IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 30, 2020, accepted November 15, 2020, date of publication November 20, 2020,
date of current version December 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3039715

Product Design Award Prediction Modeling: Design Visual Aesthetic Quality Assessment via DCNNs

JIANFENG WU[1], BAIXI XING[1], HUAHAO SI[2], JIAN DOU[2], JIN WANG[2], YUNING ZHU[3], AND XIAOJIAN LIU[1]

[1]Institute of Industrial Design, Zhejiang University of Technology, Hangzhou 310023, China
[2]Hangzhou Dianzi University, Hangzhou 310018, China
[3]College of Design and Architecture, Zhejiang University of Technology, Hangzhou 310023, China

Corresponding author: Baixi Xing (xingbaixi@zjut.edu.cn)

This work was supported in part by the Natural Science Foundation of Zhejiang Province of China under Grant LY19F020047.

ABSTRACT A visual aesthetic is a crucial determinant of product design evaluation. Through the analysis of image features, not only can we evaluate the aesthetic level, but also we can reveal the whole quality of the design proposal. We assume that it could be a potential pattern to predict the ultimate success of the proposal in product design that a visual aesthetic can be a cue for award classification modeling. Consequently, we conduct investigation on a dataset of over 10,003 design submissions in a design competition held once a year from 2008 to 2018 in order to manifest the assumption. Due to the remarkable performance of deep convolutional neural networks (DCNNs), we compare seven deep learning methods to explore an optimal model for design award prediction based on product image analysis. The result of the experiments indicates that the proposed method achieves comparative accuracy in design award classification result predication, with the optimal classification accuracy of 70.79% using the SEFL-ResNet (Squeeze and Excitation – Focal Loss – ResNet) method.

INDEX TERMS Computational aesthetics, product design, image analysis, deep convolutional neural networks.

I. INTRODUCTION

A high aesthetic quality of product design appearance can promote commercial sales tremendously. Consequently, a large amount of investment is put into visual aesthetic design and testing of new products before they are put on the market [1], [2], since visually appealing commodities and aesthetically pleasing packaging can attract consumers to purchase from offline and online retailers. According to the research of Bloch *et al.*, visual aesthetics were considered as the top three critical attributes in product choice [3]. Aesthetics gives the product competitiveness and establish differentiation beyond basic attributes, which makes visually appealing products rival other products with similar functionality [3], [4]. Generally, aesthetics has been regarded as a significant strategy for product research and development, marketing, and brand promotion [5]–[8]. Presently, the evaluation of product design mainly relies on the manual judgment from design experts, which is inevitably influenced by subjective preference.

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

Why are some products so popular that they can transcend their simple commodity category? Why can some product design schemes be chosen as winners among all the candidates' proposals? Can we model the design award prediction by machine learning methods? Various factors are affecting the awarding result of a design proposal. However, visual aesthetics is proved to be critical in product design evaluation by statistical analysis and user study [9]–[11]. We put forward a hypothesis that there could be a potential pattern to predict the successfulness of the proposal in product design that a visual aesthetic can be a critical cue for classification modeling.

The visual aesthetic level of product design scheme was explained by the fact that whether it is awarded in the competition, since excellent design proposal usually is also outstanding in visual aspect. Image aesthetics analysis can provide approach to predict the product design aesthetic level. In fact, image aesthetics computing has attracted many

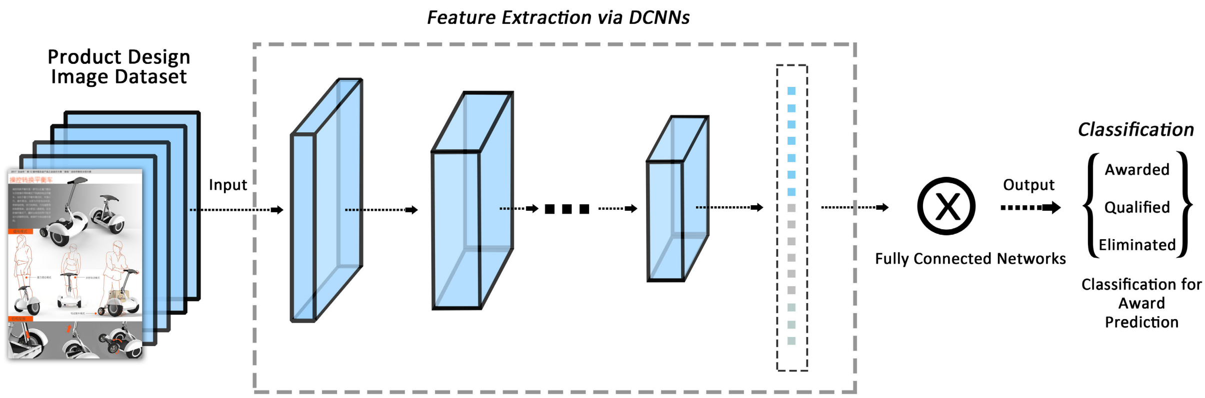

FIGURE 1. **Illustration of design award prediction modeling procedure using deep neural networks architecture.**

researchers' attention around the world [12]–[16]. Here, the relationship of visual aesthetic and image features were empirically estimated via computational algorithms in this study. We trained an aesthetic computing model based on a dataset of 10,003 product design images classified as ground-truth categories, which has 835 samples of *awarded* class (high aesthetic level), 4,990 samples of *qualified* class (middle aesthetic level) and 4,178 samples of *eliminated* class (low aesthetic level). Image features were extracted by DCNNs. We applied ResNet-50, InceptionNet, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet (Squeeze and excitation – focal loss – ResNet) to predict the classification of the design images. The result suggests that the deep learning method offers good utility and great potential to evaluate product aesthetic design automatically.

The main contributions of this research are concluded as follows. (1) A total of 10,003 product design images with ground-truth award category annotations were collected from a design competition held from 2008 to 2018 to form the novel database. (2) Multiple deep learning methods were compared to explore the optimal model for design award prediction, including ResNet-50, InceptionNet, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet. The best result was obtained by SEFL-ResNet with an average accuracy of 70.79% in classification. (3) The ratio in the total design samples in the *awarded*, *qualified* and *eliminated* class was 1: 6: 5. To train our aesthetic model with an imbalanced training dataset, re-sampling approach was studied in this work as imbalanced learning. The roadmap of this study is illustrated in Fig. 1.

This article is structured as follows: in Section 2, we review the related studies in the existing research literature; Section 3 introduces the applied DCNNs of ResNet-50, InceptionNet, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet including the methodology framework and feature analysis process; Section 4 provides the experiment procedure, including dataset construction and product aesthetic classification modeling; Section 5 presents the discussion of the experimental results; and Section 6 concludes the study and suggests the potential research direction and opportunities for further study.

II. RELATED WORKS

Various machine learning methods and statistical methods have been used in aesthetic computing, which have been proved effective by existing studies. The existing literatures have provided us with valuable classification methods, feature processing procedure and analysis of crucial factors influencing aesthetic judgments. Most of the research is based on the existing standard databases, such as AVA, Datta, and CUHKPQ etc., which studied the classification problem and score prediction of aesthetics level through statistical methods and machine learning algorithms. Specifically, the existing visual-aware product recommendation systems rarely take the aesthetic aspect into consideration. In conclusion, related research on product aesthetic assessment can be summarized in two parts: (1) Product visual appearance aesthetic quality computing and key design elements analysis; (2) Aesthetic image feature analysis and image aesthetic evaluation with various machine learning algorithms.

A. PRODUCT AESTHETIC MODELING

Here we have extensively reviewed the related works of aesthetic evaluation for product design from various aspects, including aesthetic factor analysis by statistical methods and aesthetic quality assessment by computer algorithms (see Table 1). W. Yu *et al.* took the aesthetic factor into the prediction of users' clothing preference. Clothing features of color (RGB, HSV, etc.) and semantic information (collar, hemline, fabric texture, and shape) were used in this study. The result shows that VRA-NMPR (Visually-aware Recommendation model- Neighbor-enhanced Multi-object Faced Personalized Ranking) outperforms VRA-MPR method 4.7% in F1-score [17]. M. Berghman *et al.* held two surveys to testify the principles in the Unified Model of Aesthetic and unity-in variety was proved to have the strongest impact on aesthetic pleasure [18]. Unity-in variety is thought

TABLE 1. Product aesthetic analysis studies.

Refer	Features	Classifier/Method	Descriptors	Dataset	Annotators	Results
W. Yu, 2018 [17]	Color(RGB, HSV) Sematic information (Collar, Hemline, Fabric texture, Shape)	CNN	High/Low	AVA(14 styles class) Amazon (23022 items)	-	VRA_NMPR outperforms VRA_MPR 4.7% in F1-score
P. Shamoi, 2019 [23]	color	Similarity measure	Preference level	Fashion looks (10,000 images)	22 participants	Preference for color schemes is influenced by preferences for the component basic colors and color harmony rating.
K. Liu, 2019 [27]	Color combination Composition Contrast Average saturation	SVM	Bad, poor, fair good, excellent	AVA	Amazon Mechanical Turk	Degree of consistency between the user studies and the system ranking result: 0.92
D. N. Anh, 2018 [28]	Golden ratio features, Non-local difference, balance of lighting	SVM Comply with golden ratio analysis	Attraction	-	547	MSE:0.01 SAD: 0.219 SSIM: 0.97
M. Cheung, 2019[30]	EEG	Statistical method	Beautiful/not beautiful	80 artistic paintings; 80 window displays of fashion collections	20(4M/16F)	Positive frontal alpha asymmetry is stimulated by beautiful commercial stimuli.
A.Burnap, 2019 [2]	HOG, shape	VAE, GAN	Aesthetic	1836 rated car design images	-	MAE： 0.372
Y. Wu, 2018 [33]	-	FKM, C-FKM, FAHP, SAM	Like/ Dislike	22 samples of an electric scooter	20 experts	Discover the crucial factors that influence the attractiveness of electric scooter design.
P. Lu, 2019 [35]	CNN features	VGG-SPP regression method	Low/High	SALICON AVA MSR-ICD FLMS PASCAL	-	IoU: 0.843 BDE: 0.029
L. Deng, 2012 [36]	Complexity order	INDSCAL ANOVA PREFMAP GPA	Webpage preference	24 homepages	Study 1: 47 Study 2: 57	Complexity and order have significant effect on customer's preference on webpage.
F. A. López, 2014 [20]	-	Conjoint analysis	Aesthetic appeal acceptance level	Stimuli: General elements of forms and shapes	Study1:20(12M/8F) Study2:20(11M/9F)	R-Square: 0.8235
W. Huang, 2012 [37]	Trendiness, complexity, emotion	Statistical method	Aesthetic preference	213 photos of chair design	60(39M/21F)	Both complexity and emotion showed inverted-U relationships with user aesthetic reference, trendiness presented a small positive linear relationship with aesthetic preference.
R. A. G. Post, 2016 [19]	-	Cronbach's alpha Pearson correlations Linear mixed model	Aesthetic appreciation	Photos: lamp(12), espresso machine(12);motocycle (12), car interior(12); table(12),USB-sticks(12)	Study1:33 Study2:36 Study3:31	Unity is found to be the dominant factor and it facilitates the appreciation of variety.

(CNN: Convolutional Neural Networks; VRA_NMPR: Visually-aware Recommendation Model-Neighbor-enhanced Multi-objective Faced Personalized Ranking; GAP: Global Average Pooling; MTL: Multi-task Learning Method; SAD: Sum of Absolute Difference; SSIM: Structural similarity; A&C: Assess the aesthetic quality and classify the image category; DNN: Deep Neural Networks; HOG: Histogram of Oriented Gradients; VAE: Variational Autoencoder; GAN: Generative Adversarial Networks; CC: Correlation Coefficient; RMSE: Root Mean Squared Error; IoU: Intersection over Union; BDE: Boundary Displacement Error; INDSCAL: Individual Difference Multidimensional Scaling; PREFMAP: Preference Mapping Analysis; GPA: Generalized Procrustes Analysis; MSDLM: Multi-scene Deep Learning Model).

to be a well-known principle of beauty. R. A. G. Post *et al.* investigated the influence of this factor on aesthetic appreciation. They found that unity is the dominant factor and it facilitates the appreciation of variety. Moreover, unity and variety both have positive effect on aesthetic appreciation. They arranged three studies to testify the factor correlation, using photos of lamps, espresso machines, car interiors, tables, and USB-sticks as the stimuli in the experiments [19]. Aesthetic appeal will follow certain rules of design. F. A. López et al found that objects in square shape, or have predominating straight lines are symmetrical, and they usually achieved greater acceptance. While objects with low color contrast that are arrhythmic with circular or curved forms prevailing, are less acceptable in the experiment. General elements of forms and shapes were selected as the stimuli for aesthetic acceptance study [20]. Aesthetic can not only be perceived in visual image, but also can be perceived in music and sound. The complexity of sound stimulus and sound preference was recognized as inverted-U pattern, which can achieve an optimal point. J. Delplanque *et al.* studied the sound complexity and its relation to aesthetic preference and confirmed this century-old pattern [21]. Apart from multimedia contents, we are also exposed to the influence of aesthetic labor in various services. S. Tsaur *et al.* provided a discussion on the linking between positive emotions, aesthetic labor and behavioral intentions. And they found that aesthetic labor has a significant effect on the other two factors [22]. P. Shamoi *et al.* applied FHSI color model and used fuzzy linguistic variables and hedges to propose a classification method of aesthetic judgment. The model can predict the preference level towards clothing colors. In their study, a total of 10,000 images with fashion looks were collected to form the fashion database. The result shows that users' preference for color schemes is influenced by preferences for the component basic colors and ratings of color harmony [23]. In the study of N. Myszkowski *et al.*, the visual aesthetic sensitivity test is designed to perform aesthetic judgments. They revised the VAST proposed by Götz in 1985 to achieve a substantially improved unidimensionality and structural validity [24]. V. Böβ proposed an innovative method for luxury yachts coating process. The aesthetic criteria were discussed and applied to create mathematical features and limitation of outer surface creation [25].

Aesthetic experience is not within the appreciation of images, but also within various multimedia contents. S. G. B. Johnson *et al.* conducted three studies to explore how people rated the similarity of simple mathematical arguments to landscape paintings, piano music, and they found participants' rating mainly relied on three dimensions for beauty judgment, including elegance, profundity and clarity [26]. Aesthetic experience is also important in website usability assessment. Y. Liu *et al.* discussed the principles of aesthetics in interface design in several aspects, including friendly design, color simplicity, proportion and symmetry [27].

Moreover, aesthetic evaluation can be also applied in visual attention region detection. D. N. Anh *et al.* presented a geometric aesthetic approach for visual attention region extraction from images, which can identify the region of interest in image. In this study, golden ratio features, balance of lighting, and non-local difference were applied to predict attraction region by SVM. The method was proved to be effective with a MSE of 0.01 in the modeling result [28]. Aesthetic perception is also existed in the art of linguistics. Metaphor displays aesthetic effect in various scenarios. Q. Yang *et al.* held a really interesting study to investigate metaphor's attractiveness, which indicated that attractiveness is positively correlated with figurativeness, imageability, romance and arousal, while it is negatively associated with familiarity. Hierarchical regression model and one-way ANOVA methods were utilized for analysis [29].

Additionally, physiological features analysis can be a way to detect human aesthetic perception on images. M. Cheung *et al.* collected 80 artistic paintings and 80 images of fashion window displays to explore the classification of beautiful/not beautiful images. They measured the human EEG responses on these image stimuli to found that positive frontal alpha asymmetry was stimulated by beautiful commercial stimuli [30].

How a product is perceived as pleasurable? P. Lin *et al.* studied this question and investigated the eight pleasure factors of product design based on Jordan's theory. A total of eight wooden crafts images were used for pleasure level rating in this study. The experimental result indicates that the most important attribute is color (49.5%), proportion (26.7%) and shape (23.8%) with conjoint analysis [31]. A product design with aesthetic improvement can increase commercial sales by 30% or more. The aesthetic value is critically important to product commercial value. The researchers applied machine learning method to augment user evaluation in new product aesthetic design process. Probabilistic variational autoencoder (VAE) and generative adversarial networks (GAN) were combined to train the image datasets of 1,836 rated car design schemes. The method performed well in the appeal aesthetic prediction, with a 38% improvement than conventional machine learning method. And new design proposals were generated by GAN for consideration, which proposed a promising method for automatic aesthetical product design [2].

Aesthetic evaluation of product design is also studied as a decision-making activity. The researchers proposed a new angle to solve the problem of aesthetic assessment. J. C. Arbeláez *et al.* used CARE to get feedback from users by their mobile devices. The study mainly developed an application to collect user aesthetic feedback by CARE, which is promising that the designers are able to obtain user's perception in a collaborative way [32]. Y. Wu *et al.* used continuous fuzzy Kano model to process the ambiguity of users' need, then FAHP (Fuzzy Analytic Hierarchy Process) was applied to compute evaluation criterion weight. Finally, the study discovers the crucial factors that influence the attractiveness of electric scooter design, which can provide reference for consumer perception analysis [33]. In the experiment conducted by C. Spence *et al.*, the results emphasize the similarity in aesthetic preferences for the horizontal / vertical alignment of visual perception of either paintings or plates of food. The results show that people prefer linear food elements. The participants were asked to rotate the food by moving the cursor around the center of the image until it achieved the best aesthetic quality [34]. An aesthetic-driven image cropping method was proposed based on a regression network. The experiment result outperforms the state-of-art methods, with an IoU (Intersection over Union) of 0.843 and BDE (Boundary Displacement Error) of 0.029 [35]. L Deng *et al.* collected 24 homepage of e-commerce websites to analyze the aesthetic quality. Complexity and order were selected as the salient aesthetic features for web pages. The study reveals that complexity and order have significant effect on customer's preference. There is a moderating effect of motivational orientation on consumer's preference for web complexity, while there is no such influence on the preference for web order [36]. Also, in the study of W. Huang et al in 2012, they revisited the relationship between novelty and aesthetic preference. A total of 213 photos of chair design were selected as stimuli. They employed three dimensions of product semantics, including trendiness, complexity and emotion. Both complexity and emotion show inverted-U relationships with user aesthetic preference, and trendiness presents a small positive linear relationship with aesthetic preference. The trendiness dimension has the greatest effect on novelty. Besides, a moderate level of novelty can trigger higher aesthetic preference, in comparison to product that are very typical or very novel [37]. C. Lo *et al.* used genetic algorithm to optimize the product form in 2015. They combined aesthetic measurements principles in the study to improve the aesthetic quality of product appearance [38]. Similarly, in the research of V. Cheutet *et al.* in 2005, fully free-form deformation features are introduced to be classified and it is proposed to create an efficient access to the desired shape in CAS/CAD systems [39]. R. A. G. Post *et al.* held studies on the aesthetic appreciation of six types of products. They found that unity is the dominant factor and it facilitates the appreciation of variety. Both unity and variety have positive effect on aesthetic appreciation [19]. The most related studies include W. Yu *et al.* [17], A. Burnap [2] and P. Lu al. [35], where they used the DCNNs methods to build aesthetic level evaluation model based on image datasets. The studies of aesthetic assessment on product appearance mainly used statistical methods and factor analysis to investigate the relationship between design factors and the product's aesthetic level. The computational model for aesthetic assessment of industrial product design based on large dataset is less studied in the literature.

B. IMAGE AESTHETIC MODELING

Apart from the aesthetic perception studies by psychological research method, the image aesthetic can be studied by subjective visual attributes analysis [40]–[44]. The image aesthetic evaluation method can provide the basic method and empirical research procedure [45]–[50], which can be utilized as references in product aesthetic evaluation. Machine learning algorithms, especially deep learning networks, were widely used in image aesthetic score prediction and aesthetic quality classification [51]–[54]. We have reviewed the existing studies of image aesthetic computing and listed the works in Table 2.

C. Zhang *et al.* proposed a CNN model for aesthetic classification. Global average pooling is employed to generate an aesthetic activation map and attribute activation map, which represents the likelihood of spatial location aesthetic quality and the likelihood of different attributes. They built the classification model to identify images of high and low aesthetic quality based on AVA dataset and achieved an accuracy of 78.87% [55]. Early in 2014, L. Guo *et al.* used SVM classifier to obtain a highest classification accuracy based on CUHK dataset. They combined the hand-crafting and semantic features in the experiment to achieve the best performance [56]. Aesthetic perception is also style-adaptive. F. Gao *et al.* combated the limitation of aesthetic annotation collecting and proposed a novel automatic aesthetic rating method. Aesthetic-aware features were extracted by CNN, and then SVM was applied to make style classification. Finally, they explored a multi-task learning method to learn the style-specific aesthetic evaluation model with an accuracy of 79.2% [57]. Y. Kao *et al.* developed an A&C CNN, which can simultaneously assess the aesthetic quality and classify the image category. The framework of A&C CNN has three specific CNNs for different categories. The classification accuracy achieved 86.2% based on AVA dataset [44]. X. Zhang *et al.* combined classification and regression method by multi-task learning framework to identify high/low aesthetic quality. This study addresses the challenges of taking fix-size patch as training sample and some neglect of ordinal questions in user aesthetic evaluation. GLFN-Net is equipped with random cropping method to extract the local fine-details information to obtain an accuracy of 82.95% in the experiment [58]. F. Lemarchand *et al.* built a cross-dataset aesthetic classifier based on brain-inspired image feature, extracting percentage distributions for orientation curvature, color and global reflectional symmetry. They conducted the experiment on the Datta dataset and AVA dataset using DNN method to achieve an accuracy of 71.63% [59]. W. Wang *et al.* proposed an image aesthetic analysis method based on Hadoop framework to achieve a better efficiency and user experience. A classification accuracy of 75.37% was achieved in the experiment, and the proposed method can save 67% of time cost when using Fair Scheduler [60]. W. Wang *et al.* introduced a multi-scene deep learning model (MSDLM) and successfully classified the image aesthetic quality with an accuracy of 92.59% based on the dataset of CUHKPQ and AVA. They referenced Alex_CNN to give MSDLM a strong descriptive ability on image information [61]. H. Lee *et al.* built and extended the DCNNs based approach, and made selection and edition on target images. They developed a new encoding scheme to generate a set of customized features based on SVM-SRBM method. The proposed method achieved a classification rate of 87.98%, which outperformed the state-of-art tech-

TABLE 2. Image aesthetic modeling studies.

Refer	Features	Classifier/Method	Descriptors	Dataset	Results
C. Zhang, 2018 [55]	CNN features	GAP, CNN	High/Low	AVA	Classification Accuracy: 78.87%
L. Guo, 2014 [56]	Mean of SMS, HSV, image contrast, Histogram of SMS, Hue, Blur, Spacial distribution of edges, color distribution, bright , etc.	SVM	High/Low	CUHK	Classification Accuracy: 86.78%
F. Gao, 2019 [57]	CNN features	CNN+MTL	High/Low	AVA	Classification Accuracy:79.2%
Y. Kao, 2016[44]	GIST, Fisher Vector, SIFF	A&C CNN	High/Low	AVA	Classification accuracy: 86.2%
X. Zhang, 2019 [58]	CNN features	Combine classification and regression method by multi-task learning framework	High/Low	AVA Photo.net dataset	Classification accuracy: 82.95%
F. Lemarchand, 2018 [59]	Brain-inspired visual features	DNN	High/Low	AVA Datta dataset	Classification accuracy: 71.63%
W. Wang, 2015 [60]	Image features	SVM Hadoop cluster, Map Reduce model	Low/High	Datta dataset	Classification accuracy: 75.37% CC: 0.79 RMSE: 0.244
W. Wang, 2016 [61]	CNN features	MSDLM	High/Low	CUHKPQ AVA	Classification accuracy: 92.59%
H. Lee, 2017 [62]	Encoding CNN features using SVM-SRBM	DCNN	High/Low	AVA	Classification accuracy: 87.98%
Y. Chen, 2017[63]	Weakly supervised semantic encoding image features	Five-layer CNN	Aesthetic score (0~1)	CHUK, PNE, AVA, Flickr image dataset (7000~9000 photos for each of 35 Flickr group)	CUHK: 88.79% PNE: 86.22% AVA: 84.65% Average Precision: 0.33
F. Gao, 2017 [64]	CNN features	DeepSim	Aesthetic quality score	CSIQ(866 color images), LIVE(982 images), LIVEMD(450), TID2013(3000 images)	PLCC:0.899
R. Datta, 2006[65]	color	SVM, Linear Regression	Pleasing/Displeasing	3581 photos	Classification accuracy: 70.12%
E. Mavridaki, 2015 [66]	Simplicity, colorfulness, sharpness, pattern composition	SVM	High/Low	CUHKPQ, CUHK, AVA	Classification accuracy: 85.02%
Y. Kao, 2017 [67]	CNN features	Multi-Task CNN	High/Low	AVA	Classification accuracy: 77.4%
Y. Tan, 2017 [68]	CNN features	Inception	High/Low	DPChallenge.com	Classification accuracy: 87.1%
X. Tian, 2015 [70]	CNN features	DCNN	High/Low	AVA CUHKPQ	Classification accuracy: 91.94%

(CNN: Convolutional Neural Networks; VRA_NMPR: Visually-aware Recommendation Model-Neighbor-enhanced Multi-objective Faced Personalized Ranking; GAP: Global Average Pooling; MTL: Multi-task Learning Method; SAD: Sum of Absolute Difference; SSIM: Structural similarity; A&C: Assess the aesthetic quality and classify the image category; DNN: Deep Neural Networks; HOG: Histogram of Oriented Gradients; VAE: Variational Autoencoder; GAN: Generative Adversarial Networks; CC: Correlation Coefficient; RMSE: Root Mean Squared Error; IoU: Intersection over Union; BDE: Boundary Displacement Error; INDSCAL: Individual Difference Multidimensional Scaling; PREFMAP: Preference Mapping Analysis; GPA: Generalized Procrustes Analysis; MSDLM: Multi-scene Deep Learning Model; SVM-SRBM: SVM-driven sparse restricted Boltzmann machine; DeepSim: deep similarity; PLCC: Pearson's linear correlation coefficient)

niques [62]. Y. Chen *et al.* proposed a CNN-based framework by calculating the textual and visual attributes with graphlet-based weakly supervised attributes learning method instructed by the corresponding textual attribute. The experiment demonstrates the effectiveness and inseparability of modeling components, including sparsity-constrained textual attributes, weakly supervised visual attributes localization and the normalized CNN training [63]. F. Gao *et al.* proposed DeepSim based on VGGnet, which can accurately predict the image quality across image datasets of CSIQ, LIVE, LIVEMD and TID2013 [64]. R. Datta *et al.* downloaded a total of 3581 photos from Photo.net to build the Datta dataset to discriminate images of aesthetically pleasing and displeasing with SVM and linear regression, achieving a classification accuracy of 70.12% [65]. In the study of E. Mavridaki et al, they combined the image features of simplicity, colorfulness, sharpness, pattern, composition with generic images to enhance the modeling based on SVM. The modeling performance achieves an accuracy of 85.02% in the experiment [66]. Y. Kao *et al.* showed that the semantic information is beneficial to aesthetic feature learning and the high-level features are important in aesthetic quality assessment [67]. Y. Tan *et al.* proposed an improved neural network in aesthetic computing of photos in 2016 [68]. Then they held the view that a single patch could not represent the whole image, and they cropped 10 patches of the images in the experiment and trained a fine-tuned network to predict the photos aesthetic level [69]. X. Tian *et al.* proposed query-dependent model using DCNN equipped with fewer parameters and fewer convolutional layers, which achieves better performance. The experimental classification accuracy of AVA dataset is 80.38% and the accuracy result on CUHKPQ dataset is 91.94% [70].

In general, image aesthetic computing has been investigated with different methods. Crucial elements for visual aesthetic perception have been analyzed with factor analysis and conjoint analysis. The existing studies can provide useful techniques and suggestions for research focus of design factors. By comparison, our work has three major contributions: (1) This study focuses on the exploration of product design aesthetic evaluation modeling, and a novel real-world product visual appearance image dataset was built in the experiment; (2) DCNNs methods were utilized to predict the product' visual aesthetic quality; (3) We trained our model based on the novel dataset of product design, which has ground-truth annotations collected from product competition award results. The comprehensive study which compared advanced machine learning algorithms and utilized implicit aesthetic annotation of *awarded/qualified/eliminated* proved to be effective for product design aesthetic assessment.

III. METHODOLOGIES

With respect to judging design aesthetic level of a large dataset efficiently, computational aesthetic method is in great demand in many scenarios, including new design concept evaluation, design competition review, and product online purchasing. And thus the computational evaluation can filter out the ones with low aesthetic level and free people from a great deal of work of preliminary reviewing. The framework of the experiment comprises two stages. We first set the collected product design proposal image into the standard size according to the requirement of each network. Then, we used multiple DCNNs methods to extract the deep feature and form the image feature set. Afterwards, we conducted aesthetic quality three-class classification of *awarded/qualified/eliminate* product designs based on the feature set to make a rough identification of high aesthetic quality proposals.

Many existing studies have used diverse DCNNs yielding satisfied results in aesthetic computing. Consequently, we extracted image features using ResNet-50, Inception-Net, MobileNet, DenseNet-201, VGG-19, and EfficientNet, and then fully connected network was applied in the training process. Meanwhile, SEFL-ResNet was proposed by improving ResNet-50 using extended SE-block. The experiment was implemented for a three-class classification to separate the design proposals of different aesthetic levels in the competition selection, emulating the procedure in the real-world design competition. Image features extraction approach and the selected methods are introduced below.

A. IMAGE DATA COLLECTION AND PRE-PROCESSING

The product aesthetic dataset is formed by the submission of design competition of hardware tools in China held from 2008 to 2018 once a year. We selected the submissions of design competition of 2008∼2011, 2017, and 2018 to form the dataset, since the archives of submissions is incomplete for 2012–2016. We classified the submissions into three categories: the awarded designs (*awarded* class), the ones entered into the primary selection (*qualified class*) and the eliminated designs (*eliminated* class), according to the list of award entries. The numbers of design proposals of each class in each year are presented in Table 3. A total of 11,507 product design images were collected in the original dataset. After removing the images with low resolution, damaged image files and repeated images, a total of 10,003 high-quality images (300 dpi) were gathered to build the database for modeling. The data acquisition for this aesthetic modeling study has been permitted by the Hardware Product Design Competition committee.

TABLE 3. Numbers of awarded designs, qualified designs and eliminated designs of the competition in each year.

Year	*awarded*	*qualified*	*eliminated*
2008	60	926	730
2009	162	540	445
2010	120	670	539
2011	209	1610	1297
2017	167	1200	973
2018	117	960	782
Total	**835**	**5906**	**4766**

B. FEATURE EXTRACTION VIA FINE-TUNED DCNNs

In this work, we compared multiple DCNN models to learn the product visual aesthetic quality classification and make a ranking prediction. We fine-tuned ResNet-50, Inception-Net, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet in the aesthetic computing tasks. For each network, we set the corresponding model parameters and choose a subset of the product aesthetic dataset for fine-tuning. Finally, we got the deep image feature as the output by utilizing a fully connected layer for the modeling construction in the next step. The specific feature extraction procedure details for each approach are described as follows:

A bilinear interpolation loss function was utilized in the feature extraction by each network. Bilinear interpolation loss function is an image preprocessing method, which is used for pixel calculation in image resizing. Interpolation works by using known data to estimate values at unknown points, which can solve the problem that the image will lose some pixel value in resizing [71]. Rectifier Linear function (ReL) was set to be the activation function in the network layers. Image features obtained in the extraction were reduced to a 512-dimentional features vector as an output for aesthetics learning.

C. DEEP NEURAL CLASSIFICATION MODELS

In this study, we implemented seven model training methods for comparison based on the product aesthetic database, including InceptionNet-V3, MobileNet, ResNet-50, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet, since, according to the existing related studies, these methods have been used in image aesthetic model exploration and have achieved great performance [35], [57], [62], [64], [72]. Here we summarize the characters and general framework of the applied deep learning networks in this section.

1) INCEPTIONNET

In pursuit of higher performance and better efficiency with less network weight, InceptionNet, proposed by Christian Szegedy *et al.* in 2016, is also a benchmark for deep learning network development [73]. It has been successfully applied to a large variety of computational tasks, including image aesthetic computing. It has less computational cost and network complexity than the standard deep convolutional networks, such as VGGNet and AlexNet. It has a total of 22 layers. The structure is formed with six layers of 3 × 3 convolutions and one layer of pooling, then it was followed by three layers of inception modules one pooling layer, and a logits layer, and the last layer is softmax as output. It implemented a combination of fewer parameters, effective regularization with batch-normalization method and label-smoothing, creating a solution of high-quality networks. The inception network pipeline is described in Table 4.

2) MOBILENET

MobileNet was presented by Andrew G. Howard *et al.* in 2017 [74]. It was developed for mobile and embedded vision application based on a streamlined framework, aiming to construct a lightweight deep learning method, renowned for its efficiency. It has two global hyper parameters which can trade off latency and accuracy to achieve the balance. It was proved to show strong modeling performance across various experiments and learning task in image analysis.

TABLE 4. InceptionNet network architecture.

Layer info	Layers specification		Output size
Conv1	Conv,3 x 3,32 x 2 Conv,3 x 2,32 x 1		[149,149,32] [147,147,32]
Conv2_x	ConvPadded,3 x 3,64 x 1 Max pool		[147,147,64] [73,73,64]
Conv3_x	Conv,3x3,80 x 1 Conv,3x3,192 x2 Conv,3x3,288 x1		[71,71,80] [35,35,192] [35,35,288]
3 xInception	Conv 1×1 64 Conv 1×1 48 Conv 5×5 64 Conv 1×1 64 Conv 3×3 96 Conv 3×3 96 Conv 1×1 32	x 3	[17,17,768]
5 x Inception	Conv 1×1 192 Conv 1×1 128 Conv 1×7 128 Conv 7×1 192 Conv 1×1 128 Conv 7×1 128 Conv 1×7 128 Conv 7×1 128 Conv 1×7 192 Conv 1×1 192	x 5	[8,8,1280]
2 x Inception	Conv 1×1 320 Conv 1×1 384 Conv 1×3 384 Conv 3×1 384 Conv 1×1 448 Conv 3×3 384 Conv 1×3 384 Conv 3×1 384 Conv 1×1 192	x 2	[8,8,2048]
Dense1_x	Ave_pool fc x 1		[2048] [1000]
Output	Dropout 3d,fc,Softmax		[3]

The framework of MobileNet has 28 layers. The architecture of MobileNet is based on depth-wise separable convolutions, while the depth-wise convolution applies a single filter to each input channel, described in Table 4. The depth-wise separable convolutions are formed with two layers, including depth-wise convolutions and point-wise convolutions. Batch-norm and Rectifier Linear function (ReL) are both used in MobileNet for both layers. Compared with the standard convolutional networks, MobileNet has much fewer parameters, less computation load and competitive accuracy, which constitute the superiority of this method. The network structure of MobileNet is shown in Table 5.

TABLE 5. MobileNet network architecture.

Layer info	Layers specification	Output size
Conv1	Conv,3 x3,32 x3	[112,112,32]
Conv2_x	[Conv_dw 3×3 32; Conv 1×1 32; Conv_dw 3×3 64; Conv 1×1 64]	[112,112,32] [112,112,64] [56,56,64]
Conv3_x	[Conv_dw 3×3 128; Conv 1×1 128; Conv_dw 3×3 128; Conv 1×1 128]	[56,56,128] [28,28,128]
Conv3_x	[Conv_dw 3×3 256; Conv 1×1 256; Conv_dw 3×3 256; Conv 1×1 256]	[28,28,256] [14,14,256] [14,14,512]
Conv4_x	[Conv_dw 3×3 512; Conv 1×1 512] x5	[14,14,512]
Conv5_x	[Conv_dw 3×3 512; Conv 1×1 1024; Conv_dw 3×3 1024; Conv 1×1 1024]	[7,7,512] [7,7,1024]
Dense1_x	Ave_pool fc x1	[1024] [1000]
Output	Dropout 3d,fc,Softmax	[3]

TABLE 6. ResNet-50 architecture pipeline.

Layer info	Layers specification	Output size
Conv1	Conv,7x7,64.stride 2 Max pool,3x3,stride 2	[112,112,64] [56,56,64]
Conv2_x	[Conv 1×1 64; Conv 3×3 64; Conv 1×1 256] x3	[56,56,256]
Conv3_x	[Conv 1×1 128; Conv 3×3 128; Conv 1×1 512] x4	[28,28,512]
Conv4_x	[Conv 1×1 256; Conv 3×3 256; Conv 1×1 1024] x6	[14,14,1024]
Conv5_x	[Conv 1×1 512; Conv 3×3 512; Conv 1×1 2048] x3	[7,7,2048]
Dense1_x	Average Pool, fc x3	[512] [128]
Output	Dropout 3d,fc,Softmax	[3]

3) RESNET-50

In 2015, Kaiming He *et al.* proposed ResNet, a residual learning network to ease the network training optimization, which was deeper than the existing networks at that stage, and won various prizes on the tasks of image detection and localization [75]. ResNet has two basic blocks, the identity block and the convolution block.

The existing DCNN network AlexNet and VGG usually are constructed to directly learn the mapping between input and output, while ResNet uses multiple layers to learn the residual representation between input and output. Experiments show that it is much easier (with faster convergence) and more effective (with higher classification accuracy can be achieved by using more layers) to learn residuals directly by this structure. However, the network training time of ResNet is relatively long, which limits its application. The network structure of ResNet-50 is presented in Table 6.

4) DENSENET-201

DenseNet was proposed by Huang *et al.* in 2018 [76]. Comparing to ResNet and InceptionNet, it is constructed in a new structure that is simple and effective. The network superiority can be concluded in several aspects, including vanishing-gradient reduction, enhancement of features delivery between layers, high efficiency in features utilization, and less network parameters. DenseNet has an innovative design in the dense block structure, so that the number of output feature maps of each layer can be really small. An implicit deep supervision is obtained, by means that each layer can get access to the gradients directly from the loss function and the original input data, so that it can reduce the gradient disappearance problem. The specification of the network layers are shown in Table 7.

5) VGG-19

VGG (Very Deep Convolutional Networks) was proposed by Karen Simonyan and Andrew Zisserman in 2015 [77]. VGG explores the relationship between the depth of a convolutional neural network and the network performance. By stacking 3×3 small convolutional kernels and 2×2 maximum pooling layers repeatedly, it can successfully construct a convolutional neural network with a depth of 16–19 layers. VGGNet uses 3×3 convolution core and 2×2 pooling core to improve performance by continuously increasing the depth of the network structure. There are five segments in VGGNet structure. In each segment, there are 3×3 convolutional kernels, which are followed by a maximum pooling layer. Afterwards, there are a total of three full connection layers and a softmax layer for the final output. The details of the network layers are shown in Table 8.

TABLE 7. DenseNet-201 layers specification.

Layer info	Layers specification	Output size
Conv	Conv,7x7, stride 2	112x112
Pooling	Conv,3x3, Max pool, stride 2	56x56
Dense Block_1	$\begin{bmatrix} Conv & 1\times1 \\ Conv & 3\times3 \end{bmatrix}$ x12	28x28
Transition_1	Conv, 1x1 2x2, average pool, stride	56x56 28x28
Dense Block_2	$\begin{bmatrix} Conv & 1\times1 \\ Conv & 3\times3 \end{bmatrix}$ x24	28x28
Transition_2	Conv, 1x1 2x2 average pool, stride	28x28 14x14
Dense Block_3	$\begin{bmatrix} Conv & 1\times1 \\ Conv & 3\times3 \end{bmatrix}$ x48	14x14
Transition_3	Conv, 1x1 2x2 average pool, stride	14x14 7x7
Dense Block_4	$\begin{bmatrix} Conv & 1\times1 \\ Conv & 3\times3 \end{bmatrix}$ x32	7x7
Output	Global average pool, 7x7 1000, fully connected, softmax	1x1

TABLE 8. VGG-19 layers specification.

Layer info	Layers specification	Output size
Conv1	Conv,3x3,64 x2 Max pool	[224,224,64] [112,112,64]
Conv2_x	Conv,3x3,128 x2 Max pool	[112,112,128] [56,56,128]
Conv3_x	Conv,3x3,256 x4 Max pool	[56,56,256] [28,28,256]
Conv4_x	Conv,3x3,512 x4 Max pool	[28,28,512] [14,14,512]
Conv5_x	Conv,3x3,512 x4 Max pool	[14,14,512] [7,7,512]
Dense1_x	Flatten fc	[25088] [1000]
Output	Dropout 3d,fc,Softmax	[3]

TABLE 9. EfficientNet architecture pipeline.

Layer info	Layers specification	Output size
Conv1	Conv,3x3 MBConv1, k3x3	[224,224,64] [112,112,32]
Conv2	MBConv6, k3x3 MBConv6, k5x5	[112,112,48] [56,56,80]
Conv3	MBConv6, k3x3 MBConv6, k5x5	[56x56,160] [28,28,224]
Conv4	MBConv6, k5x5	[14,14,384]
Dense1	MBConv6, k3x3 Ave_pool fc	[7,7,640] [7,7,2560] [1000]
Output	Dropout 3d,fc,Softmax	[3]

6) EFFICIENTNET

EfficientNet was proposed by M. Tan *et al.* in 2019, which balancing network specification for better performance [78]. EfficientNet is a new scaling method, which uses a simple and efficient composite coefficient to enlarge the network from three dimensions: depth, width and resolution. It does not scale the network dimensions arbitrarily. It can obtain the best set of parameters (coefficients) based on neural structure retrieval technology. EfficientNet is achieves a fast computation speed and significant model performance comparing to the existing networks. The network specification of EfficientNet is presented in Table 9.

7) SEFL-RESNET

Squeeze-and-Excitation-Focal Loss-ResNet is an improved ResNet-50 network with a squeezed excitation block and focal loss function. In the experiment, we constructed SEFL-ResNet aiming to pursue an optimal accuracy in image aesthetic assessment task, considering the model performance and efficiency. SEFL-ResNet has the advantage of Squeezed-and-Excitation block inserted as a unit in the network structure, which can recalibrates channel-wise feature responses adaptively and improve the accuracy with little additional computational cost.

Focal loss function facilitates the network with an improved cross entropy solution, in order to solve the data imbalance problem [79]. The standard cross entropy method can control the weight (w_{pn}) of positive and negative samples.

However, it cannot control the weight (w_{ed}) of samples that are easy to classify and difficult to classify. Consequently, focal loss was proposed by T. Lin et al, which can adjust both w_{pn} and w_{ed}. The focal loss F_{loss} is specified as follows:

$$F_{loss}(k_t) = -(1 - k_t)^{\alpha} \log(k_t) \quad (1)$$

In which, α is the focusing parameter, $\alpha >= 0$. And $(1 - k_t)^{\alpha}$ is the modulating factor, which is set to reduce the weight of easy negatives and make the model concentrate on the samples which are difficult to classify. In the proposed SEFL-ResNet structure, extended SE-block with double fully connected networks is inserted between layers of conv2_x, conv3_x, conv4_x, conv5_x and Dense1_x, in order to solve the problem of exploiting channel dependencies with squeeze operation and fully capture channel-wise dependencies with excitation operation [80]. Here we use global average pooling as squeeze operation. Then three fully connected layers are adopted as a bottleneck structure to construct the correlation of channels and output the weights. In the improved Squeeze-and-Excitation block structure, a fully connected network layer was added to extend the SE block processing, which enhances the resolution of network to features. We minimize the number of feature dimensions to be 1/16 of the input dimensions and activate the features with ReL to be as many as the input ones. After that, sigmoid is used to obtain normalized weights in scale of 0∼1. Finally, a scale operation is used to add the normalized weights to the features of each channel. The output feature map is then operated with a three-layer fully connected network, in which average pooling, dropout and softmax are adopted between each layer to prevent over fitting problem. The specific network structure of SEFL-ResNet is presented in Table 10 and Fig. 2.

TABLE 10. SEFL-ResNet architecture pipeline.

Layer info	Layers specification	Output size
Conv1	Conv,7x7,64.stride 2 Max pool,3x3,stride 2	[112,112,64] [56,56,64]
Conv2_x	$\begin{bmatrix} Conv & 1\times1 & 64 \\ Conv & 3\times3 & 64 \\ Conv & 1\times1 & 256 \\ fc,[16,256] \end{bmatrix}$ x3	[56,56,256]
Conv3_x	$\begin{bmatrix} Conv & 1\times1 & 128 \\ Conv & 3\times3 & 128 \\ Conv & 1\times1 & 512 \\ fc,[32,512] \end{bmatrix}$ x4	[28,28,512]
Conv4_x	$\begin{bmatrix} Conv & 1\times1 & 256 \\ Conv & 3\times3 & 256 \\ Conv & 1\times1 & 1024 \\ fc,[64,1024] \end{bmatrix}$ x6	[14,14,1024]
Conv5_x	$\begin{bmatrix} Conv & 1\times1 & 512 \\ Conv & 3\times3 & 512 \\ Conv & 1\times1 & 2048 \\ fc,[128,2048] \end{bmatrix}$ x3	[7,7,2048]
Dense1_x	Average Pool, fc x3	[512] [128]
Output	Dropout 3d,fc,Softmax	[3]

IV. EXPERIMENTS

The aim of this research is to build the optimal computational model of assessing the design aesthetic quality and find the relationship between image features and image aesthetic level. We evaluate the selected DCNNs algorithms that have potential in aesthetic modeling based on the Hardware Product Design Competition image dataset.

In the first experimental section, we first divided the dataset randomly into two sections, so that 75% of the images were set as the training set, and 25% were set as the testing set. We adopted modeling performance indices of ACC and loss to evaluate the result of different methods. The experiment procedure is described in Fig. 3.

A. DATASETS

1) PRODUCT DESIGN IMAGE COLLECTION

We collected a total of 10,003 design submission proposals from the design competition called "Hardware Product Design Competition," which is held once a year in China. The submissions from the year 2008∼2011, 2017, and 2018 were gathered to form the product aesthetic database. The image samples are shown in Fig. 4. The submission proposals were annotated as three classes, *awarded* designs, *qualified* designs and *eliminated* designs, according to the competition result of the year. The design proposal images constructed a real-world dataset, which were all evaluated by expert judges from industrial design companies and renowned design colleges in China. The award results are relatively subjective to be the implicit index for product visual aesthetic quality.

It is worth mentioning that we noticed the aesthetic style of product design is changing over time. There is a notable improvement in the overall product design quality of the submissions year by year, owing to the development of computer design assistant tools and the improvement of design ability in the industry. Consequently, it might cause some influence in the classification modeling.

2) IMAGE FEATURE DATABASE

The image features were extracted by DCNN networks. The feature dataset of ResNet-50 method was constituted with 100,352 dimensional CNN features. Meanwhile, a total of 25,088 features were extracted to form the database for VGG-19 model building. For the InceptionNet modeling, 131,072 features were extracted, in the MobileNet modeling process, 50,176 features were obtained in the experiment, and 94,084 dimensions of features were got by DenseNet-201. Additionally, 125,440 features were got by EfficientNet

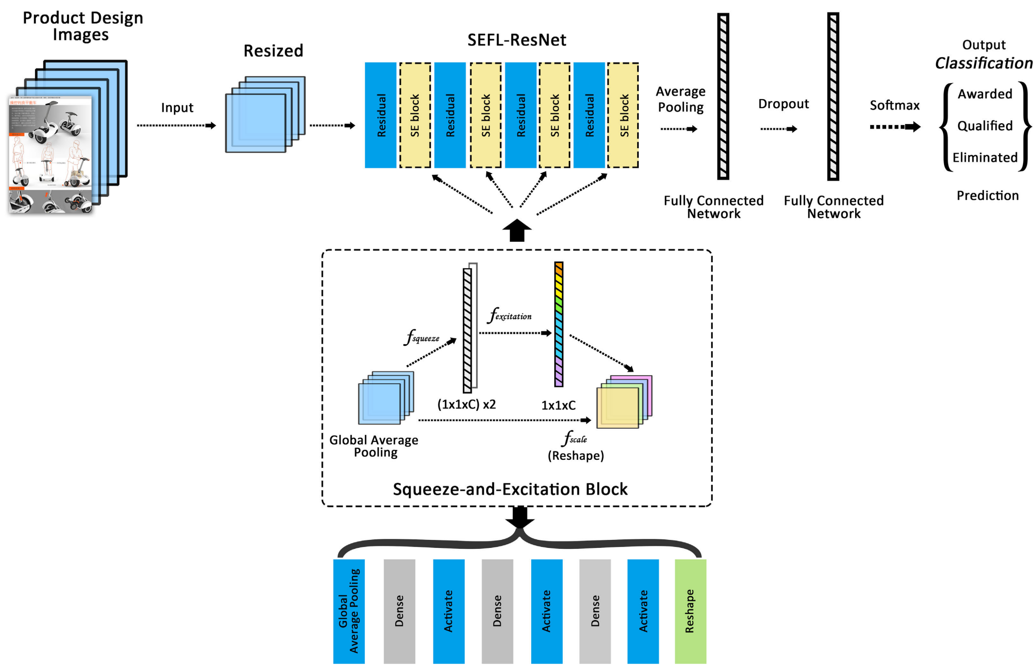

FIGURE 2. SEFL-ResNet network architecture.

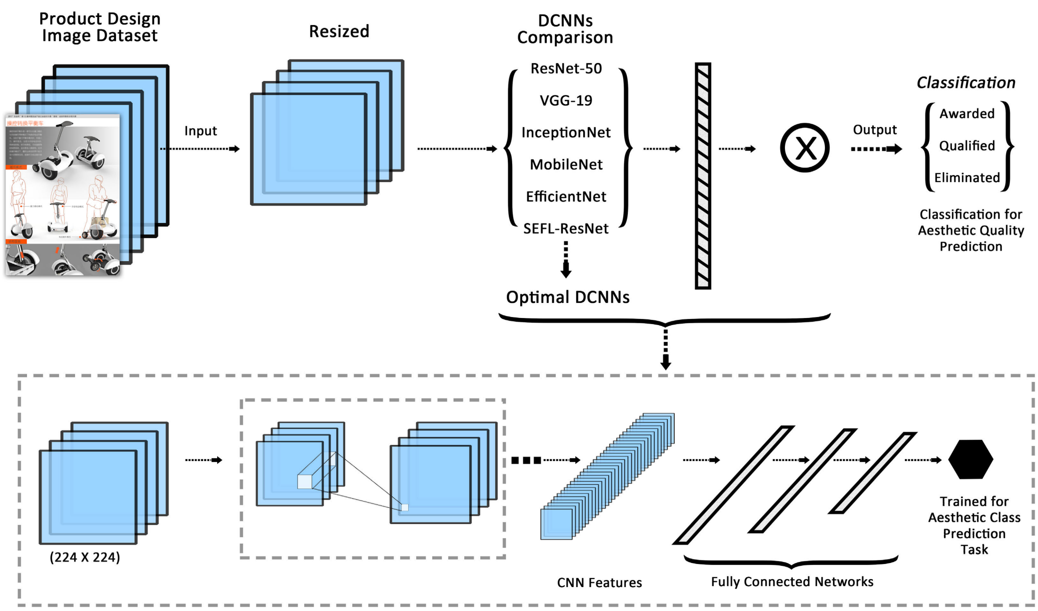

FIGURE 3. DCNNs models comparison procedure for product design aesthetic evaluation.

and 802,816 features were obtained by SEFL-ResNet. Of the dataset, 75% were used for model training and 25% were applied for model testing.

B. PRODUCT AESTHETIC CLASSIFICATION MODELING

We explored the aesthetic classification model based on the product design aesthetic dataset. Specifically, we split the dataset into two parts, so that 7,502 submitted images were used for training, 2,501 submitted images were used for testing.

1) CLASSIFICATION MODEL CONSTRUCTION

The submitted images were resized to a standard size for each network. The DCNNs networks were trained based on the datasets. The specific network training parameters of each method were described as follows:

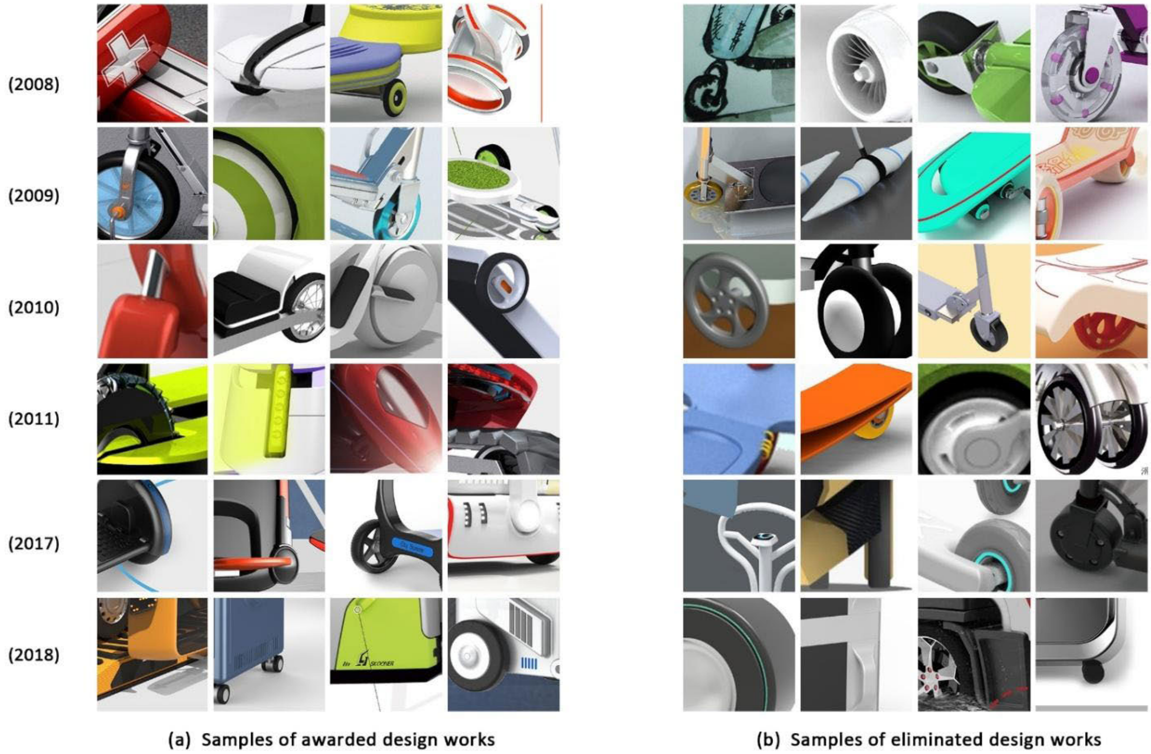

FIGURE 4. Fig (a) Awarded product design submissions layouts from the Hardware Tools Design Awards in each year and Fig (b) exhibits the eliminated submission layouts in the same year for comparison. Image cropping was conducted in the exhibition due to design copyright protection.

In the training experiment of InceptionNet-V3, MobileNet, ResNet-50, VGG-19, DenseNet, EfficientNet and SEFL-ResNet, each model was trained for 100 epochs with a batch size of 64. The network learning rate was 1e-5, and decreased step-wise by a factor of 0.5, with a patience of 9. The min_lr was set as 1e-12. The factor value affects the step size of the decline of learning rate. And the value of patience will be adjusted when the accuracy is no longer improved. In the training process, fully connected network was applied for classification. Each model was conducted with 10-fold cross validation.

2) CLASSIFICATION BASED ON IMBALANCED DATASET

It is worth noting that this product aesthetic dataset is highly skewed, that the ratio of *awarded/qualified/eliminated* designs is approximately 1: 6:5. This circumstance of data imbalance is a direct result of the preferential mechanism in the design competition, therefore only a small proportion was selected as the winners. The imbalanced dataset learning is quite challenging. Based on the theoretical literature on this issue, several effective approaches in the model training proved effective, including imbalanced learning, down-sampling, up-sampling and weighted-loss. Applying imbalanced learning method, there will be no modification on the original dataset. Whereas in training modeling with the down-sampling method, seeing that the total number of awarded ones is 835, the images of the awarded class were enriched to be as many as the qualified class in the up-sampling approach. Finally, we applied re-sampling method that combines the up-sampling and down-sampling approaches in imbalanced classification.

3) IMAGE PREPROCESSING AND DATA ENHANCEMENT

Specifically, the optimized up-sampling method adopts heuristic techniques. We used image generator to generate enhanced and standardized data of the awarded class by Keras [81]. Image sample enhancement generator creates modified samples after each iteration, aiming to achieve data enhancement result. Data enhancement can avoid the defect of simply replicate data and generate samples of different variations (rotation, shearing, zooming, etc.) in each epoch. The experimental parameters of the image generator setting are introduced below.

rescale=1/255,
featurewise_std_normalization=True,
rotation_range=40,
width_shift_range=0.2,
height_shift_range=0.2,
cval=0,
shear_range=0.2,
zoom_range=0.2,
horizontal_flip=True,
fill_mode= 'nearest'.

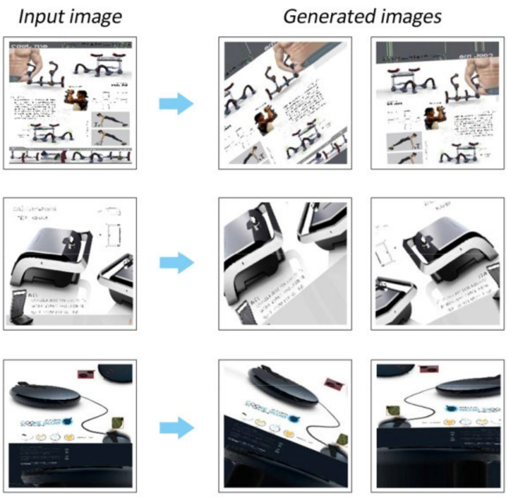

FIGURE 5. Generated image samples by image enhancement.

The generated image samples after enhancement are presented in Figure 5.

In the mean time, we conducted down-sampling to *qualified* and *eliminated* classes. Then we removed the redundant samples and noise samples to make a more balanced dataset for modeling. Finally, the proportion of each class in the obtained balanced dataset is close to 1:1:1.

4) MODEL PERFORMANCE EVALUATION METHOD

Finally, several model performance metrics were reported to evaluate the model performance, including classification accuracy and loss value. Classification accuracy (ACC) [82] and loss is the widely used indices in classification performance evaluation. Specifically, ACC refers to the proportion of correct samples predicted by the prediction model. In this experiment, ACC is calculated as:

$$accuracy = \frac{P_{awarded} + P_{qualified} + P_{eliminated}}{N_{total}} \tag{2}$$

in which, awarded P is the number of corrected classified awarded samples, qualified P is the number of corrected classified qualified samples, and eliminated P is the number of corrected classified samples in eliminated class.

Model accuracy is the measurement used to determine which method is the best at identifying patterns variables or features based on the training data. The better a method can generalize to 'unseen' data, the better insights it can produce. The learning speed will be very slow at the beginning of model training, when it is learning with gradient descent method. Consequently, the loss function of classification problem, neither the classification error rate nor the mean square error is the most suitable one. Here we used categorical crossentropy as the loss value for model evaluation. One-hot encoding method was applied to record the labels. For a dataset samples of M categories, then the labels set is labels= $(1, 2, \ldots, m)$. Therefore, if the label of the i_{th} sample is m, it is set as y_i, $m = 1$.

$$\begin{aligned} L_{\log}(Y, P) &= -\log \Pr(Y|P) \\ &= -\frac{1}{N}\sum_{i=0}^{N-1}\sum_{m=0}^{M-1}(y_{im})\log(P_{im}) \end{aligned} \tag{3}$$

Categorical-crossentropy is used to evaluate the difference between the real distribution and the probability distribution obtained by training model [83]. It describes the distance between the actual output (probability) and the expected output (probability).

The smaller the value of categorical-crossentropy is, the closer the two probability distributions are. It is indicated that SEFL-ResNet model performed better than the other fine-tuned DCNNs. The modeling results comparison is shown in Table 11.

V. RESULTS AND DISCUSSION

Experiment conducted on the product aesthetic dataset indicates that the proposed method is comparative due to its classification performance. Researchers have sought a computational way of aesthetic judgment in many previous studies; however, the user preference and subjective perception increase the difficulty of aesthetic modeling. In this study, we performed the aesthetic modeling as a classification problem of predicting image aesthetic level. An overall classification accuracy of 70.79% is thus obtained in model validation experiment, which is consistent with the results in the existing works.

A. EXPERIMENTAL RESULTS COMPARISON AND ANALYSIS OF CLASSIFICATION MODELING

In the existing studies, ResNet-50, InceptionNet, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet are considered good performers as DCNNs methods for aesthetics modeling. DenseNet and EfficientNet were applied in the experiment due to its good performance in image classification and superior efficiency. Comparison experiments were carried out to explore the best model. The detailed modeling results are: InceptionNet (ACC = 67.51%, loss = 2.1408); MobileNet (ACC = 68.31%, loss = 2.7063); ResNet-50 (ACC = 68.94%, loss = 1.5318); DenseNet-201 (ACC = 70.69%, loss = 1.9827); VGG-19 (ACC = 70.14%, loss = 1.4785); EfficientNet (ACC = 67.51%, loss = 2.1638); SEFL-ResNet (ACC = 70.79%, loss = 0.6836). As a result, the best result is achieved by SEFL-ResNet.

The specific accuracy, parameters setting and FLOPs of the models are presented in Table 11. Moreover, Fig.6 illustrates the loss during the applied networks training process, and Fig.7 presents the accuracy in the models training procedure.

The classification model was built based on the 75% of dataset of submission, and then it was tested on the rest 25% samples in the dataset. We compared the performance of four types of DCNNs to reveal that InceptionNet and EfficientNet obtained the lowest classification accuracy of 67.51% in verification experiment, and SEFL-ResNet achieved the

TABLE 11. Product image aesthetic classification accuracy.

DCNNs.	Results					
	ACC	*loss*	*Parameters setting*	*Total Parameters*	*Trainable Parameters*	*GFLOPs*
InceptionNet	67.51%	2.1408	lr = 1e-5, training_iters = 7502, batch_size = 64	21,808,931	21,774,499	3.64e+02G
MobileNet	68.31%	2.7063	lr = 1e-5, training_iters = 7502, batch_size = 64	3,256,515	3,234,627	72.7G
ResNet-50	68.94%	1.5318	lr = 1e-5, training_iters = 7502, batch_size = 64	23,643,011	23,589,891	4.94e+02G
DenseNet-201	70.69%	1.9827	lr = 1e-5, training_iters = 7502, batch_size = 64	18,373,827	18,144,771	5.49e+02G
VGG-19	70.14%	1.4785	lr = 1e-5, training_iters = 7502, batch_size = 64	3,256,515	3,234,627	2.5e+03G
EfficientNet	67.51%	2.1638	lr = 1e-5, training_iters = 7502, batch_size = 64	68,419,159	65,103,312	6.66e+02G
SEFL-ResNet	**70.79%**	**0.6836**	**lr = 1e-5, training_iters = 7502, batch_size = 64**	**4,377,875**	**4,356,627**	**9.85e+03G**

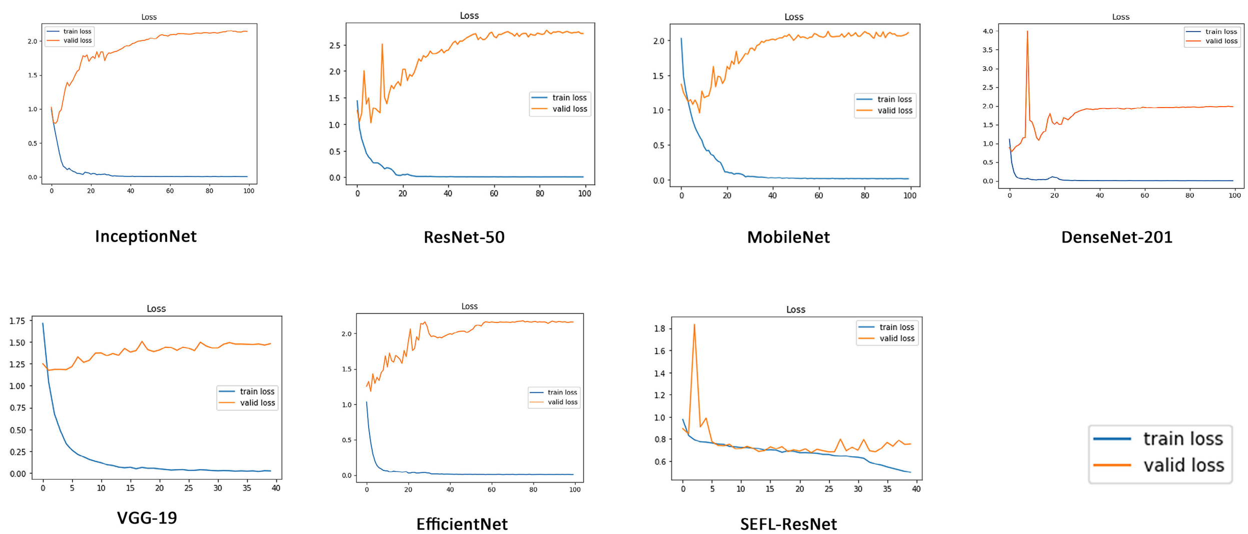

FIGURE 6. Loss during Networks training process for PDA dataset.

lowest loss value of 0.6836 and highest accuracy of 70.79% among all the selected networks in this experiment.

Recall that the optimal aesthetic binary (High/Low) classification accuracy in existing literatures is in a range of 70%∼80%, only a few studies can achieve an accuracy of 90% for binary classification task, see Table 2. The three-class classification result of the proposed method has achieved a satisfactory result, which corroborated the reported level of accuracy.

In the training experiment, the optimal classification performance becomes stable after 100 epochs of training. In the model performance evaluation, confusion matrix can provide additional evidence to measure the performance, see Fig. 8. In multiple classification problems, confusion matrix

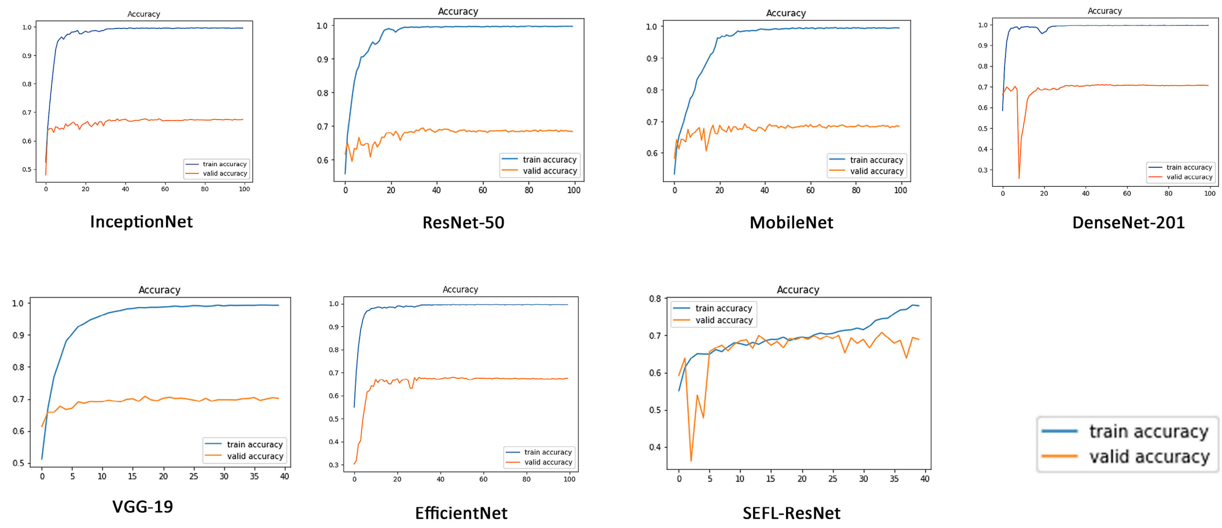

FIGURE 7. Accuracy during Networks training process for PDA dataset.

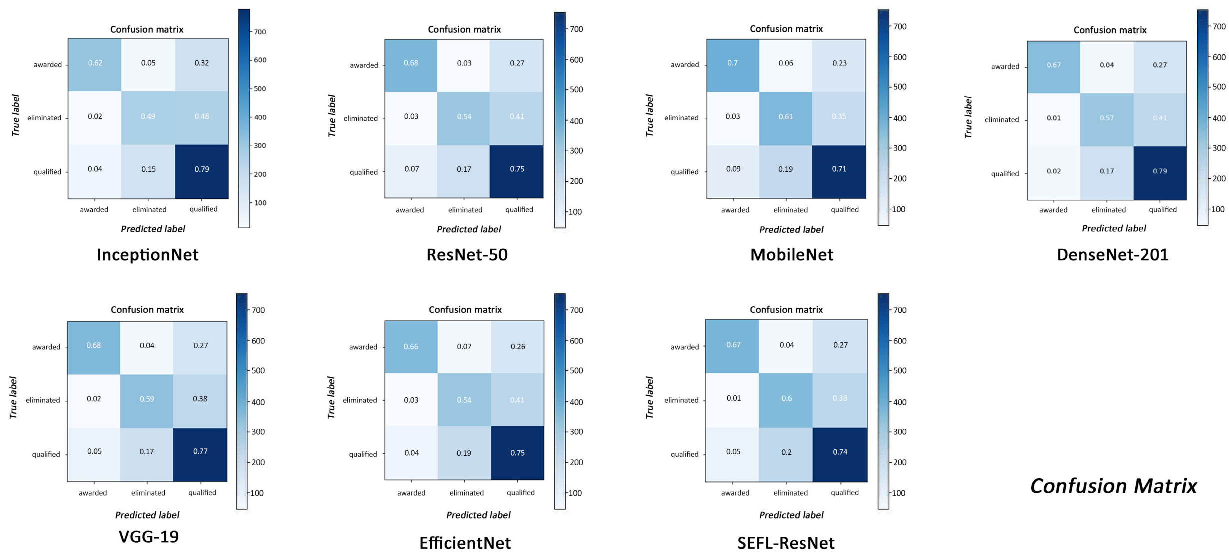

FIGURE 8. Confusion matrix for InceptionNet, ResNet-50, MobileNet, DenseNet-201, VGG-19, EfficientNet and SEFL-ResNet trained on PDA dataset.

is often used as the final evaluation method. It prevents the detailed classification performance of each category, which indicates the balance between the precision and recall.

Specifically, in the confusion matrix, labels in the abscissa are the prediction labels of *awarded/qualified/eliminated*, and labels in the ordinate are the true labels. For example, in Fig.8 the value of (*awarded, awarded*) in the matrix for SEFL-ResNet is 0.67, which indicates that the testing images of *awarded* class (true label) are predicted to be *awarded* (predicted label) by SEFL-ResNet with a probability of 0.67. After that, in the block of (*eliminated, awarded*), the value of 0.04 means that the testing images of *awarded* class are predicted to be *eliminated* class with a probability of 0.04. The higher the probability value is, the higher the corresponding classification accuracy is. According to the information in the confusion matrixes of the seven networks, SEFL-ResNet shows a relatively balanced prediction probability of each class, with an accuracy of 0.67 for the awarded class, 0.6 for the *eliminated* class and 0.74 for the *qualified* class.

B. IMBALANCED CLASSIFICATION METHOD

In fact, a hybrid method will be the optimal method to handle the data imbalance problem. Adjustment and exploration is to be made on the dataset construction and the network structure.

Here, we used the re-sampling method to balance the dataset. By increasing the numbers of training samples in awarded class and reducing the numbers of samples in qualified class, the samples distribution becomes more balanced. Consequently, the recognition rate of awarded class is improved and the generalization ability of the model is more optimal. The most direct way of up-sampling method is to copy the samples in the rare-class, but it often leads to over learning problem with poor performance of classification of rare-class.

The optimized up-sampling method adopts heuristic techniques. We used image generator to enrich data source of the *awarded* class. In the meantime, we conducted down-sampling to *qualified* and *eliminated* classes. Then we removed the redundant samples and noise samples to make a more balanced dataset for modeling. Finally, the proportion of each class in the obtained balanced dataset is close to 1:1:1.

In summary, the existing image aesthetic datasets were rated by public annotators, whereas the personal preference will influence the perception of aesthetic. The dataset collected from design competitions used award results as the ground-truth annotation, which was rated by design experts in a professional perspective, providing a promising dataset for the related research field. Besides, the method scheme can be applied in intelligent design evaluation process as a useful design assistant. The product design aesthetic evaluation with the proposed method can obtain an effective result that is consistent with human judgment.

VI. CONCLUSION

Currently, it is still challenging to predict the visual successfulness of a product scheme before a new design concept is put on the market. Designers and marketing professionals are still carrying heavy works in evaluation and analysis in pre-design and development [84], and it might be influenced by empirical and subjective factors. An increasing numbers of intelligent design methods have been applied in design industry [85], and deep learning approaches were proved to be effective in understanding product perception pattern and social manufacturing paradigm [86], [87]. We were unaware that visual appearance can be assessed with efficient aesthetic computational method based on large design schemes database as an auxiliary way. There are much discussion and debate around the rationality of deep learning method for automatic evaluation problems. However, it is also proved effective in many aesthetics level evaluation study. Here we set the product aesthetics level evaluation as a classification problem for computation. In the exploration of giving computer intelligent visual aesthetics cognition ability, deep learning methods provide possible avenue in product visual aesthetics quality evaluation.

Therefore, the current study proposed an exploration focusing on computing of the product visual aesthetic quality using the DCNNs method. Our methodology is proved to be effective on product aesthetic classification prediction based on the novel aesthetic database of product design proposals. We mainly included deep image features as the cue for aesthetic evaluation. According to the experiments, the proposed model result suggests that design proposal aesthetic quality is highly related to overall design standard, and it can be a determinant in predicting the level of the design scheme.

Specifically, a real-world design aesthetic dataset collected from works of design competitions with ground-truth annotations were built. Experiment conducted on the product design aesthetic dataset demonstrates the effectiveness of SEFL-ResNet compared to InceptionNet, MobileNet, DenseNet-201, ResNet-50, EfficientNet and VGG-19. A final aesthetic prediction is achieved with an accuracy of 70.79%. This study builds the connection between product design visual aesthetic quality and design scheme image features.

It is an interdisciplinary study and the research is related to design art and computer science. There are several avenues for the further study. Firstly, we plan to separate the content of a design proposal, for instance, design notes, figure of product main view, figures of product details. A comprehensive measurement of the aesthetic quantification of text and figures can be studied. Secondly, an efficient design evaluation system for design competitions can be developed to make a primary screening among the submissions, which can provide some ranking suggestion for reviewers. Thirdly, a fusion of multi-channel physiological signals responded to aesthetic stimuli and image features can be used to study personalized aesthetics preference modeling. The method is promising to evaluate the design scheme in the conceptual design stage and in assessing the product market positioning from the visual appearance aspect.

REFERENCES

[1] J. H. Hertenstein, M. B. Platt, and R. W. Veryzer, ''The impact of industrial design effectiveness on corporate financial performance*,'' *J. Product Innov. Manage.*, vol. 22, no. 1, pp. 3–21, Jan. 2005.

[2] A. Burnap, J. R. Hauser, and A. Timoshenko, ''Design and evaluation of product aesthetics: A human-machine hybrid approach,'' 2019, *arXiv:1907.07786*. [Online]. Available: http://arxiv.org/abs/1907.07786

[3] P. H. Bloch, ''Seeking the ideal form: Product design and consumer response,'' *J. Marketing*, vol. 59, no. 3, pp. 16–29, Jul. 1995.

[4] N. Crilly, J. Moultrie, and P. J. Clarkson, ''Seeing things: Consumer response to the visual domain in product design,'' *Design Stud.*, vol. 25, no. 6, pp. 547–577, Nov. 2004.

[5] D. A. Aaker and K. L. Keller, ''Consumer evaluations of brand extensions,'' *J. Marketing*, vol. 54, no. 1, pp. 27–41, Jan. 1990.

[6] T. M. Karjalainen and D. Snelders, ''Designing visual recognition for the brand,'' *J. Product Innov. Manage.*, vol. 27, no. 1, pp. 6–22, 2010.

[7] R. P. Jindal, K. R. Sarangee, R. Echambadi, and S. Lee, ''Designed to succeed: Dimensions of product design and their impact on market share,'' *J. Marketing*, vol. 80, no. 4, pp. 72–89, Jul. 2016.

[8] H. Cho, S. Hasija, and M. Sosa, *How Important is Design for the Automobile Value Chain*, vol. 13477. Social Science Electronic Publishing, 2016.

[9] D. Law, M.-C. Cheung, J. Yip, K.-L. Yick, and C. Wong, ''Scoliosis brace design: Influence of visual aesthetics on user acceptance and compliance,'' *Ergonomics*, vol. 60, no. 6, pp. 876–886, Jun. 2017.

[10] G. Hou and G. Lu, ''The influence of design proposal viewing strategy: Design aesthetics and professional background,'' *Int. J. Technol. Des. Edu.*, vol. 29, no. 3, pp. 543–564, May 2019.

[11] P. Nanda, J. Bos, K. Kramer, C. Hay, and J. Ignacz, ''Effect of smartphone aesthetic design on users' emotional reaction: An empirical study,'' *TQM J.*, vol. 20, no. 4, pp. 348–355, Jun. 2008.

[12] D. Nishi, S. Gitam, and K. M. Suman, ''Image aesthetics assessment using multi channel convolutional neural networks,'' *Proc. Int. Conf. Comput. Vis. Image Process.*, Sep. 2018, pp. 15–24.

[13] G. Malu, R. Bapi, and B. Indurkhya, "Learning photography aesthetics with deep CNNs," in *Proc. Modern Artif. Intell. Cogn. Sci. Conf.*, 2017, pp. 1–9.

[14] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "Predicting image aesthetics with deep learning," in *Advanced Concepts for Intelligent Vision Systems*, 2016, pp. 117–125.

[15] H. Roy, T. Yamasaki, and T. Hashimoto, "Predicting image aesthetics using objects in the scene," in *Proc. Int. Joint Workshop Multimedia Artworks Anal. Attractiveness Comput. Multimedia - MMArt&ACM*, 2018, pp. 14–19.

[16] G. Viswanatha Reddy, S. Mukherjee, and M. Thakur, "Measuring photography aesthetics with deep CNNs," *IET Image Process.*, vol. 14, no. 8, pp. 1561–1570, Jun. 2020.

[17] W. Yu, X. He, J. Pei, L. Xiong, J. Liu, and Z. Qin, "Visually-aware recommendation with aesthetic features," *IEEE Trans. Knowl. Data Eng.*, to be published, doi: 10.1007/978-981-10-7512-4_49.

[18] M. Berghman and P. Hekkert, "Towards a unified model of aesthetic pleasure in design," *New Ideas Psychol.*, vol. 47, pp. 136–144, Dec. 2017.

[19] R. A. G. Post, J. Blijlevens, and P. Hekkert, "To preserve unity while almost allowing for chaos': Testing the aesthetic principle of unity-in-variety in product design," *Acta Psychologica*, vol. 163, pp. 142–152, Jan. 2016.

[20] F. A. López and J. A. Marzal, "Quantified study of the aesthetic appeal of the formal conceptual elements in new products design through conjoint analysis," in *Design Computing and Cognition*, 2014, pp. 151–163.

[21] J. Delplanque, E. De Loof, C. Janssens, and T. Verguts, "The sound of beauty: How complexity determines aesthetic preference," *Acta Psychologica*, vol. 192, pp. 146–152, Jan. 2019.

[22] S.-H. Tsaur, H.-F. Luoh, and S.-S. Syue, "Positive emotions and behavioral intentions of customers in full-service restaurants: Does aesthetic labor matter?" *Int. J. Hospitality Manage.*, vol. 51, pp. 115–126, Oct. 2015.

[23] P. Shamoi, A. Inoue, and H. Kawanaka, "Modeling aesthetic preferences: Color coordination and fuzzy sets," *Fuzzy Sets Syst.*, vol. 395, pp. 217–234, Sep. 2020, doi: 10.1016/j.fss.2019.02.014.

[24] N. Myszkowski and M. Storme, "Measuring 'good taste' with the visual aesthetic sensitivity test-revised (VAST-R)," *Personality Individual Differences*, vol. 117, pp. 91–100, Oct. 2017.

[25] V. Böß, B. Denkena, M.-A. Dittrich, and R. Kenneweg, "Mathematical description of aesthetic criteria for process planning and quality control of luxury yachts," *Procedia CIRP*, vol. 79, pp. 478–483, Jan. 2019.

[26] S. G. B. Johnson and S. Steinerberger, "Intuitions about mathematical beauty: A case study in the aesthetic experience of ideas," *Cognition*, vol. 189, pp. 242–259, Aug. 2019.

[27] Y. Liu and Q. Zhang, "Interface design aesthetics of interaction design," in *Proc. HCII*, 2019, pp. 270–290.

[28] D. N. Anh, "Image saliency in geometric aesthetic aspect," in *Information Systems Design and Intelligent Applications* (Advances in Intelligent Systems and Computing), 2018, pp. 495–505.

[29] Q. Yang, Z. Gao, and Y. Li, "Factors contributing to the aesthetic attractiveness of metaphors in a complimentary context," *Lingua*, vol. 217, pp. 69–79, Jan. 2019.

[30] M.-C. Cheung, D. Law, J. Yip, and C. W. Y. Wong, "Emotional responses to visual art and commercial stimuli: Implications for creativity and aesthetics," *Frontiers Psychol.*, vol. 10, pp. 1–10, Jan. 2019.

[31] P. Lin, M. Yeh, and H. Lin, "Design for aesthetic pleasure," in *Proc. HCII*, 2019, pp. 519–530.

[32] J. C. Arbeláez and G. Osorio-Gómez, "Crowdsourcing augmented reality environment (CARE) for aesthetic evaluation of products in conceptual stage," *Comput. Ind.*, vol. 99, pp. 241–252, Aug. 2018.

[33] Y. Wu and J. Cheng, "Continuous fuzzy kano model and fuzzy AHP model for aesthetic product design: Case study of an electric scooter," *Math. Problems Eng.*, vol. 2018, pp. 1–13, Sep. 2018.

[34] C. Spence, J. Youssef, C. Michel, and A. Woods, "Assessing the aesthetic oblique effect in painting and plating," *Int. J. Gastronomy Food Sci.*, vol. 17, Oct. 2019, Art. no. 100168.

[35] P. Lu, H. Zhang, X. Peng, and X. Peng, "Aesthetic guided deep regression network for image cropping," *Signal Process., Image Commun.*, vol. 77, pp. 1–10, Sep. 2019.

[36] L. Deng and M. S. Poole, "Aesthetic design of e-commerce Web pages—Webpage complexity, order and preference," *Electron. Commerce Res. Appl.*, vol. 11, no. 4, pp. 420–440, Jul. 2012.

[37] W. Huang and L. Chen, "Effects of novelty and its dimensions on aesthetic preference in product design," *Int. J. Des.*, vol. 6, pp. 81–89, Aug. 2012.

[38] C.-H. Lo, Y.-C. Ko, and S.-W. Hsiao, "A study that applies aesthetic theory and genetic algorithms to product form optimization," *Adv. Eng. Informat.*, vol. 29, no. 3, pp. 662–679, Aug. 2015.

[39] V. Cheutet, C. E. Catalano, J. P. Pernot, B. Falcidieno, F. Giannini, and J. C. Leon, "3D sketching for aesthetic design using fully free-form deformation features," *Comput. Graph.*, vol. 29, no. 6, pp. 916–930, Dec. 2005.

[40] Y. Deng, C. C. Loy, and X. Tang, "Image aesthetic assessment: An experimental survey," *IEEE Signal Process. Mag.*, vol. 34, no. 4, pp. 80–106, Jul. 2017.

[41] L. Zhang, Y. Gao, R. Zimmermann, Q. Tian, and X. Li, "Fusion of multichannel local and global structural cues for photo aesthetics evaluation," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1419–1429, Mar. 2014.

[42] S. Ma, J. Liu, and C. Chen, "A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 722–731.

[43] M. Nishiyama, T. Okabe, I. Sato, and Y. Sato, "Aesthetic quality classification of photographs based on color harmony," in *Proc. CVPR*, Jun. 2011, pp. 33–40.

[44] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," 2016, *arXiv:1604.04970*. [Online]. Available: http://arxiv.org/abs/1604.04970

[45] X. Tang, W. Luo, and X. Wang, "Content-based photo quality assessment," *IEEE Trans. Multimedia*, vol. 15, no. 8, pp. 1930–1943, Dec. 2013.

[46] C. Li, A. C. Loui, and T. Chen, "Towards aesthetics: A photo quality assessment and photo selection system," in *Proc. Int. Conf. Multimedia - MM*, 2010, pp. 25–29.

[47] W. Wang, W. Zhao, C. Cai, J. Huang, X. Xu, and L. Li, "An efficient image aesthetic analysis system using Hadoop," *Signal Process., Image Commun.*, vol. 39, pp. 499–508, Nov. 2015.

[48] Z. Dong, X. Shen, H. Li, and X. Tian, "Photo quality assessment with DCNN that understands image well," in *Proc. Int. Conf. Multimedia Modeling*, 2015, pp. 535–542.

[49] L. Marchesotti, N. Murray, and F. Perronnin, "Discovering beautiful attributes for aesthetic image analysis," *Int. J. Comput. Vis.*, vol. 113, no. 3, pp. 246–266, Jul. 2015.

[50] T. V. Nguyen, S. Liu, B. Ni, J. Tan, Y. Rui, and S. Yan, "Towards decrypting attractiveness via multi-modality cues," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 9, no. 4, pp. 1–20, Aug. 2013.

[51] L. Mai, H. Jin, and F. Liu, "Composition-preserving deep photo aesthetics assessment," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 497–506.

[52] W.-H. Kim, J.-H. Choi, and J.-S. Lee, "Subjectivity in aesthetic quality assessment of digital photographs: Analysis of user comments," in *Proc. 23rd ACM Int. Conf. Multimedia - MM*, 2015, pp. 983–986.

[53] X. Lu, Z. Lin, X. Shen, R. Mech, and J. Z. Wang, "Deep multi-patch aggregation network for image style, aesthetics, and quality estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 990–998.

[54] S. Kong, X. Shen, Z. Lin, and R. Mech, "Photo aesthetics ranking network with attributes and content adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 662–679.

[55] C. Zhang, C. Zhu, X. Xu, Y. Liu, J. Xiao, and T. Tillo, "Visual aesthetic understanding: Sample-specific aesthetic classification and deep activation map visualization," *Signal Process., Image Commun.*, vol. 67, pp. 12–21, Sep. 2018.

[56] L. Guo, Y. Xiong, Q. Huang, and X. Li, "Image esthetic assessment using both hand-crafting and semantic features," *Neurocomputing*, vol. 143, pp. 14–26, Nov. 2014.

[57] F. Gao, Z. Li, J. Yu, J. Yu, Q. Huang, and Q. Tian, "Style-adaptive photo aesthetic rating via convolutional neural networks and multi-task learning," *Neurocomputing*, vol. 395, pp. 247–254, Jun. 2020, doi: 10.1016/j.neucom.2018.06.099.

[58] X. Zhang, X. Gao, W. Lu, Y. Yu, and L. He, "Fusion global and local deep representations with neural attention for aesthetic quality assessment," *Signal Process., Image Commun.*, vol. 78, pp. 42–50, Oct. 2019.

[59] F. Lemarchand, "Fundamental visual features for aesthetic classification of photographs across datasets," *Pattern Recognit. Lett.*, vol. 112, pp. 9–17, Sep. 2018.

[60] P. Wang, Z. Lin, and R. Mech, "Learning an aesthetic photo cropping cascade," in *Proc. IEEE Winter Conf. Appl. Comput. Vis.*, Jan. 2015, pp. 448–455.

[61] 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.

[62] H.-J. Lee, K.-S. Hong, H. Kang, and S. Lee, "Photo aesthetics analysis via DCNN feature encoding," *IEEE Trans. Multimedia*, vol. 19, no. 8, pp. 1921–1932, Aug. 2017.

[63] Y. Chen, Y. Hu, L. Zhang, P. Li, and C. Zhang, "Engineering deep representations for modeling aesthetic perception," *IEEE Trans. Cybern.*, vol. 48, no. 11, pp. 1–13, Dec. 2018.

[64] F. Gao, Y. Wang, P. Li, M. Tan, J. Yu, and Y. Zhu, "DeepSim: Deep similarity for image quality assessment," *Neurocomputing*, vol. 257, pp. 104–114, Sep. 2017.

[65] R. Datta, "Studying aesthetics in photographic images using a computational approach," in *Computer Vision—ECCV* (Lecture Notes in Computer Science), vol. 3, 2006, pp. 288–301.

[66] E. Mavridaki and V. Mezaris, "A comprehensive aesthetic quality assessment method for natural images using basic rules of photography," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2015, pp. 887–891.

[67] Y. Kao, R. He, and K. Huang, "Deep aesthetic quality assessment with semantic information," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1482–1495, Mar. 2017.

[68] Y. Tan, Y. Zhou, G. Li, and A. Huang, "Computational aesthetics of photos quality assessment based on improved artificial neural network combined with an autoencoder technique," *Neurocomputing*, vol. 188, pp. 50–62, May 2016.

[69] Y. Tan, P. Tang, Y. Zhou, W. Luo, Y. Kang, and G. Li, "Photograph aesthetical evaluation and classification with deep convolutional neural networks," *Neurocomputing*, vol. 228, pp. 165–175, Mar. 2017.

[70] X. Tian, Z. Dong, K. Yang, and T. Mei, "Query-dependent aesthetic model with deep learning for photo quality assessment," *IEEE Trans. Multimedia*, vol. 17, no. 11, pp. 2035–2048, Nov. 2015.

[71] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in C: The art of scientific computing*, 2nd ed. New York, NY, USA: Cambridge Univ. Press, 1992, pp. 123–128.

[72] R. He and J. McAuley, "VBPR: Visual Bayesian personalized ranking from implicit feedback," 2015, *arXiv:1510.01784*. [Online]. Available: http://arxiv.org/abs/1510.01784

[73] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-ResNet and the impact of residual connections on learning," 2016, *arXiv:1602.07261*. [Online]. Available: http://arxiv.org/abs/1602.07261

[74] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861

[75] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1–12.

[76] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269, doi: 10.1109/CVPR.2017.243.

[77] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015, pp. 1–14.

[78] M. Tan Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 1–11.

[79] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[80] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.

[81] N. Ketkar, "Convolutional neural networks," in *Deep Learning with Python*. Springer, 2017, pp. 63–78.

[82] G. Zheng, S. Li, G. Székely, and F. Crestani, *Statistical Shape and Deformation Analysis-Methods Implementation and Applications*. Amsterdam, The Netherlands: Elsevier, 2017.

[83] C. Strobl, A.-L. Boulesteix, and T. Augustin, "Unbiased split selection for classification trees based on the gini index," *Comput. Statist. Data Anal.*, vol. 52, no. 1, pp. 483–501, Sep. 2007.

[84] Y.-C. Tang, J.-J. Huang, M.-T. Yao, J. Wei, W. Li, Y.-X. He, and Z.-J. Li, "A review of design intelligence: Progress, problems, and challenges," *Frontiers Inf. Technol. Electron. Eng.*, vol. 20, no. 12, pp. 1595–1617, Dec. 2019.

[85] L. Sun, P. Chen, W. Xiang, P. Chen, W. Gao, and K. Zhang, "SmartPaint: A co-creative drawing system based on generative adversarial networks," *Frontiers Inf. Technol. Electron. Eng.*, vol. 20, no. 12, pp. 1644–1656, 2019.

[86] J. Leng, Q. Chen, N. Mao, and P. Jiang, "Combining granular computing technique with deep learning for service planning under social manufacturing contexts," *Knowl.-Based Syst.*, vol. 143, pp. 295–306, Mar. 2018.

[87] J. Leng and P. Jiang, "A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm," *Knowl.-Based Syst.*, vol. 100, pp. 188–199, May 2016.

JIANFENG WU received the Ph.D. degree in digital art and design from Zhejiang University, Hangzhou, China, in 2008. He worked as a Postdoctoral Researcher with the College of Computer Science and Technology, Zhejiang University, from 2008 to 2010. He is currently an Associate Professor with the Institute of Industrial Design, Zhejiang University of Technology. His research interests include design study, affective computing, human–computer interaction, and ergonomic study.

BAIXI XING received the bachelor's degree from the Nanjing University of Aeronautics and Astronautics, and the Ph.D. degree in digital art and design from Zhejiang University, Hangzhou, China, in 2014. She worked as a Postdoctoral Researcher with the College of Computer Science, Zhejiang University, from 2015 to 2018. She is currently an Assistant Professor with the Institute of Industrial Design, Zhejiang University of Technology. Her research interests include affective computing and multimedia retrieval. At present, she is focusing in multimodal emotion recognition and cross media retrieval. She is also interested in the research of human–computer interaction and user experience design.

HUAHAO SI is currently a Graduate Student with the School of Media and Design, Hangzhou Dianzi University. His research interests include affective computing, multimedia information analysis, and music information retrieval.

JIAN DOU is currently a Graduate Student with the School of Media and Design, Hangzhou Dianzi University. His research interests include data mining, multimedia emotion analysis, and music information retrieval.

JIN WANG is currently a Graduate Student with the School of Computer Science and Technology, Hangzhou Dianzi University. His research interests include digital design and art, multimedia retrieval, and data mining.

YUNING ZHU is currently an Assistant Professor and a Senior Designer with the College of Design and Architecture, Zhejiang University of Technology. His research interests include industrial design, service design, and human–computer interaction.

XIAOJIAN LIU received the Ph.D. degree from the Industrial Design Institute, Northwestern Polytechnical University, Xian, China, in 2005. He worked as a Postdoctoral Researcher with the College of Computer Science and Technology, Zhejiang University, from 2006 to 2008. He is currently a Professor with the Institute of Industrial Design, Zhejiang University of Technology. His research interests include intelligent design methods and the reservation of intangible cultural heritage.