Journal of Computer Technology and Applied Mathematics
Journal Home: http://jctam.suaspress.org | CODEN: JCTAAD
Vol. 1, No. 4, 2024 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

A Systematic Review of Computer Vision-Based Virtual Conference Assistants and Gesture Recognition

LIN, Weikun [1*]

[1] Shandong University of Science and Technology, China

**** LIN, Weikun is the corresponding author, E-mail: Welton.lin2233@gmail.com***

Abstract: In the process of introducing gesture recognition, it is essential to explore its technical background and implementation methods. Gesture recognition algorithms based on deep learning perform exceptionally well when processing real-time video streams. These algorithms can extract gesture features and classify them to identify user intentions. For instance, analyzing gesture images using Convolutional Neural Networks (CNN) can effectively enhance recognition accuracy and real-time performance. Additionally, combining optical flow methods with object detection techniques allows for real-time tracking of user hand movements, leading to more precise recognition results. Factors such as changes in ambient lighting, cluttered backgrounds, and the diversity of user gestures can all impact recognition accuracy. Therefore, researchers need to continuously optimize algorithms to improve the robustness and adaptability of the system. At the same time, when designing virtual conference assistants, the user interface's friendliness and usability should also be considered, enabling users of varying technical skill levels to use the system with ease.

Keywords: Gesture Recognition, Computer Vision, Deep Learning.

Disciplines: Computer Science. **Subjects:** Computer Vision.

DOI: https://doi.org/10.5281/zenodo.13889718 **ARK:** https://n2t.net/ark:/40704/JCTAM.v1n4a04

1 INTRODUCTION

In an era of rapid digital development, virtual meetings have become an important way for people to communicate and collaborate. However, traditional virtual meetings have many limitations in terms of interactivity and engagement. To address these issues, computer vision-based virtual conference assistants have emerged. These assistants not only enhance meeting efficiency but also improve user experience, which is especially important in the context of the growing prevalence of remote work and online learning. Computer vision, as a significant branch of artificial intelligence, aims to enable computers to "see" and understand and analyze image and video data. In virtual meetings, computer vision technology can be used for gesture recognition, allowing for natural interaction. This interaction method enables users to convey commands through gestures rather than solely relying on voice or text input, thereby enhancing the fluidity and interactivity of the meeting. For example, users can share screens, switch slides, or adjust volume with simple gestures, operations that often require complex mouse maneuvers or multiple clicks in traditional virtual meeting software.

By effectively integrating these technologies, a more natural and efficient meeting experience can be achieved, promoting the ongoing development of remote work and learning. In the future, as computer vision technology continues to advance, we have reason to believe that virtual conference assistants will become an essential part of modern office and educational settings, further facilitating communication and collaboration among individuals.

2 2. THEORETICAL FOUNDATIONS

2.1 Computer Vision

Computer vision is a significant branch of artificial intelligence that aims to enable computers to simulate human visual systems, allowing them to understand and process image and video data. With the rapid development of deep learning technology, the application scenarios of computer vision are continually expanding, significantly driving innovation and development in related industries such as healthcare, autonomous driving, security monitoring, and virtual reality. Image classification is the process of recognizing the categories of objects within an image, typically utilizing Convolutional Neural Networks (CNN) for feature extraction and classification. Object detection not only identifies the category of objects but also determines their locations within the image, with commonly used techniques including YOLO (You Only Look Once) and SSD (Single Shot MultiBox Detector). Image segmentation aims to divide an image into different regions for a better understanding of its structure, while pose estimation focuses

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

on detecting key points of a person or object to help understand their posture and movements.

In virtual conference assistants, computer vision technology is particularly crucial. It enables gesture recognition and facial expression analysis by analyzing users' image inputs in real time, enhancing user interaction experiences. The efficiency of virtual meetings relies not only on stable network connections and clear audio-video quality but also on the application of computer vision technology, which allows users to perform operations through natural gestures, such as turning pages, adjusting volume, and sharing screens, making meetings more fluid. Deep learning, especially Convolutional Neural Networks (CNN), performs exceptionally well in the field of image processing. CNNs can automatically learn features without manual extraction, reducing development complexity. By training on large-scale annotated datasets, CNN models can efficiently handle various image tasks. Moreover, the rise of transfer learning has made efficient learning on small datasets possible, further promoting the widespread use of computer vision technology.

2.2 Gesture Recognition Technology

Gesture recognition technology is an important application in the field of computer vision, aiming to understand user intentions by analyzing hand movements. This technology not only enhances the naturalness of human-computer interaction but also provides innovative solutions for scenarios such as virtual meetings, augmented reality, and smart homes. The implementation process of gesture recognition typically includes several key steps: data acquisition, preprocessing, feature extraction, and classification recognition. Modern smart devices are generally equipped with high-definition cameras capable of capturing subtle changes in gestures. Additionally, depth cameras (such as Kinect) and infrared sensors are widely used, providing richer spatial information to improve gesture recognition accuracy.

The acquired image data usually requires preprocessing to remove noise and enhance image quality. Preprocessing steps include graying, binarization, denoising, and image scaling. This process helps to highlight key features of the gestures, laying the groundwork for subsequent feature extraction. The purpose of feature extraction is to extract crucial visual features from gesture images for classification and recognition. Common feature extraction methods include shape features, color features, and texture features. In recent years, deep learning, particularly Convolutional Neural Networks (CNN), has become the mainstream method for feature extraction. CNNs can automatically learn high-level feature representations, significantly improving gesture recognition performance. Common classification algorithms include Support Vector Machines (SVM), Random Forests, and Neural Networks. In recent years, deep learning models, such as Long Short-Term Memory (LSTM) networks and Temporal Convolutional Networks (TCN), have also been widely applied in dynamic gesture recognition, effectively handling time series data.

2.3 Deep Learning

Deep learning is a subfield of machine learning that employs multilayer neural networks to simulate the way the human brain processes information, enabling efficient analysis and learning of complex data. In recent years, the rapid development of deep learning has led to significant achievements in various fields, including computer vision, natural language processing, and speech recognition, particularly excelling in image and video processing tasks, thus driving innovation and application in technologies such as gesture recognition.

Convolutional Neural Networks (CNN): Widely used in image processing tasks, CNNs effectively reduce the number of parameters and improve training efficiency and performance through local connections, weight sharing, and pooling operations. They perform exceptionally well in tasks such as image classification, object detection, and image segmentation.

Recurrent Neural Networks (RNN): Suitable for processing sequential data, such as text and time series. RNNs allow information to be transmitted over time steps through recurrent connections, capturing temporal features of the data. However, traditional RNNs face the problem of gradient vanishing when dealing with long sequences.

Long Short-Term Memory Networks (LSTM): An improvement over RNNs, LSTMs effectively address long-distance dependency issues. By introducing gating mechanisms, LSTMs control the forgetting and retention of information, thus performing excellently in many natural language processing tasks.

Generative Adversarial Networks (GAN): A deep learning framework used to generate new data, consisting of a generator and a discriminator. The generator attempts to create realistic data, while the discriminator assesses the similarity between the generated data and real data. GANs have widespread applications in image generation, style transfer, and more.

2.4 Human-Computer Interaction

Human-Computer Interaction (HCI) is the field that studies the interaction between humans and computer systems, aiming to improve user experience and system usability. With the rapid development of computer technology, the scope of HCI research has continually expanded, evolving from initial command-line interfaces to today's graphical user interfaces (GUIs), touch interfaces, voice interactions, and gesture-based interaction methods. The rise of gesture recognition technology has provided a new interaction mode for HCI, making interactions between humans and computers more natural, intuitive, and efficient.

Gesture recognition technology has brought revolutionary changes to human-computer interaction,

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

offering a more natural mode of interaction. Users can control computers through gestures instead of relying on a mouse and keyboard. They can interact through gestures to turn pages, mute, or adjust volume, enhancing the fluidity and engagement of meetings. In augmented reality (AR) and virtual reality (VR) environments, gesture recognition technology allows users to interact directly with virtual objects, making immersive experiences feel more authentic. Users can control appliances through gestures, such as turning lights on and off or adjusting the temperature, thereby enhancing the convenience and intelligence of everyday life.

Recent advancements in artificial intelligence (AI), particularly in Large Language Models (LLMs), have greatly contributed to advancements in various sectors. For applications in virtual meeting assistants, detecting AI-generated content is critical to maintaining the integrity of interactions. Mo et al. (2024) introduced a transformer-based deep learning algorithm to detect AI-generated text, which could ensure more reliable communication in such virtual environments[1]. This detection mechanism is essential in maintaining the authenticity of interactions when virtual meeting assistants generate or summarize content based on user input.

In governance, LLMs have proven to improve operational efficiency and enhance public trust. Qin and Li (2024) discussed how AI, particularly LLMs, streamline decision-making processes in government while reducing bureaucratic inefficiencies[2]. This highlights the potential of AI-enhanced virtual meeting systems to promote faster and more transparent interactions in administrative and corporate settings.

In the healthcare domain, LLMs provide precision-driven support. Qin and Li (2024) explored how domain-specific language models assist medical professionals in processing clinical data during telemedicine, which could also extend to virtual healthcare meetings by facilitating accurate real-time data interpretation[3]. For gesture-based virtual meeting assistants, this technology would improve communication by enabling more natural, gesture-driven interactions between doctors and patients.

Additionally, high-precision measurement technologies using AI, as described by Qu (2024), are vital for enhancing motion tracking accuracy, particularly in gesture recognition systems[4]. For virtual meeting assistants relying on real-time hand gesture tracking, these advancements are crucial for improving user experience and system performance.

Wang (2016) emphasized the role of AI in the development of information science and electronic engineering, stressing how machine learning innovations contribute to the growth of systems like virtual assistants[5]. This underlines the importance of AI-driven solutions in enabling seamless human-computer interaction in virtual meetings. Wang et al. (2024) proposed an improved UNet model for brain tumor segmentation using ASPP-coordinate attention mechanisms, which could be adapted to improve gesture recognition by accurately detecting and responding to complex hand movements in real-time[6]. Furthermore, Wu (2024) introduced the MPGAAN model, which excels in heterogeneous network classification, offering insights into how AI can classify gesture-based inputs in virtual meeting environments with complex user commands[7]. In the realm of real-time interaction, temporal awareness is key. Chen et al. (2024) introduced the TemporalMed model, designed to handle time-aware responses in large language models, which is highly applicable to virtual meeting systems that must process continuous inputs such as hand gestures over time[8].

Moreover, in ensuring reliability during interactions, hallucination detection—as described by Chen et al. (2023)—can be integrated into virtual assistants to prevent errors in gesture recognition and interpretation, ensuring smoother meeting experiences[9]. Given the complexity of real-time gesture recognition, incorporating such a detection system would be crucial to avoiding misinterpretations.

Humor response systems also offer new dimensions to virtual meeting assistants. Chen et al. (2024) developed a large-scale humor response dataset, which enables virtual assistants to interpret human emotions and reactions in meetings more effectively[10]. Similarly, named entity recognition (NER) models, as explored by Guo and Roth (2021), could enhance virtual assistant capabilities by improving understanding of specific gesture commands in low-resource environments[11].

Finally, meta-tuning LLMs for better lexical understanding, as discussed by Guo, Xu, and Ritter (2024), could extend gesture recognition systems to handle more diverse language and communication styles, further enhancing the functionality of virtual assistants in multicultural and multilingual environments[12].

3 3. LITERATURE REVIEW

3.1 Computer Vision (CV)

As a significant branch of artificial intelligence, computer vision has made remarkable progress in recent years and is widely applied across various fields, including healthcare, agriculture, and environmental monitoring. Zhao et al. (2024) systematically explored the applications of Convolutional Neural Networks (CNNs) in computer vision in their review, emphasizing the critical role of deep learning methods in tasks such as image classification, object detection, and image generation[13]. This provides essential background for understanding current computer vision technologies.

In the agricultural sector, Ghazal et al. (2024) investigated the application of computer vision in smart agriculture and precision farming, discussing technologies and applications in crop monitoring, pest and disease recognition, and resource optimization. Their research indicates that CV technologies can significantly enhance

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

agricultural productivity and promote sustainable development[14].

The applications in the healthcare domain are also continually expanding. Khang et al. (2024) examined the potential of computer vision within the medical ecosystem, particularly in disease detection, imaging analysis, and surgical assistance[15]. Schmidt et al. (2024) further reviewed tracking and mapping technologies in medical computer vision, highlighting the importance and challenges in medical image analysis[16].

Simultaneously, ethical considerations surrounding computer vision technologies are gaining increasing attention. Waelen (2024) explored ethical issues related to computer vision, particularly in terms of power distribution and social impact, urging researchers to consider their social responsibilities when advancing technological applications[17].

Goldblum et al. (2024) conducted a large-scale comparative study evaluating the performance of various pre-trained models across different computer vision tasks. This study provides empirical evidence for model selection, facilitating deeper understanding and optimization of deep learning models[18]. Additionally, the research by Shedrawi et al. (2024) demonstrated the practical utility of deep learning and computer vision technologies in managing coastal fisheries in the Pacific region, showcasing the application of CV in environmental protection and resource management[19].

3.2 Virtual Meeting Assistant

Virtual assistants have gradually become an important support tool for users in education, work, and daily life. Recent studies indicate that virtual assistants not only enhance user satisfaction and efficiency but also promote in-depth discussions on human-computer collaboration. Öncü and Süral (2024) investigated the use of virtual assistants in open educational environments, exploring user satisfaction levels. The results showed that virtual assistants play a significant role in providing personalized support and optimizing learning experiences, particularly in the context of remote education, where user satisfaction was significantly improved. This research provides valuable insights into understanding the application of virtual assistants in the education sector[20].

In daily life, Pradhan et al. (2024) proposed the concept of the "Nova Virtual Assistant," emphasizing the importance of user needs, preferences, and expectations. Through the analysis of user feedback, the researchers found that the design of virtual assistants should focus more on personalized user needs to enhance their practical effectiveness. This study offers a practical framework for developing virtual assistants that align with user expectations[21].

Katic et al. (2024) conducted a performance analysis of virtual assistants in the Croatian language environment, demonstrating the impact of language on virtual assistant interaction effectiveness. The study indicated that linguistic accuracy and cultural adaptability are key factors in enhancing user experience, providing a reference for the development of virtual assistants in other linguistic contexts[22]. In meeting and collaboration scenarios, Kreamer et al. (2024) explored how individual differences affect user participation in chats and verbal interactions during virtual meetings. This research reveals the potential of virtual assistants in promoting engagement and communication efficiency, emphasizing that design should consider the diverse needs of users to achieve better interaction outcomes[23]. Cabrero-Daniel et al. (2024) focused on the application of human-AI collaboration in agile development, proposing a customized LLM (large language model) meeting assistant. The study demonstrated that a tailored AI assistant could effectively improve meeting efficiency and engagement, providing better support for team collaboration[24].

3.3 Gesture Recognition

The application of gesture recognition technology is increasingly widespread in the field of computer vision, involving multiple data modalities and deep learning methods. Recent studies indicate that gesture recognition plays an important role not only in multimedia applications but also shows great potential in fields such as healthcare, education, and human-computer interaction. Shin et al. (2024) conducted a systematic and structural evaluation of gesture recognition methods, exploring gesture recognition technologies across different data modalities. This study provides significant insights into the current state and future development directions of the gesture recognition field[25].

Kapitanov et al. (2024) introduced a gesture recognition image dataset named HaGRID, aimed at providing researchers with high-quality training data to improve the performance of gesture recognition models. The release of this dataset contributes to advancing research in the field, particularly in dynamic gesture recognition[26].

Rastgoo et al. (2024) explored multimodal zero-shot dynamic gesture recognition, demonstrating how to combine various data sources (such as video, depth images, etc.) to enhance recognition accuracy. This research highlights the advantages of multimodal learning in tackling complex gesture recognition tasks[27]. Rahim et al. (2024) proposed a deep learning-based three-stream hybrid model for dynamic gesture recognition, further promoting technological advancements in gesture recognition. This model effectively enhances the recognition capabilities of dynamic gestures by combining multiple input streams[28].

Mohammadi et al. (2024) developed a diverse gesture recognition dataset focusing on different types of gestures, providing researchers with a rich data resource for model training and evaluation[29]. Eddy et al. (2024) examined the application of big data in electromyographic (EMG) control,

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Computer Technology and Applied Mathematics
Journal Home: http://jctam.suaspress.org | CODEN: JCTAAD
Vol. 1, No. 4, 2024 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

proposing a zero-shot EMG discrete gesture recognition method based on a multi-user model. This research illustrates the potential of integrating biosignal processing with gesture recognition, especially in control and human-computer interaction[30].

Karsh et al. (2024) introduced a modified Inception V3 network (mIV3Net) for gesture recognition, demonstrating the optimized application of convolutional neural networks for specific tasks[31]. Miah et al. (2024) explored gesture recognition in multicultural sign languages, employing a combination of graphical and deep learning networks. This research emphasizes the value of gesture recognition in promoting cultural diversity and inclusivity[32].

3.4 Literature Review

Research on gesture recognition technology in the field of computer vision is continually deepening, with the literature showcasing the effectiveness of diverse data modalities and deep learning methods in various application scenarios. Shin et al. (2024) provided a systematic evaluation of gesture recognition, offering researchers an important framework and highlighting the critical role of multimodal data in improving recognition accuracy. This methodological review lays the groundwork for subsequent research, enabling researchers to make more effective comparisons between different data sources.

In terms of dataset construction, the HaGRID dataset proposed by Kapitanov et al. (2024) provides researchers with rich training resources, particularly in the field of dynamic gesture recognition. The release of such datasets not only promotes the development of gesture recognition models but also advances empirical research on gesture recognition performance. Additionally, the diverse gesture recognition dataset designed by Mohammadi et al. (2024) underscores the importance of introducing diversity in model training, assisting researchers in enhancing model generalization capabilities in practical applications.

At the same time, the multimodal zero-shot dynamic gesture recognition study by Rastgoo et al. (2024) demonstrates how to combine various data sources to improve model robustness, thereby expanding the application scenarios of traditional gesture recognition. Furthermore, the three-stream hybrid model proposed by Rahim et al. (2024) showcases the advancements and application potential of deep learning in dynamic gesture recognition. These studies not only improve recognition accuracy but also provide new perspectives for understanding complex gestures.

In terms of technical optimization, Karsh et al. (2024) introduced the mIV3Net model, which further advances the application of convolutional neural networks in gesture recognition. This innovative approach reflects the flexibility and adaptability of deep learning models, allowing for optimization tailored to specific tasks to meet diverse user needs.

Moreover, Miah et al. (2024) emphasized the potential of gesture recognition technology in promoting cultural exchange and inclusivity through their research on multicultural sign languages, demonstrating the social value of gesture recognition. The literature provides a comprehensive view of the gesture recognition field, covering aspects from dataset construction to technical optimization, and from multimodal learning to cultural adaptation, all highlighting the diversity and complexity of this area. Future research can further explore the integration of these technologies in practical applications, focusing on their performance in different cultural and linguistic contexts to promote the broader application and development of gesture recognition technology.

4 CONCLUSION

The rapid development of gesture recognition technology in the field of computer vision highlights its importance across various application scenarios. Through the analysis of existing literature, it is evident that researchers have made significant advancements in dataset construction, deep learning model optimization, and multimodal learning. These studies not only enhance the accuracy and robustness of gesture recognition but also broaden its application potential in fields such as education, healthcare, and human-computer interaction.

Particularly, research focused on dynamic gestures and multicultural sign languages demonstrates the social value of gesture recognition technology in facilitating cultural exchange and enhancing human-computer interaction. Future research should continue to emphasize the integration of technology and application, exploring ways to optimize gesture recognition systems within different cultural contexts to achieve broader social benefits. Additionally, as technology continues to advance, the integration of new sensors and data collection methods will further enhance the accuracy and practicality of gesture recognition. Ongoing research in this field will provide stronger support for intelligent interaction and human-computer collaboration, driving continuous innovation and development of related technologies.

As a significant application of computer vision, gesture recognition showcases immense potential and value. In education, healthcare, and daily life, gestures serve as a natural and intuitive means of interaction, significantly improving user experience and work efficiency. By combining deep learning with multimodal data, recognition accuracy can be notably enhanced, especially when processing dynamic gestures. This technology not only meets personalized user needs but also fosters communication between different cultures. Research on multicultural sign languages highlights the importance of cultural adaptability in gesture recognition, helping to break down language barriers and enhance social inclusivity. Therefore, future

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

research should continue to explore how gesture recognition technology can be integrated with various application scenarios to achieve wider societal benefits.

Despite the significant advancements made thus far, the accuracy and robustness of gesture recognition in complex environments still require improvement. Future studies should focus on the diversity of datasets and the optimization of deep learning models to address gesture variations in different contexts. The integration of novel sensor technologies, such as combining gesture recognition with electromyographic (EMG) control, may open up new application areas. In smart homes and assistive devices, gesture recognition technology can provide more convenient and intuitive control methods, further enhancing user experience.

In summary, gesture recognition is not only a technological innovation but also a vital force driving social progress. We should actively monitor developments in this field and look forward to the possibilities and value it can bring us in the future.

ACKNOWLEDGMENTS

The authors thank the editor and anonymous reviewers for their helpful comments and valuable suggestions.

FUNDING

Not applicable.

INSTITUTIONAL REVIEW BOARD STATEMENT

Not applicable.

INFORMED CONSENT STATEMENT

Not applicable.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

CONFLICT OF INTEREST

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

PUBLISHER'S NOTE

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

AUTHOR CONTRIBUTIONS

Not applicable.

ABOUT THE AUTHORS

LIN, Weikun

Shandong University of Science and Technology, China.

REFERENCES

[1] Mo, Y. ., Qin, H., Dong, Y., Zhu, Z., & Li, Z. (2024). Large Language Model (LLM) AI Text Generation Detection based on Transformer Deep Learning Algorithm. International Journal of Engineering and Management Research, 14(2), 154–159. https://doi.org/10.5281/zenodo.11124440

[2] Hao Qin, & Zhi Li. (2024). A Study on Enhancing Government Efficiency and Public Trust: The Transformative Role of Artificial Intelligence and Large Language Models. International Journal of Engineering and Management Research, 14(3), 57–61. https://doi.org/10.5281/zenodo.12619360

[3] Hao Qin, & Li, Z. (2024). Precision in Practice: Enhancing Healthcare with Domain-Specific Language Models. Applied Science and Engineering Journal for Advanced Research, 3(4), 28–33. https://doi.org/10.5281/zenodo.13253336

[4] Qu, M. (2024). High Precision Measurement Technology of Geometric Parameters Based on Binocular Stereo Vision Application and Development Prospect of The System in Metrology and Detection. Journal of Computer Technology and Applied Mathematics, 1(3), 23–29. https://doi.org/10.5281/zenodo.13366612

[5] Wang, D. (Ed.). (2016). Information Science and Electronic Engineering: Proceedings of the 3rd International Conference of Electronic Engineering and Information Science (ICEEIS 2016), January 4-5, 2016, Harbin, China. CRC Press.

[6] Wang, Z., Chen, Y., Wang, F., & Bao, Q. (2024). Improved Unet model for brain tumor image segmentation based on ASPP-coordinate attention mechanism. arXiv preprint arXiv:2409.08588.

[7] Wu, Z. (2024). MPGAAN: Effective and Efficient Heterogeneous Information Network Classification.

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Computer Science and Technology Studies, 6(4), 08-16

[8] Y. Chen, J. Zhao, Z. Wen, Z. Li, and Y. Xiao, Temporalmed: Advancing medical dialogues with time-aware responses in large language models, in Proceedings of the 17th ACM International Conference on Web Search and Data Mining, 2024, pp. 116–124.

[9] Y. Chen, Q. Fu, Y. Yuan, Z. Wen, G. Fan, D. Liu, D. Zhang, Z. Li, and Y. Xiao, Hallucination detection: Robustly discerning reliable answers in large language models, in Proceedings of the 32nd ACM International Conference on Information and Knowledge Management, 2023, pp. 245–255.

[10] Yuyan Chen, Yichen Yuan, Panjun Liu, Dayiheng Liu, Qinghao Guan, Mengfei Guo, Haiming Peng, Bang Liu, Zhixu Li, and Yanghua Xiao. 2024e. Talk funny! a large-scale humor response dataset with chain-ofhumor interpretation. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 38, pages 17826–17834.

[11]Guo, R., & Roth, D. (2021, August). Constrained labeled data generation for low-resource named entity recognition. In Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021 (pp. 4519-4533).

[12] Guo, R., Xu, W., & Ritter, A. (2024, August). Meta-Tuning LLMs to Leverage Lexical Knowledge for Generalizable Language Style Understanding. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 13708-13731).

[13] Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., & Parmar, M. (2024). A review of convolutional neural networks in computer vision. Artificial Intelligence Review, 57(4), 99.

[14] Ghazal, S., Munir, A., & Qureshi, W. S. (2024). Computer vision in smart agriculture and precision farming: Techniques and applications. Artificial Intelligence in Agriculture.

[15] Khang, A., Abdullayev, V., Litvinova, E., Chumachenko, S., Alyar, A. V., & Anh, P. T. N. (2024). Application of Computer Vision (CV) in the Healthcare Ecosystem. In Computer Vision and AI-Integrated IoT Technologies in the Medical Ecosystem (pp. 1-16). CRC Press.

[16] Schmidt, A., Mohareri, O., DiMaio, S., Yip, M. C., & Salcudean, S. E. (2024). Tracking and mapping in medical computer vision: A review. Medical Image Analysis, 103131.

[17] Waelen, R. A. (2024). The ethics of computer vision: an overview in terms of power. AI and Ethics, 4(2), 353-362.

[18] Goldblum, M., Souri, H., Ni, R., Shu, M., Prabhu, V., Somepalli, G., ... & Goldstein, T. (2024). Battle of the backbones: A large-scale comparison of pretrained models across computer vision tasks. Advances in Neural Information Processing Systems, 36.

[19] Shedrawi, G., Magron, F., Vigga, B., Bosserelle, P., Gislard, S., Halford, A. R., ... & Andrew, N. L. (2024). Leveraging deep learning and computer vision technologies to enhance management of coastal fisheries in the Pacific region. Scientific Reports, 14(1), 20915.

[20] Öncü, S. E., & Süral, İ. (2024). Leveraging AI for Enhanced Support: Satisfaction Levels of Users Utilizing Virtual Assistant in Open Education. Asian Journal of Distance Education, 19(1).

[21] Pradhan, R., Jain, D., Agrawal, U., Sharma, M., Singh, D., & Sharma, D. K. (2024, March). Nova Virtual Assistant to Enhance Daily Life: On Perspective of User Needs, Preferences, and Expectations. In 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO) (pp. 1-5). IEEE.

[22] Katic, M. A., Miletic, E., & Candrlic, S. (2024). PERFORMANCE ANALYSIS OF VIRTUAL ASSISTANTS-A CASE STUDY BASED ON THE CROATIAN LANGUAGE. In INTED2024 Proceedings (pp. 7305-7315). IATED.

[23] Kreamer, L. M., Rogelberg, S. G., Tankelevitch, L., & Rintel, S. (2024). Virtual voices: Exploring individual differences in chat and verbal participation in virtual meetings. Journal of Vocational Behavior, 152, 104015.

[24] Cabrero-Daniel, B., Herda, T., Pichler, V., & Eder, M. (2024, May). Exploring Human-AI Collaboration in Agile: Customised LLM Meeting Assistants. In International Conference on Agile Software Development (pp. 163-178). Cham: Springer Nature Switzerland.

[25] Shin, J., Miah, A. S. M., Kabir, M. H., Rahim, M. A., & Al Shiam, A. (2024). A Methodological and Structural Review of Hand Gesture Recognition Across Diverse Data Modalities. IEEE Access.

[26] Kapitanov, A., Kvanchiani, K., Nagaev, A., Kraynov, R., & Makhliarchuk, A. (2024). HaGRID--HAnd Gesture Recognition Image Dataset. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 4572-4581).

[27] Rastgoo R, Kiani K, Escalera S, et al. Multi-modal zero-shot dynamic hand gesture recognition[J]. Expert Systems with Applications, 2024, 247: 123349.

[28] Rahim, M. A., Miah, A. S. M., Akash, H. S., Shin, J., Hossain, M. I., & Hossain, M. N. (2024). An advanced deep learning based three-stream hybrid model for dynamic hand gesture recognition. arXiv preprint arXiv:2408.08035.

[29] Mohammadi, Z., Akhavanpour, A., Rastgoo, R., & Sabokrou, M. (2024). Diverse hand gesture recognition dataset. Multimedia Tools and Applications, 83(17), 50245-50267.

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.

Journal of Computer Technology and Applied Mathematics
Journal Home: http://jctam.suaspress.org | CODEN: JCTAAD
Vol. 1, No. 4, 2024 | ISSN 3007-4126 (Print) | ISSN 3007-4134 (Online) | ISSN 3007-4142 (Digital)

[30] Eddy, E., Campbell, E., Bateman, S., & Scheme, E. (2024). Big data in myoelectric control: Large multi-user models enable robust zero-shot emg-based discrete gesture recognition. Frontiers in Bioengineering and Biotechnology, 12, 1463377.

[31] Karsh, B., Laskar, R. H., & Karsh, R. K. (2024). mIV3Net: modified inception V3 network for hand gesture recognition. Multimedia Tools and Applications, 83(4), 10587-10613.

[32] Miah, A. S. M., Hasan, M. A. M., Tomioka, Y., & Shin, J. (2024). Hand gesture recognition for multi-culture sign language using graph and general deep learning network. IEEE Open Journal of the Computer Society.

[33] Luo, M., Zhang, W., Song, T., Li, K., Zhu, H., Du, B., & Wen, H. (2021, January). Rebalancing expanding EV sharing systems with deep reinforcement learning. In Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence (pp. 1338-1344).

[34] Luo, M., Du, B., Zhang, W., Song, T., Li, K., Zhu, H., ... & Wen, H. (2023). Fleet rebalancing for expanding shared e-Mobility systems: A multi-agent deep reinforcement learning approach. IEEE Transactions on Intelligent Transportation Systems, 24(4), 3868-3881.

[35] Fan, H., Li, K., Li, X., Song, T., Zhang, W., Shi, Y., & Du, B. (2019). CoVSCode: a novel real-time collaborative programming environment for lightweight IDE. Applied Sciences, 9(21), 4642.

[36] Tao Y. SQBA: sequential query-based blackbox attack, Fifth International Conference on Artificial Intelligence and Computer Science (AICS 2023). SPIE, 2023, 12803: 721-729.

[37] Tao Y. Meta Learning Enabled Adversarial Defense, 2023 IEEE International Conference on Sensors, Electronics and Computer Engineering (ICSECE). IEEE, 2023: 1326-1330.

[38] Yiyi Tao, Yiling Jia, Nan Wang, and Hongning Wang. 2019. The FacT: Taming Latent Factor Models for Explainability with Factorization Trees. In Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR'19). Association for Computing Machinery, New York, NY, USA, 295–304.

[39] Yiyi Tao, Zhuoyue Wang, Hang Zhang, Lun Wang. 2024. NEVLP: Noise-Robust Framework for Efficient Vision-Language Pre-training. arXiv:2409.09582.

[40] Sun, Y., & Ortiz, J. (2024). Machine Learning-Driven Pedestrian Recognition and Behavior Prediction for Enhancing Public Safety in Smart Cities. Journal of Artificial Intelligence and Information, 1, 51-57.

Copyright © 2024 The author retains copyright and grants the journal the right of first publication.
This work is licensed under a Creative Commons Attribution 4.0 International License.