognition, but also guarantee the privacy of speech content.

As early as 2003, Intel developed the lip recognition software Audio Visual Speech Recognition (AVSR), which allowed developers to develop computers that can perform lip recognition.

In 2016, Google's DeepMind collaborated with Oxford University on a project to develop an artificial intelligence system with lip reading capabilities. They selected nearly 5,000 hours of TV programs from the British BBC to provide a database for this AI system to read the lip test. These TV programs include "Evening News" and "BBC Breakfast and Question Time", which contain a total of 118,000 sentences. Only need to identify the speaker's lips, the system can accurately decipher the entire sentence. In the quiz to decipher 200 randomly selected fragments from the database, the AI lip reading system defeated professional lip readers. The accuracy rate of professional lip readers is only 12.4%, while the accuracy rate of AI systems is 46.8%.

In China, Sogou officially launched a new technology of lip recognition, which uses machine vision to recognize the speaker's lip movements to interpret what the speaker said and enhance the far-field voice interaction function. Sogou demonstrated this technology at the Wuzhen Internet Conference. According to Sogou, through the end-to-end deep neural network technology for lip sequence modeling, and thousands of hours of real lip data training, the vocabulary created by the lip recognition system has more than 100,000 words. General lip recognition has reached an accuracy rate of 60%, while it can reach an accuracy rate of 90% in vehicle and smart home scenarios.

2.3.4 Relational reasoning[18]

The main idea of relational inference technology is to use the knowledge existing in the knowledge graph to automatically infer the missing relationship between entity pairs. The knowledge graph is the most effective way to express relationships. In layman's terms, the knowledge graph is a relational network obtained by connecting all kinds of information together. The knowledge graph provides the ability to analyze problems from a "relationship" perspective.

In the early days, nearly 96% of the characters in the Freebase knowledge base lacked information about brothers or sisters, about 75% of the character objects lacked nationality information, and about 68% of the character objects lacked professional information. These information themselves are difficult to obtain through public Internet data. Relational reasoning technology can effectively supplement and improve this situation and improve the knowledge coverage of the knowledge base.

At present, many scholars at home and abroad attach great importance to the research of relational reasoning technology. As the topic of the WSDM Cup2017 competition held by top academic conferences in the field of Internet search and data mining, Triple Scoring Task as a challenge task set by the organizing committee is a typical relationship reasoning task. With the rapid development of artificial intelligence, related research results have been successfully applied in the current mainstream knowledge graph project, and Google's Knowledge Vault and Carnegie Mellon University's NELL project are the most representative.

Relational reasoning technology is also widely used in many aspects, such as intelligent question answering, personalized recommendation, assisted decision-making, etc. These applications will enhance our ability to control technology, so that people can understand and operate complex systems in society, making humans and machines The interaction will be more natural, predictable and emotional.

3 SYSTEM WORKFLOW

The system architecture is shown in Figure 1. The system consists of three major components, i.e., graphical user interface, artificial intelligence powered backend, and content database.

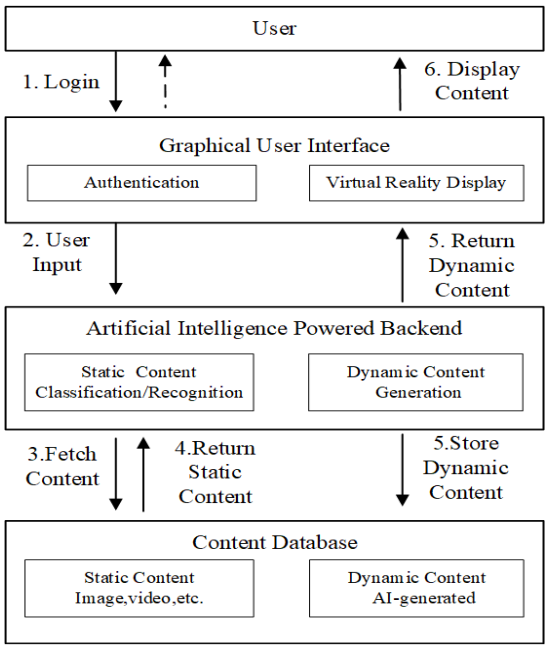

Figure 1. System structure of the proposed AI-powered digital media platform

3.1 System Modules

The module of graphical user interface performs the biometric identification function. If the user fails to pass the login check, the interface would generate the corresponding error info and prevent the user from using this system. The graphical user interface is also responsible for the content display via virtual reality technologies.

The module of artificial intelligence powered backend is responsible for searching and generating digital media content. By leveraging the state-of-the-art artificial intelligence technologies as we discuss in the last section, this module not only searches existing static content in the database, but also creates dynamic content based on user's input, e.g., command, emotion, movement, etc.

The module of content database is responsible for storing the digital media content. It not only stores the static multimodal content, e.g., text, audio, video, but also stores the dynamic content generated by artificial intelligence functions, in case they would be accessed again. To store a large volume of digital content, distributed data centers and cloud computing would be considered for store and transmit these content.

3.2 System Workflow

The workflow of the system can be summarized as follows:

Step 1. Login: After the users enter the system, the module of graphical user interface will conduct the iris certification. The user with a successful login would be proceeded to the functions of this system. Otherwise, an error info would be provided to the user.

Step 2. User Input: The user can make some interaction with the system with multimodal input, including but not limited to audio command, hand gesture, head and body movement, etc.

Step 3. Fetch Content: Based on the user's input, the module of artificial intelligence powered backend would fetch the related static content, including text, images, videos, etc.

Step 4. Return Static Content. The module of content database would return the searched static content to the module of artificial intelligence powered backend. Then these static content would become the basics of building dynamic content for a real-time interaction with users.

Step 5. Return&Store Dynamic Content. The AI-generated dynamic content would both be returned to the graphical user interface and be stored in the content database in case it would be used for multiple times.

Step 6. Display Content. Through virtual reality and other technologies, the content that the user wants to see is presented to the user through the interface, and the process ends.

4 APPLICATION SCENARIOS

After explaining the system architecture, we discuss the possible application scenarios of the AI-powered digital media platform.

4.1 Education

Education is a great market. In China alone, by 2017, the number of k12 students in primary and high schools had exceeded 183 million, and the number of teachers nationwide had reached 20 million. These numbers will continue growing in the future, bring both challenges and opportunities for AI-based education products. The future of education must be a high degree of integration of artificial intelligence techniques and campus activities. The physical campus will become more and more intelligent. Most of the basic educational facilities will become more intelligent, which is convenient for students to learn efficiently. The school teaching method will also change, from the traditional ways to the an interactive and immersive way. Many tech companies are already working with schools to bring smart educational devices to schools.

With machine translation techniques, a global virtual classroom can be set up for those who speak different languages or may have visual or auditory disabilities. Subtitles can be generated in real time for what the teacher says. It also makes it possible for students who are too ill to attend school, or need to study at different levels, or whose schools lack the specific subjects they need. Artificial intelligence can help break down barriers between schools and traditional grades.

4.2 Companion

Parents who are struggling to help their children with algebra and are busy at home with their children's homework or exam preparation must be excited to see the potential of artificial intelligence. Thanks to artificial intelligence, tutoring and learning programs are becoming more and more advanced, and soon they will be more accessible and able to cope with a variety of learning styles.

The application value of emotional artificial intelligence can go deep into every aspect of people's daily work and life. Combined with emotional functions, our AI-powered digital media platform can be widely used in business scenarios based on video acquisition because it has the characteristics of continuous, real-time, remote and non-contact acquisition of physiological and psychological information.

4.3 Healthcare

By 2050, nearly one in four people will be over 60, twice as many as now. In addition, the number of people aged 80 and over will quadruple. This demographic shift has opened up new prospects for AI technology in the daily medical management of the elderly and has become a useful tool for medical professionals and institutions to treat the elderly. The elderly care market is expected to exceed $5.5 billion by 2022 and will be one of the most important supporting roles for AI in the future society. AI

systems can also track the health of the elderly and proactively identify risks. By using machine learning algorithms to analyze historical data, AI programs can also discover unknown and predicted associations, such as daily habits and unusual sleep habits.

4.4 Cultural Heritage

Historical relics are a valuable historical and cultural heritage for human beings. Artificial intelligence has great application value in spreading Chinese culture, and museums are the sacred space where culture and history can be guaranteed and spread. The technical museums closely connected with the times will only become more Dazzling, the artificial intelligence system can start from the scene and the tourists, change the sensory experience of the visitors, change the viewing mode of the antiques that are now appreciated through the glass window, and improve the relevance of museum exhibits and tourists. Let the audience visit the exhibition, experience cultural relics, and acquire knowledge in an unprecedented way. With the help of artificial intelligence relationship processing technology, we will be able to achieve refined management of information such as museum collections and visitors, and provide ubiquitous services to visitors.

Our AI-powered digital media platform has prepared graphics restoration and virtual reality technology to help the museum reproduce the damaged cultural relics and bring the disappeared cultural relics back to life. In addition, the system can be used to calculate the voices of people from all over the world and synthesize a speech library, so as to save different national languages, pass the achievements of human civilization, let more people understand Chinese culture, and make "Chinese stories" more exciting.

4.5 Public welfare services

The emergence of artificial intelligence has promoted the development of public welfare undertakings and benefited the public.The artificial intelligence program uses the calculation of big data to help people find the object of donation, help donors determine which are the most urgent needs, and propose the most effective way to meet these needs through charity, so as to achieve a reasonable match between supply and demand.

The AI-powered digital media platform system uses artificial intelligence to compare the archive photos stored in the lost population database with the current photos of the rescued, found victims, and lost children, and uses machine learning algorithms to compare 27 Different facial features are analyzed and compared to determine whether the two face images are the same person. It is especially suitable for scenes where children have lost a lot of changes in their backs for many years.

5. CONCLUSION

In this paper, we investigated the current development of iris authentication, virtual reality, artificial intelligence technologies and relational reasoning technology, proposed a AI-powered digital media platform and explained its application process. We analyzed the feasibility of this platform based on these technologies. Based on the functionality of this platform, we proposed potential application scenarios including education, companion, healthcare., cultural Heritage and public welfare services. Based on our discussion about the technical background, we conclude that current technology development is capable of supporting the AI-powered digital media platform that we propose. Our discussion in this paper will inspire the implementation of similar platforms.

ACKNOWLEDGEMENT

The author(s) received no financial support for the research, authorship, and/or publication of this paper. There is no conflict of interest in this paper.

REFERENCES

[1] Burdea G C, Coiffet P. Virtual reality technology[M]. John Wiley & Sons, 2003.

[2] Beck M, Crié D. I virtually try it… I want it! Virtual Fitting Room: A tool to increase on-line and off-line exploratory behavior, patronage and purchase intentions[J]. Journal of Retailing and Consumer Services, 2018, 40: 279-286.

[3] Perret J, Vander Poorten E. Touching virtual reality: a review of haptic gloves[C]//ACTUATOR 2018; 16th International Conference on New Actuators. VDE, 2018: 1-5.

[4] Bhattacharyya D, Ranjan R, Alisherov F, et al. Biometric authentication: A review[J]. International Journal of u-and e-Service, Science and Technology, 2009, 2(3): 13-28.

[5] Valdes-Ramirez D, Medina-Pérez M A, Monroy R, et al. A Review of Fingerprint Feature Representations and Their Applications for Latent Fingerprint Identification: Trends and Evaluation[J]. IEEE Access, 2019, 7: 48484-48499.

[6] Soni Y S, Somani S B, Shete V V. Biometric user authentication using brain waves[C]//2016 International Conference on Inventive Computation Technologies (ICICT). IEEE, 2016, 2: 1-6.

[7] Tulceanu V. Comprehensive brainwave authentication using emotional stimuli[C]//2012 Proceedings of the 20th European Signal Processing Conference (EUSIPCO). IEEE, 2012: 1772-1776.

[8] Nguyen K, Fookes C, Jillela R, et al. Long range iris recognition: A survey[J]. Pattern Recognition, 2017, 72: 123-143.

[9] Jiang W, Zhang L. Edge-SiamNet and Edge-TripleNet: New Deep Learning Models for Handwritten Numeral Recognition[J]. IEICE Transactions on Information and Systems, 2020, 103(3): 720-723.

[10] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(12): 2481-2495.

[11] Jiang W, Zhang L. Geospatial data to images: A deep-learning framework for traffic forecasting[J]. Tsinghua Science and Technology, 2018, 24(1): 52-64.

[12] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[13] Jiang W. Applications of deep learning in stock market prediction: recent progress[J]. arXiv preprint arXiv:2003.01859, 2020.

[14] Jiang W. Time series classification: nearest neighbor versus deep learning models[J]. SN Applied Sciences, 2020, 2(4): 1-17.

[15] Zhu X, Qian Y, Zhao X, et al. A deep learning approach to patch-based image inpainting forensics[J]. Signal Processing: Image Communication, 2018, 67: 90-99.

[16] Babu R V, Tom M, Wadekar P. A survey on compressed domain video analysis techniques[J]. Multimedia Tools and Applications, 2016, 75(2): 1043-1078.

[17] Lombardi L. A survey of automatic lip reading approaches[C]//Eighth International Conference on Digital Information Management (ICDIM 2013). IEEE, 2013: 299-302.

[18] Han Minghao. Research on Relational Reasoning Algorithm Based on Knowledge Graph [D]. University of Electronic Science and Technology of China, 2018: 1-20