Application of Multimodal Fusion in Creative Packaging Design

Xue Dong[1] and Xiaofeng Zhao[2]

[1]School of Art and Design, Shandong Women's University, Ji'nan 250300, China,
dongxue@sdwu.edu.cn
[2]Teaching Affairs Office, Shandong Women's University, Ji'nan 250300, China,
zhxf@sdwu.edu.cn

Corresponding author: Xue Dong, dongxue@sdwu.edu.cn

Abstract. In the context of the digital age, dynamic design is developing rapidly. Dynamics have the characteristics of quickly attracting attention, enhancing interactive fun, and intuitively and quickly transmitting information. Multi-modal fusion technology (MFT) is one of the means to achieve the dynamic development of packaging design. The emergence of MFT technology has created new opportunities for the development of the packaging industry and has also brought new interactive modes and meanings to product packaging. In terms of interactive content, MFT content is not limited by spatial entities and can accommodate massive information resources. In terms of interactive mode, the packaging uses MFT technology as a medium to transform the dimensions and forms of interaction with consumers from a relatively simple two-dimensional visual interaction to a rich three-dimensional multi-sensory interaction. This article explores how to integrate MFT technology with packaging in the context of the digital age by analyzing existing packaging in the market and the latest relevant achievements of universities. The experimental evaluation emphasized the powerful ability of the algorithm to significantly improve the creativity and personalized quotient of packaging design while enhancing their functional practicality.

Keywords: Computer-Aided Design; Multimodal Fusion; Creative Packaging Design
DOI: https://doi.org/10.14733/cadaps.2025.S3.39-51

1 INTRODUCTION

In recent years, multimodal technology has gradually become known to people, but it is still in the development stage, and its application in many fields is still under research and exploration. Utilize the unlimited carrying capacity of multimodal information and apply it to camera packaging decoration. Users can obtain richer and more detailed dynamic product information and usage guidance through multimodal technology. Enable consumers to have a 3D interactive experience through the multimodal virtual instructions on the packaging, making it more vivid and convenient for them to master camera usage skills. Ultimately, the camera packaging interaction will be transformed from two-dimensional to three-dimensional, and the packaging form will be transformed from static to dynamic, designing a more interactive and user-friendly camera packaging [1]. At the same time, practicality, environmental friendliness, and coordination with production technology need to be taken into account. The rapid development of computer technology, especially the emergence of CAD software, has brought revolutionary changes to the field of packaging design [2].

In the current development process, online shopping not only greatly facilitates consumers' purchasing behaviour, but also promotes product packaging design as a key factor in attracting consumers' attention and increasing sales volume. With the continuous improvement of people's acceptance of MFT technology, some regions have begun to use this technology in supermarkets [3]. Each product has a special logo placed on it, and consumers only need to scan it with their mobile phone to display detailed product information. Consumers can see more accurate product information on their mobile devices. The transformation of this consumption pattern has brought about transformative development in supermarket shopping, which is the interactive experience of information hiding and autonomous selection brought by AR technology to people [4]. The advantage of using MFT technology to expand information-carrying capacity effectively is that its developers integrate and summarize information from products in supermarkets. The dense text descriptions and icons on traditional packaging often sacrifice the overall aesthetic of the design, while MFT technology transforms this lengthy information into AR interactive content. This interaction is not only highly autonomous but also greatly enriches consumers' sense of participation and exploration, making the process of obtaining product information a fun and surprising journey [5]. Specifically, the application of MFT technology significantly optimizes the spatial layout and visual presentation of packaging design. When people scan the packaging pattern of Heinz tomato sauce, the label on the packaging turns into a small book, providing detailed cooking methods for consumers to use the product more conveniently and quickly. The limited packaging display area objectively limits the information-carrying capacity of the packaging. Excessive information will occupy most of the packaging decoration space, sacrificing visual communication to convey textual information. The application of MFT technology cleverly solves the problem of a decrease in the artistic beauty of packaging. Heinz tomato sauce packaging is an intelligent packaging that integrates AR technology. Product producers hope to provide consumers with detailed instructions, pairing suggestions, and cooking techniques for the product, but for traditional printed packaging, this is almost impossible to achieve [6]. At the same time, it maintains a simple and beautiful visual display effect of decoration.

Dynamic packaging decoration has become an effective means to attract consumers, and MFT technology is used to complete the dynamic design of packaging decoration [7]. Apply MFT technology to camera packaging to complete digital instruction manuals and dynamic packaging decoration design. Consumers only need to directly scan the identification pattern on the packaging or remove the identification card on the packaging to obtain a digital interactive manual. This facilitates consumers to carry and read digital manuals while using grayscale cards to measure exposure and determine light ratio, achieving integration of photography equipment and packaging and more effective utilization of packaging [8]. This packaging enables consumers who have product information but cannot fully master the usage skills to interact with the product. Transforming two-dimensional flat decoration into three-dimensional dynamic decoration can attract consumers while effectively expanding the information-carrying capacity of packaging. On the other hand, the development of mobile terminal devices has provided conditions for dynamic design, which has become a new trend in design development in recent years. At this point, the digital instruction manual for camera packaging that integrates MFT technology can effectively increase the information-carrying capacity of the packaging and maximize the information dimension. In addition, the back of the special part that can be removed will be designed as an 18% neutral grayscale card [9].

This interdisciplinary research method provides a broader perspective and abundant resources for the study in this article and also promotes the intersection and integration between different disciplines. The introductory section of this article delves into the expansive macro backdrop and profound implications of the research, firmly establishing a robust groundwork for subsequent investigations. Proceeding further, the text pivots to spotlight the expansive avenues for CAD and

MFT utilization within creative packaging design, unveiling the untapped synergy when these technologies intertwine. Our delving into this realm encompasses the meticulous crafting of an innovative MFT algorithm, innovatively incorporating DL principles. The article meticulously outlines the algorithm's construction, detailing each step of its process and operational nuances. To substantiate its practical value and viability, we meticulously orchestrated a series of experiments, ensuring a thorough validation. Closing the discourse, we consolidate our key insights and innovations in a comprehensive conclusion, offering a clear perspective on the study's essence. Furthermore, we cast a vision towards future research trajectories, offering strategic recommendations aimed at fostering the sustained growth and advancement of this dynamic field.

2 RELATED WORK

Multimodal fusion manufacturing provides new ideas for environmentally friendly packaging design with its efficient use of materials and flexible design. Using AR technology (a modality) to transform packaging design from static images to dynamic, interactive 3D models, allowing consumers to preview the actual packaging effect before shopping, and even simulate the unboxing experience. By combining multimodal fusion techniques, Uzelac et al. [10] comprehensively utilize various modal information such as CAD models, user feedback videos, and trend analysis on social media to jointly shape packaging designs that are both environmentally friendly and highly creative. The multimodal fusion in creative packaging design can also enhance product interactivity and user experience. For example, by analyzing user preference data (one pattern) and material sustainability information (another pattern), design packaging that is both popular and uses recyclable materials. In the context of additive manufacturing, multimodal fusion can help designers optimize packaging material selection and structural design more accurately, reducing unnecessary material usage and EC. Meanwhile, by combining other methods such as sound and touch, a more immersive shopping experience can be created, enhancing consumers' memory points and willingness to purchase products. Designers can design packaging structures that are both lightweight and sturdy while reducing material waste and production costs. Multimodal fusion also promotes the rapid collection and iterative design of market feedback. By analyzing the mechanical properties, cost, environmental impact (multiple modes), and specific packaging requirements (such as protection, portability, etc.) of different materials. By collecting data through various channels such as social media, online surveys, and user comments, Vasileiadis et al. [11] timely understand consumers' feedback on packaging design, including preferences, improvement suggestions, etc. These data can serve as input for subsequent design, guiding designers to continuously optimize packaging design, and improve market acceptance and customer satisfaction.

In the rapidly changing market environment, enterprises not only pursue rapid innovation but also increasingly value enhancing market competitiveness by improving design creativity and attractiveness. In this context, Wang et al. [12] not only established a general parameterized design cognitive model but also explored the application of multimodal fusion in creative packaging design, aiming to further accelerate and optimize the design process. These multimodal data can inspire designers to think from different dimensions, promoting the collision and fusion of creativity. Designers can more accurately position their design direction and create packaging works that meet market demand and have unique personalities. In some studies, parameterized design cognitive models integrate stimuli from multiple modalities (such as visual images, textual descriptions, user feedback videos, social media trends, etc.), providing designers with richer and more diverse sources of inspiration. This means that designers can not only input their creativity through traditional hand-drawn sketches, but also directly utilize AR/VR technology (visual and interactive modes), sound description (auditory mode), and even tactile feedback (tactile mode) to construct and adjust their designs. To accelerate the creative implementation process, the form generation design technique developed by Wang et al. [13] has been further expanded to support multimodal input. By analyzing users' colour preferences for product packaging (visual form) and emotional feedback (textual form). This multimodal input method greatly enriches the means of design expression, enabling creativity to be quickly and accurately transformed into concrete 3D models. The research results indicate that the model can quickly generate 3D sketch plans that meet expected goals, which not only meet functional and aesthetic requirements but also inspire designers with more creative inspiration. Through a series of case studies combining multimodal fusion elements, this study validates the effectiveness and practicality of the established model in creative packaging design.

In the early stages of creative packaging design, Yang et al. [14] not only relied on traditional market research and customer interviews, but also actively integrated multimodal data sources such as social media, online comments, and user-generated content (UGC). For example, using AR/VR technology, designers and potential customers can preview the 3D effects of packaging designs in a virtual environment, while combining haptic feedback devices to simulate real touch, making the design more intuitive, easy to iterate and optimize. In addition, sentiment analysis tools have been introduced to provide a scientific basis for design decisions by analyzing users' emotional responses under different design stimuli. Zhu et al. [15] created packaging designs that meet market demand and have unique personalities by comprehensively analyzing multimodal information such as visual preferences, colour combinations, and material textures. In the process of transforming concepts into 3D CAD models, this study fully utilizes the complementarity of multimodal data. By maintaining consistency in the brand's visual identity system (VI) in multimodal presentation, such as colour, font, pattern, etc., the recognition and overall sense of the product line have been improved. By utilizing advanced 3D rendering technology, design sketches are transformed into highly realistic virtual models, further improving the visualization effect and communication efficiency of the design. Meanwhile, according to the usage scenarios and target user groups of different products, flexibly adjust the focus and expression of multimodal information to meet the needs of different segmented markets.

In the vast field of packaging design and optimization, research is constantly advancing with unprecedented depth and breadth, from foundational design theories to cutting-edge AI algorithm applications. The wave of technological innovation is constantly expanding the boundaries of design, injecting infinite possibilities and creativity into the field of packaging design. Looking ahead to the future, with the continuous leap of technology, we have reason to believe that the packaging design industry will witness more disruptive breakthroughs and unprecedented innovations, leading the industry towards a new era of more intelligent and sustainable development.

3 THE APPLICATION OF CAD AND MFT IN CREATIVE PACKAGING DESIGN

Currently, research in this field is still in the exploratory and experimental stage in various countries around the world. China's three major network operators have confirmed that people will be able to use 5G networks by 2020. Intelligent packaging technology is a product of continuous social development, based on multidisciplinary technology. So how the packaging industry can use this technology as an opportunity to change the traditional packaging development model and adapt to the new requirements of new media conditions is a topic worth studying. The development of intelligent packaging is in the early stage of vigorous rise, driven by the deep integration and innovative application of cross-disciplinary technologies. This process not only greatly enhances the attractiveness and affinity of the packaging, but also gives consumers an unprecedented shopping and usage experience, making the entire process more autonomous, convenient, and fun. Transforming packaging from simple containers or protective layers into an intelligent platform that integrates functionality, interactivity, and personalization. In the multidimensional development path of intelligent packaging, information intelligent packaging stands out with its unique charm and has become one of the most forward-looking and promising fields for development. This technological integration not only improves the efficiency and fun of information acquisition but also deepens consumers' understanding and emotional connection with the product, creating a unique competitive advantage for the brand. As an innovative technology that actively responds to the future development needs of the packaging industry, intelligent packaging technology is gradually reshaping traditional packaging design concepts. The application of AR technology in packaging, as an advanced form of information intelligent packaging, demonstrates its strong potential and broad application prospects. Information intelligent packaging is not limited to traditional information transmission but utilizes cutting-edge technologies such as augmented reality (AR) to present product information, usage instructions, brand stories, etc. to consumers in a richer, intuitive, and interactive way. This unprecedented interactive experience not only greatly enhances the attractiveness and fun of packaging, but also promotes deep interaction between products and consumers, shaping a more three-dimensional and vivid image for the brand. By scanning specific labels on the packaging, consumers can instantly enter a world where virtual and real are intertwined, experiencing the functionality, effects, and even the production process and cultural stories behind the product firsthand.

Figure 1: Application example of CAD in packaging design.

Figure 1 shows an example of the application of CAD in packaging design. For the product packaging studied in this project, due to the limitations of the display surface size of the product packaging itself and the inconvenience of carrying and reading the product manual, CAD technology is used to hide the instruction information in a virtual scene. The application of CAD technology makes packaging smarter, information exchange more convenient, and truly maximizes the dimension of information. The construction of scenes and models is a key step in AR implementation. Scene design needs to consider environmental factors, lighting conditions, and the degree of integration between virtual elements and the real world to create a visual effect that is both realistic and harmonious. The creation of models relies on high-precision 3D modelling technology to ensure that virtual objects can be perfectly presented in terms of details, proportions, and materials. In addition, the writing of animation and interactive logic is also indispensable, as they give life to the model, enabling it to respond to user actions and generate corresponding feedback. Consumers can obtain a more vivid interactive experience and learn the basic usage methods of the product more quickly and conveniently. Among them, dynamic vision is more able to attract consumers' attention. In the packaging design process, in addition to graphic visual design, dynamic visual expression design is also needed. CAD technology is a multidisciplinary and comprehensive image recognition technology. The dynamic visual design also needs to be based on the different features of each product, emphasizing the expression of product characteristics and playing a more vivid role in promotion and advertising.

4 COMBINING DL WITH MFT ALGORITHM

4.1 Application of DL in MFT

The application of DL in MFT is becoming increasingly widespread, and the development of this field greatly promotes the performance improvement of machine learning (ML) models and the ability to handle complex problems. The core of the MFT strategy lies in integrating data features from different types, such as images, text, audio, etc., to eliminate heterogeneity differences between them, thereby improving the generalization ability and accuracy of the model. In MFT, federated architecture is an important implementation approach. This architecture achieves the fusion of multimodal features by mapping data representations of different modalities into a shared semantic subspace (as shown in Figure 2). This mapping process enables data that originally belonged to different modalities to be compared and operated on in the same dimension, thus fully utilizing the complementary information between different modalities. The implementation of joint architecture relies on various technologies in DL, such as convolutional neural networks (CNN), recurrent neural networks (RNN), and self-attention mechanisms.

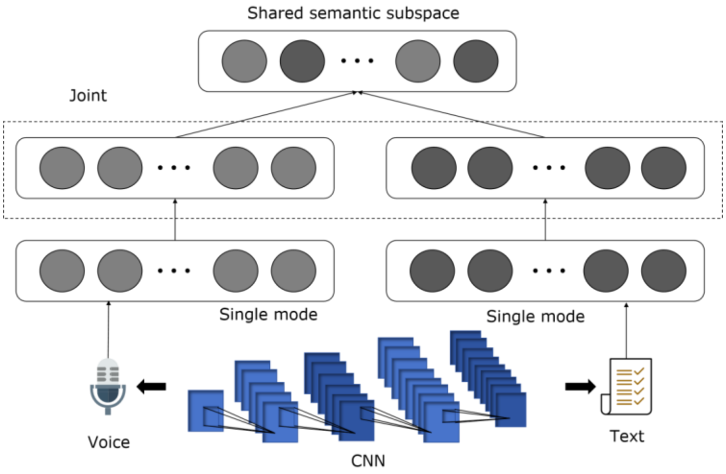

Figure 2: Schematic diagram of joint fusion architecture.

Early fusion involves feature fusion in the early stages of data processing, which can capture low-level correlation information between different modalities but may increase the complexity and computational cost of the model. Late-stage fusion combines the predicted results of different modalities in the model output stage, with each modality processed independently. The model training is simple, but it may not fully capture the interaction information between modalities. Hybrid fusion combines the advantages of early and late fusion, performing multiple fusions at different stages to more flexibly capture multi-level modal relationships.

4.2 Algorithm Principle

The essence of multimodal joint architecture lies in the feature "fusion". The basic method is a direct connection (additive joint), which builds a shared semantic space in each hidden layer, aggregates single modal feature vectors, and achieves efficient fusion, as shown in equation (1), fully demonstrating the significance of multimodal information integration.

$$z = f \ w_1^T v_1 + w_2^T v_2 + \cdots + w_n^T v_n \tag{1}$$

Among them, z is the output result in the shared semantic subspace, v the input of each single modality, w the weight, and the subscript represent different modalities. By mapping f, all submodalities are semantically transformed into the shared subspace.

The cross-modal similarity method maintains the similarity structure inside and outside the modality through similarity measurement, minimizes the cross-modal distance between semantically similar objects, and maximizes the distance between semantically different objects. To achieve effective integration of vision and text, the intermodal ranking method is introduced, which represents the matching embedding vectors of vision and text in the form of $(v,t) \in D$, where D is the joint embedding space. The objective function of the fusion process is reflected through a carefully designed loss function f, which aims to optimize the fusion effect. The specific expression is shown in equation (2), accurately characterizing and optimizing the correlation between visual and textual information.

$$\begin{aligned} f = \sum_{v}\sum_{t^-}\max\left(0, \alpha - S(v,t) + S(v,t^-)\right) + \\ \sum_{t}\sum_{v^-}\max\left(0, \alpha - S(t,v) + S(t,v^-)\right) \end{aligned} \tag{2}$$

Among them, a is the edge, S the similarity measurement function, t^- the embedding vector that does not match v, v^- is the embedding vector that does not match t, and t^- v^- randomly selected samples.

As one of the most widely used DL models in the field of image recognition, CNN's architecture essence lies in integrating multiple carefully designed convolutional and sampling layers. The main output of convolutional layers can be ingeniously expressed as a series of carefully filtered and transformed feature maps, which not only profoundly reveal the intrinsic structure and patterns of the image, but also provide a rich and powerful information foundation for subsequent processing.

$$D_{i,k} = \theta\left(\sum_{b=1}^{s-1} h_{b,k} v_{b+i}^{T} + a_k\right) \tag{3}$$

In the fusion process of images and text, we set X_i, Y_i the image and text features extracted after careful preprocessing, $i = 1 \sim N$. Based on the assumption that the dimensions of these feature vectors remain consistent, one of the most intuitive methods is to use a direct superposition strategy, which adds the feature vectors of the two and then uses convolution operations to learn adaptive weight allocation. This process is carried out without changing the original basic network architecture. In the fusion step, to improve the consistency and comparability of features, we perform regularization on the embedded image features $S(X_i)$ and text features $S(Y_i)$ from these two branches. Afterwards, the fused visual feature $f(X_i)$ and text feature $g(Y_i)$ can be accurately calculated using the following formula:

$$g(Y_i) = W_T^{fuse} \otimes S(Y_i) + b_T^{fuse} \tag{4}$$

$$f(X_i) = W_I^{fuse} \otimes S(Y_i) + b_I^{fuse} \tag{5}$$

Among them, W_I^{fuse}, W_T^{fuse} is the fusion weight learned in multiple training sessions, b_I^{fuse}, b_T^{fuse} is the corresponding image and text bias vector, and $\otimes$ represents the inner product operation.

By ingeniously transforming the time-domain information in audio data into a frequency-domain representation, we can generate a frequency-domain graph. This transformation process not only reveals the rich characteristics of audio signals in the frequency dimension but also greatly promotes the efficiency and accuracy of audio samples in the training process. Specifically, Mel Frequency Cepstral Coefficients (MFCC), as an efficient audio feature extraction method, deeply reflects the nonlinear characteristics of human auditory perception in its relationship with frequency f, further enhancing the performance and robustness of audio recognition systems.

$$f_{mel}(f) = 2595 \times \log 10\left(1 + \frac{f}{700}\right) \tag{6}$$

$$X\ k\ = \sum_{i=0}^{n-1} x\ i\ e^{-\sqrt{2\pi}\frac{k}{n}}, 0 \le k \le n \tag{7}$$

In the formula, $x\ i$ is the input speech signal, and n is the number of Fourier transform points.

After in-depth learning of image features and audio information, we obtained a high-level image feature set $P = \ p_1, p_2, \cdots, p_t$ that contains rich semantic information; And the corresponding high-level speech feature set $V = \ v_1, v_2, \cdots, v_t$. To capture complementary information across modalities more comprehensively, we cleverly concatenate these two sets of high-level feature vectors to construct a multimodal fusion feature vector D .

$$D = \left[P, V\right] \tag{8}$$

5 EXPERIMENTS ANALYSIS AND DISCUSSION

To verify the effectiveness of the algorithm proposed in this article, experiments will be conducted next. Figure 3 visually presents the comparison results of the image recognition performance between our algorithm and traditional ML algorithms in the field of creative packaging design. As can be seen from the figure, compared to traditional ML algorithms, the multimodal fusion algorithm designed in this paper demonstrates significant advantages in image recognition accuracy. This improvement is mainly due to the powerful feature extraction and representation capabilities of DL, as well as the effective application of multimodal fusion strategies. The advantages of this algorithm are particularly evident in the complex application scenarios of creative packaging design.

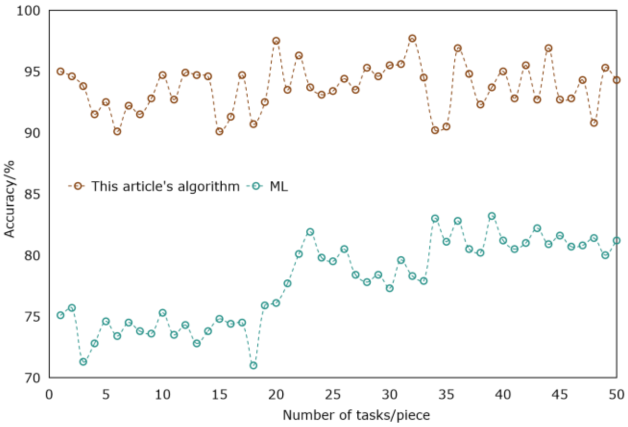

Figure 3: Comparison of recognition accuracy.

In the in-depth analysis of Figure 4, we compared in detail the task processing efficiency of the innovative multimodal fusion algorithm proposed in this paper with traditional machine learning algorithms in addressing the challenges of creative packaging design. It not only fundamentally accelerates the design process, allowing designers to complete preliminary designs in a shorter amount of time, but also greatly improves the efficiency of design iterations. Compared to traditional ML methods, our carefully designed multimodal fusion algorithm has demonstrated overwhelming advantages in task execution speed, manifested in significantly reduced processing time. This time advantage creates valuable thinking space for designers, allowing them to focus more on stimulating and optimizing their creativity rather than being constrained by lengthy calculation processes. The core driving force behind this leap in time efficiency can be attributed to the precise optimization strategy of deep learning models and the high efficiency of multimodal information fusion mechanisms. In the highly time-sensitive field of creative packaging design, the fast response capability of the algorithm proposed in this article is particularly important.

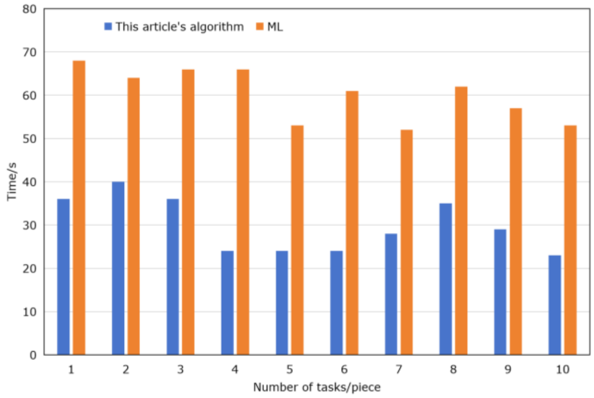

Figure 4: Comparison of task processing time.

Under the detailed analysis in Figure 5, the recall advantage of our algorithm in creative packaging design tasks is demonstrated, which is particularly significant compared to traditional machine learning algorithms. It significantly surpasses traditional methods, demonstrating its strong recognition ability and wide applicability in complex design tasks. From the data trends presented in the chart, our carefully designed multimodal fusion algorithm demonstrates excellent performance in terms of recall rate. It means that our algorithm can more comprehensively and deeply capture the subtle intentions of designers and the diverse needs of consumers, effectively reducing the omission and misjudgment of design elements or concepts. This high degree of precision not only directly improves the overall quality and market satisfaction of design works but also provides designers with rich materials and sources of inspiration, inspiring their infinite creativity and imagination. It also indicates its enormous potential in promoting the digital transformation of the design industry, improving design production efficiency, and optimizing user experience. In the field of creative packaging design, which emphasizes both creativity and practicality, the importance of a high recall rate is self-evident. Furthermore, from a broader perspective, multimodal fusion algorithms have shown outstanding performance in recall rate.

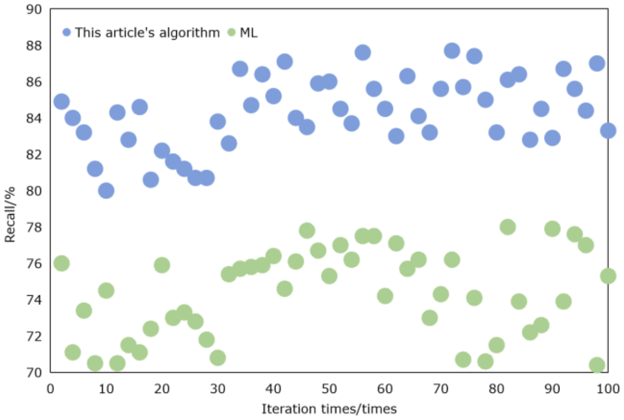

Figure 5: Comparison of recall rates.

In sharp contrast to Figure 6, the algorithm proposed in this paper demonstrates unparalleled advantages in improving user satisfaction in the field of creative packaging design. This algorithm can deeply understand the core requirements of design tasks and keenly capture users' subtle personalized preferences. By integrating information from multiple modalities such as images, text, and even user behaviour, a more diverse and highly fitting design space is constructed that meets user expectations. This achievement is not only a strong proof of its progressiveness in technology but also a direct reflection of its positive contribution to improving the user experience. Behind this significant advantage is the unique ability of multimodal fusion algorithms to play a crucial role. The data in the figure reveals the fact that compared to traditional machine learning algorithms, the creative designs generated by our innovative multimodal fusion algorithm have significantly improved user satisfaction. This comprehensive and deep-level information processing capability enables our algorithm to create design works that are both creative and personalized, thus winning high recognition and satisfaction from users. In addition, the introduction of DL models greatly enhances the creativity and adaptability of algorithms. It can automatically learn and optimize the combination and layout of design elements, generating design works that are both aesthetically pleasing and innovative.

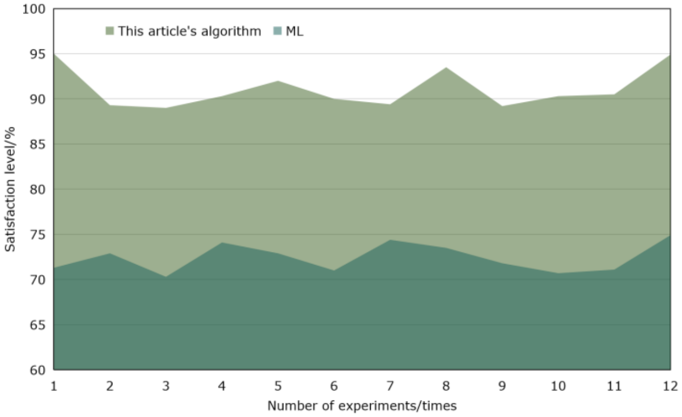

Figure 6: Satisfaction comparison.

Figure 7 profoundly underscores the superiority of the algorithm introduced herein in design accuracy within the realm of creative packaging design, contrasting it with conventional ML algorithms. As a pivotal metric, design accuracy quantifies the congruence between algorithmic outputs and design objectives, thereby influencing the exactitude and dependability of design outcomes. The algorithm presented demonstrates marked enhancements in this regard. This high-precision implementation is mainly attributed to the powerful data analysis and pattern recognition capabilities of DL technology, as well as the comprehensive integration and utilization of information through multimodal fusion strategies. In the practical application of creative packaging design, high design precision means that algorithms can more accurately capture the designer's intentions, reduce design errors and deviations, and improve the success rate and market acceptance of design works.

Figure 8 shows in detail the significant improvement in brand recognition of our algorithm in creative packaging design compared to traditional ML algorithms. Brand recognition is an important indicator to measure whether a design work can effectively convey brand information and enhance brand impression, which is crucial for enhancing brand value and market competitiveness. The graph provides a clear illustration that the algorithm featured in this article exhibits remarkable proficiency in the realm of brand recognition. This advantage is mainly due to the efficient processing capability of DL models for complex data and the deep integration of multi-source information through multimodal fusion strategies. The DL model can deeply explore brand features in the image, text and other data, including brand logos, colour styles, copywriting expressions, etc., to accurately identify and strengthen brand elements in the design. Multimodal fusion further enriches the expression of brand information, enabling design works to maintain consistency and coherence with the brand in multiple dimensions, such as visual and textual aspects.

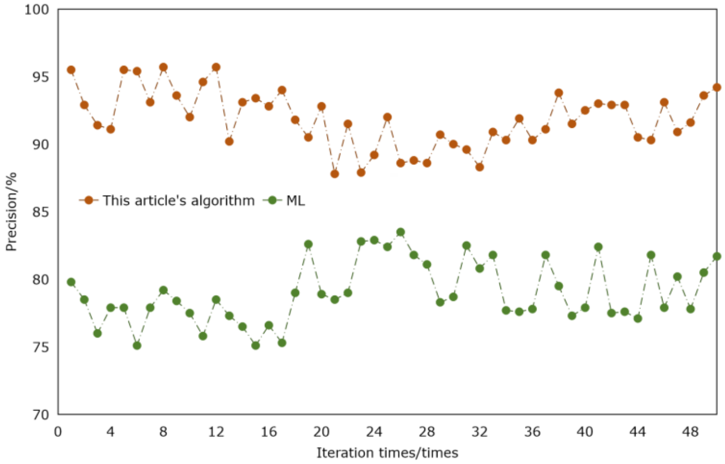

Figure 7: Comparison of design precision.

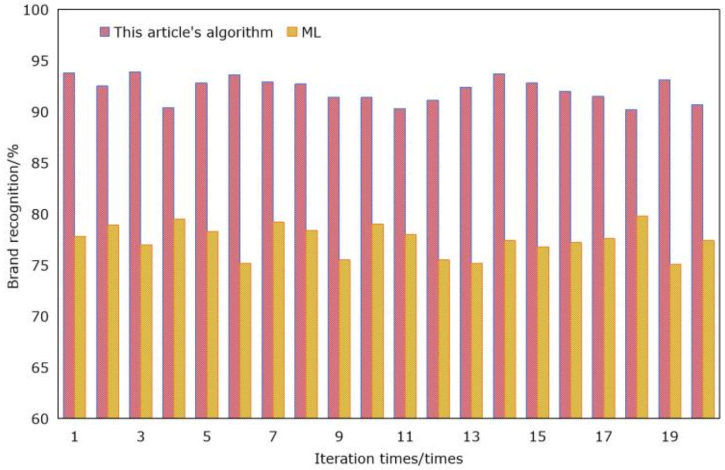

Figure 8: Comparison of brand recognition.

6 CONCLUSIONS

This article deeply explores the innovative application of multimodal fusion in the field of creative packaging design under the CAD technology framework, and innovatively proposes a deep fusion deep learning (DL) multimodal integration algorithm. At the same time, it also brings consumers an unprecedented and unique visual feast, redefining the aesthetic standards and experiential levels of packaging design. This milestone achievement has injected unprecedented vitality and transformative power into the field of packaging design. The packaging design concept born from this is not only filled with infinite imagination but also precisely meets the diversified needs of the current market. It not only greatly broadens the designer's creative vision and boundaries, but also sparks unprecedented creativity. This algorithm, with its excellent data manipulation and fusion capabilities, cleverly integrates multiple modalities of input information such as images, text, and colour tones in a complex and sophisticated manner, achieving seamless integration and deep fusion between creative design inspiration and market demand. The full verification of experimental data irrefutably demonstrates the outstanding effectiveness of the algorithm in enhancing packaging design creativity and personalization.

Xue Dong, https://orcid.org/0009-0004-9451-2695
Xiaofeng Zhao, https://orcid.org/0009-0009-9322-5906

REFERENCES

[1] Bahlau, J.; Lee, E.: Designing moulded pulp packaging using topology optimization and superimpose method, Packaging Technology and Science, 35(5), 2022, 415-423. https://doi.org/10.1002/pts.2639
[2] Blösch, P.-A.; Shea, K.: Enhancing creative redesign through multimodal design heuristics for additive manufacturing, Journal of Mechanical Design, 143(10), 2021, 102003. https://doi.org/10.1115/1.4050656
[3] Brenes, A.; Marín, R.-G.; López, G.: Improving packaging design using virtual reality in the market research process, Multidisciplinary Digital Publishing Institute Proceedings, 31(1), 2019, 12. https://doi.org/10.3390/proceedings2019031012
[4] Fu, L.; Xu, M.: Network assisted agricultural product packaging design based on virtual reality, Computer-Aided Design and Applications, 20(13), 2023, 124-135. https://doi.org/10.14733/cadaps.2023.S13.124-135
[5] Georgakarakou, C.; Riskos, K.; Tsourvakas, G.; Yfantidou, I.: What features of green product packaging are more eye-catching? An eye-tracking exploratory study about organic agricultural products, International Journal of Technology Marketing, 14(2), 2020, 93-124. https://doi.org/10.1504/IJTMKT.2020.110124
[6] Idumah, C.-I.; Zurina, M.; Ogbu, J.; Ndem, J.-U.; Igba, E.-C.: A review on innovations in polymeric nanocomposite packaging materials and electrical sensors for food and agriculture, Composite Interfaces, 27(1), 2020, 1-72. https://doi.org/10.1080/09276440.2019.1600972
[7] Park, J.; Park, M.; Choi, D.-S.; Jung, H.-M.; Hwang, S.-W.: Finite element-based simulation for edgewise compression behavior of corrugated paperboard for packaging of agricultural products, Applied Sciences, 10(19), 2020, 6716. https://doi.org/10.3390/app10196716
[8] Qiu, Y.; Zhang, Y.: Aided Product Modeling Design Based on Virtual Reality Technology, International Journal of New Developments in Engineering and Society, 5(3), 2021, 10-25. https://doi.org/10.25236/IJNDES.2021.050302
[9] Rodríguez, P.-L.; Mayuet, P.-F.; Gámez, A.-J.: Custom design of packaging through advanced technologies: A case study applied to apples, Materials, 12(3), 2019, 467. https://doi.org/10.3390/ma12030467
[10] Uzelac, O.; Mijatović, M.-D.; Lukinović, M.: The role of branding agricultural products in better market valorization, Economics of Agriculture, 69(2), 2022, 613-625. https://doi.org/10.5937/ekoPolj2202613U
[11] Vasileiadis, T.; Tzotzis, A.; Tzetzis, D.: Combining product and packaging design for increased added value and customer satisfaction, Journal of Graphic Engineering and Design, 10(2), 2019, 5-15. https://doi.org/10.24867/JGED-2019-2-005
[12] Wang, E.; Liu, Z.; Gao, Z.; Wen, Q.; Geng, X.: Consumer preferences for agricultural product brands in an E-commerce environment, Agribusiness, 38(2), 2022, 312-327. https://doi.org/10.1002/agr.21732
[13] Wang, K.; Song, Y.; Sheng, H.: Energy efficiency design for eco-friendly additive manufacturing based on multimodal attention fusion, Journal of Manufacturing Processes, 79(1), 2022, 720-730. https://doi.org/10.1016/j.jmapro.2022.05.010

[14] Yang, W.; Su, J.; Qiu, K.: Supporting computer-aided product form design research with a cognitive model of the creative process, Multimedia Tools and Applications, 81(15), 2022, 21619-21639. https://doi.org/10.1007/s11042-022-12119-4

[15] Zhu, Z.; Shen, Q.; Gao, Z.: Consumer choices in agricultural markets with multitier collective labels and private brands, Agribusiness, 38(4), 2022, 905-922. https://doi.org/10.1002/agr.21747