Evaluation and Design Method for Product Form Aesthetics Based on Deep Learning

AIMIN ZHOU[ID], **HONGBIN LIU**[ID], **SHUTAO ZHANG**[ID], **AND JINYAN OUYANG**
School of Design Art, Lanzhou University of Technology, Lanzhou 730050, China
Corresponding author: Shutao Zhang (zhangsht@lut.edu.cn)

ABSTRACT Currently, evaluations of products from aesthetics are mostly carried out with knowledge expressions of aesthetic features as tools, achieving remarkable results. However, obtaining a large aesthetic feature vocabulary is a challenge because of the experience of researchers and the comprehension abilities of subjects. In addition, due to manual feature extraction, the sample sizes of experimental dataset are generally small, leading to results with poor generalization. To address this problem, a method of aesthetic evaluation and form design for products based on deep learning was proposed. First, a crawler tool was used to collect the front images of cars with corresponding appearance ratings, and a dataset was constructed with users' intuitive and simple ratings as the labels. A deep convolutional neural network (CNN) was designed, and a grading threshold was used as the classification basis. During the process of training the network, batch normalization and other methods were used to optimize the network, and good classification effects were achieved. Based on the above work, an adversarial neural network was used for the aesthetic design of a product form, a shape sketch of an automobile front face was generated, the proposed evaluation model was used to evaluate it, and the result obtained was excellent. These results show that the method used in this study can correctly evaluate product form aesthetics and then generate a design scheme with a high aesthetic level, thereby providing powerful technical support for the intelligent design of product forms.

INDEX TERMS Aesthetic evaluation, aesthetic design, product form, deep learning.

I. INTRODUCTION

In the era of the aesthetic economy, users are paying increasing attention to the spiritual functionality of products and to the aesthetic and emotional experiences brought by enjoyable technology. In the field of mature technology, the functional technology of a product is merely the base requirement for entering the market. The functional gaps between products from different manufacturers are decreasing, and homogenization is common. At this time, the aesthetic quality of a product form becomes the key decision-making factor regarding consumer behavior. Aesthetically attractive products can give consumers good aesthetic experiences and put them in a happy mood. Good aesthetics can also improve the added value of products and enhance the competitiveness of enterprises. Therefore, a good product design should have excellent aesthetic quality. This requires designers to understand the law of aesthetic cognition. By carrying out accurate product aesthetic positioning and information transmission

The associate editor coordinating the review of this manuscript and approving it for publication was Luigi De Russis[ID].

in a given design, the design of the product can be better recognized by users.

In the information age, people are willing to share their interests, hobbies, opinions on products, emotional tendencies and so on in shopping networks or on related forums. Many users comment on the appearance of products, and the extraction of such big network data can solve the small sample size problem. At the same time, to a certain degree, deep learning methods can be used to solve problems that are difficult for traditional feature extraction methods. Deep learning methods can be established without artificially extracted features, and they use many original sample data. In addition, they make full use of hidden layers to perform step-by-step and in-depth studies of abstract image information, providing comprehensive and direct access to image characteristics. According to the characteristics of input data, automatic learning in image classification, object recognition, face recognition and other fields has made breakthrough progress. Among such methods, the greatly improved accuracy of convolutional neural networks (CNNs) on classification tasks has attracted extensive attention. On the basis of this research,

if the aesthetic evaluation problem is transformed into an image classification problem, a threshold aesthetic value can be used as the classification label, and a deep learning method can be used to learn the morphological characteristics of product samples, realize the prediction of the aesthetic values of unfamiliar samples and achieve the purpose of aesthetic evaluation. In this paper, the front face of an automobile is taken as the research object, and a product from an aesthetic evaluation method based on deep learning is proposed.

According to the goal of utilizing the evaluation of a user's aesthetic demand to produce actively positive guidance for designing an aesthetic product form based on an existing product form design, a form of aesthetic evaluation is used as the research object. This approach is taken to enable people to quantify aesthetic cognition in the form of product information, to build an aesthetic evaluation model, and to develop an aesthetic product form design method through this evaluation model. However, there are still inevitable problems faced by the computational aesthetics method when constructing an aesthetic formula. This design method still has some unavoidable problems, such as the design of experiments, the selection and calculation of the beauty index, and knowledge representation, and there are great difficulties and workloads. With the introduction of generative adversarial networks [1] (GANs), their powerful computing power and graphic derivation ability can generate clearer and more diverse images based on sample data. The advantage of a GAN over the traditional genetic algorithm is that it does not need to manually determine the feature points, and it can automatically learn the features in a given sample. Furthermore, GAN has a stronger generalization ability and generates better and more realistic images.

In this context, based on the aesthetic evaluation method of product form, this paper applies deep learning technology to the field of aesthetic evaluation and industrial design and constructs an aesthetic evaluation model and aesthetic design system of product form. It is expected that the use of advanced technology can reduce the manual workload, improve the efficiency of aesthetic evaluation, and generate a product design scheme with higher aesthetic value that can meet people's aesthetic needs and expectations more quickly.

II. RELATED WORK

A. AESTHETIC EVALUATION

The aesthetic evaluation of a product form is a process of aesthetic cognition in which the aesthetic subject (human) compares, judges and evaluates the aesthetic value of the aesthetic object (product form) according to their own aesthetic needs and aesthetic standards. It is an important means of inspecting the quality of a product form design and guiding the design process. In view of the research on the aesthetic evaluations of products, at present, there are two main approaches for performing aesthetic evaluation: subjective evaluation and objective evaluation. For example, Kang [2] introduced electric field mechanics into the field of aesthetic evaluation

and proposed a color aesthetic evaluation method based on the combination of form and color. Ranjan [3] proposed a computational model for predicting web aesthetics based on the linear kernels of support vector machines. Kobayashi [4], based on the analysis of the relationship between customer sensibility and aesthetic elements, proposed a method to support design processes with aesthetics and explored an optimal aesthetic design approach with a genetic algorithm. Orsborn [5] took cars as research objects. He created experimental samples based on the characteristics and attributes of car designs. Aesthetic evaluation results were obtained through standard deviation (SD) investigation. Finally, he constructed a model for the relationships between the preferences of aesthetic subjects and the characteristics of car designs by using a utility function. Roussos [6] proposed four groups of aesthetic standards for the form, material, color and simplicity of a product. Through investigation and experimentation, the Platts decision matrix method was used to conduct an aesthetic evaluation of the product. Zhou [7] proposed an optimization design method for product forms in the field of multimodal transportation based on quantitative aesthetic evaluation to obtain an aesthetic product form scheme. The aesthetic characteristic system, which is consistent with the aesthetic principle and the Gesar principle, was established.

However, there are still many problems to be solved in the existing aesthetic evaluation methods. First, the subjective evaluation method mainly relies on a variety of survey methods to obtain basic data or experts giving subjective weights to build an evaluation model. The scoring data of the proposed method are the scores provided by users according to their own aesthetic experiences and preferences, which are simple and intuitive. The scoring data are capable of reflecting the basic cognition and subjective will of decision makers in terms of aesthetic issues with strong explanatory power. However, participants have different evaluation criteria, subjective preferences and emotions, which may easily lead to evaluation results with low reliability. To obtain more reliable data, a large number of subjects would be necessary, rendering the data collection period in this scenario excessively long and incurring higher costs. In addition, when the number of experimental samples is large, an excessively lengthy evaluation process easily causes fatigue and reduces the reliability of the obtained evaluation results. Second, the objective evaluation method is mainly based on the thought process of computing aesthetics, and an aesthetic evaluation index-based system is used. In this system, the built product form and the index quantitatively describe the aesthetic degree of the product. Compared with the subjective evaluation method, the objective evaluation method uses an experimental computer program, and its experimental process is easier to control in terms of index calculation. Furthermore, it provides accurate data information and has a strong mathematical theoretical basis, remaining unaffected by the interference of man-made factors. However, the objective evaluation of experimental design, feature extraction and selection methods and the calculation of aesthetics in such problems results

in considerable difficulty and a large workload. This method is less dependent on the designer's personal experience, knowledge reserves and experimental process. In addition, the method samples data from experiments, so the quality and the quantity of the acquired experimental sample determine the evaluation results. For example, when the sample size is too small, the evaluation results are underrepresented, and their generalizability is poor. In contrast, when the sample is too large, the resulting workload may be too large, making the evaluation difficult to achieve.

B. CONVOLUTIONAL NEURAL NETWORKS

In 2012, the deep CNN model (AlexNet) proposed by Krizhevsky [8] achieved a very high evaluation accuracy on the standard dataset provided, breaking the highest record at that time. At this point, the study of neural networks entered a new era, yielding an upsurge in neural network research. In 2014, the Visual Geometry Group (VGG) network designed by Simonyan [9] won first place for the location task and second place for the classification task. In 2014, the GoogLeNet model developed and designed by Szegedy [10] used the novel inception structure as the basic module for cascading, thereby improving the computational efficiency of the algorithm. In 2015, He Kaiming [11] and his team proposed a special CNN, a residual neural network (ResNet), which can easily reach hundreds or even thousands of layers and complete training within an acceptable time frame, thus greatly improving the accuracy of image recognition.

The powerful feature extraction abilities and computing power of these models enable many researchers to apply them in the field of aesthetic evaluation. For example, Wang [12] and others proposed a scene depth model to realize the automatic learning of aesthetic characteristics. Based on the existing neural network designs, such models add a scene convolution layer that consists of a network of multiple sets of descriptors such that the model has a comprehensive image aesthetic learning ability. Suchecki [13] analyzed the aesthetic values of input images through a deep CNN, classified the input images, and finally evaluated the aesthetic values of digital photos to help optimize a photographer's workflow. Lemarchand [14] proposed an image aesthetic evaluation method. This method is based on the discoveries of psychology and neuroscience and built across datasets of aesthetic classifiers and a set of effective features extracted from the image for learning purposes. Then it classified images according to their aesthetic ratings through the analysis of the characteristics of images and their aesthetic differences. It enabled people to observe the aesthetic preference of a two-dimensional static scene. Dahal [15] and others developed a global view, partial view, style and CNN-based model of semantic information and alternately used image content and semantics to guide the model. Finally, according to the pairs of image aesthetic quality and relative ranking scores, the research results could be applied to enhance automatic photo management tools or other image editing software.

Zhang [16] proposed a Chinese ink painting-based aesthetic evaluation model utilizing deep learning with a CNN to determine the aesthetic characteristics of Chinese ink and washes for learning purposes. The model relies on the manual fusion of art expert knowledge characteristics. Finally, a comprehensive aesthetic evaluation model was established. They provided a reference for evaluations based on the learning of Chinese painting aesthetic frameworks. Moreover, they also probed into the characteristics of handmade art to determine the extent to which they can help with feature prediction based on learning the aesthetic views of human beings. Zhao [17] uses global attributes of graphs to represent various aesthetic aspects and uses gate units to combine composition features and miscellaneous aesthetic features for aesthetic prediction based on CNNs. Mikhailava [18] contributed to experiments using CNNs for aesthetic evaluation of images, especially for food plates. Such assessments can be of benefit to professional and amateur food makers, restaurant critics, photographers and travelers. Wu [19] reveals the overall quality of the design through the analysis of the image features. It is assumed that visual aesthetics can be used as a clue for the modeling of prize classification, and this hypothesis is proven by the design competition submissions. Finally, based on deep CNN (DCNN) analysis of product images, the optimal model of design award prediction is constructed.

These aesthetic evaluation studies using deep learning are mostly carried out from the perspective of the overall quality of images, such as the composition and color of images. Few studies have been conducted on the aesthetic quality of product form. From this point of view, the content of our study is novel.

C. GENERATIVE ADVERSARIAL NETWORK

As a new type of generative model, generative adversarial networks have become a new research hotspot in the field of deep learning and artificial intelligence technology, showing great application and development prospects in image and visual computing, speech language processing, information security and other fields.

In various applications, GANs have attained many significant achievements. Jaiswa [20] proposed a new method based on a GAN for anime character design, learning the characteristics and features from training image datasets and combining them to create new features and to build a new image of the training dataset. This method can not only help artists and designers preview new and unique cartoon images but also prevent any copyright infringement behaviors. Sagawa [21] used a deep convolution GAN (DCGAN) to generate face images from corresponding features after setting "smiles" and other features, and the author achieved excellent results. Ito [22] used a conditional GAN (CGAN) to generate a new Raman image based on a Raman image dataset. In recent years, some researchers have tried to apply GANs to product designs. For example, Radhakrishnan [23] proposed a GAN-based intelligent vehicle design model, which created new unseen design schemes through the sketches of a car design

studio, thereby improving design efficiency. Kularatne [24] proposed a fashion design method combining the expertise of fashion designers and pattern makers by using a GAN on the basis of existing fashion styles to generate new fashion design schemes using existing clothes.

III. METHOD

The purpose of this study is to conduct a product form aesthetic evaluation and generate new design schemes according to the evaluation results. To simplify the problem, this study is defined as a binary classification task; that is, the images in the dataset are classified as aesthetic or unaesthetic, and the goal of the study is to identify whether the sample images are aesthetic or not. Then, we select the aesthetically pleasing dataset for image generation, hoping to generate a new aesthetically pleasing design scheme through a DCGAN. We finally evaluate the scheme through the constructed aesthetic evaluation model. The specific process is shown in Figure 1.
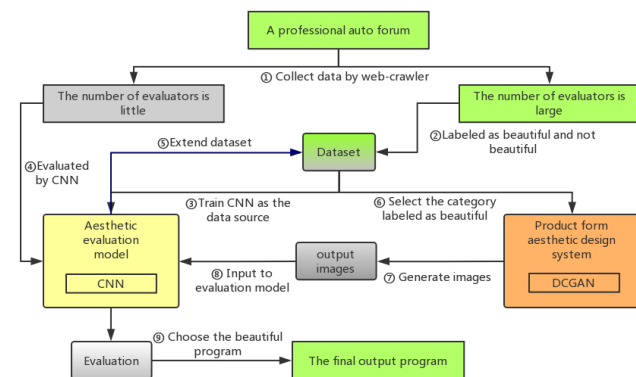

FIGURE 1. Overall flow chart of the study.

A. DEEP CONVOLUTIONAL NEURAL NETWORK

A CNN has been one of the core algorithms in the field of image recognition for a long time, as it has stable learning performance when sufficient data are available [25]. A CNN, through a local receptive field and shared weights, reduces the need for large training weights and reduces the computational complexity of the network. At the same time, a pooling operation makes the network have certain invariance to partial transformations of the input, such as translation invariance and scale invariance, and improves the generalization ability of the network. This enables the realization of a deep network model. In addition, a CNN can directly input the original data into the network for learning purposes, avoiding the influences of human factors on data processing.

In this paper, a CNN model is designed on the basis of the VGG network and named MeiduNet. The network input data size is $256 \times 256 \times 3$; the network has 6 convolution layers; the convolution kernel sizes are 3×3; the numbers of feature maps are 32, 64, 64, 128, 128, and 128; the activation functions are rectified linear units (ReLUs); each pooling

layer after a convolution layer is a max pooling layer with two fully connected layers; the numbers of neurons are 512 and 2; and the last output layer is a softmax layer, which provides the classification of the output image. To improve network performance, a batch normalization (BN) layer and dropout layer are added to the network. The network structure is shown in Figure 2.

A normalization layer is added after each convolution layer and the first fully connected layer with the purpose of standardizing the sample feature distribution and improving the learning speed of the neural network. A dropout [8] layer is also added; in this way, negative phenomena such as overfitting and gradient vanishing that may occur during the network training process are eliminated, and the effects of these layers are verified by a control experiment.

B. DEEP CONVOLUTION GENERATION ADVERSARIAL NETWORK

The original GAN is a generative model proposed by Goodfellow *et al.* in 2014. Its core idea comes from the two-person zero-sum game in game theory. The basic GAN model consists of a generator (G) and a discriminator (D) in its structure. The generator is a model that reconstructs the "initial random noise" continuously according to the pixel probability density distribution of the training image until the pixel probability density distribution of the generated image approaches that of the training image. The task of the generator is to make the generated image "look like the real image". The function of the discriminator is to distinguish between the generated image and the training image and constantly enhance its own sensitivity during the training process.

The DCGAN [26] is a learning model that combines a GAN and a CNN. Its basic principle is similar to that of a GAN, except that G and D in the classical GAN are replaced by two improved CNNs. The basic framework is shown in Figure 3.

Keras is used to build the DCGAN model layer by layer, and the specific construction process and parameter settings are shown in Figure 4.

The network is iterated to generate design types; because the generator and the discriminator use the convolution of the neural network structure, a BN layer is added after each convolution to improve the stability of the network. In addition, the generator uses the ReLU activation function, and the discriminator uses the leaky-ReLU activation function, which avoids model collapse in practice. By removing all pooling layers and replacing the deconvolution layers with upsampling layers, the computational cost is reduced and the model learning speed is improved.

C. OPTIMIZATION METHODS

In the process of CNN training, the input data are usually standardized. However, their mean value and standard deviation will change as the input data are passed step by step in the hidden layer. This phenomenon is called the covariant drift phenomenon, which is considered to be one of the reasons
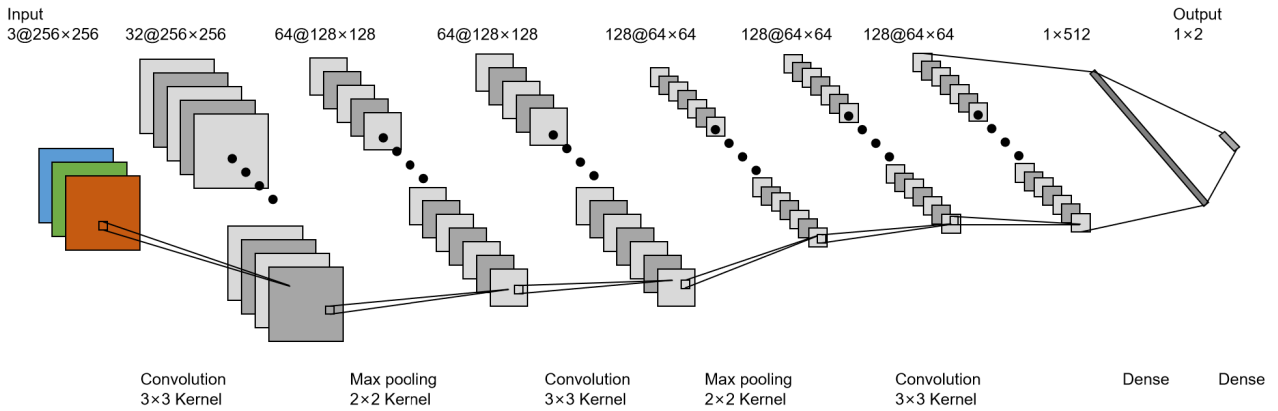
Input
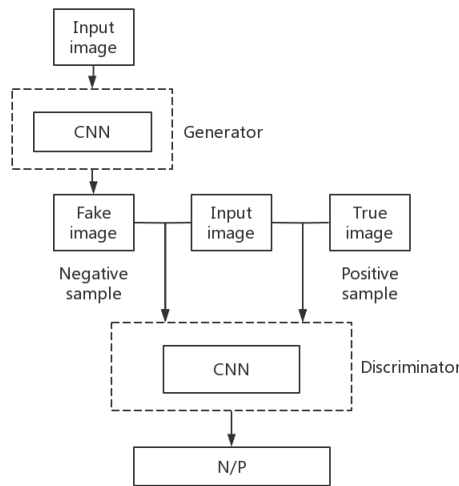3@256×256   32@256×256   64@128×128   64@128×128   128@64×64   128@64×64   128@64×64   1×512

Output
1×2

Convolution
3×3 Kernel

Max pooling
2×2 Kernel

Convolution
3×3 Kernel

Max pooling
2×2 Kernel

Convolution
3×3 Kernel

Dense   Dense

FIGURE 2. MeiduNet structure.

FIGURE 3. Schematic diagram of the DCGAN.

FIGURE 4. DCGAN construction process and parameter settings.

for the gradient disappearance of the deep network. The introduction of a BN layer solves this problem by adding a series of parameters. The BN layer overcomes this phenomenon because it allows the network to learn to recover the distribution of features that the original network was trying to learn.

In the process of training, too many parameters will cause overfitting. Dropout was first introduced by Hinton in 2012. It can improve the performance of the neural network by preventing the coaction of the characteristic detector to effectively suppress the phenomenon of overfitting. In other words, when propagating forward, the activation value of a certain neuron stops working with a certain probability p. This makes the model more generalizable, as it does not excessively rely on some local features. By reducing the coadaptive relationship between neurons in this way, the updating of weights is no longer dependent on the joint action of implicit nodes with fixed relationships. This mechanism forces the network to learn more robust features, reducing the possibility of overfitting.
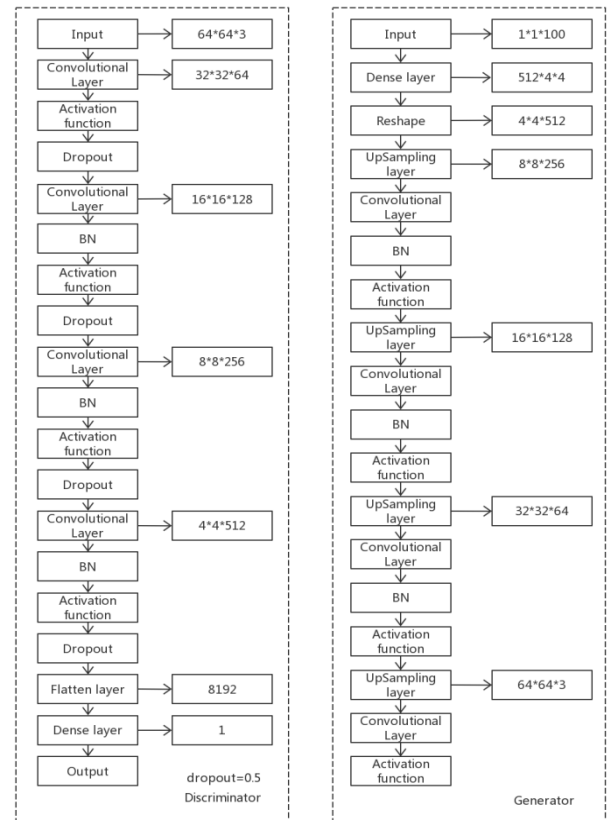
IV. EXPERIMENT AND ANALYSIS

In this section, first, a database with aesthetic labels is constructed for training purposes; second, the evaluation indexes of the CNN aesthetic evaluation model are set up; third, the aesthetic evaluation training model of the control group is set up, and the performance of the model is analyzed and compared with those of other models; fourth, the aesthetic evaluation model is tested with test samples; fifth, the product scheme generation model of the DCGAN is trained, and car front view sketches are generated; and sixth, the effect

drawing is created according to the sketch, and whether the scheme is beautiful is evaluated through the aesthetic evaluation model.

The experiments are all carried out on a Win10 64-bit computer with an Intel i7-7700HQ CPU and a 4G NVIDIA GTX1050 GPU, and the dataset is trained based on the TensorFlow 1.12.0+Keras2.2.4 deep learning framework.

A. THE DATASET

This study uses car front view images for the aesthetic evaluation study because there is no existing dataset that contains aesthetic-grade car front view images. To validate the proposed method, through the use of a web crawler, car front view photos and their appearance scores are collected. Then, the images with extreme scores are eliminated, resulting in a total of 750 images with rating labels. The images and ratings come from a website called AutoHome, which has more than 1 billion daily visitors and whose large user base makes the feedback more objective. Its users are distributed across all age groups, which makes the rating data more comprehensive.

According to their median scores, the collected images are divided into two categories: 0 and 1. Category 0 is marked as "low", and category 1 is marked as "high", as shown in Figure 5. The number of images in each category is 375. All images are randomly divided: 600 in each training set, 300 in each category, 150 in each validation set, and 75 in each category at a 4:1 ratio. There is no crossover between the datasets after division.

(a) Labeled "low" (b) Labeled "high"

FIGURE 5. The results of database image classification.

The purpose of this study is to evaluate the beauty of the front face of the car through deep learning. To eliminate the interference caused by redundant factors, background cutout processing is carried out on the images in the dataset. The results are shown in Figure 6.

FIGURE 6. Image background processing effect.

For machine learning tasks, such as image classification, machine translation, and text-to-speech translation, the number of samples available for training is critical to achieving high performance. In machine learning, data enhancement is often used to expand the sample size to prevent the overfitting of the samples in the training set during the process of network training from leading to poor training results. Due to the small number of samples collected in this paper, data enhancement processing is carried out on the original dataset. Data enhancement refers to the process of introducing visual invariance into the dataset through preprocessing and amplifying the training data. Specifically, new images are generated by changing the angle rotation, size scaling, position translation and so on for each image. In this paper, the sample size of the original dataset is expanded ten times through data enhancement. The results are shown in Table 1.

TABLE 1. Data enhancement results.

Original plot	Generated images with data enhancement

Finally, the sample size of the dataset is expanded 10 times, and the resulting data distribution is shown in Table 2.

TABLE 2. Distribution of the dataset.

Image Type	Training	Validation	Total
Low	3000	750	3750
High	3000	750	3750
Total	6000	1500	7500

B. EVALUATION INDEXES FOR THE AESTHETIC EVALUATION MODEL

In deep learning, the actual category of an image is called the positive category, while other categories are called negative categories. Then, the image classification model classifies samples in the following four ways. True positive samples (TP) predicted by the model are the numbers of samples that are actually positive and judged to be positive. True negative samples (TN) are negative samples predicted by the model that are actually negative. False positive (FP) samples

are positive samples predicted by the model that are actually negative, and false negative samples (FN) are negative samples predicted by the model that are actually positive. When evaluating the performances of deep learning models, the following four indicators are usually adopted:

1) Accuracy (Acc) represents the proportion of the number of correctly classified test cases to the total number of test cases. The calculation formula is as follows:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

2) Precision (Pre), also called the precision ratio, represents the proportion of the number of positive examples that are correctly classified to the number of positive examples. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \quad (2)$$

3) Recall (Rec), also known as the recall rate, represents the proportion of the number of correctly classified positive cases to the number of actual positive cases. The calculation formula is as follows:

$$R = \frac{TP}{TP + FN} \quad (3)$$

4) Comprehensive evaluation index (F1-score, F1): This index is based on the harmonic average of the recall rate and precision, so it is the comprehensive evaluation indicator. The calculation formula is as follows:

$$F1 = \frac{2}{\frac{1}{P} + \frac{1}{R}} = \frac{2 \times P \times R}{P + R} \quad (4)$$

In general, higher accuracy indicates better performance of the classification model. However, if the number of a certain type of samples in the data set is much larger than that of other types of samples, and these small number of samples are the focus of the research, it will be very one-sided to evaluate the performance of the model only by the level of accuracy. In order to evaluate the model performance more comprehensively and accurately, the precision and recall are introduced. They are generally applied in the evaluation of dichotomous models, and it is required that the sample sizes of the two categories are similar and sufficient. In this paper, the data used in the experiment contains two types of samples, they are beautiful and unbeautiful car front face images, and the number of them is equal. Sometimes there is a contradiction between the precision and the recall, so it is necessary to use the comprehensive evaluation index to evaluate the model by combining the values of precision and recall. To sum up, in order to evaluate the model objectively, we established four indicators to evaluate performance of the model by combining the types and amounts of data used in the experiment.

C. TRAINING AND COMPARATIVE ANALYSIS OF THE AESTHETIC EVALUATION MODEL

For MeiduNet training, we set up two groups of control experiments: (1) To reduce variables, we choose a network

with a similar structure but different network layers to train the same dataset and compare the results. Therefore, AlexNet is selected as a benchmark model in this paper to verify the performance of MeiduNet. (2) Whether to apply the BN layer and the dropout layer. After training the MeiduNet network, the BN layer and the dropout layer are closed. This group of control experiments verifies whether the optimization method has an obvious optimization effect.

The parameter setting of the CNN has a direct impact on the results, and the parameter setting of the whole training process refers to the literature [27]. Each convolution layer and the fully connected layer of the first layer are randomly initialized with a Gaussian distribution with a mean of 0 and a standard deviation of 0.01. The dropout probability is set to 0.5, which is 50% of the total number of randomly paused neurons. Set the batch size to 16; that is, the number of images learned in each batch is 16. The learning rate is set to 0.01.
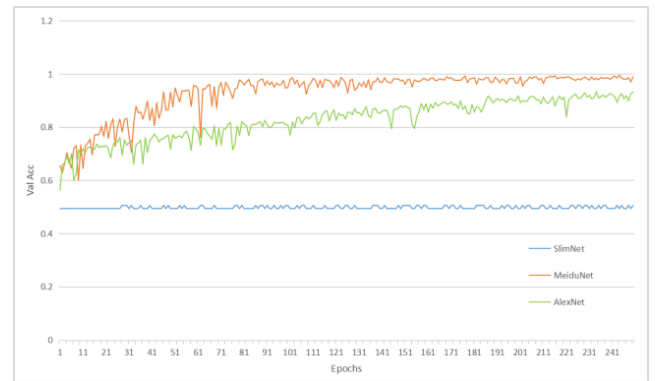

FIGURE 7. Verification set accuracy line diagram.

First, AlexNet was trained, and its parameters were only modified to be consistent with MeiduNet network parameters without modification of the network structure. Then, 250 epochs were trained to obtain evaluation index values of the model. Then, MeiduNet was trained with the same number of epochs. Then, to verify the functions of the BN and dropout layers, they were closed in MeiduNet, the model at this time was named SlimNet, and the model was trained. Finally, the curves of the validation set accuracy (Val ACC) and validation set loss function (Val Loss) of the final three models are calculated, as shown in Figures 7 and 8. The model evaluation indexes are shown in Table 3. These charts are compared to verify whether the evaluation network model designed in this paper has better performance.

In these charts, the model accuracy line of SlimNet fluctuates near 0.5 without increasing later on, the loss function converges prematurely, it suffers serious overfitting phenomena, and its accuracy is only 49.4%. In the two-category case, the accuracy is lower than 50%; this shows that for the utilized datasets, the network cannot meet the classification requirements of research. After adding BN and dropout layers, the negative effects are improved, and the classification performance of the model is also improved.
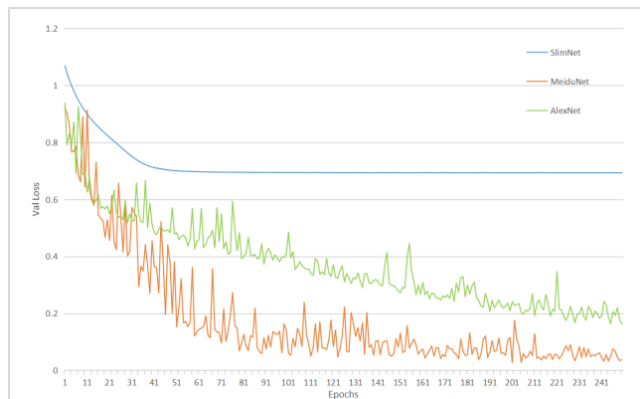
FIGURE 8. Line diagram of the verification set loss functions.

TABLE 3. Comparison of evaluation indexes.

Network Type	Acc (%)	Pre (%)	Rec (%)	F1(%)
SlimNet	49.4	24.0	49.0	33.0
MeiduNet	98.9	99.0	99.0	99.0
AlexNet	93.2	93.0	93.0	93.0

After verifying the influence of the optimization method of the network, in the lateral control experiment, the performance of the network made with another AlexNet structure is similar to that of the proposed network trained on the same dataset. The depths of the two networks are different. AlexNet has five convolution layers, which has one fewer convolution layer than MeiduNet, and all connection layers, which are the same as in MeiduNet, do not repeat the training process. As seen from Figures. 7 and 8, the values of the validation set loss functions of the AlexNet and MeiduNet models show gradually decreasing trends and tend to 0. In contrast, the accuracies on the validation set show gradually increasing trends and tend to 1, which proves that these networks can complete the task of classifying the datasets. The graph indicates that MeiduNet converges faster than AlexNet under the same number of epochs. The accuracy rate of MeiduNet is 98.9%, which is 5.7% higher than that of AlexNet. The control experiment proves that MeiduNet has better performance than AlexNet.

Through the experiment, the following conclusions can be drawn:

1) The product form aesthetic evaluation model constructed based on a CNN is feasible and performs well. As shown in Table 4, the accuracies of both MeiduNet and AlexNet are higher than 93%, and MeiduNet even reaches 98.9%. Even under the background of binary classification tasks, such an accuracy rate is very high.

2) The network depth of the CNN has an influence on the classification accuracy of the model. When the other parameters are basically the same, MeiduNet has a deeper network

TABLE 4. Prediction results of the MeiduNet model for images.

depth than AlexNet, and the final accuracy rate increases by 5.7%. It can be seen from this that an increase in network depth can improve model performance to some extent. However, some studies have pointed out that when the depth of the network deepens to a certain extent, the recognition accuracy declines with increasing depth. Therefore, the depth of MeiduNet that will lead to a performance decline needs to be further verified in subsequent studies.

D. VERIFICATION OF THE AESTHETIC EVALUATION MODEL

The front view images of cars that do not cross with the samples in the dataset used above are collected, the backgrounds are selected, and the MeiduNet model with the best performance is selected for aesthetic evaluation. Table 4 shows some of the evaluation results output by the model.

To verify the accuracy of the evaluation results in Table 4, a questionnaire survey is conducted. Since the number of predicted classifications in the experiment is 2, the number of classifications in the questionnaire for evaluation is also 2, namely, "0 – low" and "1 – high". Finally, the questionnaire results are compared to verify whether MeiduNet's aesthetic evaluation of the image sample of the car front view is accurate.

TABLE 5. The results of a questionnaire survey.

	1	2	3	4	5	6	7	8
0 (%)	84.85	84.85	66.67	63.64	45.45	33.33	27.27	24.24
1 (%)	15.15	15.15	33.33	36.36	54.55	66.67	72.73	75.76

A questionnaire survey is conducted to verify the prediction results of the eight samples in Table 5. The results are shown in Table 5. The majority of people evaluate serial numbers 1-4 as 0, namely, "low". Serial numbers 5-8 are evaluated by most people as 1, namely, "high". This result is consistent with the evaluation result of the MeiduNet model on the image, which proves that the aesthetic evaluation

model constructed in this paper can evaluate and predict the beauty of the front view of a car.

E. PRODUCT GENERATION MODEL TRAINING AND RESULTS

In the experiment regarding the product form aesthetic design model based on a GAN, to generate some product schemes for aesthetic evaluation, the image labeled ''high'' in the dataset is selected as the training data of the DCGAN with a size of $64 \times 64 \times 3$. The output of the sketch scheme of the front view of the car has a size of $64 \times 64 \times 3$.

Before the training, some parameters of the model were set according to reference [28]. The whole model was optimized by the stochastic gradient descent algorithm, and a batch size of 128 and a learning rate of 0.002 were set. All convolutional layers in the generator and discriminator are initialized with a Gaussian distribution with a mean of 0 and a standard deviation of 0.02. In addition, the parameter k is also set. In the actual training process of DCGAN, due to the great difference between the initial generated data and the real sample, the discriminator is very likely to win the confrontation with the generator after the first training. This results in the vanishing gradient of the generator. Therefore, to address the case of confrontation between the two, discriminator training requires the training of k (k > 1) generators to avoid optimal realization of the discriminator in advance during the training process. This ensures that the generator and the discriminator have the necessary confrontation ability. Since the k value is different in the training of different databases, it is necessary to adjust the k value according to the actual training situation in the process of the experiment. After many training iterations, the k value is adjusted to 10. When the model has been trained, the results are shown in Figure 9.

FIGURE 9. The generated image results.

As shown in Figure 11, at the beginning of training, the generated image is a mass of random pixels in disarray. After training for 2000 epochs, the disorderly pixels begin to converge toward the center, presenting the form of a central object and blank background. After 4000 epochs of training, the general outline of the car can be seen vaguely. Later, after 6000 epochs of training, some details of the front face, such as the air intake grille, front windshield, and rearview mirror position, can be viewed. Finally, after 8000 epochs of training, a sketch of the front face of the car with obvious features is obtained.

After 8000 rounds of network training, the experimental results are obtained, as shown in Figure 10. The results show the outputs after the training iterations from among a group

FIGURE 10. Schemes generated after 8000 epochs.

of 25 total car face sketch plans. Based on the distributions of the pixels, the sketches are consistent with the pixel locations in the image datasets. Some solutions are relatively clear; the details can be seen, and the sketches can be used as sketch plans to guide the product design process. However, some details are more obscure and have been generated with poor quality. Nevertheless, the experiment proves the feasibility of the proposed model. In terms of the overall shape, there are differences among the generation schemes, such as the width, height and feature locations, which reflect the diversity of the generation schemes. Some of these shapes are suitable for sport utility vehicles (SUVs), and some are suitable for cars. This situation is related to the given real image types in the dataset. These real image samples are mainly SUV and car images, so the generation scheme is basically tailored to these two types of models. In a few generation schemes, the front windshield and rearview mirror are not clear. An analysis of the original image data indicated that the reason for omission of some of the front windshield data in color images and the borders and rearview mirrors in the 64×64 format is that there are too few pixels. This sparsity leads to the network having difficulty when learning these features.

F. AESTHETIC EVALUATION OF THE GENERATION SCHEME

There are still a certain number of noise blocks in the car front face sketch obtained in the experiment, and these affect its aesthetic evaluation. Therefore, the generated car front view sketch is further drawn to obtain superior renderings, and the aesthetic evaluation model is used. The result is shown in Figure 11.

After the evaluation of the generated sketch by the evaluation model, the result obtained is ''high'', that is, beautiful. The usability of the aesthetic evaluation model is verified above, proving that the model can produce a more correct aesthetic evaluation of the front face of the car. Therefore, the image generated this time is beautiful and can be used as one of the alternatives in the design.

FIGURE 11. Result of manual refinement.

V. DISCUSSION

In the research on product form aesthetic evaluation, most studies discuss traditional subjective evaluation and objective evaluation methods. In recent years, some new aesthetic evaluation methods using deep learning have also emerged. To some extent, these methods solve the problems of traditional methods, such as small sample sizes, difficulty in feature extraction and large amounts of manual calculations. However, the most commonly used method in these studies is still to establish aesthetic indicators, which take multiple image words as the evaluation criteria for product form aesthetics. Therefore, it is difficult to guarantee whether the subjects can accurately understand these indicators and make correct judgments. Hence, in this study, the feedback of users in terms of the aesthetic degree of a product form is simplified to the question of "aesthetic or not aesthetic", that is, the most intuitive feelings of users. The responses are then quantitatively presented in the form of ratings. This enables the exploration of a method of product form aesthetic evaluation based on the simplest and direct aesthetic feelings of users as the evaluation standard. Taking this as a starting point, an image dataset of car front faces with aesthetic ratings is established, and a model for the aesthetic evaluation and sketch generation of car front face shapes is constructed with a deep learning method.

During the dataset construction stage, a crawler tool is used to collect the most intuitive and simple images and the corresponding scoring provided by users on the aesthetics of the automobile in an automobile forum. The collected images are evenly divided into two categories of equal quantity, with the purpose of enabling the model to learn the characteristics of the samples better and to have better classification performance. At the same time, to better conform to the era of user aesthetics, only the images of car front faces from between 2018 and 2020 are collected. Therefore, the dataset is not large enough, the aesthetic threshold can only be set as "beautiful" or "not beautiful", and the aesthetic evaluations of users cannot be further refined. Data enhancement expands the sample size to a certain extent and effectively prevents the problem of overfitting, but there is no substantial increase in the richness of the automobile categories. It is expected that in subsequent studies, the data volume can be expanded, the diversity of the samples can be improved, and more evaluation levels can be added to make the evaluation more detailed.

In the experimental stage of the aesthetic evaluation, the deep CNN is optimized, and the positive effects of the BN layer and dropout layer on network training are verified. In addition, a comparison between the developed MeiduNet and the classic AlexNet network reveals that the indicators of MeiduNet are better than those of AlexNet; the accuracy is 98.9%, which is a 5.7% increase over that of AlexNet. The abilities of the models to predict unknown samples are evaluated, and finally, through a questionnaire survey, the network evaluation results are found to be consistent with the evaluation results of most people. These findings prove that the model can effectively make aesthetic evaluations on car front faces that are in accordance with the aesthetic ideals of users.

In the product form aesthetic design stage, to generate a product form with high aesthetics, the sample data evaluated as "beautiful" are used as the dataset instead of all the datasets. The DCGAN is applied for image generation, and the powerful feature extraction ability of the CNN is used to improve the network performance of the GAN. After 8000 epochs, relatively clear and obvious features of the car front face are obtained in the sketch. After a small amount of manual depiction, the constructed aesthetic evaluation model is used to predict the aesthetic degree of the sketch. The result shows that the probability of beauty is 99.28%, which proves that the sketch scheme has a high aesthetic degree. However, there are also some shortcomings: limited by the sample size and hardware conditions, the resolution of the generated image is only 64×64, which makes it difficult to directly generate a clear and real car front image. It can only be used as a sketch and as a design scheme through artificial secondary processing, which increases the required human workload to a certain extent. The product samples used to generate the scheme are products that have been sold in the past three years. Whether the scheme generated based on this dataset can meet the aesthetic needs of users in the next few years still needs further research.

In this paper, we use the car front face image as the experimental sample for research. However, this does not mean that the proposed method is only applicable to the aesthetic evaluation and design of car front faces. As long as there are enough images and matching appearance scores, the method proposed in this paper can be used to complete the aesthetic evaluation and design of car front faces. However, some points need to be noted in the selection of samples. First, the chosen product images enjoy widespread public popularity and familiarity so that users can more accurately score the shape of such products. Therefore, the car front face was chosen as the research sample in this paper. Second, the chosen products are easier to score, facilitated by having dedicated websites to review them. This article uses a professional automobile evaluation website to obtain the car front face image and appearance rating. Without this convenience for obtaining images and samples, it may cost researchers considerable time and labor. Finally, the sample ratings should come from a single source, that is, from the same group.

For example, when using multiple sources of data and their evaluation groups and criteria are not consistent, then it is very likely to diminish the utility of the composed dataset. Therefore, computational verification must be performed to determine their suitability for combined use in a dataset.

VI. CONCLUSION

1) A method of aesthetic evaluation of product form based on a CNN was proposed. The method has been proven to be able to give correct and aesthetic evaluations to unfamiliar car front images. Compared with traditional aesthetic evaluation methods, this method saves time and effort. It does not require manual extraction of features. It also does not need to develop relevant evaluation criteria based on experience and professional knowledge. The simpler and more intuitive evaluation method is more commensurate with the aesthetic cognitive approach taken by ordinary people.

2) A product form aesthetic design method based on a DCGAN was proposed, and it realized the generation of a car front view sketch, assisted by a small amount of manual drawing. After evaluating the aesthetic evaluation model, the result was a car front face shape with a high aesthetic degree. Compared with the product design method using a genetic algorithm, the image generated by this method is more realistic.

3) This paper verifies the feasibility of deep learning technology in the field of aesthetic evaluation and industrial design through experiments. More advanced technology enables designers to evaluate product solutions more quickly and make more correct design decisions. The design scheme generated at the same time can guide the designer to complete the product design more efficiently. However, the method proposed in this paper also has limitations. This method only studies the automobile form from a single perspective, so it ignores the aesthetic features of other perspectives. Whether it can form a beautiful design entirety in the final application to vehicle design still needs further verification. Moreover, the research of this paper focuses on verifying the feasibility of deep learning in the fields of aesthetic evaluation and industrial design but does not address innovation of deep learning algorithms. This facet is expected to be studied from a more comprehensive perspective in subsequent studies. We will innovate deep learning algorithms to achieve better results.

REFERENCES

[1] I. J. Goodfellow, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, Oct. 2014, vol. 63, no. 10, pp. 139–144.

[2] W. Kang, S. Qin, and Q. Zhang, "Computer-aided color aesthetic evaluation method based on the combination of form and color," *Math. Problems Eng.*, vol. 2015, no. 1, Jan. 2015, Art. no. 153103.

[3] R. Maity and S. Bhattacharya, "A model to compute webpage aesthetics quality based on wireframe geometry," in *Proc. IFIP Conf. Hum.-Comput. Interact.* Cham, Switzerland: Springer, 2017, pp. 85–94.

[4] M. Kobayashi, T. Kinumura, and M. Higashi, "A method for supporting aesthetic design based on the analysis of the relationships between customer kansei and aesthetic element," *Comput.-Aided Des. Appl.*, vol. 13, no. 3, pp. 281–288, Jan. 2016.

[5] S. Orsborn, J. Cagan, and P. Boatwright, "Quantifying aesthetic form preference in a utility function," *J. Mech. Design*, vol. 131, no. 6, pp. 1–10, Jun. 2009.

[6] L. Roussos and A. Dentsoras, "Formulation and use of criteria for the evaluation of aesthetic attributes of products in engineering design," in *Proc. 19th Int. Conf. Eng. Design, Hum. Behav. Design (ICED)*, Seoul, South Korea, 2013, pp. 547–556.

[7] A. Zhou, J. Ouyang, J. Su, and S. Zhang, "Multimodal optimisation design of product forms based on aesthetic evaluation," *Int. J. Arts Technol.*, vol. 12, no. 7, pp. 128–154, Jul. 2020.

[8] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. 26th Annu. Conf. Neural Inf. Process. Syst. (NIPS)*, Lake Tahoe, NV, USA, 2012, pp. 1097–1105.

[9] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *Comput. Sci.*, vol. 140, no. 9, pp. 1–14, Sep. 2014.

[10] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, Jun. 2015, pp. 1–9.

[11] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[12] W. Wang, M. Zhao, L. Wang, J. Huang, C. Cai, and X. Xu, "A multi-scene deep learning model for image aesthetic evaluation," *Signal Process., Image Commun.*, vol. 47, pp. 511–518, Sep. 2016.

[13] M. Suchecki and T. Trzciski, "Understanding aesthetics in photography using deep convolutional neural networks," in *Proc. Signal Process.: Algorithms, Architectures, Arrangements, Appl. (SPA)*, Poznan, Poland, Sep. 2017, pp. 149–153.

[14] F. Lemarchand, "Fundamental visual features for aesthetic classification of photographs across datasets," *Pattern Recognit. Lett.*, vol. 112, no. 9, pp. 9–17, Sep. 2018.

[15] B. Dahal and J. Zhan, "USRRM: Pairwise ranking and scoring images using its aesthetic quality," *IEEE Access*, vol. 7, pp. 141171–141178, Sep. 2019.

[16] J. Zhang, Y. Miao, J. Zhang, and J. Yu, "Inkthetics: A comprehensive computational model for aesthetic evaluation of Chinese ink paintings," *IEEE Access*, vol. 8, pp. 225857–225871, 2020.

[17] L. Zhao, M. Shang, F. Gao, R. Li, F. Huang, and J. Yu, "Representation learning of image composition for aesthetic prediction," *Comput. Vis. Image Understand.*, vol. 199, Oct. 2020, Art. no. 103024.

[18] V. Mikhailava, E. Pyshkin, and V. Klyuev, "Aesthetic evaluation of food plate images using deep learning," in *Proc. 22nd Int. Conf. Adv. Commun. Technol. (ICACT)*, YeongChang, South Korea, Feb. 2020, pp. 285–289.

[19] J. Wu, B. Xing, H. Si, J. Dou, J. Wang, Y. Zhu, and X. Liu, "Product design award prediction modeling: Design visual aesthetic quality assessment via DCNNs," *IEEE Access*, vol. 8, pp. 211028–211047, Dec. 2020.

[20] D. P. Jaiswal, S. Kumar, and Y. Badr, "Towards an artificial intelligence aided design approach: Application to anime faces with generative adversarial networks," *Procedia Comput. Sci.*, vol. 168, no. 2, pp. 57–64, Feb. 2020.

[21] Y. Sagawa and M. Hagiwara, "Face image generation system using attribute information with DCGANs," in *Proc. 2nd Int. Conf. Mach. Learn. Soft Comput. (ICMLSC)*, Phu Quoc Island, Viet nam, 2018, pp. 109–113.

[22] Y. Ito, W. Shimoda, and K. Yanai, "Food image generation using a large amount of food images with conditional GAN: Ramengan and recipeGAN," in *Proc. 4th Joint Workshop Multimedia Cooking Eating Activities Multimedia Assist. Dietary Manage.*, Stockholm, Sweden, 2018, pp. 71–74.

[23] S. Radhakrishnan, V. Bharadwaj, and V. Manjunath, "Creative intelligence–automating car design studio with generative adversarial networks (GAN)," in *Proc. 2nd Int. Cross-Domain Conf. Mach. Learn. Knowl. Extraction (CD-MAKE)*, Hamburg, Germany, 2018, pp. 160–175.

[24] S. D. M. W. Kularatne, A. N. I. Nelligahawatta, D. Kasthurirathna, and S. A. Wickramage, "Deep learning based apparel product development system," in *Proc. From Innov. Impact (FITI)*, Colombo, Sri lanka, Nov. 2019, pp. 1–6.

[25] M. Egmont-Petersen, D. D. Ridder, and H. Handels, "Image processing with neural networks—A review," *Pattern Recognit.*, vol. 35, no. 10, pp. 2279–2301, Oct. 2002.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *Int. Mach. Learn. Soc. (IMLS)*, vol. 1, no. 1, pp. 448–456, Jul. 2015.

[27] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis.*, Zurich, Switzerland, 2014, pp. 818–833.

[28] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. 4th Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, 2016, pp. 1–16.

AIMIN ZHOU received the Ph.D. degree in mechanical manufacturing and automation from Lanzhou University of Technology, Lanzhou, China, in 2020. In 2002, he joined the Department of Industrial Design, Lanzhou University of Technology, where he is currently an Associate Professor. He specializes in industrial design, computational aesthetics, and intelligent design.

HONGBIN LIU is currently pursuing the degree in industrial design with the School of Design and Art, Lanzhou University of Technology. He specializes in industrial design, computational aesthetics, and intelligent design.

SHUTAO ZHANG received the Ph.D. degree in mechanical manufacturing and automation from Lanzhou University of Technology, Lanzhou, China, in 2014. In 2014, he joined the Department of Industrial Design, Lanzhou University of Technology, where he is currently an Associate Professor. For the last decade, he has been working on industrial design, Kansei engineering, and cognitive thinking. As a Project Leader, he presides over the National Natural Science Foundation of China that supports this article.

JINYAN OUYANG received the master's degree in design art from Tsinghua University, Beijing, China, in 2007. She is currently an Associate Professor with the Department of Product Design, Lanzhou University of Technology. Her recent research interests include product design, design aesthetics, and Kansei engineering.

• • •