INDERSCIENCE PUBLISHERS
Linking academia, business and industry through research

International Journal of Data Science

ISSN online: 2053-082X - ISSN print: 2053-0811
https://www.inderscience.com/ijds

The analysis of image recognition for tourism cultural and creative products visual design based on deep learning

Wei Zhang, Yang Yu, Enhong Liu

DOI: 10.1504/IJDS.2025.10072738

Article History:

Received:	07 April 2025
Last revised:	17 June 2025
Accepted:	26 June 2025
Published online:	16 January 2026

The analysis of image recognition for tourism cultural and creative products visual design based on deep learning

Wei Zhang*

Jilin Province Research Center for Cultural Tourism Education
and Enterprise Development,
The Tourism College of Changchun University,
Changchun, 130000, China
Email: ccdxlyxyzw@163.com
*Corresponding author

Yang Yu

School of Artificial Intelligence,
The Tourism College of Changchun University,
Changchun, 130000, China
Email: 450341758@qq.com

Enhong Liu

School of Tourism and Culture,
The Tourism College of Changchun University,
Changchun, 130000, China
Email: 1007743933@qq.com

Abstract: This work explores consumer demand and preferences for cultural and creative tourism products to enhance the effectiveness of visual design and market competitiveness. A comprehensive model integrating the deep convolutional neural network (DCNN) and deep belief network (DBN) is developed using deep learning technology. This model aims to extract both the underlying features of product images and the semantic features of consumers, thereby providing data support to optimise product design. The results indicate that the constructed model achieves a prediction accuracy of 98.5% and a recall rate of 98.2% in product image recognition, demonstrating its effectiveness in capturing consumer demand characteristics.

Keywords: deep learning; image analysis; product visual design; cultural and creative products; DBN; deep belief network.

Reference to this paper should be made as follows: Zhang, W., Yu, Y. and Liu, E. (2025) 'The analysis of image recognition for tourism cultural and creative products visual design based on deep learning', *Int. J. Data Science*, Vol. 10, No. 7, pp.114–135.

Biographical notes: Wei Zhang received the Master's degree from JiLin Normal University, China. Now, she works in The Tourism College of Changchun University. Her research interests include tourism education, tourism product design and smart tourism.

Yang Yu received the Master's degree from Changchun University, China. Now, she works in The Tourism College of Changchun University. She is committed to advancing research in the field of deep learning.

Enhong Liu received the Master's degree from Jilin University of Arts, China. Currently, she works at the School of Tourism Culture, Changchun University Tourism College, specialising in Cultural Industry Management. Her research interests include tourism cultural and creative product design, new media communication, and related fields.

1 Introduction

With the continuous development of internet technology, the demand for data across various industries is steadily increasing, making data analysis vital in modern design and production processes (Mortati et al., 2023; Chatterjee et al., 2024; Ogundipe et al., 2024). As a key medium connecting consumers with cultural significance, cultural and creative products require not only aesthetic considerations in design but also a deep understanding of market demands and consumer behaviour (Chaudhuri et al., 2024). Therefore, leveraging advanced technologies to improve design quality and enhance product market appeal has become a core challenge in the cultural and creative industries (Abbas et al., 2024). Image has become indispensable tools in various fields, including industry, computing, military, and biomedical sectors. As the volume of image data continues to grow, researchers increasingly rely on extracting relevant information from images to guide their work, leading to successful applications across many areas. However, its use in product visual design remains somewhat underdeveloped (Lee et al., 2017; Ran et al., 2024).

In the era of big data, deep learning (DL) is an important represents a significant milestone in machine learning, consistently driving scientific research forward (Li et al., 2018). DL techniques have proven highly effective in analysing images and videos, with widespread applications across various devices, yielding promising results in multiple domains (Kato et al., 2016). In recent years, deep convolutional neural networks (DCNN) have been applied in fields such as healthcare, military, industrial inspection, and remote sensing. Their exceptional feature extraction and automated analysis capabilities provide efficient solutions for complex data processing (Li and Wang, 2022). Particularly in image analysis, DL can extract multi-level feature representations from raw data through multi-layer neural networks, offering more precise data support for decision-making processes (Han et al., 2022). However, despite the significant achievements of DCNN in these areas, its application in designing cultural and creative products remains in its early stages.

As a sub-sector of the cultural and creative industry, tourism-related cultural and creative products are facing increasingly diverse market demands due to the growth of the tourism industry. Consumers are setting higher standards for product visual design, which must reflect cultural significance while incorporating functionality and market competitiveness (Sun, 2022). However, traditional design methods often rely on designers' experience and intuition, making it difficult to effectively capture potential consumer needs. This presents an opportunity for the introduction of DCNN (Lu et al., 2022). By analysing video image data from user interviews, DCNN can extract visual features from multiple layers and perspectives, optimising the product design process, reducing costs, and enhancing design quality (Wang, 2024).

This work aims to explore the application of DCNN in the visual design of tourism-related cultural and creative products. It adopts an integrated approach combining DCNN and deep belief network (DBN) for image recognition and feature extraction. Specifically, it merges DCNN's efficient feature extraction capabilities with DBN's strong representation power in unsupervised learning to create a cohesive model. This model analyses the deep features of cultural and creative product images, improving recognition accuracy and processing efficiency. The goal is to provide new technical support for the design process of cultural and creative products, enhancing their market adaptability and human-centred design to better meet consumer demands.

The innovation of this work is evident in several key aspects: First, it is the first to apply DCNN, particularly the combination of DCNN and DBN, to the visual design of tourism-related cultural and creative products, filling a technical gap in this field. Second, by optimising the network structure and training strategies, it improves the model's recognition accuracy for cultural and creative product images, leading to higher-quality design outputs. Lastly, this work not only focuses on technical implementation but also explores the practical value of DL in the design process of cultural and creative products, promoting the development of more intelligent, human-centred design processes and aiming to bring new opportunities and challenges to the cultural and creative industry.

2 Literature review

2.1 Image recognition

Image analysis has become a critical tool for researchers across various fields to extract valuable information. Wen et al. utilised number theory-based (Wen et al., 2018) image analysis to study historical climate change. They discovered that the greyscale of spruce tree rings in Western Sichuan was primarily influenced by early spring high temperatures, exhibiting a negative correlation. Naylor preprocessed and extracted key parameters from images of tea leaves, applying SVM to classify the extracted data as a parameter set (Naylor, 2018). This enabled preliminary identification of tea quality through traceability analysis, demonstrating the efficacy of image analysis (Wang et al., 2017). Fei et al. (2019) used computer processing technology to analyse images in webpage design, enhancing the timeliness, aesthetics, and overall quality of web content.

2.2 DL

Bankhead et al. (2017) pointed out that DL had become the preferred method for analysing medical images. Their review highlighted the significant contributions of DL technologies in the medical field, demonstrating effective applications in analysing images of the nervous system, retina, heart, muscles, and skeleton. Primpke et al. (2017)

introduced the use of single-layer and DCNN in remote sensing data analysis, proposing the application of greedy layer-wise unsupervised pre-training and efficient algorithms for unsupervised learning of sparse features. The results showed that a single-layer convolution network could extract strong discriminant features only when the perceptual domain covered nearby pixels, while DCNN was capable of capturing more abstract and complex features, ultimately enhancing computational efficiency. Ye found that DL-based image retrieval systems effectively addressed the 'semantic gap' present in traditional content-based image retrieval systems by comparing low-level retrieval technologies with DL methods for extracting meaningful image features and similarity matching. Ding discussed the application of DL in medical image analysis in detail from the perspectives of medical image segmentation, medical image recognition, medical image recognition and computer-aided diagnosis (Ye et al., 2019; Ding et al., 2019). The findings indicated that deep learning technology could effectively extract valuable information from medical images, assisting doctors in diagnosing and treating patient conditions.

2.3 DL and visual design

DL has nearly become the standard in computer vision research, with applications in face recognition, image recognition, video recognition, pedestrian detection, and large-scale scene recognition. Product development, a complex process, must be tailored to the specific company, product, and designers involved. In this process, designers' creative ideas and expertise are highly individualised, typically expressed through rule-based approaches. Machine learning, particularly DL, has proven effective in recognising patterns and extracting knowledge from complex datasets. Krahe et al. (2020) proposed that autoencoder networks are well-suited for transforming various 3D input data, such as point clouds, into compact latent representations. Point clouds, a common representation of 3D objects, can be derived from multiple 3D data formats. The goal of the proposed DL method was to identify product family-specific design patterns from potential representations and use the extracted knowledge to automatically generate new object representations that meet different product feature specifications. Burnap et al. used DL to enhance human judgement in designing and testing new product aesthetics. Their model combined a probabilistic variational autoencoder with a generative adversarial network (GAN), ensuring the system met both the company's modelling assumptions and product design requirements (Burnap et al., 2021; Quan and Lu, 2024; Yuan et al., 2025).

Deep neural network (DNN) technology has been applied to image retrieval, utilising its non-linear mapping, self-learning, and adaptive capabilities to enable the network to learn visual features directly from the image content, similar to the human brain. The features used in retrieval consist of two parts: the underlying features learned from the images during pre-training and the category information output by the fine-tuning model, which is derived from labelled data and aligns more closely with human descriptions of the image. Consequently, the retrieval features encompass both low-level visual features and high-level semantic features.

Product visual design must integrate market demand, product functionality and expandability, user reputation, and other factors. Previous studies primarily relied on imitation and feedback loops, with limited use of relevant data to inform the design process. Although DCNN has achieved significant results in various fields, its application in cultural and creative product design remains relatively rare. Some research has begun to explore how to leverage DCNN to enhance the design process's intelligence. For example, Belhi et al. (2023) investigated the application of DL in generating artistic works, using DCNN to create culturally characteristic visual works. Their findings highlighted the potential of DL in the cultural and creative field. Additionally, Gaber et al. (2023) examined the role of DL in product design, emphasising the importance of automated feature extraction and design optimisation. They noted that DL could help designers better understand consumer preferences, thereby enhancing product competitiveness in the market.

However, despite preliminary explorations of DL in cultural and creative product design, there is still a lack of systematic research in this area. The current literature lacks an in-depth analysis of the innovative applications of DCNN in this specific field, particularly in combining DCNN and DBN for image recognition and feature extraction. Therefore, this work aims to fill this gap. It explores the specific applications of DCNN in the design of tourism-related cultural and creative products and investigates its potential contributions to optimising the design process, reducing costs, and improving product competitiveness (Ma et al., 2024).

Despite preliminary explorations of DL in cultural and creative product design, systematic research in this area remains lacking. The current literature does not provide an in-depth analysis of innovative applications of DCNN particularly in combining DCNN and DBN for image recognition and feature extraction. This work aims to address this gap by exploring the applications of DCNN in designing tourism-related cultural and creative products and investigating its potential to optimise the design process, reduce costs, and enhance product competitiveness (Ma et al., 2024).

3 Proposed method

3.1 Image analysis in cultural and creative products

Image analysis is vital in various fields, especially in designing and developing cultural and creative products. This technology uses mathematical models and image processing techniques to analyse visual features, extracting valuable information to enhance market appeal and consumer experience. Image analysis examines not only the subject's content but also its interpretation, analysis, and recognition. This process is essential for designers to understand consumer needs better during product development (Pelt and Sethian, 2018). The main steps in image analysis include input, segmentation, recognition, and interpretation. Consequently, pattern recognition and computer vision are closely linked to image analysis and computer science. Figure 1 illustrates a typical hierarchical image analysis model.

In image analysis, sensor input converts real-world scenes into representations suitable for computer processing. Segmentation extracts objects and their components from the scene, composed of image primitives. Recognition assigns names to the identified objects in the segmented image, classifying them based on shape and greyscale information through decision theory and structural methods. Alternatively, a set of image models for known objects can be constructed, allowing for comparison and matching of the objects to be recognised against these models. Interpretation employs appropriate techniques to identify the objects present in the scene and their relationships.

Figure 1 Hierarchical image analysis model

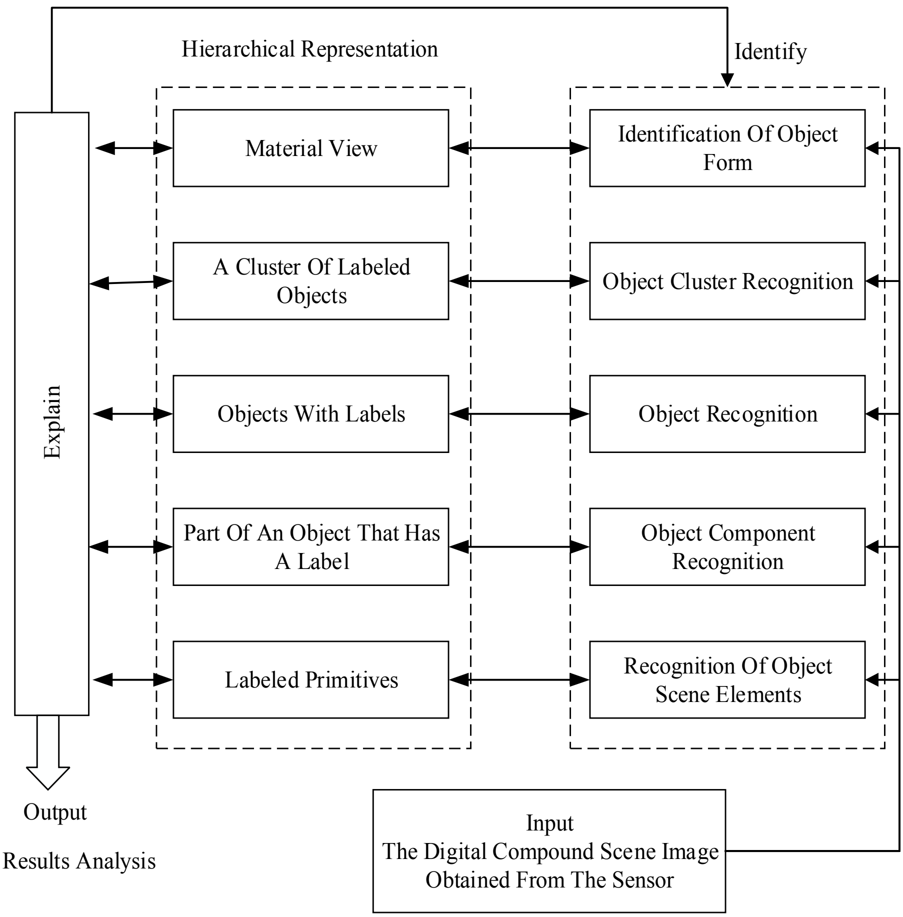


Image analysis technology has been widely applied and has achieved promising results in industries such as manufacturing, medicine, bioinformatics, and the military. However, its potential in art and cultural and creative product design remains underexplored, with hidden information within images often insufficiently analysed. By utilising DCNN, it is possible to better extract and analyse the visual features of cultural and creative products. This technology enhances image recognition accuracy and helps designers gain deeper insights into the relationship between a product's market positioning and consumer preferences, fostering more human-centred and intelligent product design. In this context, image analysis technology offers a new perspective for designing cultural and creative products, enabling designers to leverage intelligent algorithms for in-depth insights into market trends and consumer demands, ultimately leading to the development of more competitive products.

3.2 Visual design of domestic cultural and creative products

In daily life, household appliances, furniture, kitchenware, bathrooms, and decorative products are closely tied to people's lives. The function, structure, appearance, and design of these appliances have become a focus of research for enterprises. In product visual design, innovation in functionality directly influences the design of specific features, while functional aspects also impact appearance design (Aykut et al., 2019; Zeng et al., 2019). The integration of technological aesthetics and functionality will be crucial for driving product sales and promoting technological advancement, ultimately enhancing people's quality of life (Biswas et al., 2019).

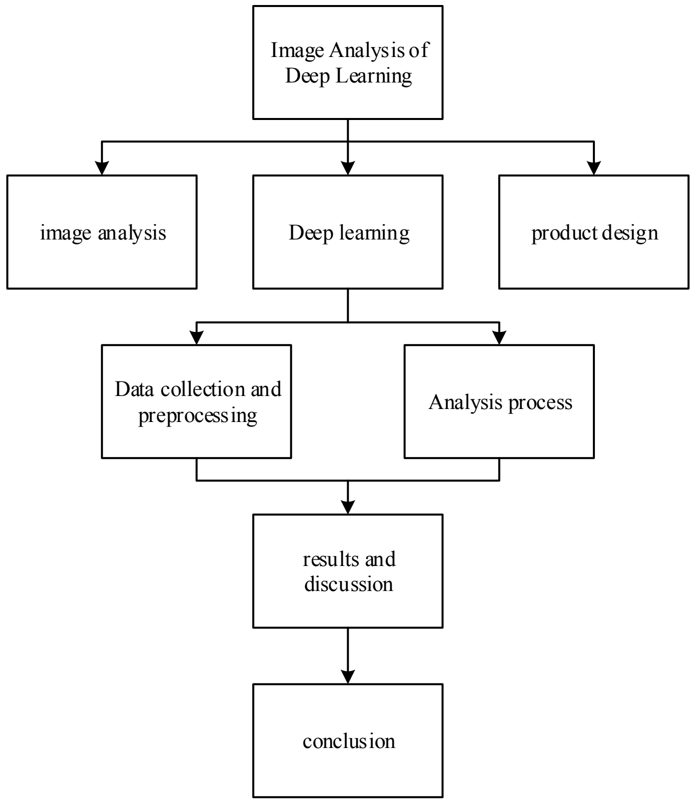
The quality and appearance of everyday items directly affect consumers' purchasing decisions. After using domestic products, many individuals often consider foreign alternatives for future purchases. This shift is largely due to the perception that some domestic products are underdeveloped or of lower quality, leading to a generalised misconception that all such products have quality issues. Consequently, this perception undermines consumer confidence in the broader category of domestic goods. In actual production, material selection and craftsmanship significantly influence product quality. Sometimes, companies choose low-cost materials to cut expenses, which can negatively affect the product's lifespan (Dabiri et al., 2019; Kan et al., 2019). The typical product visual design process includes several stages: user research, demand analysis, conceptual design, functional design, process design, prototype design, and more. User research is the first step in the user-centred product visual design process, effectively helping to understand users and align their goals, needs, and commercial purposes. It benefits both product design and consumers (Shen et al., 2019), creating a win-win scenario. For product visual design, user research can save time and resources, reduce costs, and lead to better products. For users, it ensures that products align more closely with real-life needs. By gaining insights into user behaviours and preferences, designers can create practical, user-friendly features that address real-world problems (Zhu et al., 2019).

This work analyses video image data obtained during user interviews to gain a deeper understanding of users' purchasing needs for tourism-related cultural and creative products. This analysis aims to provide crucial support for the visual design of cultural and creative products. Through this data-driven approach, the goal is to achieve more human-centred and intelligent designs, ultimately enhancing the market competitiveness of the products.

3.3 DBN

The advantage of DL is its ability to use unsupervised or semi-supervised feature learning and efficient hierarchical feature extraction algorithms instead of manual feature acquisition, as shown in Figure 2.

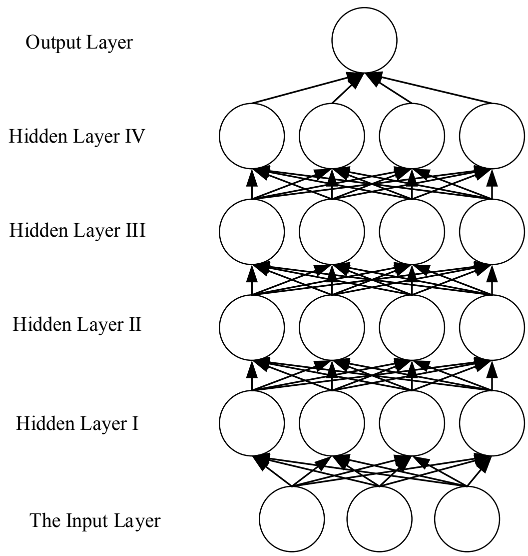
Figure 2 Systematic process diagram



DBN algorithm is a kind of neural network, which can be used in unsupervised learning and supervised learning (Xu, 2022). This algorithm is actually a probability generation model, which establishes a joint distribution between observation data and their corresponding labels (Luo et al., 2021). By training the weights among the neurons, the whole neural network can generate training data based on maximum likelihood (Hatvani et al., 2018). DBN not only identifies features and classifies data but also generates new data. As a highly practical learning algorithm, it boasts a broad range of applications and significant scalability, making it suitable for tasks such as handwriting recognition, speech recognition, and image processing in the field of machine learning. Figure 3 displays a typical DBN model.

The DBN algorithm is a type of neural network used in both unsupervised and supervised learning (Xu, 2022). It functions as a probability generation model, establishing a joint distribution between observation data and their corresponding labels (Luo et al., 2021). By training the weights of the neurons, the neural network can generate training data based on maximum likelihood (Hatvani et al., 2018). DBN not only identifies features and classifies data but also generates new data. As a highly practical learning algorithm, it offers a wide range of applications and significant scalability, making it suitable for tasks such as handwriting recognition, speech recognition, and image processing in machine learning. Figure 3 displays a typical DBN model.

Figure 3 DBN model



The DBN consists of two main components: the multi-layer Boltzmann perceptron and the feedforward backpropagation network (Zhao, 2019). The multi-layer Boltzmann perceptron is primarily used for pre-training the network, while the feedforward backpropagation network refines the network built on the Restricted Boltzmann Machine (RBM). Figure 4 illustrates the structure of the RBM model.

Figure 4 RBM model

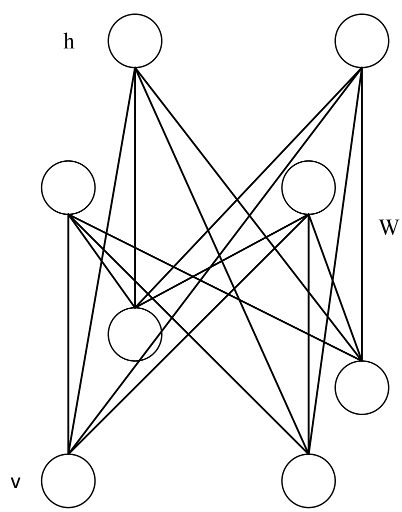


Figure 4 shows that there are no connections between the hidden layer nodes or the visible layer nodes. Energy functions are often used to describe the state of the entire system, as shown in equation (1). A more orderly network or a more concentrated probability distribution corresponds to lower energy levels, while a more disordered network or a less centralised probability distribution results in higher energy levels. Therefore, the energy function reaches its minimum value when the network is in its most stable state.

$$E(v,h|\theta) = -\sum_{i=1}^{n}\sum_{j=1}^{m} W_{ij} h_i v_j - \sum_{j=1}^{m} b_j v_j - \sum_{i=1}^{n} a_i h_i \tag{1}$$

In equation (1), m denotes the number of visible nodes, and n denotes the number of hidden nodes. $\theta = \{W, b, a\}$ represents the parameters of the RBM model. Vector V denotes the state of visible nodes, and vector h denotes the state of hidden nodes. b_j and A_j denote the offset of visible node j and the offset of visible node i respectively, and W_{ij} denotes the connection matrix between visible node j and hidden node I. When the model parameters are given, the joint probability distribution can be obtained (Luo et al., 2019).

$$P(v,h|\theta) = \frac{e^{-E(v,h|\theta)}}{Z(\theta)} \tag{2}$$

$$Z(\theta) = -\sum_{v,h} e^{-E(v,h|\theta)} \tag{3}$$

In the equation, $P(v,h|\theta)$ is a Boltzmann function and $Z(\theta)$ is a normalised function, thus the probability of the visible node V can be obtained.

$$P(v|\theta) = \frac{1}{Z(\theta)} \sum_{h} \exp\left(-E(v,h|\theta)\right) \tag{4}$$

In practical research, the distribution of state conditions between hidden layer nodes and visual nodes is independent, as shown in equations (5) and (6).

$$P(v|h) = \prod_{i=1}^{n} P(v_i|h) \tag{5}$$

$$P(h|v) = \prod_{j=1}^{m} P(h_i|v) \tag{6}$$

When the state of the visual node is provided, the activation probability of the jth hidden layer node is given in equation (7), where σ the sigmoid activation function is represented. Equation (8) defines this function. Due to the symmetry of all hidden layer nodes and the model, the activation probability of the visual node is obtained as shown in equation (9).

$$P(h_i = 1|v,\theta) = \sigma\left(b_j + \sum_{j} v_i W_{ij}\right) \tag{7}$$

$$\sigma(x) = \frac{1}{1+\exp(-x)} \tag{8}$$

$$P\left(v_i = 1 \middle| h, \theta\right) = \sigma\left(a_i + \sum_j h_j W_{ij}\right) \tag{9}$$

On this basis, it is also necessary to maximise the logarithmic likelihood of RBM on the training set to get its parameter θ:

$$\theta^* = \arg\max_{\theta} \sum_{t=1}^{T} \log P\left(v^t \middle| \theta\right) \tag{10}$$

When using the maximum likelihood function to train the values on the data, the update criteria of W_{ij}, b_j and A_j are as follows:

$$\Delta W_{ij} = \varepsilon\left(\left\langle v_j h_i \right\rangle_{data} - \left\langle v_j h_i \right\rangle_{recon}\right) \tag{11}$$

$$\Delta a_i = \varepsilon\left(\left\langle h_i \right\rangle_{data} - \left\langle h_i \right\rangle_{recon}\right) \tag{12}$$

$$\Delta b_j = \varepsilon\left(\left\langle v_j \right\rangle_{data} - \left\langle v_j \right\rangle_{recon}\right) \tag{13}$$

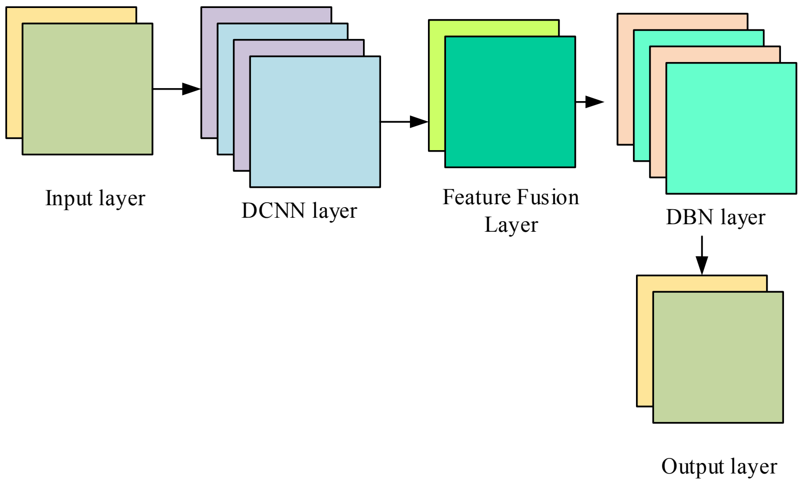
3.4 The use of DL integrated models in image recognition of cultural and creative products

The convolutional neural network (CNN) has three crucial characteristics: local connection, weight sharing, and sub-sampling in time or space (Zhu and He, 2017). These features enable the CNN to process a small region of the image, known as the local receptive field, as input, subsequently passing the information through each layer in sequence (Wei et al., 2019). Each layer identifies the most significant features of the data using a specific digital filter. This characteristic allows the CNN to possess a certain degree of translation, scaling, and distortion, which can enhance the accuracy and adaptability of recognition. Based on the CNN, an accurate and effective image evaluation model is established. Furthermore, a GAN is employed to create an emotional product design scheme tailored to specific product images, aligning with enterprise requirements for rapid product iteration.

To achieve a more accurate and efficient evaluation of cultural and creative product images, this work has developed an integrated model (Figure 5) that combines DCNN and DBN. The model aims to fully leverage the advantages of DCNN to improve the accuracy and efficiency of product image recognition.

Figure 5 illustrates the structure of the integrated model, which includes the following components. The input layer receives image data. The DCNN layer consists of multiple convolutional layers, pooling layers, and fully connected layers, responsible for extracting high-level features from the images. The convolutional layers efficiently capture features such as edges and textures through local perception and weight-sharing mechanisms. The feature fusion layer merges the high-level features extracted by the DCNN with the low-level features captured by the DBN, providing a more comprehensive representation of the image information. The DBN layer learns deep features by stacking multiple RBMs, further refining and optimising the feature representation. The output layer produces the final classification or regression results based on task requirements.

Figure 5 DL integrated model (see online version for colours)



The training process of the integrated model is divided into two stages: First, the DBN undergoes unsupervised pre-training, learning the low-level features of images layer by layer. This process utilises unlabelled data to learn the underlying structure of images through a generated probabilistic model. The goal of this stage is to obtain better initial weights for the model, thereby improving the fine-tuning effect in subsequent steps. Next, the DCNN is connected to the DBN, followed by fully supervised fine-tuning. In this phase, labelled training data are used to optimise the entire integrated model, employing a cross-entropy loss function or other suitable loss functions to guide model learning. The network weights are updated using the backpropagation algorithm to enhance recognition accuracy.

During image recognition, the integrated model first uses the DCNN for initial feature extraction of product sample images. The DCNN enhances this process through key characteristics such as local connections, weight sharing, and spatial downsampling. Its structure includes multiple convolutional and pooling layers; the convolutional layers extract low-level features like edges and textures, while the pooling layers reduce feature dimensions via downsampling, retaining essential information. As the number of network layers increases, the model learns higher-level features such as product shape, colour, and context. Each convolutional layer is followed by a ReLU activation function to introduce non-linearity, enhancing the model's expressive capability.

Once feature extraction is complete, the extracted features are input into the DBN for further optimisation. The DBN effectively captures complex relationships between features through a combination of unsupervised pre-training and supervised fine-tuning across layers. It constructs inter-layer probabilistic graphical models based on the features extracted by the DCNN, facilitating effective feature reconstruction and abstraction. This process improves the model's generalisation ability, resulting in more stable and accurate recognition performance across different product images.

The integrated model includes optimisations for data preprocessing and feature selection to ensure high-quality input data and model adaptability. Given the success of VGGNet in image recognition, VGGNet 16 is used for fine-tuning product image recognition. VGGNet 16 has 16 weighted layers, including 13 convolutional layers and 3 fully connected layers. The network input is the product sample image after preprocessing, followed by a convolution operation between the convolutional kernels and the input image to extract features. Convolution is the primary operation for feature extraction, while the pooling layer enhances translation and size invariance of the features. The network architecture consists of repeated convolution and pooling operations; initial convolutional layers extract low-level features, while deeper layers capture more abstract feature mappings.

Each convolutional layer is followed by a rectified linear unit (ReLU) activation function to increase the network's nonlinearity. Compared to traditional activation functions like sigmoid and tanh, ReLU offers advantages such as higher computational efficiency, avoidance of gradient saturation, and faster convergence. The equation is as follows.

$$f(x) = \max(0, x) \tag{14}$$

The output calculation equation of the *j*th neuron in the *i*th layer is as follows:

$$h_{i,j} = ReLU\left(\sum_{k=1}^{n} W_{i-1,k} X_{i-1,k} + b_{i-1}\right) \tag{15}$$

$W_{i-1,k}$ is the *k*th weight of the *i*–1th layer, $X_{i-1,k}$ is the input of the *k*th neuron of the *i*–1th layer, and b_{i-1} is the bias term of the *i*–1 layer.

Softmax is used as a classifier to classify product images. The calculation equation is as follows:

$$P_p = \frac{1}{1 + \exp(-h_{FC3})} \tag{16}$$

P_p is the probability output of product image, and h_{FC3} is the output of the fully connected layer FC3. According to the probability of the product image, the output of the specific product image of the input sample image can be obtained after the output.

The whole process of network training is to find the minimum value of softmax loss function $J(W)$. First, the experimental samples in the dataset are used as input and propagated forward through the network to calculate the loss function $J(W)$. Then, the random gradient descent algorithm is used for back propagation. The derivative $\frac{\partial J(W)}{\partial W_i}$ of the loss function for each layer weight is calculated. *i* is the number of layers of the network. The weight of each layer is updated and the equation is as follows.

$$W_{i+1} = W_i - \alpha \frac{\partial J(W)}{\partial W_i} \tag{17}$$

Among them, α is the super parameter learning rate, which is used to continuously reduce the loss function to minimise the error of product image recognition.

The cross-entropy is typically used as the loss function when the Softmax function serves as the activation function for the output node. Softmax is utilised for multi-class classification, allowing for the output of classification probabilities. The category with the highest probability corresponds to the predicted class of the input image. TensorFlow provides a unified interface that simultaneously implements the Softmax function and the cross-entropy loss function, effectively managing exceptions related to unstable values. When using the TensorFlow DL framework, it is advisable to use this unified interface to avoid separating the Softmax function and cross-entropy loss function.

Accuracy and recall are employed as evaluation metrics for the performance of the DL model, and the calculation equations are:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \tag{19}$$

TP indicates that the sample is positive and the prediction result is positive; FP indicates that the sample is negative and the prediction result is positive; TN means that the sample is negative and the prediction result is negative; FN indicates that the sample is positive and the prediction result is negative.

Through this multi-level, multi-dimensional feature extraction and optimisation strategy, this work aims to provide intelligent support for designing cultural and creative products, enhancing their market competitiveness. The integrated model demonstrates excellent performance in image recognition tasks and provides a robust data foundation for subsequent emotional product design solutions. This data-driven approach significantly supports product innovation and market adaptability in the cultural and creative industries, promoting the sustainable development of related products.

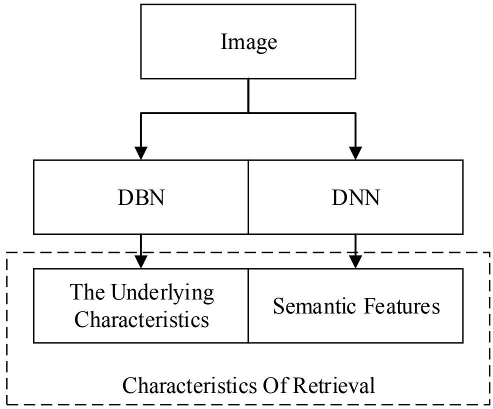
4 Experiments

4.1 Data collection and data preprocessing

To analyse consumers' needs in the visual design process of products, this work focuses on tourism cultural and creative products. By examining video and image data from consumer interviews, it aims to collect data beneficial for product design. The primary methods of data collection include online surveys and offline interviews, ensuring representation from different age groups and backgrounds. This work extracts consumers' views and expectations regarding cultural and creative products from the interviews and identifies the key factors they consider when selecting such products.

In analysing video and image data from consumer interviews, it is essential to extract feature information, categorised into two main types: underlying image features and semantic features. Underlying image features refer to the basic characteristics of the images, while semantic features capture subjective descriptions of consumer demands for products. The underlying image features are derived from a DBN trained on a substantial dataset of images. In contrast, semantic features are extracted using a DNN. Since the DBN serves as the foundational model for the DNN during implementation, more DBN can be obtained in the training process, saving time and improving analysis efficiency. Figure 6 presents the flow chart of image feature extraction.

Figure 6 Image feature extraction process



This work selects 30 categories of images from video data to analyse consumers' preferences and needs for tourism cultural and creative products in depth. Among these categories, there are 1560 labelled images, averaging 20–100 images per category. These images are carefully chosen to represent various tourism cultural and creative products, such as souvenirs, handicrafts, and design items. Additionally, 1105 unlabelled images are available for further unsupervised learning and feature extraction. During data preprocessing, each image is cleaned and cropped to ensure it contains only one main object, removing background clutter and unnecessary items. The final database includes 1560 labelled images, ensuring data quality and consistency. The labelling information for these images encompasses product categories, consumers' emotional responses, and preference characteristics, providing a foundation for subsequent feature extraction and deep learning model training.

The broader the dataset coverage, the more features will be extracted during model training, enhancing the training effectiveness. Therefore, it is crucial first to process the original image data to meet model training requirements and improve adaptability. Following this, text data undergoes cleaning and organisation. Ultimately, image preprocessing is and left-right rotation.

4.2 Parameter settings

The setting of the main parameters in the entire experimental process includes the parameter settings for both the DBN and DNN. The number of input neurons for the DBN is 64×64, the number of output neurons is 3740, and the number of neurons in the hidden layer is 200. The number of network layers is typically 3, 4, or 5, and the number of iterations is 5000. The mean square error is set to 0.01, and the learning rate is set to 0.001.

The parameter settings for the Back Propagation (BP) neural network are as follows: the number of input neurons is 64×64, the number of output neurons is 3740, the number of hidden layer neurons is 200, the number of network layers is 3, and the number of iterations is 5000. The mean square error is set to 0.01, and the learning rate is set to 0.05. When the reconstruction error decreases, the default initial value remains unchanged. When the reconstruction error is unchanged or increasing, half of the current learning rate is taken as the new learning rate. When the learning rate falls below one in

10,000, training is stopped. When setting the initial values, the offset of the hidden layer and output node is set to zero, and the activation probability of the visual layer node is set to Pi.

In the pre-training of deep networks, the size of datasets varies, affecting the network's learning characteristics. A larger dataset allows the network to learn better features, improving retrieval results, while a smaller dataset may hinder the learning of useful features, complicating the retrieval process. Therefore, selecting appropriate datasets for network pre-training is crucial.

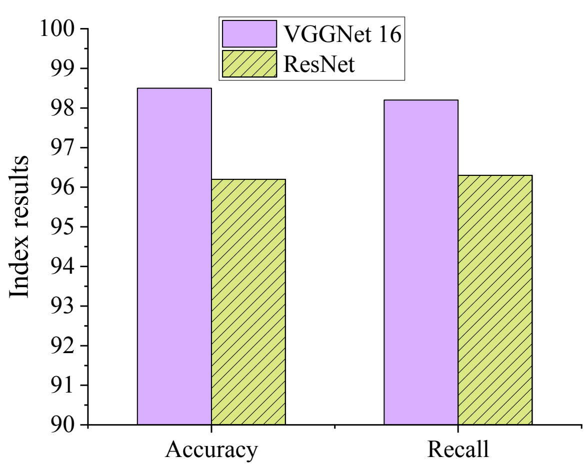
DBN eliminates the need for feature extraction and can directly process original images, aligning more closely with human cognitive habits. During the pre-training of the hidden layers in a DBN, parameters are assigned purposefully rather than randomly, as in BP neural networks, resulting in better parameter quality. Additionally, using unlabelled data significantly reduces the manual labelling workload, requiring only a portion of labelled data to fine-tune the pre-trained network. This approach combines low-level features and high-level semantics for image retrieval, more accurately reflecting human retrieval behaviour than using low-level features alone.

4.3 Performance testing of the integrated model for image recognition

This section evaluates the performance of the proposed integrated model for product image recognition. The experiment uses the TensorFlow framework, an open-source software library that employs a data flow diagram for numerical calculations. Data flow calculations are represented by a directed graph composed of a set of nodes.

When testing the model trained with the VGGNet 16 network structure, the prediction accuracy is 98.5%, and the recall rate is 98.2%. In contrast, the model trained with a 20-layer ResNet network structure achieves a prediction accuracy of 96.2% and a recall rate of 96.3%. The comparison of training and test results (Figure 7) indicates that training with the VGGNet 16 network structure leads to rapid convergence and better detection performance compared to the ResNet network.

Figure 7 Comparison of image recognition accuracy and recall between VGGNet 16 and ResNet (see online version for colours)



Data from feature extraction indicate that consumers primarily focus on price, functionality, and aesthetic design when selecting cultural and creative products (Figure 8). Figure 9 illustrates consumers' expectations for the visual design of these products.

Figure 8 Factors influencing consumers' selection of cultural and creative products

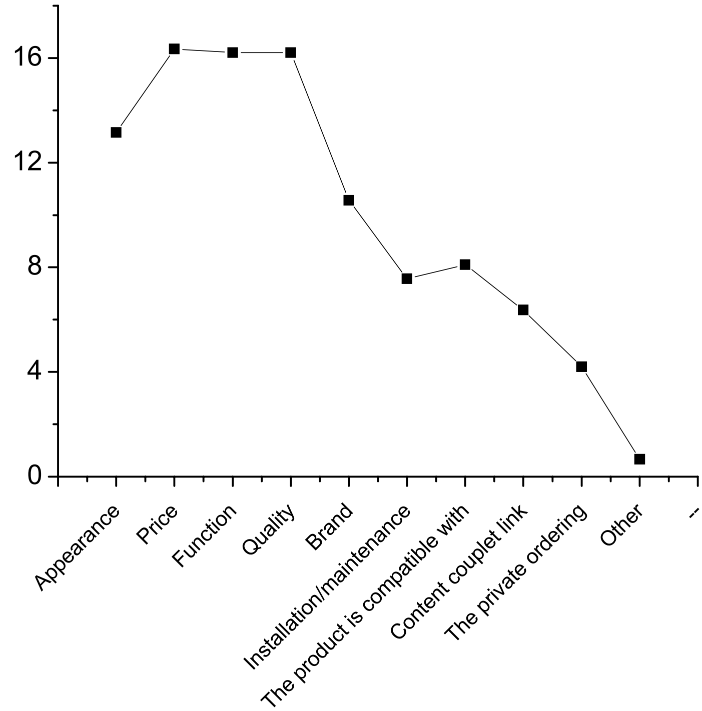
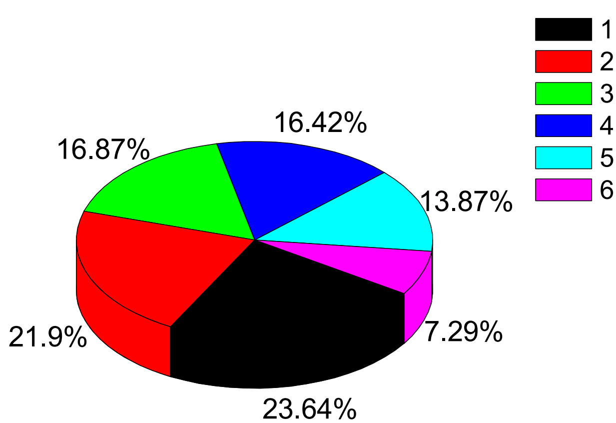


Figure 9 Consumers' expectations for the visual design of cultural and creative products (see online version for colours)



Among them, 1: panel colour collocation novel, fashionable; 2: appearance design innovation; 3: energy saving and environmental protection; 4: more functions, easy to use; 5: low price; 6: other.

Using tourism souvenirs and handicrafts as examples, Tables 1 and 2 present the image analysis and retrieval results for these products. Within each category, only a selection of representative images is included for a detailed analysis of their market performance and consumer preferences.

Table 1 Image analysis and retrieval of tourism souvenirs

Product type	*Average rating*	*Colour preference*	*Design innovation*	*Price range (CNY)*	*Purchase intention* (%)
Postcards	4.5	Bright, Vivid	4.2	5–20	85
Keychains	4.0	Classic, Simple	4.0	10–30	70
Refrigerator magnets	4.3	Colourful, Fun	4.1	15–40	75
Magnetic bookmarks	4.1	Fresh, Natural	3.9	8–25	60
Handmade souvenirs	4.7	Unique, Personalised	4.5	30–100	90

Table 2 Image analysis and retrieval of handicrafts

Product type	*Average rating*	*Material preference*	*Design complexity*	*Price range (CNY)*	*Purchase intention* (%)
Ceramic products	4.6	Ceramic	4.3	50–150	80
Fabric products	4.2	Pure Cotton, Linen	4.0	30–80	75
Wooden products	4.5	Solid Wood	4.2	40–120	85
Jewelry	4.8	Metal, Gemstones	4.6	100–500	90
Hand carvings	4.4	Natural Wood	4.5	80–200	70

Table 1 presents the image analysis results for various tourism souvenirs, including postcards, keychains, and refrigerator magnets. Analysing these images identifies the design elements and colour combinations most favoured by consumers, reflecting their preferences for the visual appeal of these products. Additionally, the image retrieval results reveal popular trends associated with these souvenirs, providing important data for subsequent product design. Table 2 showcases the image analysis of different types of handicrafts, including ceramics, textiles, and wooden products. This analysis uncovers consumer focal points, such as unique shapes, exquisite craftsmanship, and the use of eco-friendly materials. Furthermore, the image retrieval results help identify popular design styles in the handicraft market, offering valuable references for designers in new product development.

In summary, the DBN in DL technology effectively analyses video image data from consumers, revealing their various product needs. These needs often encompass aspects such as appearance, colour matching, and functionality, providing essential data for product visual design. This capability enables designers to better understand consumer requirements and create products that align more closely with their preferences.

Emotional design in products can enhance users' experiences. By transforming users' vague emotional needs into concrete design elements, product images can guide the emotional design process. However, creating an accurate mathematical model for product image recognition using traditional methods poses significant challenges. The CNN-based product image recognition model overcomes these limitations by automatically extracting features, enabling the development of a more precise recognition system. This advancement offers new ideas and methods for product image research, establishing a communication bridge between designers and users, which helps companies better meet consumer needs and expand market share.

To optimise the design process, a comprehensive model based on DL is proposed, combining DCNN with DBN for precise identification of product images and in-depth analysis of consumer demands. This optimised model allows companies to quickly capture consumer aesthetic trends and functional requirements, shorten the product design cycle, and enhance design efficiency. Additionally, by integrating user research data and employing data-driven design decisions, product design aligns more closely with consumers' actual preferences. The work also emphasises the importance of appearance design for tourism cultural and creative products by enhancing their visual appeal. By addressing consumers' concerns regarding colour, material, innovation, and practicality, a strategy is proposed to organically integrate innovative design with cultural elements. This approach enhances the artistic and cultural value of the products while improving their practicality and interactivity through technological empowerment, satisfying consumers' dual demands for functionality and emotional experience. Finally, this process, which combines market research, user feedback, and advanced design technology, enhances the competitiveness of the company's products and helps establish unique brand positioning in the market, attracting more consumers and further expanding market share.

5 Conclusions

This work explores the significant role of product visual design in enhancing consumer purchasing intentions and market competitiveness by analysing consumer demands and preferences for tourism cultural creative products. An innovative DL model is constructed, combining DCNN with DBN to extract underlying features of product images and consumer semantics. This approach improves product image recognition accuracy and provides designers with valuable data to better meet consumer needs. Furthermore, the work conducts an in-depth analysis of how key factors such as price, functionality, and appearance impact consumer purchasing decisions, providing empirical evidence for product design and enhancing market competitiveness. Despite achieving certain results, this work has limitations, such as a relatively small dataset that may affect the model's generalisation ability. Additionally, it primarily focuses on specific types of tourism cultural and creative products, lacking comprehensive coverage of all related categories. Future studies should expand the sample range for more representative results. Future research directions should include

1 enlarging the dataset to improve the model's accuracy and adaptability

2 exploring other DCNNs (such as GANs) to optimise product image generation and recognition

3 examining long-term changes in consumer behaviour to investigate how product design can enhance consumer loyalty.

In summary, this work offers new perspectives and methodologies for designing tourism cultural and creative products, providing important references for enterprises in product innovation and market competition.

Acknowledgements

This work was supported by Social Science Research Project of The Education Department of Jilin Province "Research on Digital Economy Enabling High-quality Integrated Development of Agricultural, Cultural and Tourism in Jilin Province" (JJKH20241595SK).

Conflicts of Interest

All Authors declare that they have no conflict of interest.

References

Abbas, J., Balsalobre-Lorente, D., Amjid, M.A., Al-Sulaiti, K., Al-Sulaiti, I. and Aldereai, O. (2024) 'Financial innovation and digitalization promote business growth: the interplay of green technology innovation, product market competition and firm performance', *Innovation and Green Development*, Vol. 3, No. 1, p.100111.

Aykut, T., Xu, J. and Steinbach, E. (2019) 'Realtime 3D 360-degree telepresence with deep-learning-based head-motion prediction', *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, Vol. 9, No. 1, pp.231–244.

Bankhead, P., Loughrey, M.B., Fernández, J.A., Dombrowski, Y., McArt, D.G., Dunne, P.D., and Hamilton, P.W. (2017) 'QuPath: open source software for digital pathology image analysis', *Scientific Reports*, Vol. 7, No. 1, p.16878.

Belhi, A., Bouras, A., Al-Ali, A.K. and Foufou, S. (2023) 'A machine learning framework for enhancing digital experiences in cultural heritage', *Journal of Enterprise Information Management*, Vol. 36, No. 3, pp.734–746.

Biswas, D., Everson, L., Liu, M., Panwar, M., Verhoef, B.E., Patki, S., and Van Helleputte, N. (2019) 'CorNET: deep learning framework for PPG-based heart rate estimation and biometric identification in ambulant environment', *IEEE Transactions on Biomedical Circuits and Systems*, Vol. 13, No. 2, pp.282–291.

Burnap, A., Hauser, J.R. and Timoshenko, A. (2021) *Design and Evaluation of Product Aesthetics: A Human-Machine Hybrid Approach*, SSRN (No. 3421771).

Chatterjee, S., Chaudhuri, R. and Vrontis, D. (2024) 'Does data-driven culture impact innovation and performance of a firm? an empirical examination', *Annals of Operations Research*, Vol. 333, No. 2, pp.601–626.

Chaudhuri, R., Chatterjee, S., Vrontis, D. and Thrassou, A. (2024) 'Adoption of robust business analytics for product innovation and organizational performance: the mediating role of organizational data-driven culture', *Annals of Operations Research*, Vol. 339, No. 3, pp.1757–1791.

Dabiri, S., Lu, C.T., Heaslip, K. and Reddy, C.K. (2019) 'Semi-supervised deep learning approach for transportation mode identification using GPS trajectory data', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 32, No. 5, pp.1010–1023.

Ding, X., Chen, Y., Tang, Z. and Huang, Y. (2019) 'Camera identification based on domain knowledge-driven deep multi-task learning', *IEEE Access*, Vol. 7, pp.25878–25890.

Fei, Y., Wang, K.C., Zhang, A., Chen, C., Li, J.Q., Liu, Y. and Li, B. (2019) 'Pixel-level cracking detection on 3D asphalt pavement images through deep-learning-based CrackNet-V', *IEEE Transactions on Intelligent Transportation Systems*, Vol. 21, No. 1, pp.273–284.

Gaber, J.A., Youssef, S.M. and Fathalla, K.M. (2023) 'The role of artificial intelligence and machine learning in preserving cultural heritage and artworks via virtual restoration', *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 10, pp.185–190.

Han, S., Shi, Z. and Shi, Y. (2022) 'Cultural and creative product design and image recognition based on the convolutional neural network model', *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, p.2586042.

Hatvani, J., Horváth, A., Michetti, J., Basarab, A., Kouamé, D. and Gyöngy, M. (2018) 'Deep learning-based super-resolution applied to dental computed tomography', *IEEE Transactions on Radiation and Plasma Medical Sciences*, Vol. 3, No. 2, pp.120–128.

Kan, S., Cen, Y., He, Z., Zhang, Z., Zhang, L. and Wang, Y. (2019) 'Supervised deep feature embedding with handcrafted feature', *IEEE Transactions on Image Processing*, Vol. 28, No. 12, pp.5809–5823.

Kato, N., Fadlullah, Z.M., Mao, B., Tang, F., Akashi, O., Inoue, T. and Mizutani, K. (2016) 'The deep learning vision for heterogeneous network traffic control: proposal, challenges, and future perspective', *IEEE Wireless Communications*, Vol. 24, No. 3, pp.146–153.

Krahe, C., Brunche, A., Jacob, A., Stricker, N. and Lanza, G. (2020) 'Deep learning for automated product design', *Procedia CIRP*, Vol. 91, pp.3–8.

Lee, J.G., Jun, S., Cho, Y.W., Lee, H., Kim, G.B., Seo, J.B. and Kim, N. (2017) 'Deep learning in medical imaging: general overview', *Korean Journal of Radiology*, Vol. 18, No. 4, pp.570–584.

Li, H., Ota, K. and Dong, M. (2018) 'Learning IoT in edge: deep learning for the internet of things with edge computing', *IEEE Network*, Vol. 32, No. 1, pp.96–101.

Li, R. and Wang, C. (2022) 'Cultural and creative product design and image recognition based on deep learning', *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, p.7256584.

Lu, X., Yang, Y., Wang, L. and Wang, H. (2022) 'Cultural and creative design of Dunhuang murals in xixia period based on deep learning', *Wireless Communications and Mobile Computing*, Vol. 2022, No. 1, p.7161593.

Luo, H., He, C., Zhou, J. and Zhang, L. (2021) 'Rolling bearing sub-health recognition via extreme learning machine based on deep belief network optimized by improved fireworks', *IEEE Access*, Vol. 9, pp.42013–42026.

Luo, J., Tang, J., So, D.K., Chen, G., Cumanan, K. and Chambers, J.A. (2019) 'A deep learning-based approach to power minimization in multi-carrier NOMA with SWIPT', *IEEE Access*, Vol. 7, pp.17450–17460.

Ma, X., Li, Y. and Asif, M. (2024) 'E-commerce review sentiment analysis and purchase intention prediction based on deep learning technology', *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp.1–29.

Mortati, M., Magistretti, S., Cautela, C. and Dell'Era, C. (2023) 'Data in design: How big data and thick data inform design thinking projects', *Technovation*, Vol. 122, p.102688.

Naylor, C.D. (2018) 'On the prospects for a (deep) learning health care system', *JAMA*, Vol. 320, No. 11, pp.1099–1100.

Ogundipe, D.O., Babatunde, S.O. and Abaku, E.A. (2024) 'AI and product management: a theoretical overview from idea to market', *International Journal of Management and Entrepreneurship Research*, Vol. 6, No. 3, pp.950–969.

Pelt, D.M. and Sethian, J.A. (2018) 'A mixed-scale dense convolutional neural network for image analysis', *Proceedings of the National Academy of Sciences*, Vol. 115, No. 2, pp.254–259.

Primpke, S., Lorenz, C., Rascher-Friesenhausen, R. and Gerdts, G. (2017) 'An automated approach for microplastics analysis using focal plane array (FPA) FTIR microscopy and image analysis', *Analytical Methods*, Vol. 9, No. 9, pp.1499–1511.

Quan, C. and Lu, B. (2024) 'Enhancing innovation management and venture capital evaluation via advanced deep learning techniques', *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp.1–22.

Ran, J., Zou, G. and Niu, Y. (2024) 'Deep learning in carbon neutrality forecasting: a study on the SSA-attention-BIGRU network', *Journal of Organizational and End User Computing*, Vol. 36, No. 1, pp.1–23.

Shen, Z., Shang, X., Zhao, M., Dong, X., Xiong, G. and Wang, F.Y. (2019) 'A learning-based framework for error compensation in 3D printing', *IEEE Transactions on Cybernetics*, Vol. 49, No. 11, pp.4042–4050.

Sun, Y. (2022) 'Design and purchase intention analysis of cultural and creative goods based on deep learning neural networks', *Computational Intelligence and Neuroscience*, Vol. 2022, No. 1, p.3234375.

Wang, J. (2024) 'Deep learning-based hand-drawn illustration in packaging design of cultural and creative products', *International Journal of Maritime Engineering*, Vol. 1, No. 1, pp.359–370.

Wang, M., Derry, S. and Ge, X. (2017) 'Guest editorial: fostering deep learning in problem-solving contexts with the support of technology', *Journal of Educational Technology and Society*, Vol. 20, No. 4, pp.162–165.

Wei, W., Dai, Q., Wong, Y., Hu, Y., Kankanhalli, M. and Geng, W. (2019) 'Surface-electromyography-based gesture recognition by multi-view deep learning', *IEEE Transactions on Biomedical Engineering*, Vol. 66, No. 10, pp.2964–2973.

Wen, C.K., Shih, W.T. and Jin, S. (2018) 'Deep learning for massive MIMO CSI feedback', *IEEE Wireless Communications Letters*, Vol. 7, No. 5, pp.748–751.

Xu, Z. (2022) 'Construction of intelligent recognition and learning education platform of national music genre under deep learning', *Frontiers in Psychology*, Vol. 13, p.843427.

Ye, W., Gu, W., Guo, X., Yi, P., Meng, Y., Han, F. and Wang, X. (2019) 'Detection of pulmonary ground-glass opacity based on deep learning computer artificial intelligence', *Biomedical Engineering Online*, Vol. 18, No. 1, pp.1–12.

Yuan, S., Navid, S.P. and Jorge, O. (2025) 'GeXSe (Generative explanatory sensor system), A deep generative method for human activity recognition of smart spaces IOT', *IEEE Sensors Journal*, Vol. 25, No. 7, pp.12407–12421.

Zeng, M., Li, M., Fei, Z., Wu, F.X., Li, Y., Pan, Y. and Wang, J. (2019) 'A deep learning framework for identifying essential proteins by integrating multiple types of biological information', *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, Vol. 18, No. 1, pp.296–305.

Zhao, Z.Q. (2019) 'Object detection with deep learning: a review', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 99, pp.1–21.

Zhu, H., Jiao, L., Ma, W., Liu, F. and Zhao, W. (2019) 'A novel neural network for remote sensing image matching', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 30, No. 9, pp.2853–2865.

Zhu, W. and He, Y. (2017) 'Green product visual design in supply chains under competition', *European Journal of Operational Research*, Vol. 258, No. 1, pp.165–180.