Review Paper

Multimodal Fusion Algorithm for Generating Transmedia Art Installations: A Review

Ewelina Kowal[1,*], Joanna Tomaszewski[2]

[1] Faculty of Electrical Engineering, Automatics, Computer Science and Biomedical Engineering, University of Silesia in Katowice, Katowice 40-007, Poland
[2] Faculty of Electrical and Computer Engineering, Wrocaw University of Science and Technology, Wrocaw 50-370, Poland
*Corresponding author: kowal@wpias.edu.pl ORCID: https://orcid.org/0009-0008-9250-659X

Abstract. With the rapid development of artificial intelligence and digital art, multimodal fusion algorithms bring new opportunities for the creation of cross-media art installations. This paper summarizes the research on the application of multimodal fusion algorithms in the generation of cross-media art installations, firstly, it introduces the types and characteristics of multimodal data, as well as common multimodal fusion algorithms, including the fusion algorithms based on the feature level, decision level and hybrid level. Then the characteristics and creative needs of cross-media art installations are elaborated, and the key role and special requirements of multimodal fusion in them are analyzed. Then we discuss in detail the application of multimodal fusion algorithms in the generation of different types of cross-media art installations, including visual and auditory, text and other modalities, and multimodal interactive art installations. Finally, we summarize the advantages and limitations of multimodal fusion algorithms in generating cross-media art installations, and look forward to the future development direction. The study shows that multimodal fusion algorithms can provide richer contents and forms for the creation of cross-media art installations, but at the same time, they also face the challenges of data processing and algorithm optimization.

Keywords: *Multimodal fusion algorithms; cross-media art installations; feature-level fusion; decision-level fusion; hybrid-level fusion*

Introduction

In the contemporary era of rapid development of science and technology, the integration of art and technology has become an important trend of innovation in the field of art [1]. As a product of the combination of art and technology, cross-media art installation integrates multiple media forms, breaks the traditional art boundaries, and brings a new aesthetic experience to the audience [2]. The rise of multimodal fusion algorithms provides strong technical support for the creation of cross-media art installations, enabling creators to more accurately integrate and process different modal information, and realizing the innovation and expansion of artistic expression [3].

The application of multimodal fusion algorithm in the field of cross-media art installation has important practical significance and potential value [4]. From the perspective of artistic creation, it expands the boundaries of artistic creation, enriches the forms and contents of artistic expression, stimulates the creative inspiration of artists, and generates more novel and unique art works [5]. From the perspective of audience experience, the integration of a variety of modal information art installations can stimulate the audience's senses in an all-round way, providing immersive, interactive experience, enabling the audience to better understand the connotation and emotional expression of the art works, and enhancing the artistic influence and impact [6]. From the perspective of cultural inheritance and development, this technology helps to present traditional cultural

elements innovatively, attract more people to pay attention to and participate in cultural construction, and promote cultural inheritance and development [7].

Multimodal fusion algorithms are widely used in cross-media art installations. In visual and auditory fusion devices, sound signals are converted into visual images or animations by algorithms to realize audio-visual synchronization [8]; in text and other modal fusion devices, semantic information of text is converted into images, sounds and other forms to realize multimodal narrative [9]; in multimodal interactive devices, corresponding feedback is generated based on information of the audience's interactive behaviors, which enhances the interactivity and interestingness [10].

The contribution of this paper is mainly reflected in the systematic sorting out and in-depth analysis of multimodal fusion algorithms [11] in cross-media art installations [12] application research. The overall structure of this paper is organized as follows. First, it provides a theoretical foundation by systematically reviewing the characteristics and types of multimodal data, and by summarizing common multimodal fusion algorithms, including feature-level, decision-level, and hybrid-level methods. This section lays the groundwork for understanding the technical principles behind cross-media art installations. Second, the study explores the definition, categorization, and creative requirements of cross-media art installations, analyzing the crucial role of multimodal fusion in enhancing artistic expression, enriching narrative structures, and improving audience interactivity. Building on this, the paper presents detailed application analyses across different categories of installations, such as visual–auditory fusion, text–modality integration, and multimodal interactive devices. Case studies are used to illustrate how fusion algorithms operate in practice and generate innovative artistic effects. Subsequently, the advantages and limitations of applying multimodal fusion algorithms in this field are critically assessed, with attention to technical complexity, data processing challenges, and the uncertainty of fusion outcomes. Finally, the paper outlines future research directions, including deeper applications of deep learning, integration of emerging computing technologies, diversification of artistic expression, enhancement of immersive interactions, and expansion of cross-sectoral applications. Through this progressive framework, the article not only synthesizes existing knowledge but also provides guidance for future innovation in art and technology integration

Multimodal Fusion Algorithm

### Types and characteristics of multimodal data

Multimodal data, which is the basis of multimodal fusion algorithms, has a rich variety of types. According to the creation perspective, image data can be divided into static pictures [13] and dynamic videos [14]. Text data is divided into original text and adapted text according to the creation source; natural language text and programming language text according to the creation language. Audio data are categorized into natural sound and synthetic sound according to the creation source.

Image data is intuitive and able to convey information and emotions through visual elements such as color, shape, and line [15]. Image data need to consider composition, light and shadow, color matching and other factors in the creation process to achieve the desired artistic effect [16]. Text data is centered on semantics and conveys thoughts and emotions through the arrangement and combination of words [17]. In cross-media art installations, textual data can be presented in a variety of forms, such as textual descriptions on the installation and interactive textual games, which can guide the audience's understanding and thinking and enrich the connotation of the art installation [18]. Audio data mainly conveys information and emotion through the frequency, amplitude and other characteristics of sound [19-20]. In art installations, audio data can be played through sound equipment, and cooperate with other modal data such as images and texts to enhance the audience's sensory experience and create an immersive art atmosphere.

Common multimodal fusion algorithms

Multimodal fusion algorithms are the key technology to realize the creation of cross-media art installations [21]. Common multimodal fusion algorithms include algorithms based on feature-level fusion [22], algorithms based on decision-level fusion [23], algorithms based on hybrid-level fusion [24], etc., which are analyzed as follows:

as illustrated in Figure 1, the three primary multimodal fusion methods—feature-level, decision-level, and hybrid-level—address different aspects of integrating multimodal data within cross-media art installations
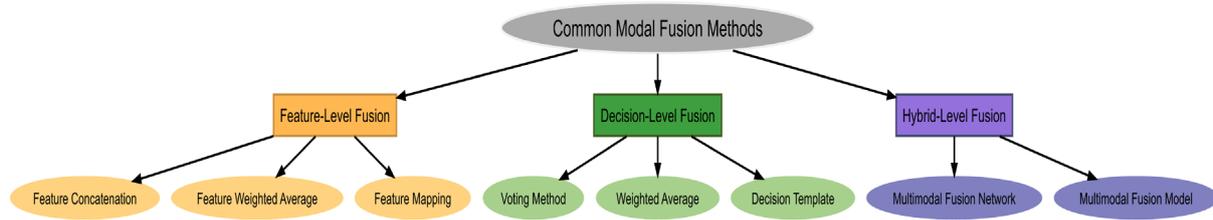
Figure 1 Common Modal Fusion Methods in Cross-Media Art Installations

Algorithms based on feature-level fusion

Algorithms based on feature-level fusion fuse data from different modalities at the feature level. The integration of multimodal information is achieved by extracting the features of each modal data and combining these features to form a fused feature representation [22]. The algorithm based on feature-level fusion is widely used in image and text fusion to generate art installations. The algorithm can fully utilize the feature information of each modality, but it also faces the challenges of large differences in feature dimensions and high dimensionality of the fused feature vectors [25]. In addition, the accuracy and effectiveness of feature extraction will also affect the fusion effect, and it is necessary to choose appropriate feature extraction methods and models. Algorithms based on feature-level fusion usually include feature concatenation [26], feature weighted averaging [27], feature mapping [28], etc., which are analyzed in Table 1.

Table 1. Comparative analysis of algorithms based on feature-level fusion

Algorithm	Advantages	Disadvantages	Applicable Scenario
Feature Concatenation	Fully utilizes features from all modalities	Large feature dimension differences High fused vector dimension	Image and text fusion art device generation
Feature Weighted Average	Highlights important features Suppresses unimportant ones	Subjective weight assignment Fusion effectiveness depends on experience	Multimodal emotion recognition art device
Feature Mapping	Maps features from different modalities to the same space for easier fusion	Difficult mapping function design High computational complexity	Visual and auditory fusion art device

Algorithms based on decision-level fusion

Algorithms based on decision-level fusion are based on processing and decision making for each modal data separately, and then fusing the decision results of each modality [23]. Commonly used fusion strategies include voting [29], weighted average [30], and decision template [31].

The voting method is a simple and direct method for decision fusion. In the task of categorizing multimodal art installations, each modality categorizes the type or theme of the installation, and then counts the number of votes for each category through the voting method and selects the category with the most votes as the final result [29]. The weighted average method synthesizes the final decision result by assigning different weights to the decision results of each modality. In the evaluation task of multimodal art installations, the results can be weighted and averaged according to the importance or reliability of each modality [30]. The decision template method utilizes a predefined decision template to match and fuse the decision results of each modality. In the control task of multimodal interactive art installations, decision templates can be designed to define the control actions corresponding to different modal decision combinations [31]. A comparative analysis of the advantages and disadvantages of algorithms based on decision-level fusion is shown in Table 2.

Table 2. Comparative analysis of algorithms based on decision-level fusion

Algorithm	Advantages	Disadvantages	Applicable Scenario
Voting Method	Simple calculation Easy to implement	May ignore inter-modality relationships Limited fusion performance	Classification tasks for multimodal art devices
Weighted Average Method	Highlights important modalities Suppresses unimportant ones	Subjective weight assignment Fusion effectiveness depends on experience	Evaluation tasks for multimodal art devices
Decision	Fully utilizes inter-modality	Difficult template design High	Control tasks for multimodal

Template Method	relationships good fusion effect	computational complexity	interactive art devices

Algorithms based on hybrid level fusion

Hybrid-level fusion combines feature-level fusion and decision-level fusion to integrate information at the feature level as well as to synthesize judgments at the decision level [24].

One implementation of feature-level and decision-level fusion is to first perform feature-level fusion on some modal data, obtain the fused features for initial decision-making, and then perform decision-level fusion of that decision result with the decision results of other modalities [32]. Multimodal fusion network is a hybrid level fusion algorithm based on deep learning. It automatically learns the association and fusion mode between multimodal data by constructing a multilayer neural network structure [33]. In the creation of intelligent cross-media art installations, the multimodal fusion network can simultaneously input multiple modal data such as images, text, audio, etc., and generate innovative and artistic installations through the automatic learning and training of the network [34]. Multimodal fusion model, on the other hand, fuses multimodal data by constructing a unified model framework [35]. In the generation of multimodal interactive art installations, the multimodal fusion model can be used to fuse multimodal information such as the audience's interactive behavioral data and the feedback data of the installation, so as to generate real-time art effects that interact with the audience [36].

According to the above analysis of the principle of the algorithm based on hybrid level fusion, its advantages and disadvantages are summarized as shown in Table 3.

Table 3. Research directions in the field of medical imaging diagnostics

Algorithm	Advantages	Disadvantages	Applicable Scenario
Feature and Decision Fusion	Combines advantages of feature and decision levels	Complex algorithm structure High computational resource demand	Complex cross-media art device generation
Multimodal Fusion Network	Automatically learns inter-modality relationships, good fusion effect	Requires lots of data and computational resources Risk of overfitting	Intelligent cross-media art device creation
Multimodal Fusion Model	Fully utilizes multimodal information Strong expressive ability	Difficult model training Poor interpretability	Generation of multimodal interactive art devices

By analyzing the characteristics and principles of the three different fusion algorithms, their algorithm complexity, fusion effect and applicable scenarios are summarized in Table 4.

Table 4. Comparative analysis of common multimodal fusion algorithms

Algorithm Type	Algorithm Complexity	Fusion Effect	Applicable Scenario
Feature-level	High	Good	Image and text fusion art device generation
Decision-level	Low	Average	Classification tasks for multimodal art devices
Hybrid-level	Highest	Best	Complex cross-media art device generation

Characteristics and Creative Needs of Transmedia Art Installations

Definition and categorization of cross-media art installations

Cross-media art installation refers to the creation of installations with unique artistic effects and interactive experiences through the combination of multiple media and technical means [37]. This kind of installation breaks the boundaries of traditional art forms, integrates visual, auditory, tactile and other sensory experiences, and aims to provide the audience with immersive art enjoyment [38].

Table 5. Classification and characteristics by media mix

Category	Characteristics	Example
Visual and Auditory Combined Device	Integrates visual and auditory elements Creates audiovisual integrated experience	Sound visualization device
Text and Other Modalities Fusion Device	Fuses text with image sound etc Achieves multimodal narrative	Historical story telling device
Interactive Multimedia Device	Emphasizes audience interaction Feedback based on audience behavior	Interactive painting device

Categorized by media combinations, cross-media art installations are divided into visual and auditory combined installations [39], text and other modal fusion installations [40], interactive multimedia installations [41], etc., with characteristics analyzed as shown in Table 5.

Categorized by display space, cross-media art installations are divided into indoor installations [42], outdoor installations [43], and so on. Indoor installations are usually displayed in indoor spaces such as art galleries, museums, and exhibition halls. These installations can make full use of indoor environmental features, such as spatial layout and light control, to create a specific artistic atmosphere. Literature [44] makes the audience feel as if they are in a virtual art world through multi-surface projection and surround sound. Outdoor installations are placed in outdoor public places, such as squares, parks and streets. Outdoor installations need to consider natural environmental factors, such as weather changes, light conditions, etc., and blend with the surrounding environment [45].

Categorized by the theme of creation, cross-media art installations are divided into narrative, abstract, and cultural installations [46]. Narrative installations are centered on telling a story or conveying a specific thematic idea. Through the integration of multiple media, the plot, characters, scenes, and other elements of the story are presented to the audience [47]. Abstract installations focus on expressing abstract concepts, emotions, or art forms. Instead of focusing on concrete narratives or images, it triggers the audience's associations and thoughts through the combination and change of elements such as colors, shapes, and sounds [48]. Cultural installations feature specific cultural elements or traditional culture. By combining traditional culture with modern technology and art forms, it provides an innovative presentation of traditional culture [49].

The key role of multimodal fusion

The key role of multimodal fusion in the creation of cross-media art installations is mainly reflected in the enhancement of artistic expression, enhancement of narrative ability, promotion of audience interaction, and improvement of the innovation and uniqueness of art installations [50], which are analyzed as follows: 1) Enhancement of artistic expression. Single-modal art works often can only communicate with the audience through one sensory channel, while multimodal fusion can stimulate the audience's multiple senses at the same time, making the art works more vivid, three-dimensional and infectious [51]; 2) Enhance the narrative ability. In narrative cross-media art installations, multimodal fusion can realize more complex and richer narrative structures and modes; 3) Promote audience interaction. Interactivity is an important feature of modern art installations, and multimodal fusion provides more possibilities and modes of interaction between viewers and installations [52]; 4) Enhance the innovation and uniqueness of art installations. Multimodal fusion provides artists with a broader creative space and more means of creative expression, enabling them to break through the limitations of traditional art forms and create innovative and unique art installations [53].

Special requirements for the creation of cross-media art installations

The special requirements for multimodal fusion algorithms in the creation of cross-media art installations mainly include efficient data processing capability, powerful real-time interaction capability, high-quality fusion effect, good scalability and customizability, etc. [54], and the specific analytical counterparts are shown in Table 6.

Table 6. Special requirements for the creation of cross-media art installations

Requirement Aspect	Specific Description	Related Algorithm or Technology
Efficient Data Processing	Quickly process large amounts of multimodal data Unify heterogeneous data	Data preprocessing algorithms Heterogeneous data conversion
Strong Real-time Interaction	Adjust fusion strategy in real time according to audience input feedback	Adaptive algorithms Real-time data processing technology
High Quality Fusion Effect	Coordinate all modality information Integrate naturally Meet artistic aesthetics	Deep learning multimodal fusion networks Attention mechanism fusion methods
Good Scalability and Customization	Flexible adjustment of modules and parameters Adapt to different needs	Modular algorithm design Software integration technology

Application Analysis

Transmedia Art Installation Integrating Vision and Hearing

Transmedia art installations that integrate visual and auditory senses are mainly based on the synchronized processing mechanism of the human brain for visual and auditory information [55]. When people perceive their surroundings, visual and auditory information are simultaneously received and integrated by the brain to form a complete understanding of the event or scene.

Cross-media art installation applications that merge vision and hearing include sound visualization installations [56], and immersive audiovisual theater installations [57]. Sound visualization installations use sound signal processing techniques and computer graphics to convert music or environmental sounds into dynamic visual patterns or animations. Figure 2 illustrates the multimodal interaction model used in visual and auditory fusion devices, highlighting the synchronized processing of visual and auditory signals to achieve dynamic artistic effects Literature [58] analyzes sound signals in the frequency domain through algorithms such as Fourier transform to extract features such as frequency and amplitude of the sound, and then controls the shape, color, size, and motion trajectory of the visual elements based on these features. The immersive audio-visual theater installation combines projection mapping technology and a three-dimensional sound system to create an immersive theater experience for the audience. During the performance, the actors' performances are precisely synchronized with the background music and sound effects through a multimodal fusion algorithm [59].
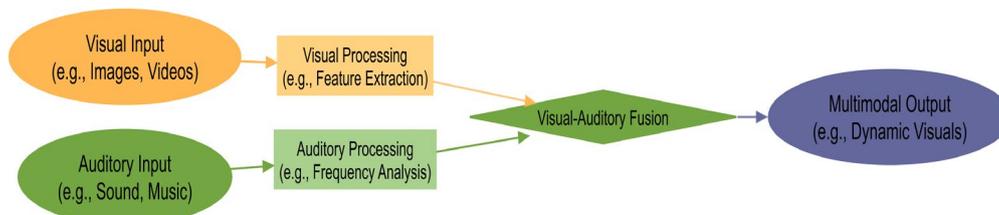

Figure 2 Multimodal Interaction Model in Visual and Auditory Fusion Devices

Transmedia Art Installation for the Fusion of Text and Other Modalities

Cross-media art installations that integrate text with other modalities are mainly based on semantic associations and narrative logic [60]. Text, as an information-intensive modality, can convey rich story content, ideological views and emotional expressions [61]. Through multimodal fusion algorithms, text is fused with other modalities such as image, sound, video, etc., which can present the information embedded in the text in a more vivid, intuitive and varied way, and at the same time provide semantic guidance and interpretation for the other modalities to enhance the narrativity and comprehensibility of the whole art installation.

Cross-media art installation applications where text is fused with other modalities include narrative multimedia installations [62] and interactive literary installations [63]. Narrative multimedia installations are often used to present historical events, cultural stories, or the context of artworks [62]. Multimodal fusion algorithms can automatically select and switch the corresponding visual and auditory materials according to the keywords or themes of the textual content to realize the coherence and logic of multimodal narratives. As shown in Figure 3, the semantic-driven fusion framework enables text to interact with other modalities, such as images and sounds, by extracting and analyzing semantic information, which enhances narrative logic and multimodal storytelling in the application of interactive literary installations, the audience can input the text through the touch screen or voice interaction, etc., and the installation generates corresponding visual and auditory feedback based on the audience's input [63].
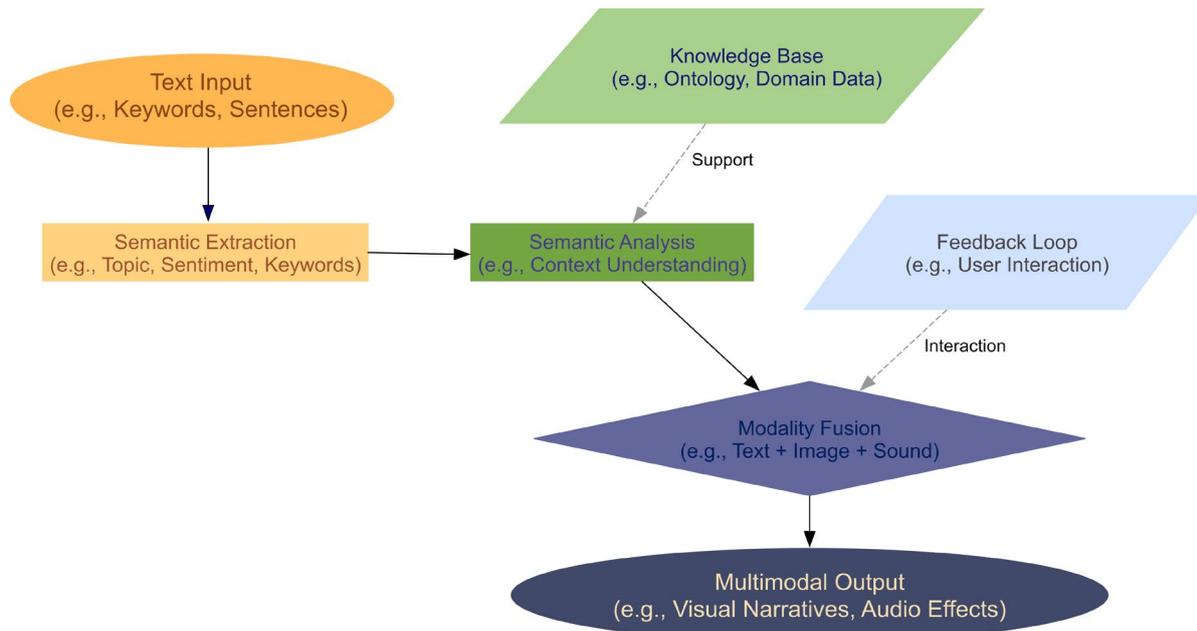
Figure 3   Semantic-Driven Framework for Text and Other Modalities Fusion

Multimodal Interactive Transmedia Art Installation

Multimodal interactive cross-media art installations are mainly based on human-computer interaction technology and multimodal fusion algorithms, which interact with the audience by capturing multimodal input information such as the audience's movements, speech, expressions, gestures, etc., processing and generating corresponding feedback in real time [64]. This kind of installation incorporates the audience into the process of art creation, making the audience a part of the artwork, thus creating a dynamic and personalized art experience.

Multimodal interactive cross-media art installation applications include interactive dance installations [65], emotional interactive installations [66], and so on. Interactive dance installations are used in stage performances or art exhibitions by installing cameras, sensors, and other devices on the stage to capture the audience's movements and dance postures, and then utilizing multimodal fusion algorithms to fuse the audience's movements with preset dance music and lighting effects [67]. Emotional interaction devices use emotion recognition technology to analyze the audience's emotional state by capturing their facial expressions through cameras and their voice tones through microphones, and then generate appropriate feedback based on the audience's emotions. A comparison of the application of multimodal fusion in different cross-media art installations is shown in Table 7.

Table 7. Comparison of multimodal fusion in different cross-media art installations

Device Type	Main Fused Modalities	Fusion Methods
Visual and Auditory Fusion Device	Visual Auditory	Sound signal processing and visual pattern generation 3D audio and projection mapping synchronization
Text and Other Modalities Fusion Device	Text Visual Auditory	Semantic-driven multimodal material selection and switching Interactive text input and visual auditory feedback generation
Multimodal Interactive Device	Visual Auditory Tactile etc	Based on sensors and interactive software for multimodal data capture and fusion Emotion recognition and feedback generation

Advantages and Limitations

Multimodal fusion algorithms bring many significant advantages to the creation of cross-media art installations, promote the innovation of artistic expression and experience, and bring about a new change in the field of art, but there are also some limitations. To sum up, in the future, we need to make efforts in various aspects such as technological breakthroughs, cost reductions, and standardization of data annotations, in order to give full

play to its potential, and to promote the creation of cross-media art installations to new heights. The advantages of multimodal fusion algorithms for generating cross-media art installations are specifically analyzed as follows:

• Enhancement of artistic expression and infectious power. The multimodal fusion algorithm can integrate information from different modalities, so that the art installation can synchronize the power in multiple sensory levels such as visual, auditory, and tactile [68];

• Expanding the boundaries of artistic creation and imagination. With the help of multimodal fusion algorithms, artists break through the limitations of traditional single media to create previously unimaginable art forms [69];

• Enhance audience participation and interactive experience. Multimodal interactive installations rely on the algorithm to capture audience movements, expressions and other feedback in real time, and dynamically adjust the behavior and presentation of the installation accordingly [70].

In interactive dance installations, when the audience dances, the installation will change the music melody and the content of the screen based on the amplitude and rhythm of their movements, making the audience part of the artistic creation [71]. This kind of deep participation and real-time interaction changes the traditional mode of art appreciation and brings a new and unforgettable art experience for the audience [72]. The limitations of multimodal fusion algorithms for generating cross-media art installations are mainly reflected as follows:

• Difficulty and high cost of technical realization. The multimodal fusion algorithm involves complex signal processing, pattern recognition, deep learning and other technologies, which requires a high level of skill from the creators [73];

• Difficulty in data acquisition and labeling. The training of multimodal fusion algorithms relies on massive and accurately labeled multimodal data. Multimodal data annotation needs to coordinate the semantic relationship between different modalities, and the inconsistency of annotation standards and the varying professionalism of annotators lead to low efficiency and unstable quality of annotation, which in turn affects the performance of the algorithm [74-75];

• Uncertainty of the fusion effect. Despite the continuous progress of the algorithm, the multimodal fusion effect is still difficult to be completely predicted and precisely controlled. The information differences and associations between different modalities are complex, and problems such as semantic mismatch and inconsistent expression of emotions may occur when fusion occurs [76-77].

Future directions

In response to the advantages and limitations mentioned above, this section summarizes future directions and trends in the application of modal fusion algorithms for generating cross-media art installations：

Technology deepening and innovation

On the one hand, deep learning technology will continue to deepen its application in multimodal fusion, and the pre-trained model will be continuously optimized and improved to enhance the understanding of the deep semantics of different modal data and the ability of correlation mining, so as to achieve a more accurate and smarter multimodal fusion. It can more accurately generate highly matched visual images and music effects based on text descriptions in the creation of art installations [78]. On the other hand, emerging technologies such as quantum computing and edge computing will be gradually integrated to provide more powerful computational support for multimodal fusion algorithms, solving the problems of real-time and high efficiency, so that complex multimodal fusion cross-media art installations can be smoothly run on ordinary devices [79].

Diversity in Artistic Creation

From the perspective of creative style, cross-media art installations will present more diverse styles in the future. Artists will continue to explore new ways of integration and creative expression, break the boundaries of traditional art styles, and create works that integrate a sense of technology, futuristic and traditional cultural elements and other styles intertwined [80]. In terms of creative themes, with the development of society and people's concern for various issues, the art installations will cover themes in more fields, such as environmental protection, artificial intelligence ethics, cosmic exploration, etc., and more comprehensively and profoundly show the complexity and importance of these themes through multimodal fusion [81].

Interactive experience upgrade

Multimodal interaction technology will be continuously upgraded to make the interaction between the audience and the art installation more natural, smooth and rich. On the one hand, the brain-computer interface technology realizes the direct interaction between the audience's brainwaves and the art installation, and the audience's emotions and thinking activities can influence the presentation form of the installation in real time [82]; on the other hand, the combination of virtual reality, augmented reality, and mixed reality technology creates a more realistic and immersive interactive environment, so that the audience feels as if they are in the art work, and can even interact and communicate with the virtual elements in the work. interaction and communication.

Accelerating cross-sectoral integration

The cross-disciplinary integration of art and technology will be further deepened, and experts in multidisciplinary fields such as computer science, electronic information, and materials science will work more closely with artists to promote the development of cross-media art installations. Material scientists have developed new materials with unique optical, electrical, and other properties [83], providing more possibilities for the presentation of art installations; electronic engineers have designed smaller, more efficient, and low-power sensors and chips to enhance the performance of multimodal data acquisition and processing.

Application Scenario Expansion

The application scenarios of cross-media art installations will continue to expand, and in addition to traditional cultural venues such as art galleries, museums, exhibition halls, etc., they will also be widely used in urban public spaces, commercial venues, educational institutions, medical institutions, etc. [84]. In urban public space, large-scale interactive cross-media art installations will become an important part of the urban landscape, providing citizens with unique leisure and entertainment and cultural and artistic experiences; in commercial venues, art installations can be used for brand displays, marketing promotions, etc., attracting consumers' attention and enhancing the brand image; in educational institutions, they can be used as a teaching tool to stimulate students' creativity and imagination; in medical institutions, they It helps to relieve patients' anxiety and promote rehabilitation treatment.

Conclusion

Multimodal fusion algorithms open up new paths and bring unprecedented opportunities for the creation of cross-media art installations. It fuses different modal information, enriches artistic expressions, stimulates artists' creativity, and provides an immersive experience for the audience. At present, although there are challenges in technical realization, data processing, etc., the future development of technology and multi-disciplinary cooperation will promote its continuous progress, expand its application, and help the creative expression of art. The combination of multimodal fusion algorithm and cross-media art installation is a model of deep integration of art and technology, which will promote the common development of art and technology, energize the cause of human culture and art, enrich the spiritual life, and promote social innovation and progress.

References

[1]   Ismail, M. M. Ben, Alfaraj, S. N., & Bchir, O. (2019). Automatic image annotation using possibilistic clustering algorithm. International Journal of Fuzzy Logic and Intelligent Systems, 19(4), 250–262. DOI:10.5391/IJFIS.2019.19.4.250

[2]   Hui, X., & Yin, Y. (2019). The Design and Research of Subway Wayfinding System Under the Impact of New Media Art. Science Discovery, 7(5), 306. DOI:10.11648/j.sd.20190705.18

[3]   Corbi, A., Santos, O. C., & Burgos, D. (2019). Intelligent framework for learning physics with aikido (martial art) and registered sensors. Sensors (Switzerland), 19(17), 3681-. DOI:10.3390/s19173681

[4]   Le, T. H., Huang, S. C., & Jaw, D. W. (2019). Cross-Resolution Feature Fusion for Fast Hand Detection in Intelligent Homecare Systems. IEEE Sensors Journal, 19(12), 4696–4704. DOI:10.1109/JSEN.2019.2901259

[5] Tomoe, H., & Kanji, T. (2019). Long-Term Knowledge Distillation of Visual Place Classifiers. 2019 IEEE Intelligent Transportation Systems Conference, ITSC 2019, 541–546. DOI:10.1109/ITSC.2019.8917199

[6] Tianyu, S. (2019). As the 'Intellectual Subject' of Artificial Intelligence – The Intrinsic Understanding of Artificial Intelligence on The Ideas of 'Time' and 'Space.' AM Journal of Art and Media Studies, 20, 125–135. DOI:10.25038/am.v0i20.328

[7] Ling, X. (2019). Model for evaluating the art education teaching quality with uncertain information. Journal of Intelligent and Fuzzy Systems, 37(2), 1967–1972. DOI:10.3233/JIFS-179258

[8] Zhang, C., Zhang, H., Qiao, J., Yuan, D., & Zhang, M. (2019). Deep Transfer Learning for Intelligent Cellular Traffic Prediction Based on Cross-Domain Big Data. IEEE Journal on Selected Areas in Communications, 37(6), 1389–1401. DOI:10.1109/JSAC.2019.2904363

[9] Zheng, Y. (2020). The use of deep learning algorithm and digital media art in all-media intelligent electronic music system. PLoS ONE, 15(10 October). DOI:10.1371/journal.pone.0240492

[10] Ning, Y., Han, Z., Zhong, L., & Zhang, C. (2020). DRAN: Deep recurrent adversarial network for automated pancreas segmentation. IET Image Processing, 14(6), 1091–1100. DOI:10.1049/iet-ipr.2019.0399

[11] Gao, Z., Zhang, H., Dong, S., Sun, S., Wang, X., Yang, G., Wu, W., Li, S., & De Albuquerque, V. H. C. (2020). Salient Object Detection in the Distributed Cloud-Edge Intelligent Network. IEEE Network, 34(2), 216–224. DOI:10.1109/MNET.001.1900260

[12] Chen, C., Li, M., Ferreira, A., Huang, J., & Cai, R. (2020). A Copy-Proof Scheme Based on the Spectral and Spatial Barcoding Channel Models. IEEE Transactions on Information Forensics and Security, 15, 1056–1071. DOI:10.1109/TIFS.2019.2934861

[13] Ma, Z., Guan, J., Chae, E., & Wu, D. (2020). How to Construct and Apply Self-learning Support System of Art Curriculum in the New Media Environment. Lecture Notes in Electrical Engineering, 590, 128–133. DOI:10.1007/978-981-32-9244-4_17

[14] Lin, Y. (2020). Research on the Innovation of Art Design Products Based on the Concept of "aI" Boundary. Journal of Physics: Conference Series, 1574(1), 012078 (5pp). DOI:10.1088/1742-6596/1574/1/012078

[15] Yang, W., Su, J., Zhang, S., Qiu, K., Su, P., & Zhang, X. (2020). Research on the intelligent design of the multidimensional form of a product under a brand cobweb constraint. International Journal of Arts and Technology, 12(4), 352–374. DOI:10.1504/IJART.2020.112648

[16] Dey, S., Sarkar, I., Chakraborty, S., & Roy, S. (2020). Depression Detection Using Intelligent Algorithms From Social Media Context - State of the Art, Trends and Future Roadmap. Journal of Xidian University, 14(8), 45–62. DOI:10.37896/jxu14.8/007

[17] Liu, S., & Li, T. (2020). AVG comprehensive practice reform of digital media art based on fuzzy theory. Journal of Intelligent and Fuzzy Systems, 38(4), 3697–3706. DOI:10.3233/JIFS-179592

[18] Li, S. (2020). Cross Border Integration of Art Design and Artificial Intelligence. Journal of Physics: Conference Series, 1624(5), 52027. DOI:10.1088/1742-6596/1624/5/052027

[19] Guleria, K., Prasad, D., Lilhore, U. K., & Simaiya, S. (2020). Asynchronous Media Access Control Protocols and Cross Layer Optimizations for Wireless Sensor Networks: An Energy Efficient Perspective. Journal of Computational and Theoretical Nanoscience, 17(6), 2531–2538. DOI:10.1166/jctn.2020.8926

[20] Kotis, K., Spiliotopoulos, D., & Papasalouros, A. (2020). Intelligent Collaborative Authoring of Place-Based, Cross-Cultural and Media-Rich Experiences. Challenges, 11(1), 10. DOI:10.3390/challe11010010

[21] Naik, M. V., Vasumathi, D., & Kumar, A. P. S. (2020). An Improved Intelligent Approach to Enhance the Sentiment Classifier for Knowledge Discovery Using Machine Learning. International Journal of Sensors, Wireless Communications and Control, 10(4), 582–593. DOI:10.2174/2210327910999200528114552

[22] Liu, L., & Wang, Y. (2021). Innovation and Entrepreneurship Practice Education Mode of Animation Digital Media Major Based on Intelligent Information Collection. Mobile Information Systems, 2021. DOI:10.1155/2021/3787018

[23] Wenhao, D. (2021). Multisensor Information Fusion-Assisted Intelligent Art Design under Wireless Virtual Reality Environment. Journal of Sensors, 2021. DOI:10.1155/2021/6119127

[24] Shi, G., Xiao, L., Chen, Y., & Zhou, S. K. (2021). Marginal loss and exclusion loss for partially supervised multi-organ segmentation. Medical Image Analysis, 70, 101979. DOI:10.1016/j.media.2021.101979

[25] Eide, S. S., Riegler, M. A., Hammer, H. L., & Bremnes, J. B. (2021). Temperature Forecasting using Tower Networks. ICDAR 2021 - Proceedings of the 2021 Workshop on Intelligent Cross-Data Analysis and Retrieval, 18–23. DOI:10.1145/3463944.3469099

[26] Ren, X., & Cui, Y. (2021). Embedded System Intelligent Platform Design Based on Digital Multimedia Artistic Design. Wireless Communications and Mobile Computing, 2021. DOI:10.1155/2021/3959199

[27] Gawas, M. A., & Govekar, S. (2021). State-of-Art and Open Issues of Cross-Layer Design and QOS Routing in Internet of Vehicles. Wireless Personal Communications, 116(3), 2261–2297. DOI:10.1007/s11277-020-07790-5

[28] Ismaiel, aml, Gad, W., Mostafa, T., & Badr, N. (2021). A Location Prediction Methods: state of art. International Journal of Intelligent Computing and Information Sciences, 21(3), 119–133. DOI:10.21608/ijicis.2021.84159.1111

[29] Gao, Z., & Lin, L. (2021). The Intelligent Integration of Interactive Installation Art Based on Artificial Intelligence and Wireless Network Communication. Wireless Communications and Mobile Computing, 2021. DOI:10.1155/2021/3123317

[30] Chen, J. C. (2021). Machine Learning-Inspired Algorithmic Framework for Intelligent Reflecting Surface-Assisted Wireless Systems. IEEE Transactions on Vehicular Technology, 70(10), 10671–10685. DOI:10.1109/TVT.2021.3110970

[31] Chen, F., Liu, X., Yu, H., & Ha, Y. (2021). CLIF: Cross-layer information fusion for stereo matching and its hardware implementation. Proceedings - IEEE International Symposium on Circuits and Systems, 2021-May. DOI:10.1109/ISCAS51556.2021.9401077

[32] Xi, M. (2022). Reconstruction of New Media Art Ecology Based on Intelligent Technology Environment. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/8084536

[33] Wang, X., & Zhu, Y. (2022). Intelligent Art Design Management Based on Wireless Communication Microprocessor and Mobile Internet. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/5012875

[34] Su, Z., Tavolara, T. E., Carreno-Galeano, G., Lee, S. J., Gurcan, M. N., & Niazi, M. K. K. (2022). Attention2majority: Weak multiple instance learning for regenerative kidney grading on whole slide images. Medical Image Analysis, 79. DOI:10.1016/j.media.2022.102462

[35] Xu, W. J., & Jin, S. (2022). Analysis of Daily Public Relations in Art Colleges in the New Media Environment. Smart Innovation, Systems and Technologies, 258, 265–273. DOI:10.1007/978-981-16-5168-7_33

[36] Li, Y., & Zhuge, W. (2022). Application of Animation Control Technology Based on Internet Technology in Digital Media Art. Mobile Information Systems, 2022(Pt.4), 4009053.1-4009053.11. DOI:10.1155/2022/4009053

[37] Xi, M. (2022). Reconstruction of New Media Art Ecology Based on Intelligent Technology Environment. Wireless Communications and Mobile Computing, 2022. DOI:10.1155/2022/8084536

[38] Chen, X., Jin, W., Wu, Q., Zhang, W., & Liang, H. (2022). A hybrid cost-sensitive machine learning approach for the classification of intelligent disease diagnosis. Journal of Intelligent and Fuzzy Systems, 43(3), 3039–3050. DOI:10.3233/JIFS-213486

[39] Sahargahi, V., Majidnezhad, V., Afshord, S. T., & Jafari, Y. (2022). An intelligent chaotic clonal optimizer. Applied Soft Computing, 115(115-), 115. DOI:10.1016/j.asoc.2021.108126

[40] Hossain, E., Nadim Kaysar, M., Jalal Uddin Joy, A. Z. M., Mizanur Rahman, M., & Wahidur Rahman. (2022). A Study Towards Bangla Fake News Detection Using Machine Learning and Deep Learning. Advances in Intelligent Systems and Computing, 79–95. DOI:10.1007/978-981-16-5157-1_7

[41] Huang, X., Shi, K., Zhou, J., Liang, Y., Liu, Y., Zhang, J., Guo, Y., & Jin, C. (2022). Development of a Machine Learning-Assisted Model for the Early Detection of Severe COVID-19 Cases Combining Blood Test and Quantitative Computed Tomography Parameters. Journal of Medical Imaging and Health Informatics, 11(11), 2747–2753. DOI:10.1166/jmihi.2021.3866

[42] Wu, Y., Ridwan, A. R., Niaz, M. R., Qi, X., Zhang, S., Alzheimer's Disease Neuroimaging Initiative, Bennett, D. A., & Arfanakis, K. (2022). Development of high quality T1-weighted and diffusion tensor templates of the older adult brain in a common space. NeuroImage, 260, 119417. DOI:10.1016/j.neuroimage.2022.119417

[43] Rathour, N., Singh, R., Gehlot, A., Vaseem Akram, S., Kumar Thakur, A., & Kumar, A. (2022). The decadal perspective of facial emotion processing and Recognition: A survey. Displays, 75, 102330. DOI:10.1016/j.displa.2022.102330

[44] Khatoon, S., Asif, A., Hasan, M. M., & Alshamari, M. (2022). Social Media-Based Intelligence for Disaster Response and Management in Smart Cities. Springer Optimization and Its Applications, 186, 211–235. DOI:10.1007/978-3-030-84459-2_11

[45] Habib Khan, N., Jamal, R., Ebeed, M., Kamel, S., Zeinoddini-Meymand, H., & Zawbaa, H. M. (2022). Adopting Scenario-Based approach to solve optimal reactive power Dispatch problem with integration of wind and

solar energy using improved Marine predator algorithm. Ain Shams Engineering Journal, 13(5). DOI:10.1016/j.asej.2022.101726

[46] Indhumathi, R., Narmadha, T. V., & kurunathan, H. (2022). Hybrid pixel based method for multimodal image fusion based on Integration of Pulse Coupled Neural Network (PCNN) and Genetic Algorithm (GA) using Empirical Mode Decomposition (EMD). Microprocessors and Microsystems, 94, 104665. DOI:10.1016/j.micpro.2022.104665

[47] Lan, H., & Zhang, P. (2022). Learning and Integrating Multi-Level Matching Features for Image-Text Retrieval. IEEE Signal Processing Letters, 29(000), 374–378. DOI:10.1109/LSP.2021.3135825

[48] Wang, P., Zheng, S., Jiang, Y., Li, C., Liu, J., Wen, C., Patronov, A., Qian, D., Chen, H., & Yang, Y. (2022). Structure-Aware Multimodal Deep Learning for Drug-Protein Interaction Prediction. Journal of Chemical Information and Modeling, 62(5), 1308–1317. DOI:10.1021/acs.jcim.2c00060

[49] Foster, J. M., & Maithel, S. K. (2022). Introduction: Surgeons establishing the landscape of contemporary clinical trials in oncology. Journal of Surgical Oncology, 125(1), 5–6. DOI:10.1002/jso.26732

[50] Wu, J., & Jia, L. (2022). Neural Network Model for Perceptual Evaluation of Product Modelling Design Based on Multimodal Image Recognition. Computational Intelligence and Neuroscience, 2022, 1665021. DOI:10.1155/2022/1665021

[51] Bala, N., Kumar, A., & Gupta, R. (2022). An Effective Multimodal Biometrics Recognition System using Hybrid Fusion Technique for Efficient Authentication. 2022 8th International Conference on Signal Processing and Communication, ICSC 2022, 277–282. DOI:10.1109/ICSC56524.2022.10009259

[52] Stahlschmidt, S. R., Ulfenborg, B., & Synnergren, J. (2022). Multimodal deep learning for biomedical data fusion: A review. Briefings in Bioinformatics, 23(2), 2. DOI:10.1093/bib/bbab569

[53] Lim, K. (2022). Expanding Multimodal Artistic Expression and Appreciation Methods through Integrating Augmented Reality. International Journal of Art and Design Education, 41(4), 562–576. DOI:10.1111/jade.12434

[54] Cao, W.-M., Wang, X., Liu, J., Wang, L., Zhang, X., Pan, J., Ye, W., Chen, Z., Zheng, Y., Shao, X., & Xu, Y. (2022). BRCANet: A deep hybrid network in predicting BRCA1/2 gene mutation of breast cancer with dynamic contrast-enhanced breast MRI. Journal of Clinical Oncology, 40(16_suppl), e13576–e13576. DOI:10.1200/jco.2022.40.16_suppl.e13576

[55] Xue, Z., Li, J., Pianetta, P., & Liu, Y. (2022). Data-Driven Lithium-Ion Battery Cathode Research with State-of-the-Art Synchrotron X-ray Techniques. Accounts of Materials Research, 3(8), 854–865. DOI:10.1021/accountsmr.2c00098

[56] Kanimozhi, P., & Aruldoss Albert Victoire, T. (2022). Oppositional tunicate fuzzy C-means algorithm and logistic regression for intrusion detection on cloud. Concurrency and Computation: Practice and Experience, 34(4), 34. DOI:10.1002/cpe.6624

[57] Kamphuis, M. E., Kuipers, H., Verschoor, J., van Hespen, J. C. G., Greuter, M. J. W., Slart, R. H. J. A., & Slump, C. H. (2022). Development of a dynamic myocardial perfusion phantom model for tracer kinetic measurements. EJNMMI Physics, 9(1), 31. DOI:10.1186/s40658-022-00458-y

[58] Unnikrishnan, R., S., S. K., & V.S., A. (2023). Efficient parameter tuning of neural foundation models for drug perspective prediction from unstructured socio-medical data. Engineering Applications of Artificial Intelligence, 123(Pt.A), 106214–106216. DOI:10.1016/j.engappai.2023.106214

[59] Singh, N. K., Singh, P., Das, P., & Chand, S. (2023). XRBi-GAC: A hybrid deep learning framework for multilingual toxicity detection. Journal of Intelligent and Fuzzy Systems, 45(1), 1409–1421. DOI:10.3233/JIFS-224536

[60] Himani Sharma, & Navdeep Kanwal. (2023). Smart Cities: A Worldwide Journey into Intelligent Urbanism and State-of-the-Art Technologies. Scientific and Technical Information Processing, 50(4), 328–355. DOI:10.3103/S0147688223040081

[61] Wang, L., Cheng, J., Song, J., Pan, X., & Zhang, C. (2023). Learning to measure infrared properties of street views from visible images. Measurement: Journal of the International Measurement Confederation, 207, 112320-. DOI:10.1016/j.measurement.2022.112320

[62] Alomari, E. A., & Mehmood, R. (2023). Smart Cities, Smarter Roads: A Review of Leveraging Cutting-Edge Technologies for Intelligent Event Detection from Social Media. International Journal of Advanced Computer Science and Applications, 14(11), 364–374. DOI:10.14569/IJACSA.2023.0141137

[63] Liu, Y., Fan, C. S., & Liu, P. X. (2023). Research on Digital Media Intelligent Art Creation Based on the Fusion of Virtual Reality and Semantic Features. Journal of Information Science and Engineering, 39(1), 55–66. DOI:10.6688/JISE.202301_39(1).0004

[64] Liu, Y. (2023). RETRACTED: Research on the development trend and application of digital media art in graphic design education (International Journal of Electrical Engineering & Education, (2021)). International Journal of Electrical Engineering and Education, 60(1_suppl), 2020–2029. DOI:10.1177/0020720920984310

[65] Gong, T., Zhou, W., Qian, X., Lei, J., & Yu, L. (2023). Global contextually guided lightweight network for RGB-thermal urban scene understanding. Engineering Applications of Artificial Intelligence, 117(Pt.A). DOI:10.1016/j.engappai.2022.105510

[66] Wang, M. K., Meng, M., Liu, J., & Wu, J. (2023). Learning Adequate Alignment and Interaction for Cross-Modal Retrieval. Virtual Reality and Intelligent Hardware, 5(6), 509–522. DOI:10.1016/j.vrih.2023.06.003

[67] Biddle, J. L., & Hibberd, L. (2023). The Artificial as an Intelligent Indigenous/Indigenizing System: The Experimental Art and Artifice of r e a. Visual Anthropology Review, 39(2), 458–474. DOI:10.1111/var.12284

[68] Shu, M., Liu, X., & Zhou, H. (2023). Interactive modeling of painting art communication structure from the perspective of integrated media. Journal of Intelligent and Fuzzy Systems, 45(6), 10053–10062. DOI:10.3233/JIFS-234284

[69] Zhang, Y. (2023). Immersive Multimedia Art Design Based on Deep Learning Intelligent VR Technology. Wireless Communications and Mobile Computing, 2023. DOI:10.1155/2023/9266522

[70] Nana, Y., & Yang, N. (2023). Retraction Note: Analysis of image processing algorithm based on bionic intelligent optimization (Cluster Computing, (2019), 22, S2, (3505-3512), 10.1007/s10586-018-2198-8). Cluster Computing, 26(Suppl 1), 159. DOI:10.1007/s10586-022-03911-w

[71] Chen, D., Su, W., Wu, P., & Hua, B. (2023). Joint multimodal sentiment analysis based on information relevance. Information Processing and Management, 60(2). DOI:10.1016/j.ipm.2022.103193

[72] Wang, W., Li, C., Li, A., Li, F., Chen, J., & Zhang, T. (2023). One-stage self-supervised momentum contrastive learning network for open-set cross-domain fault diagnosis. Knowledge-Based Systems, 275. DOI:10.1016/j.knosys.2023.110692

[73] Xu, W. (2023). Data Extraction and Intelligent Analysis of Aesthetic Laws Accepted by VR Digital Media Art on Film Art Based on NVIDIA GPU Hardware. International Journal of Science and Engineering Applications, 111–113. DOI:10.7753/ijsea1203.1041

[74] Saghir, S., Imenes, K., & Schiavone, G. (2023). Integration of hydrogels in microfabrication processes for bioelectronic medicine: Progress and outlook. Frontiers in Bioengineering and Biotechnology, 11(1), 15. DOI:10.3389/fbioe.2023.1150147

[75] Liu, L., Chang, J., Zhang, P., Ma, Q., Zhang, H., Sun, T., & Qiao, H. (2023). A joint multi-modal learning method for early-stage knee osteoarthritis disease classification. Heliyon, 9(4), 14. DOI:10.1016/j.heliyon.2023.e15461

[76] Zhang, D., Xu, C., & Li, S. (2023). Heuristic multi-modal integration framework for liver tumor detection from multi-modal non-enhanced MRIs. Expert Systems with Applications, 221. DOI:10.1016/j.eswa.2023.119782

[77] Chen, Z., Zhu, M., Chen, S., Lu, L., Tang, H., Hu, X., & Ji, C. (2023). Discriminative feature fusion for RGB-D salient object detection. Computers and Electrical Engineering, 106, 108579-. DOI:10.1016/j.compeleceng.2023.108579

[78] Mousa-Pasandi, M., Liu, T., Massoud, Y., & Laganière, R. (2023). RGB-LiDAR fusion for accurate 2D and 3D object detection. Machine Vision and Applications, 34(5). DOI:10.1007/s00138-023-01435-w

[79] Fan, T., Wang, H., & Hodel, T. (2023). Multimodal knowledge graph construction of Chinese traditional operas and sentiment and genre recognition. Journal of Cultural Heritage, 62, 32–44. DOI:10.1016/j.culher.2023.05.003

[80] Curtin, A., Topoglu, Y., Alshehri, S., Woodburn, M., Martin, L., Suri, R., & Ayaz, H. (2023). Evaluating the restorative impact of nature through multimodal mobile sensing of neural, physiological, and behavioral activity in ambulatory settings. Neuroergonomics and Cognitive Engineering, 102. DOI:10.54941/ahfe1003015

[81] Niknam, G., Molaei, S., Zare, H., Pan, S., Jalili, M., Zhu, T., & Clifton, D. (2023). DyVGRNN: DYnamic mixture Variational Graph Recurrent Neural Networks. Neural Networks, 165(000), 596–610. DOI:10.1016/j.neunet.2023.05.048

[82] Aslam, A., Sargano, A. B., & Habib, Z. (2023). Attention-based multimodal sentiment analysis and emotion recognition using deep neural networks. Applied Soft Computing, 144, 144. DOI:10.1016/j.asoc.2023.110494

[83] Kim, H., Kang, H., Choi, H., Jung, D., & Hong, T. (2023). Human-building interaction for indoor environmental control: Evolution of technology and future prospects. Automation in Construction, 152. DOI:10.1016/j.autcon.2023.104938

[84] Liu, Z., Cheng, Q., Song, C., & Cheng, J. (2023). Cross-scale cascade transformer for multimodal human action recognition. Pattern Recognition Letters, 168, 17–23. DOI:10.1016/j.patrec.2023.02.024