Computational model for predicting user aesthetic preference for GUI using DCNNs

Baixi Xing[1], Huahao Si[2], Junbin Chen[3], Minchao Ye[3*], Lei Shi[4]

[1] *Institute of Industrial Design, Zhejiang University of Technology, Hangzhou, China*

[2] *Hangzhou Dianzi University, China*

[3] *China Jiliang University, China*

[4] *Durham University, United Kingdom*

Abstract

Visual aesthetics is vital in determining the usability of the graphical user interface (GUI). It can strengthen the competitiveness of interactive online applications. Human aesthetic preferences for GUI are implicit and linked to various aspects of perception. In this study, an aesthetic GUI image database was constructed with 38,423 design works collected from Huaban.com, a popular social network website for art and design sharing, collection, and exhibition in China. The numbers of user collection and likes of each design work were used as the annotation to represent user preference levels. Deep convolutional neural networks were applied to evaluate the aesthetic preferences of GUIs, based on a large dataset of user interface design images with the ground-truth annotations. The experimental result indicated the feasibility of the proposed method, with a mean squared error (MSE) of 0.0222 for user *collection* prediction and an MSE of 0.0644 for user *likes* prediction in the best model performance of Squeeze-and-Excitation-VGG19 networks (SE-VGG19). This study aims to build a large aesthetic image database, and to explore a practical and objective evaluation model of GUI aesthetics.

Keywords: Computational aesthetics; GUI aesthetic assessment; DCNNs

---

* Corresponding Author: Minchao Ye (Email: yeminchao@cjlu.edu.cn)

1. Introduction

A graphical user interface (GUI) provides the graphic medium of online visual environments, often presenting a user's first impression of a product or experience. Excellent GUI design not only provides visually pleasing experiences but also improves the usability and acceptability of an interface to a certain extent [1]. Aesthetics is a crucial element of an interactive system among all the user experience elements [2-4]. Although aesthetic preferences can vary among people, notably beautiful user interfaces clearly have a relatively high scoring consistency [5]. Therefore, apart from a specialized or personalized style in GUIs, general aesthetic qualities can be evaluated using computational methods. This scenario is the basis of our study on aesthetic computing and analysis of visual factors that affect user perception of an interface.

Beauty is associated with order [6], rules of design [7], and laws of nature [8]. It provides us with the possibility of using computational ways to formulate visual aesthetics. Interfaces with high-quality visual appearance are the reflection of product quality and competitiveness. The GUI captures user attention and promotes group cohesion. These notions have been supported by extant studies that showed that aesthetically designed GUI can reduce perceived complexity and cognition burdens while improving interface efficiency. However, GUI design provides various features for different functions, which increases its complexity. Thus, general design principles and guidelines need to be used to measure the fuzzy content of aesthetics [9-11]. However, the existing methods are inadequate to evaluate various aspects of GUIs .

Computational modeling provides a promising method of evaluating the aesthetic quality of GUIs, and it can lead to more objective judgments based on large samples. However, few studies have investigated this in sufficient depth. Design styles consistently change and trend over time. Therefore, invariant models are clearly not applicable for GUI aesthetics evaluation. A flexible and iterative aesthetic evaluation method is instead needed for adapting to design trends over time.

In this paper, we create a visual aesthetic evaluation model for GUI design. Our research makes three contributions to the field. We show that user collection and likes reflect user ratings of GUI aesthetic design. A total of 38,423 GUI images are collected for aesthetic modeling. User likes and collection information are collected as labels for these images [12][13]. We apply advanced computational models (i.e., ResNet-50, InceptionNet-V3, MobileNet, VGG-19 and SE-VGG19) to explore optimal results in our experiment. Our optimal experimental result achieves a mean-square error (MSE) of 0.0222 with SE-VGG19 for user *collection* prediction and an MSE of 0.0644 for user *likes* prediction. The research roadmap is presented in Figure 1.

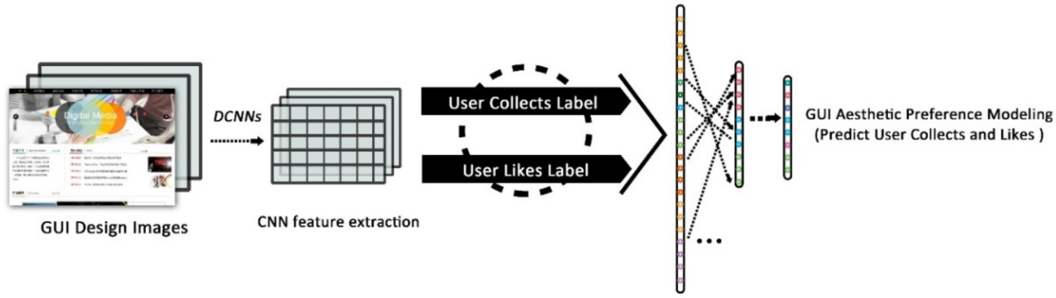

Figure 1. Research roadmap of GUI aesthetic preference modeling

The remainder of this article provides details of our study, constructed as follows: Section 2 provides a review of the related theoretical and application literatures; Section 3 introduces the computational models of ResNet-50, VGG-19, SE-VGG19, InceptionNet-V3, and MobileNet; Section 4 presents the results and discussion of the empirical results; Section 5 presents the results and discussion of the empirical results; and at last, Section 6 provides a conclusion and opportunities for further studies.

2. Related Works

According to our review of computer-related aesthetic works, evaluative judgments of aesthetic quality turn out to be predictable when using various methods based on subjective statistical image properties and user perception cues. Research has focused on intelligent aesthetic assessment of multimedia design, relying on two aspects: database construction, i.e., collections of multimedia data or aesthetic stimuli, and method selection, i.e., statistical methods or machine-learning algorithms. By identifying the determinant factors of image features, subjective scales of aesthetic ratings and effective aesthetic learning methods have been developed, providing us with useful methodologies and promising directions. A general review of related works using detailed information of databases, methods, experiment specifications, and results are presented in this section.

2.1 Aesthetic modeling for websites

Most studies have examined GUI features using statistical methods (e.g., analysis of variance), correlations, and machine learning approach. They provided key measures based on subjects' aesthetic ratings and quantitative website characteristics (e.g., color, proportion, space, symmetry, and balance). In recent years, aesthetic assessment research based on machine-learning algorithms has emerged. Many studies have explored aesthetic models based on GUI image data using various advanced algorithms, presenting novel methods and implications for this domain. A review of aesthetic assessment modeling for GUI research is presented in Table 1.

(1) GUI aesthetic assessment computing based on machine-learning methods

Machine-learning algorithms are widely applied to multimedia processing and pattern recognition. They are also applicable to aesthetic quality prediction and classification tasks. R. Chen et al. developed a fuzzy-rule-based approach to predict webpage aesthetics. Their proposed method had better predictive ability with less training time compared with a linear regression model [14]. Maity et al. used a support vector machine (SVM) to predict webpage aesthetics quality based on 11 features associated with image aesthetic and six associated with text aesthetics. They reported a study of 83 subjects' aesthetic rating on 250 images. Their model achieved a root-MSE (RMSE) of 0.68 in aesthetic image prediction and an RMSE of 0.58 in text aesthetics measurement. In their validation experiment, the proposed method achieved an RMSE of 0.79 [16]. Miniukovich et al. tested a best-fit regression model on a dataset of 300 selected webpages. They applied contrast as a weighted sum of contour pixels using the Canny algorithm. Overall, 62 (22 F, 40 M) subjects were recruited in the rating experiment. Their model optimally performed and explained 42% of the rating variance [15]. Another aesthetic computing study of on-screen images by R. Maity et al. achieved an MSE of 0.03 in rating prediction based on a non-linear regression model with an SVM classifier [4]. Later in 2017, they investigated a feature set of 13 independent webpage characters using SVM modeling to predict the aesthetic score with an accuracy of 90% [17].

(2) GUI aesthetic assessment studies based on statistical methods

Many studies conducted factor analysis for website aesthetic perception, aiming to discover crucial factors of aesthetics. For example, Maity et al. conducted GUI text aesthetic quality prediction in 2016, achieving a result of 87% accuracy with their prediction model six features of text were extracted to create the database, including chromatic contrast, luminance contrast, font size, letter spacing, line height, and word spacing. A group of 50 participants were invited to the rating experiment of 15 text samples. The statistical model was implemented in the data analysis [21]. Tuch et al. studied the role of VC (Visual Complexity) and PT (Prototypicality) in aesthetic judgment of websites. They applied the visual analog scale in the rating experiment. In the three-way ANOVA analysis, VC and PT were set as the dependent sample factors, while presentation time was set as the independent factor. The result demonstrated that websites with low VC and high PT were perceived as highly appealing, and VC and PT would affect aesthetic perception even within 17 ms. Moreover, VC affects perceived beauty more strongly, while the effect of PT was relatively blunted in the condition of a high VC level [18]. Huang et al. collected 3306 color combinations for the study of rating consistency of aesthetic preference for icon-background color combinations. They reported that the effect of gender on aesthetic preference was significant in the cognition. Finally, 30 color combinations with high rating consistency were obtained for the GUI skin design [5]. In order to discover principle determinants of GUI aesthetic, Ngo et al. assumed 14 detailed aesthetic measures based on the study of 79 undergraduate students' aesthetic cognition for five interfaces. They used mathematical methods to formulate the computation of each aesthetic determinants in the study, including measure of balance, equilibrium, symmetry, sequence, cohesion, unity, proportion, simplicity, density, regularity, economy, homogeneity, rhythm, order and complexity. These measures provided an evaluative metric and design guidelines for interface design and automated layout generation [19]. However, the interface design style and users' aesthetic preference are changing, there might be more facets of determinants of GUI in present interface design, such as interaction and colorfulness. Shamoi et al. explored aesthetic judgment modeling based on harmony and user preference level (high/low). A total of 10,000 web pages from popular fashion websites were utilized to rate color harmony and users' preference score in the experiment. Color features of FHSI (Fuzzy Hue Saturation Intensity) were extracted for the analysis. They used 2AFC (two-alternative forced choice method) and rating method in the study. Personal taste and domain-specific knowledge were applied as variables for aesthetic preference prediction. In the testing experiment, 73.3% agreement for the harmony test and 72% agreement for the preference test were achieved as a result [20]. Jylhä et al. collected 68 game app icons in 17 categories and invited 569 participants for icon aesthetic rating. In their study, app icon successfulness was predicted via icon aesthetic quality. The result showed that aesthetically pleasing character of icon can lead to more clicks, downloads and purchases [22]. Robins et al. studied aesthetics and credibility relation in website design. It is interesting to find in this work that websites with a higher aesthetic level would be perceived as more credible. In the experiment, 20 subjects were recruited to compare 21 pairs of website pages (one was the original web page, the other one was the web page with less aesthetic treatment) [23]. The experiment procedure could be improved in several aspects, including the number of website samples could be expanded and websites of different aesthetic levels could be compared in a further study. In the study of T. Lavie et al. in 2004, they listed 25 aesthetic items, seven usability items, six playfulness items, five pleasure items and five service quality items for aesthetic evaluation. Finally, classical and expressive factors were proved to explain 55.7% of the total variance of websites aesthetics quality based on the factor analysis. A confirmatory factor analysis of 384 users was conducted for cross validation [24]. In 2006, Tractinsky et al. applied statistics method to explore the relation between website aesthetics and users' attractiveness rating. They found that aesthetic dimensions of website (classical and expressive) were associated with attractiveness to users, and visual aesthetic was proved to be important in users' evaluation of websites [27]. They confirmed the result of the study of Lavie et al. in certain extent. Casey et al. implemented experiments to ensure sufficient variability in respondent's rating of aesthetic quality and social presence. A total of 181 subjects were asked to rate aesthetic score for eight web survey interfaces in a scale of 1-10. In the result of ANOVA analysis, aesthetic quality was proved to be clearly implicated in the web survey response process [25]. Seckler et al. explored the relationship between objective and subjective factors of aesthetic perception. Their study could help designers to target specific facets of visual aesthetics. A total of 144 screenshots from six categories of websites were collected as the stimuli. Then they analyzed the effects of two objective structural factors (vertical symmetry, and visual complexity) and three objective color factors (hue, saturation and brightness) on subjective aesthetic perception (simplicity, diversity, colorfulness and craftsmanship) [26]. In the research of Bauerly et al., symmetry, balance and the number of visual groups were discussed. In the rating experiment, high symmetric images of websites were preferred by users. Moreover, they also discovered that increas the number of groups in a web page caused a decrease in users' aesthetic scoring [28]. S. Park et al. invited 12 expert web designers and 418 subjects in an aesthetic cognition experiment to identify critical factors of web pages. A total of 278 terms of adjectives were considered in the rating. The experimental result showed that user perception variability is closely related to their website aesthetic fidelity [29]. Zheng et al. extracted low-level image features (color intensity, texture and entropy) and features of symmetry, balance and equilibrium to explore the correlation between image statistics and users' aesthetics and affective judgment of websites. The result indicated that balance was correlated with the largest number of aesthetic dimensions, while symmetry was less important than balance, and equilibrium is the least important character [30]. Schaik et al. enrolled 125 subjects to score aesthetic value for 62 English local-government websites. Based on an ANOVA analysis, the result showed that the context was a pivotal factor effecting users' perception stability [31]. In the study of Porat et al. in 2012, emotion status was added as a factor in users' perception on websites. Multiple research methods were conducted in the experiment, including focus groups, operationalization issues and preliminary studies and factor analysis. They suggested that the aesthetics and usability of web store influence the emotion of users, and the emotion will affect users' attitudes towards the web store. Additionally, among all the emotion states, pleasure affects the attitude most [32]. In the recent study of Lin et al., they also pointed that beauty in nature has common rules, such as golden ratio and silver ratio. These principles were strictly proportional, harmonious with rich aesthetic value. Specifically, in the study of Lin in 2012, the ratio of graphics to text in web design is discussed. The empirical results presented that the ratio of 3:1~ 1:1 would give the users the best feeling of ease-to-use and clear-to-follow, while the websites with a ratio > 3:1 will give the users the fanciest appearance [33]. Lin investigated how web homepage aesthetic quality affects users' satisfaction. They concluded three key elements based on the research of job-hunting websites, including body color, layout style and presentation form of advertisement. Specifically, layout style was significantly influencing aesthetic formality [34]. Moshagen et al. proposed VisAWI (Visual aesthetics of website inventory), using factor analysis method to conclude 18 items for visual aesthetics of websites. They found that simplicity, colors and proportion are the most important aspects among all the items [35]. In the review study of aesthetic emotion by I. Schindler et al., they reached a conclusion that aesthetic emotion is intertwined with aesthetic judgment [2]. N. Tractinsky made an investigation into the Automated Teller Machine interface. ANOVA analysis was applied to reveal that the aesthetic of the ATM system affected the post-use perceptions of both aesthetics and usability [36]. There are also multiple concepts of design aesthetics proposed in existing studies, which can be the guideline for GUI aesthetic computing. For example, APID (Aesthetic Pleasure in Design), and UMA (Unified Model of Aesthetics) were proposed in study of Berghman et al., and Garrido-Possauner et al. applied the scale for Spanish speaking countries [37]. Ciesielski et al. assumed that certain features were related to high aesthetic quality and tried to identify them. Finally, the optimal classification accuracy was around 70% and the most relevant features are obtained. Specifically, wavelet and texture features were the key features for the photographic images, while color-based features were most important for the abstract images [38].

Table 1. Aesthetic assessment modeling for website

Refer	Features	Classifier/Method	Descriptors	Dataset	Annotators	Results
R. Chen, 2016 [14]	Color, Structure, Complexity, Texture	Fuzzy rule-based model	Aesthetic rating	Webpage images	N/A	Better predictive ability than linear regression
R. Maity, 2015 [16]	Image: color contrast, hue, saturation, value, smoothness, aspect ratio, unique colour, sharpness, rule of third-hue, rule of	SVM	Aesthetic rating	250 webpages for training, 150 for validation	185	Image aesthetic prediction RMSE: 0.68 Text aesthetic prediction RMSE: 0.58 Validation interface aesthetic prediction RMSE: 0.79

	third-saturation, rule of third-value Text: font size, letter spacing, word spacing, line height, luminance contrast, chromatic contrast					
A.Miniukovich, 2015 [15]	Visual clutter, color variability, contour congestion, figure-ground contrast, layout quality, symmetry, prototypicality, ease of grouping	Regression model	Appealing score	300 webpages	62(22F/40M)	The model performed well and explained 42% of rating variance.
R. Maity, 2017 [17]	Density, economy, rhythm, simplicity, balance, cohesion, equilibrium, homogeneity, proportion, regularity, sequence, symmetry and unity	ANOVA SVM	Aesthetic score	52 webpages	100	Accuracy :90%
R. Maity, 2015 [4]	Object distribution or layout geometry related features, Content or image related features	Non-linear regression model with a SVM classifier	Aesthetic score	80 on-screen images	100	MSE: 0.03
A. N. Tuch, 2012 [18]	N/A	ANOVA	Visual complexity (VC)(High/Low); prototypicality (PT) (High/Low)	119 screenshots of websites	59(45F/14M)	Effect of PT is less pronounced than the one of VC. Websites with low VC and high PT were perceived as highly appealing.
S. M. Huang, 2012 [5]	N/A	ANOVA	Aesthetic preference rating	3306 color combinations	36(18M/18F)	30 color combinations were obtained for interface skin design.
D. C. L. Ngo, 2003 [19]	N/A	Mathematical method	N/A	5 interface images	79	Propose 14 detailed aesthetic measures as an evaluative metrics for interface aesthetics assessment
P. Shamoi, 2019 [20]	Color (Fuzzy, Hue, Saturation, Intensity)	2AFC Rating method	Aesthetic preference rating	10000 images of fashion websites	22	73.3% agreement for harmony test 72% agreement for preference test
R. Maity, 2016 [21]	Chromatic contrast, Luminance contrast, font size, letter spacing, line height, word spacing	Statistical model	Aesthetic quality rating	15 GUI text samples	50(25M/25F)	Prediction accuracy: 87%
H. Jylhä, 2019 [22]	N/A	Statistical methods	Semantic differential scale (22 adjective pairs)	68 game app icons	569	Aesthetically pleasing and good quality app icons lead to more

						clicks, downloads and purchases
D. Robins, 2007 [23]	N/A	Statistical methods	Aesthetics Credibility	21 pairs of website pages	20(14F/6M)	High aesthetic level is associated with a higher credibility ranking.
T. Lavie, 2004 [24]	N/A	Factor analysis	Aesthetics quality	One website	145(61F/84M)	Classical factor and expressive factor explained 55.7% of the total variance.
T. W. Casey, 2014 [25]	N/A	ANOVA	Aesthetic quality	8 web survey interfaces	181(115M/142F)	Aesthetic quality is clearly implicated in the web survey response process.
M. Seckler, 2015 [26]	Objective structural factors, objective color factors, subjective aesthetic perception factors	Factor analysis	Aesthetic quality	144 website screenshots	194	Websites of high symmetry, low complexity, blue hue, medium brightness or medium and high saturation are perceived with the highest score of aesthetics.
N. Tracinsky, 2006 [27]	N/A	Statistical methods	Attractive score	50 webpages	40	Aesthetic dimensions of web pages (classical and expressive) are associated with users' attractiveness rating.
M. Bauerly, 2006 [28]	Symmetry number of groups balance	Statistical methods	Aesthetic score	30 web pages	16	High symmetric images are preferred by users. Increasing the number of groups in a web page causes a decreasing in aesthetic scoring.
S. Park, 2004 [29]	N/A	Statistical methods	278 terms of adjectives	12 web pages	12 expert web designers 418 subjects(203M/215F)	User perception variability is closely related to their web page aesthetic fidelity.
X. S. Zheng, 2009 [30]	Low-level image features (color intensity, texture, entropy)	Statistical methods	Aesthetic Affective level	30 web pages	22(16F/16M)	Aesthetic correlation ranking: Balance >symmetry>equilibrium
P. V. Schaik, 2009 [31]	N/A	ANOVA	Aesthetic value	62 English local-government web sites	125(105F/20M)	The context is a pivotal factor effecting users' perception stability
T. Porat, 2012 [32]	N/A	Focus Groups Operationalization issues and preliminary studies Factor analysis	Aesthetics Usability Emotion	Web store pages	327	Design characters of web store(aesthetics and usability) influence the emotion of users, and the emotion will affect their attitudes towards the web store.
Y. C. Lin, 2012 [33]	Ratio of graphics to text	Grey relational analysis (GRA) Neural network Factor analysis	ease-to-use clear-to-follow	33 web pages	30	"Hyperlink style" element achieved the highest GRA value 0.66. Ratio of graphics to text 3:1~1:1 will give the users the best feeling of

						ease-to-use and clear-to-follow, when the ratio > 3:1 pages have the fanciest appearance.
W. Liu, 2016 [34]	Body color Layout style Presentation form of advertisements	ANOVA	Emotion (PAD) Ease of use	9 designed Job-hunting demo websites	7	Layout style significantly influences aesthetic formality.
M. Moshagen, 2010 [35]	Simplicity, diversity, colorfulness, craftsmanship	Factor analysis	Aesthetic level	100 websites	Experiment 1: 300/Experiment 2: 512	Propose VisAWI (Visual aesthetics of website inventory) Simplicity, colors, proportion are most important facets for visual aesthetics
N. Tractinsky , 2000 [36]	N/A	ANOVA Correlation analysis	Aesthetic quality usability	ATM system	132	The aesthetic quality of the ATM system affected the post-use perceptions of both aesthetics and usability.
V. Ciesielski, 2013 [38]	Colour features computed from the whole image , wavelet/texture features, colour features based on sub regions	OneR, J48, RandomForeset, SMO	Aesthetic score	Datta dataset	N/A	Accuracy: 70%

(LSBoost: Least-squares Boosting; BAG: Bagged Tree Ensembles; RF: Random Forests; POI: Place of Interest; SDAL: Semi-supervised Deep Active Learning; DNN: Deep Neural Networks; CNN: Convolutional Neural Networks; BDE: Boundary Displacement Error; mAP: mean Average Precision, Adjusted R^2: Adjusted R-square)

2.2 Aesthetic modeling for images

Works evaluating image aesthetics using computational methods were reported in [39-44]. A review of highly cited studies of image aesthetic assessment modeling is shown in Table 2. Aesthetic assessment studies have also been widely conducted for image datasets. Early in 2009, C. Li et al. used global and local image features to predict aesthetic quality by means of Naïve Bayes and adaptive boosting methods, achieving an error rate of 33.8% [45]. Then, W. Jiang et al. conducted an experiment to classify photos according to their aesthetic value. They collected 450 photos from Flickr, Kodak Picture of the Day, and studied observers. Their study reported that 85% of images had no or small class misplacements via the Diff-Rank Boost regression method [46]. Moorthy et al. predicted visual aesthetic judgments of consumer videos on a YouTube dataset comprising 160 video segments, achieving a prediction accuracy of 73%. Each video was rated by 16 subjects on an aesthetic quality scale of (−2, 2). In their work, frame-level, micro-shot, and video-level features were extracted for prediction modeling, including actual frame rate, motion features, sharpness/focus of the region of interest, colorfulness, luminance, and color harmony [47]. P. Lu et al. took advantage of the intrinsic structural properties of conditional random fields to build a color harmony model of images. They applied a Gated Convolutional Neural Network (GCNN) for the aesthetic quality assessment and achieved an AUC of 0.876 based on the AVA dataset [48]. Y. Tan et al. collected 47,304 photos from DPChallenge.com to predict aesthetic scores (high/low) based on deep-CNN (DCNN), reaching a classification accuracy of 87.10% [49]. W. Wang et al. built a photo aesthetic evaluation model based on the Datta dataset (3,140 images) using an SVM algorithm, achieving a classification accuracy of 82.4%. Additionally, a total of 16 novel features were obtained, including those of processing complexity, image color, texture, rules-base, color templates, dark channel, and depth of field, information theory-based features, and image complexity [50]. W. Wang et al. presented a multi-scene deep-learning model to realize automatic aesthetic image-feature learning. A scene convolutional layer was designed, giving the model a comprehensive aesthetic learning capacity. They applied their proposed model to the CHUKPQ and AVA datasets to realize competitive performance in aesthetic score prediction [51]. C. Zhang et al. presented an end-to-end CNN for aesthetic classification and a sample-specific classification method, achieving an accuracy of 78.87% with the AVA dataset. They also pointed out which image areas supported the aesthetic prediction in the study [52]. An adaptive-layout-aware multi-patch DCNN for photo aesthetic assessment was developed by S. Ma et al. This method achieved an accuracy of 82.5% based on the AVA dataset, outperforming other state-of-the-art methods [53]. W. Yu et al. applied the aesthetic assessment model to clothing recommendations based on a feature set of CNN extracted features, scale-invariant feature transforms, and color histogram features. They proposed a brain-inspired deep-structure neural network to build model using AVA and Amazon datasets. The modeling results outperformed other methods that used single-modal features (e.g., CNN-only or aesthetic-only) [54]. In the survey of M. Kucer et al, they performed a baseline comparison of four CNN models (VGG16, VGG19, ResNe50, InceptionNet), including CUHKPQ, HiddenBeauty, AVA and Kodak. Therefore, a result improvement of up to 2.2% can be achieved by fusing CNN features and multiple hand-crafted features.

Table 2. Aesthetic assessment modeling for images

Refer	Features	Classifier/Method	Descriptors	Dataset	Annotators	Results
Y. Tan, 2017 [49]	image	DCNN	Aesthetic score (High/Low)	47304 images from DPChallenge.com	More than 100 users	Classification accuracy: >87.10%
W. Wang, 2016 [50]	Color, color templates, texture, rule-based features, dark channel, depth of field, etc.	SVM	Aesthetic score (High/Low)	Datta dataset (3140 images)	>10 users	Classification accuracy: 82.4%
P. Lu, 2019 [48]	Color	Gated CNN	Aesthetic score	AVA	N/A	AUCs: 0.875
A. K. Moorthy, 2010 [47]	Video low level features	SVM	Aesthetic Value (High/Low)	160 consumer videos from YouTube (15 second segment)	33	Aesthetic prediction accuracy:73%
W. Wang, 2016 [51]	image	Multi-scene deep learning	Aesthetic score	CHUKPQ AVA	N/A	Accuracy: CUHKPQ: 91.69%~94.92% AVA: 84.88% & 76.94%
C. Zhang, 2018 [52]	image	End-to-end CNN	Aesthetic score	AVA	N/A	Accuracy: 78.87%
S. Ma, 2017 [53]	image	Adaptive layout-aware multi-patch DCNN	Aesthetic score	AVA	N/A	Accuracy: 82.5%
W. Yu, 2018 [54]	CNN features, SIFT features and color histogram	Brain-inspired deep structure pretrained neural network	Aesthetic score	AVA and Amazon dataset	N/A	Outperforms CNN feature based method and Aesthetic feature based method 506% and 8.79% on Recall@50.

C. Li, 2009 [45]	Global and local features	Naïve Bayes, Adaptive boosting	Aesthetic quality (high/low)	100 paintings from google image search	42 subjects	Classification error rate: 33.8%
W. Jiang, 2010 [46]	image	Diff-RankBoost regression method	Aesthetic quality (high/low)	450 images from Fllickr, Kodak picture of the day, etc.	30 subjects	85% images have no or small class misplacement
O. Wu, 2016[43]	Image ((structural, local visual, global visual, and functional)	Structural SVM and Multitask Fusion Learnin	Aesthetic score	1000 webpages	Multi-users	Testing error: 0.2743
X. Lu, 2015 [44]	Image (global, local image features)	Double-column DCNN	Aesthetic score	AVA and IAD	N/A	Accuracy:75.42%
L. Zhang, 2014 [55]	image features	Embedded algorithm	Aesthetic score(High/Low)	AVA CUHK PNE	N/A	Accuracy:90.3%
N. Murray, 2012 [56]	Image (SIFT, LBP, Color)	SVMs	Aesthetic score Low/High	AVA 250,000 images	Hundreds of amateur and professional photographers	Mean average precision: 53.85%
M. Kucer, 2018 [57]	Image hand-designed features and CNN features	VGG16, VGG19, ResNe50, InceptionNet	Aesthetic score Low/High	CUHKPQ HiddenBeauty, AVA, Kodak	N/A	An improvement of up to 2.2% can be achieved by fusing CNN features and hand-crafted features

(LSBoost: Least-squares Boosting; BAG: Bagged Tree Ensembles; RF: Random Forests; POI: Place of Interest; SDAL: Semi-supervised Deep Active Learning; DNN: Deep Neural Networks; CNN: Convolutional Neural Networks; BDE: Boundary Displacement Error; mAP: mean Average Precision, Adjusted R^2: Adjusted R-square)

2.3 User preference modeling for images

In prevalence of social media, users will actively create and interact with massive user-generated contents with their own preference and comments. Among all the user-generated data, image is one of the major multimedia modalities that show user aesthetic preference. Here we reviewed some related works of user preference modeling in different manners, in order to provide experimental method for user aesthetic preference prediction via *collection* and *likes* count. The related works of likes modeling can be concluded in three aspects: (1) *Like* prediction via user character analysis. The research predicted whether a user would like the image based on multiple user characters expressed in social media; (2) Likes prediction based on likes counts and image features via machine learning methods; (3) User *likes* prediction based on users' social cues, including social media data of *follows*, *groups* and content semantics data, such as *interest* and *user comments.* Some of these related works are listed below.

Research of Likes is quite important in understanding user preference in social media data and developing outstanding personalized service. Many researchers have addressed this topic in their study in various aspects. D. Lee proposed tutorials of"Likeology" in *WWW'15* conference [58] and *WebSci'2016* conference [59]. The tuorials present a comprehensive overview of Likes study in social media, including topics of Likes modeling, prediction of the evolution of Likes, and how to aggregate Likes. S. C. Cuntuku et al. combined image visual and tags textual features to predict user likes based on a dataset of 60,000 images crawled from Flickr. They tried to predict which user would have liked an image. They found that the fusion of text and visual features can further boost the prediction performance with nAUC of 80% [60]. A. Khosla et al. investigated the topic of "what makes an image in social network". A dataset of 2.3 million images from Flickr was used to predict the normalized view count of images. The experimental result achieves a rank correlation of 0.81 based on both image content and social cues using support vector regression method [61]. S. Ohsawa et al. conducted likes prediction modeling to discover the mutual influences of Likes among related entities. They constructed a dataset of 20 million pages with 30 billion likes from Facebook. They achieved a prediction root mean square of 0.810 in the experiment [62]. J. Y. Jang et al. proposed a study to explore the question of "How do teens use and engage in Instagram compared to adults? ". They detected age information based on the analysis of textual and facial features and presented a study of 27K users in Instagram. The study shows that teens tend to have more likes than adults, but teens post fewer photos than adults [63]. They also conducted a study to present like activities analysis based on a dataset of 20 million users and their 2 billion like activity in Instagram. They defined Like Network in this study and measured the trend of a like network [64]. M. Kosinski et al. used logistic linear regression method to predict user psychodemographic profiles from Facebook likes. The prediction result achieves a correlation score r of 0.43, which is very close to the test-retest reliability for Openness ($r=0.5$) [65].

Table 3. Image user preference modeling

Refer	Features	Classifier/Method	Descriptors	Dataset	Annotators	Results
S. C. Cuntuku et al., 2015 [60]	Visual and textual features	CNN	User likes	60,000images from Flickr	300 users	User liking prediction accuracy: nAUC of 80%
A. Khosla et al.,2014[61]	Image deep learning features	SVR	View counts	2.3 million images from Flickr	400K users	View counts prediction accuracy: rank correlation of 0.81
S. Ohsawa et al., 2013[62]	Features of related entities	SVR	User likes	20 million pages from Facebook	-	User likes prediction accuracy: root mean square of 0.810
J. Y. Jang et al., 2015 [63]	Textual and facial features	Textual pattern recognition algorithm and Face ++	Age information	User profiles in Instagram	27K users	teens tend to have more likes than adults, but teens post fewer photos than adults
J. Y. Jang et al., 2015 [64]	User information	statistical method	User likes	User information of 20 million users and their 2 billion likes	20 million users	Like network is formed and developed by both followers and random users.
M. Kosinski et al., 2013 [65]	User information	logistic linear regression	User likes	Users' Facebook likes	58000 users	The prediction result achieves a correlation score r of 0.43, which is very close to the test-retest reliability for Openness (r=0.5)

In conclusion, aesthetic assessment via statistical methods and machine learning algorithms has been well discussed in existing studies. However, intelligent aesthetic judgment using the ground truth annotation of implicit user preference cues has not been fully investigated. In this work, we applied the information of UI image collects and likes as a cue for aesthetic assessment, which is a new perspective for evaluating the user aesthetic preference level. While image aesthetic score (High/Low) was used as annotation for aesthetic classification in most aesthetic assessment studies**.** Consequently, the research scheme using massive social media data and deep learning algorithms can be a promising direction with great potential.

3. Methodologies

DCNNs have emerged as effective tools for aesthetic computing. As depicted in Fig 3~Fig 7, collected GUI images were zoomed to standard size according to the requirement of each network without cropping. Then, the input images were put into the first convolutional layer filters. We processed the input images using four deep-learning networks having different architectures. The output produced a distribution over the *likes* and *collection* labels to indicate user aesthetic preference. The general pipeline of our proposed method comprises two steps: feature extraction by DCNN and aesthetic evaluation modeling exploration. To find the optimal model, we compared the modeling performance of several DCNNs (i.e., ResNet-50, VGG19, SE-VGG19, InceptionNet-V3, and MobileNet) to assess aesthetic quality by predicting the number of *likes* and *collection* of GUI works. The detailed method is presented as follows.

3.1 GUI image data collection

Digital image producers have created many GUI works, resulting in an image database large enough for aesthetic assessment. We collected design images from Huaban.com, a well-known website for UI designers in China, to form our GUI database. Because visual appearance is presumably the main factor affecting the numbers of *likes* and *collection*, the user ground-truth labels of *likes* and *collection* for each work were collected to form a real-world index for the archive, which can indicate the aesthetic quality of the GUI images [56][67]. Consequently, the number of *likes* and *collection* should help predict the levels of aesthetics. A total of 38,423 works was collected for this experiment. The images were collected in a format of 72 dpi, and the input images for DCNNs were resized to standard size without cropping. The distribution map of the samples with *collection* and *like* counts in the dataset is illustrated in Fig. 2. The abscissa value is the *collection* counts of images and the ordinate value is the *likes* counts. It is shown that a few samples have high degree of dispersion. In order to solve the data dispersion problem, we conducted the annotation normalization using $Z_{\log}$. The specific annotation normalization method is introduced in section 4.

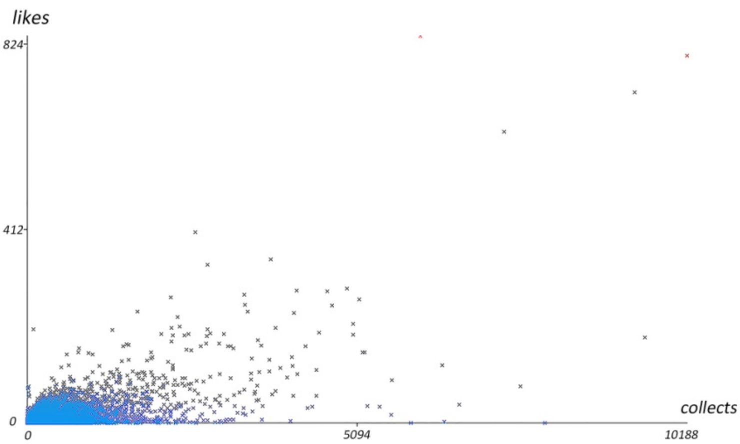

Figure 2. Distribution map of the samples with collection and like counts in the dataset.

3.2 Feature Extraction

CNN image features can express rich image information for fuzzy content, and so they are widely used to explore aesthetic computing (see Tables 1 and 2). Aesthetic GUI image features were extracted using corresponding deep-learning approaches. Because each GUI part has its own functionality, and image cropping can influence the aesthetic perception of the interface, we did not crop images to maintain its integrity.

The input GUI images were zoomed to 299 × 299 pixels for InceptionNet-V3. The input images were resized to 224 × 224 pixels, when they were fed into MobileNet, VGG-19, and ResNet-50. An output vector of 1000 features was obtained in the feature extraction. It was reduced to a 512-dimensional feature vector with a fully connected network. The specific feature extraction process is described as follows.

- MobileNet

A total of 1024 dimensions were extracted. Then, the feature set was reduced to a 1000-dimensional feature vector having a fully connected network. At last, an output vector of 512-dimensions was obtained.

- InceptionNet

The input GUI images were zoomed to 299×299 pixels via bilinear interpolation. They were then fed to the first convolutional layer filters. An output vector of 2048 features was obtained, which was reduced to a 512-dimensional feature vector using the next step.

- VGG-19

The input GUI images were fed into VGGNet and zoomed to 224×224 pixels. With the VGG-19 feature extraction, a total of 25,088 dimensions were obtained. Then, after feature dimension reduction, a total of 1,000 features were used for the output vector, which was further reduced to a 512-dimentional feature vector with a fully connected network.

- ResNet-50

We used ResNet-50 to extract a total of 2048 feature dimensions. A total of 1000 features were obtained for the output vector, which was reduced to a 512-dimentional feature vector with a fully connected network.

- SE-VGG19

The input GUI images were fed into VGG19 network and resized to 224×224 pixels. With the VGG-19 feature extraction and Squeeze Excitation Scale, a total of 25,088 dimensions were obtained. Then, after feature dimension reduction, a total of 1,000 features were used for the output vector.

3.3 Algorithms

In this study, we applied Python to build the DCNN model for the image processing experiments. Four types of DCNNs were implemented for the aesthetic prediction, including MobileNet, InceptionNet-V3, VGG-19, SE-VGG19, and ResNet-50. A brief description of these algorithms is presented below.

(1) Fully Connected Neural Network

A fully connected neural network is essentially a multilayer neural network that connects inputs and outputs. Specifically, it is applied in the last steps of a DCNN to obtain the final classification results. The fully connected network has advantages of reliability, low latency, and large network throughput. It can produce state-of-the-art accuracy for image processing, speech recognition, and video analysis. It can also achieve successful learning results with large model sizes [53].

(2) CNN

CNN is a class of deep neural networks and is most commonly applied to analyzing visual imagery. CNNs are regularized versions of multilayer perceptrons. A multilayer perceptron usually refers to fully connected networks, where each neuron in one layer is connected to all neurons in the next layer. The full connectedness of these networks makes them prone to overfitting. Typical regularization methods include adding magnitude measurements of weights to the loss function. However, CNNs take a different approach toward regularization; they take advantage of the hierarchical pattern in data and assemble more complex patterns using smaller and simpler patterns. Therefore, on the scale of connectedness and complexity, CNNs are at the lower extreme.

(3) MobileNet

MobileNet is an efficient DCNN used for vision applications and image processing. It is built on a streamlined architecture of light weight with depth-wise separable convolutions. Global hyperparameters are introduced to balance the latency and accuracy of the model, which can allow the model to select the best size for the specific application. It has shown good performance across various applications and scenarios. The architecture of MobileNet is presented in Table 3. It has a total of 28 layers. Each layer is followed by a BatchNorm and a Rectified Linear Unit (ReLU). Average pooling was implemented to reduce the spatial resolution before the final fully connected layer with nonlinearity, and it was fed into a softmax layer for classification outputs. Depth-wise convolutions are conducted using down-sampling. Compared with the other convolutional networks, MobileNet has fewer parameters and presents better computing efficiency. The detailed structure of MobileNet is demonstrated in Fig 3, and the specific network construction is presented in Table 4.

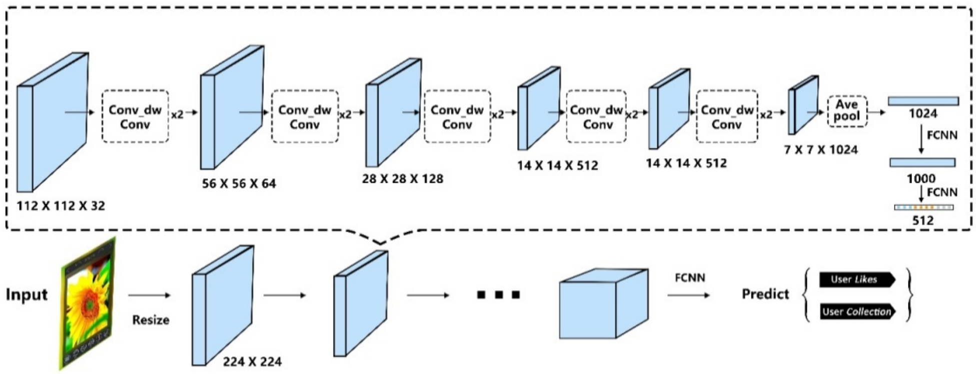

Figure 3. MobileNet network architecture.

Table 4. The Architecture of MobileNet

Layer name	Layers	Output size
Conv1	Conv,3x3,32 × 3	[112,112,32]
Conv2_x	$\begin{bmatrix} Conv\_dw & 3\times3 & 32 \\ Conv & 1\times1 & 32 \\ Conv\_dw & 3\times3 & 64 \\ Conv & 1\times1 & 64 \end{bmatrix}$	[112,112,32] [112,112,64] [56,56,64]
Conv3_x	$\begin{bmatrix} Conv\_dw & 3\times3 & 128 \\ Conv & 1\times1 & 128 \\ Conv\_dw & 3\times3 & 128 \\ Conv & 1\times1 & 128 \end{bmatrix}$	[56,56,128] [28,28,128]
Conv3_x	$\begin{bmatrix} Conv\_dw & 3\times3 & 256 \\ Conv & 1\times1 & 256 \\ Conv\_dw & 3\times3 & 256 \\ Conv & 1\times1 & 256 \end{bmatrix}$	[28,28,256] [14,14,256] [14,14,512]
Conv4_x	$\begin{bmatrix} Conv\_dw & 3\times3 & 512 \\ Conv & 1\times1 & 512 \end{bmatrix}\times5$	[14,14,512]

Conv5_x	$\begin{bmatrix} Conv\_dw & 3\times3 & 512 \\ Conv & 1\times1 & 1024 \\ Conv\_dw & 3\times3 & 1024 \\ Conv & 1\times1 & 1024 \end{bmatrix}$	[7,7,512] [7,7,1024]
Dense1_x	Ave_pool fc × 1	[1024] [1000]
Output	fc, Sigmoid	[512]

(4) InceptionNet-V3

DCNNs are the mainstream method of large image recognition computing tasks and have achieved optimal performance. The principles of the network are characterized into four aspects. It can first avoid representational shortcomings of feature loss caused by compression during pooling. It can more easily process higher dimensional representations locally. It can promote spatial aggregation with less loss but faster learning. Finally, the depth and width of the network can be better balanced to maximize network performance. The structure of an InceptionNet was reduced from the traditional 7×7 convolution to a three 3×3 convolution. Inception part was reduced to a 17×17 grid having 768 filters. This is followed by five inception modules as portrayed in the structure. Then, it was reduced to an 8×8×1280 grid having an output filter in a size of 2,048. The specific map of network structure is described in Fig 4, and the framework structure is presented in Table 5.

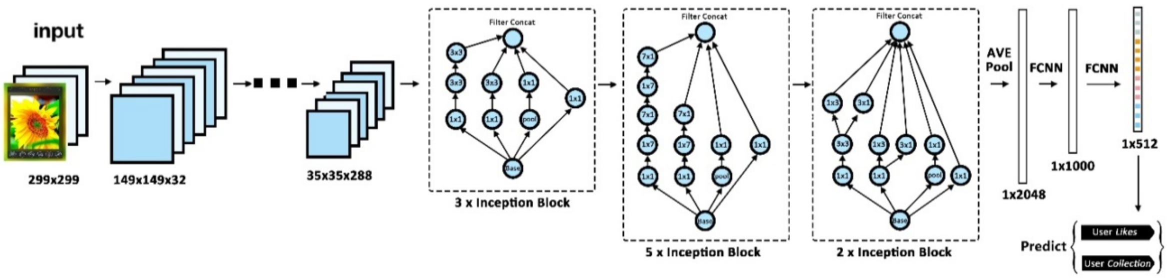

Figure 4. InceptionNet-V3 network architecture.

Table 5. The Architecture of InceptionNet-V3.

Layer name	Layers	Output size
Conv1	Conv,3x3,32 × 2 Conv,3x2,32 × 1	[149,149,32] [147,147,32]
Conv2_x	ConvPadded,3x3,64 × 1 Max pool	[147,147,64] [73,73,64]
Conv3_x	Conv,3x3,80 × 1 Conv,3x3,192 × 2 Conv,3x3,288 × 1	[71,71,80] [35,35,192] [35,35,288]
3 × Inception	$\begin{bmatrix} Conv & 1\times1 & 64 \\ Conv & 1\times1 & 48 \\ Conv & 5\times5 & 64 \\ Conv & 1\times1 & 64 \\ Conv & 3\times3 & 96 \\ Conv & 3\times3 & 96 \\ Conv & 1\times1 & 32 \end{bmatrix}\times3$	[17,17,768]

5 × Inception	$\begin{bmatrix} Conv & 1\times1 & 192 \\ Conv & 1\times1 & 128 \\ Conv & 1\times7 & 128 \\ Conv & 7\times1 & 192 \\ Conv & 1\times1 & 128 \\ Conv & 7\times1 & 128 \\ Conv & 1\times7 & 128 \\ Conv & 7\times1 & 128 \\ Conv & 1\times7 & 192 \\ Conv & 1\times1 & 192 \end{bmatrix} \times 5$	[8,8,1280]
2× Inception	$\begin{bmatrix} Conv & 1\times1 & 320 \\ Conv & 1\times1 & 384 \\ Conv & 1\times3 & 384 \\ Conv & 3\times1 & 384 \\ Conv & 1\times1 & 448 \\ Conv & 3\times3 & 384 \\ Conv & 1\times3 & 384 \\ Conv & 3\times1 & 384 \\ Conv & 1\times1 & 192 \end{bmatrix} \times 2$	[8,8,2048]
Dense1_x	Ave_pool	[2048]
	fc × 1	[1000]
Output	fc, Sigmoid	[512]

(5) VGG-19

For large-scale image recognition, VGG-19 has achieved significant accuracy improvements in the depth of 16–19 weight layers [67], and it has been effective in various image computational aesthetic research [57][68][69]. It uses 3 × 3 convolution filters. During network training, the image input to the convolutional network is set to 224 × 224 pixels of convolutional layers followed by three fully connected layers. The final layer is a softmax layer, and a 25,088-dimensional feature vector was obtained by the fully connected network for classification in the network output. The VGG-19 network architecture is described in Fig 5. ReLU is set as the hidden layer for the activation function. Convolutional network configurations are shown in Table 6.

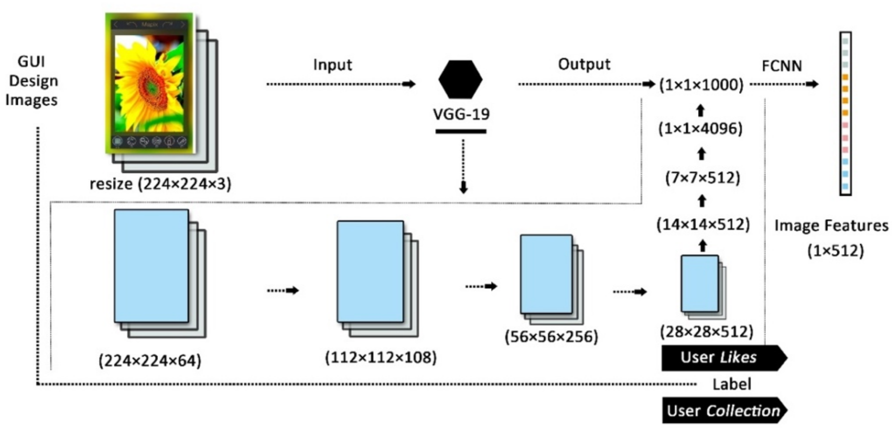

Figure 5. VGG-19 network architecture

Table 6. Convolutional network configuration of VGG-19.

Layer type	Layers	Output size

Conv1	Conv,3x3,64 × 2 Max pool	[224,224,64] [112,112,64]
Conv2_x	Conv,3x3,128 × 2 Max pool	[112,112,128] [56,56,128]
Conv3_x	Conv,3x3,256 × 4 Max pool	[56,56,256] [28,28,256]
Conv4_x	Conv,3x3,512 × 4 Max pool	[28,28,512] [14,14,512]
Conv5_x	Conv,3x3,512 × 4 Max pool	[14,14,512] [7,7,512]
Dense3_x	Flatten fc × 3	[25088] [1000]
Output	fc, Sigmoid	[512]

(6) ResNet-50

ResNet-50 [70][71] is proposed to extract multimodal features of GUI image features for its good performance in image analysis and aesthetic computing [56]. The input image is rescaled to 224 × 224 pixels to feed the DNN. A ResNet architecture with 50 layers was applied as the DCNN to extract the features. A fully connected network with four layers was applied to obtain the output vector. The dimension of image features was reduced by the fully connected network for the classification model. For feature processing, optimized methods were applied, including standard feature normalization shifting, rotation, zooming, and nearest-fill. We used 2,048 dimensional features and the normalized correlation value to train the fully connected network for 300 epochs. ReLU was set to provide the activation functions of the fully connected networks. In the output layer, softmax was applied as the activation function. During training, binary cross entropy was used for the loss function, and Adam was used as the optimizer. An overview of the feature extraction process is given in Table 7, and the architecture of ResNet-50 is presented in Fig 6.

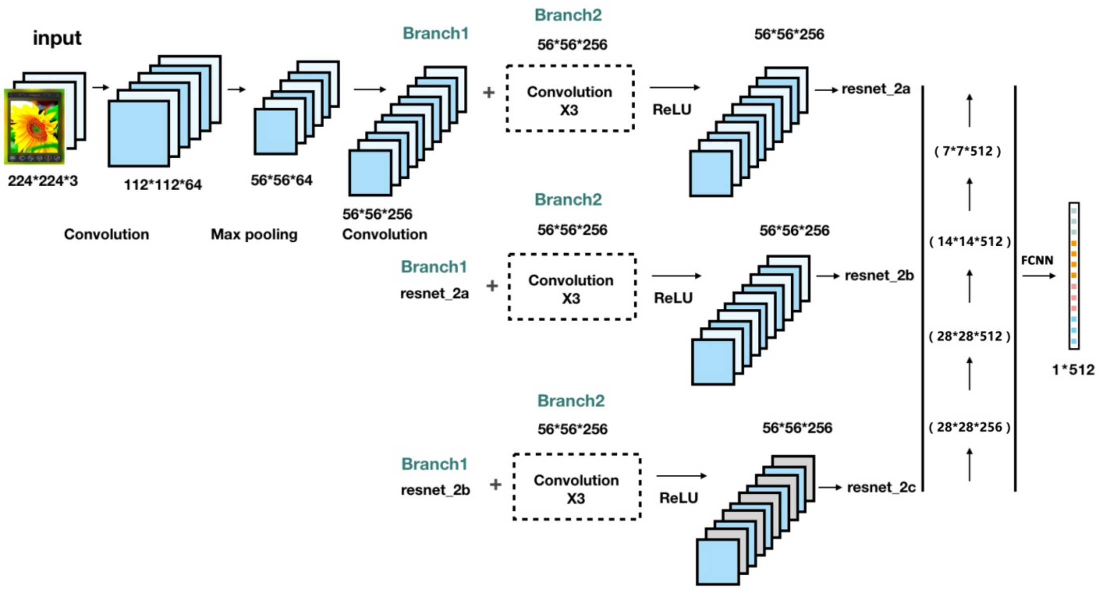

Figure 6. ResNet-50 network architecture.

Table 7. ResNet-50 architecture and feature extraction process.

Layer name	Layers	Output size
Conv1	Conv,7x7,64.stride 2 Max pool,3x3,stride 2	[112,112,64] [56,56,64]
Conv2_x	$\begin{bmatrix} Conv & 1\times 1 & 64 \\ Conv & 3\times 3 & 64 \\ Conv & 1\times 1 & 256 \end{bmatrix} \times 3$	[56,56,256]
Conv3_x	$\begin{bmatrix} Conv & 1\times 1 & 128 \\ Conv & 3\times 3 & 128 \\ Conv & 1\times 1 & 512 \end{bmatrix} \times 4$	[28,28,512]
Conv4_x	$\begin{bmatrix} Conv & 1\times 1 & 256 \\ Conv & 3\times 3 & 256 \\ Conv & 1\times 1 & 1024 \end{bmatrix} \times 6$	[14,14,1024]
Conv5_x	$\begin{bmatrix} Conv & 1\times 1 & 512 \\ Conv & 3\times 3 & 512 \\ Conv & 1\times 1 & 2048 \end{bmatrix} \times 3$	[7,7,2048]
Dense3_x	Average Pool, fc× 1	[2048] [1000]
Output	fc, sigmoid	[512]

(7) SE-VGG19

Squeeze-and-Excitation-VGG19 is an improved VGG19 network with a Squeezed-and-Excitation block (SE block) [72] at the last layer of VGG19. Here we constructed SE-VGG19 to pursue an optimal result in user aesthetic preference modeling. SE-VGG19 has a Squeezed-and-Excitation block added as a unit at the end of VGG19 network structure, in order to learn the feature weight according to the loss function through the network. Consequently, SE block can increase the weight of effective feature map, and reduce the weight of the unvalued features. SE block can achieve higher model accuracy with great computational efficiency. Specifically, we applied global average pooling as squeeze operation. After that three fully connected layers are utilized to explore the correlation of channels and output weights. We used sigmoid to normalize weights in scale of 0~1. Then, a scale operation was applied to give the normalized weights to features of each channel. Three-layer fully connected network was applied on the output features. In order to prevent over fitting problem, average pooling, dropout and softmax are adopted in the operation. End-to-end training is adopted for SE-VGG19, making feature extractor and predictor to be trained in a unified model. The specific network structure of SE-VGG19 is presented in Table 8 and Fig. 7.

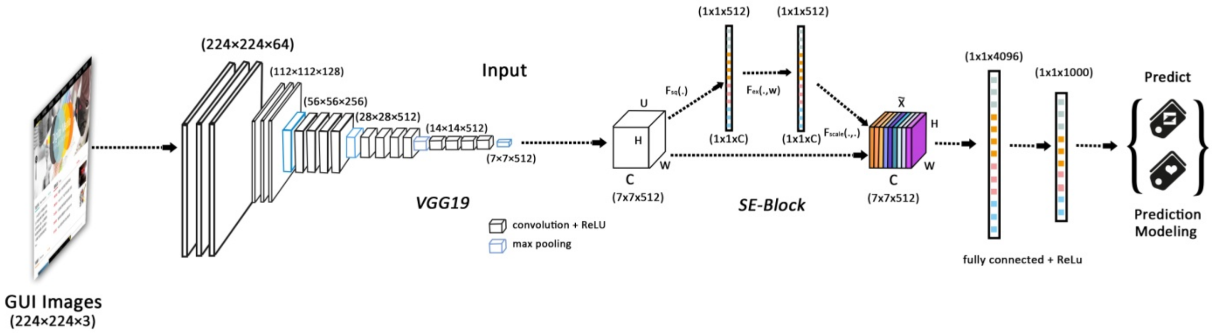

Figure 7. SE-VGG19 network architecture

Table 8. Convolutional network configuration of SE-VGG19.

Layer type	Layers	Output size
Conv1	Conv,3x3,64 × 2 Max pool	[224,224,64] [112,112,64]
Conv2_x	Conv,3x3,128 × 2 Max pool	[112,112,128] [56,56,128]
Conv3_x	Conv,3x3,256 × 4 Max pool	[56,56,256] [28,28,256]
Conv4_x	Conv,3x3,512 × 4 Max pool	[28,28,512] [14,14,512]
Conv5_x	Conv,3x3,512 × 4 Max pool	[14,14,512] [7,7,512]
SE-block	Squeeze Excitation Scale	[1,1,512] [1,1,512] [7,7,512]
Dense3_x	Flatten fc × 3	[25088] [1000]
Output	fc	[1]

4. Experiments

GUI images are suitable for homogeneous comparisons of aesthetics. We collected GUI images from HUABAN.com from 2000 to late 2019 to build our GUI image database of 38,423 images rated by website visitors. Fig 8 demonstrates the image samples for the database. The GUI works were annotated using data on visitor *likes* and *collection*. This labeling information reflects the aesthetic preferences of the users, which can then be used to label the aesthetic model [46].

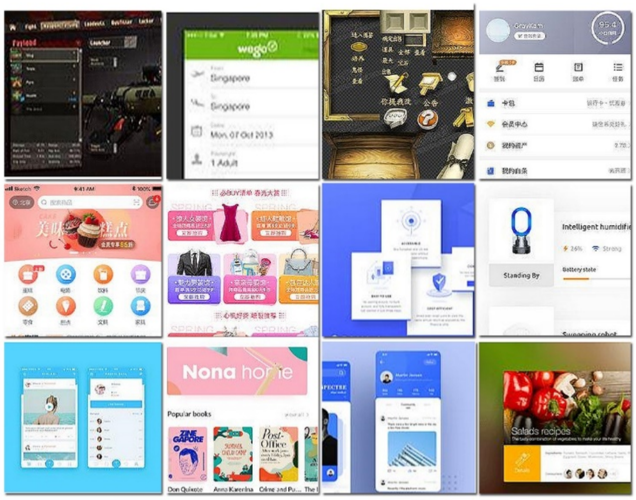

Figure 8. Image samples for the GUI aesthetic dataset.

To evaluate aesthetic prediction performance, we adopted five model indices: MSE, RMSE, Mean Absolute Error (MAE), R^2 and R- adjusted. The method that achieved the lowest MSE is regarded as the optimal model, indicating a high precision result. Moreover, the loss value is presented to demonstrate the modeling details. The general pipeline of GUI aesthetic modeling procedure is presented in Fig 9.

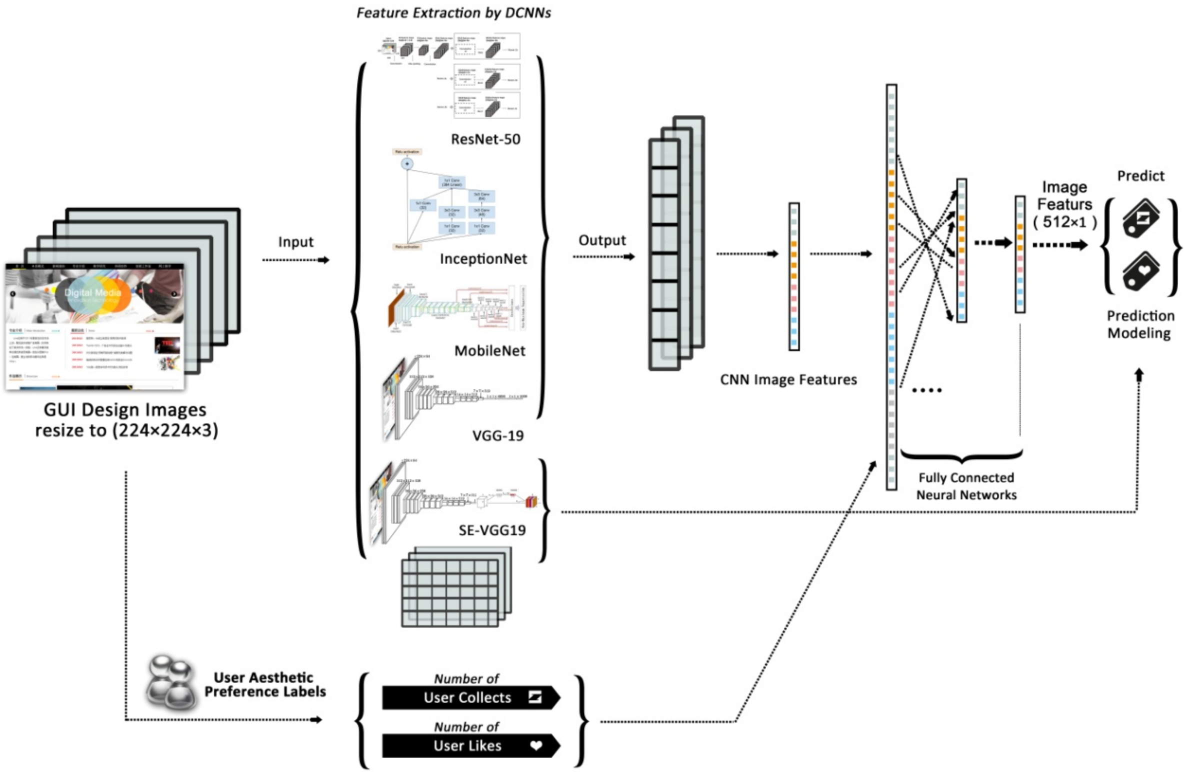

Figure 9. Experimental procedure for GUI aesthetic preference modeling.

4.1 GUI dataset construction

Huaban.com is a popular social media website for image sharing in China. It is introduced in Huaban.com that more than 1 million professional designers and life style designers use Huanban browser collection tool to collect more than two million inspiring images from the internet every day. Since Huaban.com was online, more than two billion collections have been collected by tens of millions of Huaban users.

According to the user tracking data provided by iresearch.cn that around 30% of Huaban users are male, and 70% are females. This might be due to the fact that women usually have an enthusiasm for collecting and sharing images, and they are like to pay more attention to the quality of life and inner feelings. Users aged from 25 to 30 constitute for around 30%, which is the highest proportion.

Most existing aesthetic datasets were built using images of land, nature, and human figures. Their image content is diverse and their content frameworks differ. Unlike GUI images, user-interface design follows certain rules of content arrangement. For instance, a website usually has a header, a banner, a text part, and a footer, presented in a relatively standard design structure. To explore an aesthetic model for GUI, we used GUI works from Huaban.com to construct a dataset suitable for GUI aesthetic quality evaluation modeling, see Fig 10. *Likes* and *collection* data of each work were obtained as ground-truth annotation, revealing the aesthetic quality of each image.

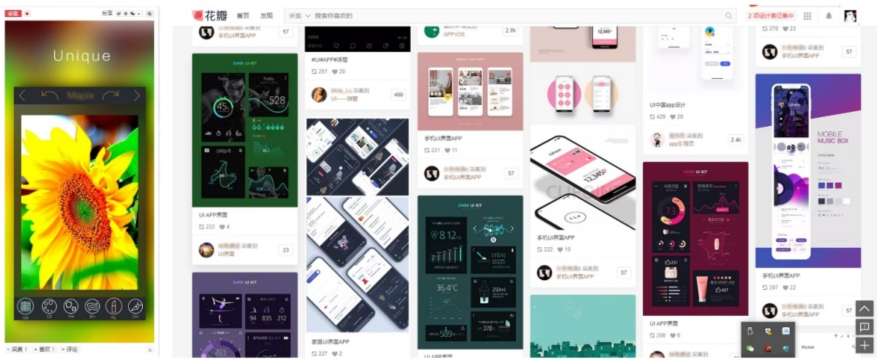

Figure 10. GUI design works exhibited at HUABAN.com. Users can give the GUI works likes (thumbs and collection as a compliment).

4.2 GUI aesthetic modeling implementation

In extant aesthetic computational studies, machine-learning algorithms have been widely applied for aesthetic score prediction and high/low aesthetic quality classification [52-62]. Additionally, multiple statistical methods have been used for aesthetic factor analysis and weight computing to discover the crucial factors influencing aesthetic perception [60]. Regarding the difficulty of manual aesthetic rating, we set user labeling information of *likes* and *collection* for GUI works exhibit in HUABAN.com as a natural annotation reflecting user aesthetic preferences. We built the dataset based on the downloaded GUI images. The images of GUI works have similar patterns of aesthetical and functional design in content frameworks and element arrangements, which is an appropriate image data resource for aesthetic computing. We conducted an experiment mainly in three sessions: (1) label standardization and sample visualization; (2) CNN image features standardization; and (3) aesthetic regression modeling exploration.

(1) Label standardization and visualization

First, the numbers of user *likes* and *collections* were used the labels of user aesthetic preference. A Z_{log} method was used for label standardization. In this experiment, from the distribution map of data samples, we can see that a few samples have high degree of dispersion. Thus we remove the points, zeros and empty data with high degree of individual dispersion and then conducted label normalization on the rest of the data. Collection labels and likes labels were normalized by Z_{log} method into a scale of 0 to 1. A Z_{log} is a standardization method that value Z was transformed by *log* function, that

$$Z_{log}=log(X_{(i)})/log(X_{max}).$$

Finally, the value of Z_{log} is adopted as the standardized label for aesthetic modeling in the next step. The label standardization process is shown in Figs 11, 12.

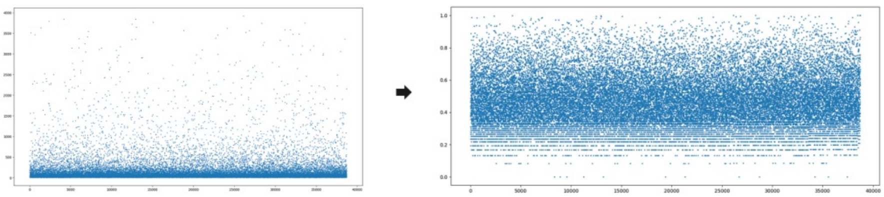

Figure 11. Label standardization process for user *collection.*

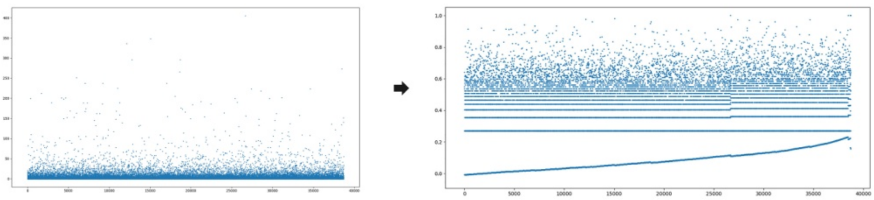

Figure 12. Label standardization process for user *likes.*

(2) CNN image features standardization
We standardized the image features extracted by CNN by subtracting the mean value and zooming to the unit variable. The centralization and zooming are carried out independently on each feature during the sample correlation calculation in the test set, and the mean value and standard deviation are saved for the standardization calculation.
(3) Aesthetic regression modeling exploration
We then used ResNet-50, VGG, SE-VGG19, InceptionNet-V3, and MobileNet to extract the CNN image features and explored the aesthetic regression modeling based on RandomForest algorithm to predict the labels of *likes* and *collection*. The network framework settings are denoted in Tables 4-8 and Figs 3-7. Specifically, we trained the model with 34,844 images and tested it with another 3,579 images. In the construction of RandomForest model, the number of estimators is set as 1000, and the parameter of random_state 42.

5. Results and Discussions

Researchers have endeavored to seek a computational means to obtain aesthetic judgment. However, user preferences and subjective perceptions increase the difficulty of aesthetic modeling. In this study, we performed aesthetic modeling as a regression problem. An overall regression accuracy for user *likes* and *collection*, represented by indices of MSE, RMSE, MAE, R^2 and R-adjusted were obtained, showing that our model outperformed is satisfied. Comparison experiments were conducted to identify the best model. The detailed modeling results for user *collection* were: InceptionNet-V3 (MSE = 0.0276, RMSE = 0.1662, MAE = 0.13, R-adjusted=0.005); MobileNet (MSE = 0.0252, RMSE = 0.1588, MAE = 0.13, R^2 =0.0229, R-adjusted=0.0005); ResNet-50 (MSE = 0.0251, RMSE = 0.1584, MAE = 0.13, R^2 =0.0268, R-adjusted=0.0007); VGG-19 (MSE = 0.0247, RMSE = 0.1571, MAE = 0.13, R^2 =0.0431, R-adjusted=0.0018); SE-VGG19 (MSE = 0.0222, RMSE = 0.1489, MAE = 0.13, R2 =0.0437, R-adjusted=0.0022). The prediction results for user like*s* were: InceptionNet-V3 (MSE = 0.0761, RMSE = 0.2760, MAE = 0.22, R-adjusted=0.008); MobileNet(MSE = 0.0694, RMSE = 0.2635, MAE = 0.21, R^2 =0.0077, R-adjusted=0.00006); ResNet-50 (MSE = 0.068, RMSE = 0.2608, MAE = 0.21, R^2 =0.0278, R-adjusted=0.0007); VGG-19 (MSE = 0.0674, RMSE = 0.2579, MAE =

0.21, R^2 =0.0361, R-adjusted=0.0013); SE-VGG19 (MSE = 0.0644, RMSE = 0.2538, MAE = 0.21, R2 =0.0368, R-adjusted=0.0016). The specific results of the comparison of the models are presented in Tables 9 and 10. The best regression accuracy of user *collection* and *likes* was achieved by SE-VGG19. Moreover, Fig 13 presents the loss during SE-VGG19 networks training process.

Table 9. User *collection* prediction for GUI aesthetic preference modeling

Algor.	Results					
	MSE	RMSE	MAE	R^2	R-adjusted	*Parameters setting*
InceptionNet-V3	0.0276	0.1662	0.13	-	0.005	Input size：299 × 299
MobileNet	0.0252	0.1588	0.13	0.0229	0.0005	Input size：224 × 224
ResNet-50	0.0251	0.1584	0.13	0.0268	0.0007	Input size：224 × 224
VGG-19	0.0247	0.1571	0.13	0.0431	0.0018	Input size：224 × 224
SE-VGG19	**0.0222**	**0.1489**	**0.13**	**0.0437**	**0.0022**	Input size：224 × 224

Table 10. User *likes* prediction for GUI aesthetic preference modeling

Algor.	Results					
	MSE	RMSE	MAE	R^2	R-adjusted	*Parameters setting*
InceptionNet-V3	0.0761	0.2760	0.22	-	0.008	Input size：: 299 × 299
MobileNet	0.0694	0.2635	0.21	0.0077	0.00006	Input size：224 × 224
ResNet-50	0.068	0.2608	0.21	0.0278	0.0007	Input size：224 × 224
VGG-19	0.0674	0.2597	0.21	0.0361	0.0013	Input size：224 × 224
SE-VGG19	**0.0644**	**0.2538**	**0.21**	**0.0368**	**0.0016**	Input size：224 × 224

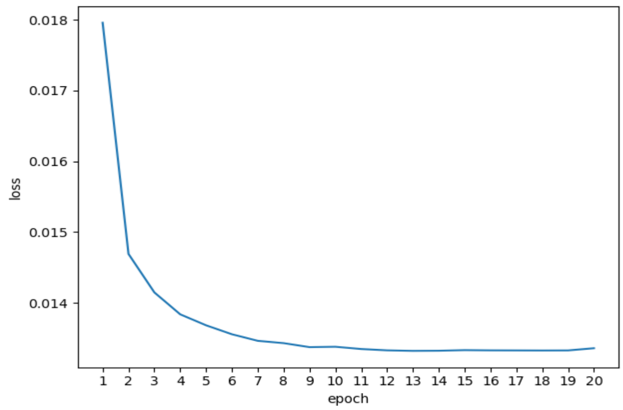

Figure 13. Loss during SE-VGG19 Networks training process

Experimentally, our proposed method exhibited good performance on a real-world GUI design dataset, especially when predicting of the numbers of user *collection* and likes. The result also showed that visual features extracted by DCNNs could be effective in aesthetic evaluation. Experiments indicated that our method is suitable to determine the aesthetic quality of GUI images.

6. Conclusion and Future Work

GUI aesthetic evaluation is a fuzzy systematic study. In this study, human perceptions of GUI layout, color, and texture were transformed as computational features for quantizing user aesthetic perceptions. We proposed a comprehensive aesthetic quality score prediction model for GUI design based on CNN features. User *likes* and *collection* of GUI design works were utilized to represent user aesthetic preference levels. This approach provided an effective computational method for intelligent aesthetic evaluation of GUI images. With an optimal DCNN, GUI aesthetics can be automatically analyzed to judge visual appearances. According to ranking prediction results, we provided an objective data-based method of GUI evaluation and design recommendation.
Specifically, we collected 38,423 GUI design works from professional GUI design communities to build the large database. The *likes* and *collection* data were transformed to ground-truth annotations reflecting GUI aesthetic quality. In view of the superior DCNN performance, VGG-19, SE-VGG19, ResNet-50, InceptionNet-V3, MobileNet networks were compared. The optimal result was achieved by SE-VGG19 with an MSE of 0.0222 for user *collection* prediction and an MSE of 0.0644 for user *likes* prediction. Consequently, our empirical study indicates that the proposed method is effective for GUI aesthetic evaluation. Because the collected data sample was relatively extensive, the database and the modeling results can be regarded as a significant contribution to the overall knowledge base and may lead to theoretical aesthetics computations in further studies.
The main drawback of the present study is that the web interfaces were usually recognized as a single layout image for analysis. The functionality and emphasis of different parts in such layouts are rarely discussed, and the weights of different parts of the layout structure were not studied. Focusing on these aspects may influence future aesthetic prediction models. A comprehensive study of both global and local features may lead to a more scientific aesthetic cognition model.
In the future, we will delve into improving the method based on an expanded dataset and testing the research scheme across different datasets. Besides, eye-tracking technique can be applied to collect user preference data of GUIs in the further study [73][74]. The obtained computational model should serve as a theoretical basis for various scenarios that can be expanded in multiple directions. For

example, applications for GUI aesthetic evaluation can be developed for independent designers and product developers. Moreover, aesthetical retrieval methods can be developed based on this model to assist design innovation, helping meet agile development requirements.

7. Acknowledgements

This study is partly supported by the Natural Science Foundation of Zhejiang Province of China (LY19F020047). Author contributions are stated as follows: Conceptualization: Baixi Xing, Huahao Si; Data curation: Baixi Xing, Huahao Si, Junbin Chen,Minchao Ye.;Methodology: Baixi Xing, Huahao Si, Junbin Chen,Minchao Ye; Writing – original draft: Baixi Xing; Writing – review & editing: Baixi Xing, Minchao Ye, Lei Shi. The authors declare that there is no conflict of interest related to this work.

Reference

1. Norman DA. Emotional Design: Why We Love (Or Hate) Everyday Things. Basic Books, U.S..2004.

2. Chien C, Kerh R, Lin K, Yu AP, 2016. Data-driven innovation to capture user-experience product design: An empirical study for notebook visual aesthetics design, Computers & Industrial Engineering. 2016; 162–173.

3. Hou G, Lu G. The influence of design proposal viewing strategy: design aesthetics and professional background, International Journal of Technology Design Education. 2019; 29: 543–564.

4. Maity R, Bhattacharya S. Is My Interface Beautiful?—A Computational Model-Based Approach, IEEE TRANSACTIONS ON COMPUTATIONAL SOCIAL SYSTEMS. 2019; 6(1): 149-162.

5. Huang S. The rating consistency of aesthetic preferences for icon-background color combinations. Applied Ergonomics. 2011; 43(1): 141-150.

6. Mccracken J, Feagin SL, Maynard P. Aesthetics. The Oxford University Press, New York, USA . 1997.

7. Cheng S, Dey A. I see, you design: user interface intelligent design system with eye tracking and interactive genetic algorithm, CCF Transactions on Pervasive Computing and Interaction. 2019; 1(3): 224-236.

8. Kruft HW. A history of architectural theory: From Vitruvius to the present. Zwemmer and Princeton architectural press, New York. 1994.

9. Liao W, Chen P. Analysis of Visual Elements in Logo Design, Proceedings of International Symposium on Smart Graphics. 2014; pp. 73-85.

10. Zhang C, Lei K, Jia J. AI Painting: An Aesthetic Painting Generation System, Proceedings of ACM Multimedia. 2018; 1231-1234.

11. Erdem AN, Halici U. Applying Computational Aesthetics to a Video Game Application Using Machine Learning, IEEE Computer Graphics and Applications. 2016; 36(4): 23-33.

12. Cui P, Liu S, Zhu W, Luan H, Chua T, Yang S. Social-Sensed Image Search, ACM Transactions on Information Systems. 2014; 32 (2): No.8.

13. Deng X, Cui C, Fang H, Nie X, Yin Y. Personalized image aesthetics assessment, Proceedings of ACM Conference on Information and Knowledge Management. 2017; pp. 2043-2046.

14. Chen R, Hua L, Xie Y, Lin T, Tang N. A Fuzzy-Rule-Based Approach for Webpage Aesthetics Modeling. Proceedings of Nicograph International. 2016, pp. 142-143.

15. Miniukovich A, Angeli AD. Computation of Interface Aesthetics, Proceedings of the 33rd Annual ACM Conference. 2015.

16. Maity R, Uttav A, Verma G, Bhattacharya S. A Non-Linear Regression Model to Predict Aesthetic Ratings of On-Screen Images, The Annual Meeting of the Australian Special Interest Group for Computer Human Interaction. 2015.

17. Maity R, Bhattacharya S. A Model to Compute Webpage Aesthetics Quality Based on Wireframe Geometry, IFIP Conference on Human-Computer Interaction. 2017; pp. 85-94.

18. Tuch AN, Presslaber EE, Stöcklina M, Opwisa K, Bargas-Avilab JA. The role of visual complexity and prototypicality regarding first impression of websites: Working towards understanding aesthetic judgments. International Journal of Human-Computer Studies. 2012; 70(11): 794-811.

19. Ngo DCL, Teo LS, Byrne JG. Modeling interface aesthetics, Information Sciences. 2003; 152: 25-46.

20. Shamoi P, Inoue A, Kawanaka H. Modeling aesthetic preferences: Color coordination and fuzzy sets, Fuzzy Sets and Systems. 2019; 1-18.

21. Maity R, Madrosiya A, Bhattacharya S. A Computational Model to Predict Aesthetic Quality of Text Elements of GUI, Procedia Computer Science. 2016; 84: 152-159.

22. Jylhä H, Hamari J. An icon that everyone wants to click on: The relationship between consumer perceptions and app icon successfulness, International Journal of Human-Computer Studies. 2019; doi: 10.1016/j.ijhcs.2019.04.004.

23. Robins D, Holmes J. Aesthetics and credibility in web site design, Pergamon Press. 2008.

24. Lavie T, Tractinsky N, 2004. Assessing dimensions of perceived visual aesthetics of web sites, International Journal of Human-Computer Studies. 2004; 60(3): 269-298.

25. Casey TW, Poropat A. Beauty is more than screen deep: Improving the web survey respondent experience through socially-present and aesthetically-pleasing user interfaces, Computers in Human Behavior. 2014; 30: 153-163.

26. Seckler M, Opwis K, Tuch AN. Linking objective design factors with subjective aesthetics: An experimental study on how structure and color of websites affect the facets of users' visual aesthetic perception, Computers in Human Behavior. 2015; 49: 375-389.

27. Tractinsky N, Cokhavi A, Kirschenbaum M, Sharfi T, 2006. Evaluating the consistency of immediate aesthetic perceptions of web pages, International Journal of Human-Computer Studies. 2006; 64(11): 1071-1083.

28. Bauerly M, Liu Y. Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics, International Journal of Human-Computer Studies. 2006; 64(8): 670-682.

29. Park S, Choi D, Kim J. Critical factors for the aesthetic fidelity of web pages: empirical studies with professional web designers and users, Interacting with Computers. 2004; 16(2): 351-376.

30. Zheng XS, Chakraborty I, Lin JJ, Rauschenberger . Correlating low-level image statistics with users' rapid aesthetic and affective judgments of web pages, Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, Boston, MA, USA, April 4-9. 2009.

31. Schaik PV, Ling J. The role of context in perceptions of the aesthetics of web pages over time, International Journal of Human-Computer Studies. 2009; 67(1): 79-89.

32. Porat T, Tractinsky N. It's a Pleasure Buying Here: The Effects of Web-Store Design on Consumers' Emotions and Attitudes, Human-computer Interaction. 2012; 27(3): 235-276.

33. Lin Y, Yeh C, Wei C. How will the use of graphics affect visual aesthetics? A user-centered approach for web page design, International Journal of Human-Computer Studies. 2013; 71(3): 217-227.

34. Liu W, Guo F, Ye G, Liang X. How homepage aesthetic design influences users' satisfaction: Evidence from China, Displays. 2016; No. S0141938216300178.

35. Moshagen M, Thielsch MT. Facets of visual aesthetics, International Journal of Human-Computer Studies. 2010; 68(10): 689-709.

36. Tractinsky N, Katz AS, Ikar D. What is beautiful is usable, Interacting with Computers. 2000; 13(2): 127-145.

37. Garrido-Possauner LM, Maya J. The Aesthetic Pleasure in Design Scale for Spanish Speaking Countries: A Method for the Cross-Cultural Implementation and Adaptation of Psychometric Scales, Cross-Cultural Design. Methods, Tools and User Experience, Proceedings of The 11th International Conference, CCD 2019, Orlando, FL, USA, July 26–31. 2019.

38. Ciesielski V, Barile P, Trist K. Finding Image Features Associated with High Aesthetic Value by Machine Learning, Evolutionary and Biologically Inspired Music, Sound, Art and Design.2013.

39. Guo F, Li M, Hu M, Li F, Lin B. Distinguishing and quantifying the visual aesthetics of a product: An integrated approach of eye-tracking and EEG, International Journal of Industrial Ergonomics. 2019; 47–56.

40. Parul N, Jeff B, K. Laurin KK, Catharine H, Jennifer I. Effect of smartphone aesthetic design on users' emotional reaction: an empirical study, The TQM Journal. 2008; 348–355.

41. Schindler I, Hosoya G, Menninghaus W, Ursula B, Valentin W, Michael E, Klaus RS. Measuring aesthetic emotions: A review of the literature and a new assessment tool, PloS One. 2007; 12(6): e0178899.

42. Temel D, AlRegib G. A COMPARATIVE STUDY OF COMPUTATIONAL AESTHETICS, Proceedings of the International Conference on Image Processing. 2014; pp. 590-595.

43. Wu O, Zuo H, Hu W, Li B. Multimodal Web Aesthetics Assessment Based on Structural SVM and Multitask Fusion Learning. IEEE TRANSACTIONS ON MULTIMEDIA. 2016; 18(6): 1062-1077.

44. Lu X, Lin Z, Jin H, Yang J, Wang JZ. Rating Image Aesthetics Using Deep Learning, IEEE TRANSACTIONS ON MULTIMEDIA. 2015; 17(11): 2021-2035.

45. Li C, Chen T. Aesthetic Visual Quality Assessment of Paintings, IEEE Journal of Selected Topics in Signal Processing. 2009; 3(2): 236-252.

46. Jiang W, Loui AC, Cerosaletti CD. Automatic aesthetic value assessment in photographic images, Proceedings of 2010 IEEE International Conference on Multimedia and Expo. 2010.

47. Anush K, Moorthy PO, Nuria O,. Towards Computational Models of the Visual Aesthetic Appeal of Consumer Videos, Proceedings of Computer Vision - ECCV 2010 - 11th European Conference on Computer Vision, Heraklion, Crete, Greece, September 5-11, 2010.

48. Lu P, Peng X, Yu J, Peng X. Gated CNN for visual quality assessment based on color perception, Signal Processing: Image Communication. 2018; 72: 105-112, doi: 10.1016/j.image.2018.12.007.

49. Tan Y, Tang P, Zhou Y. Photograph aesthetical evaluation and classification with deep convolutional neural networks, Neurocomputing. 2016; 228: 165-175.

50. Wang W, Cai D, Wang L, Huang Q, Xu X, Li X. Synthesized computational aesthetic evaluation of photos, Neurocomputing. 2016; 172: 244-252.

51. Wang W, Zhao M, Wang L, Huang J, Cai C, Xu X. A multi-scene deep learning model for image aesthetic evaluation, Signal Processing: Image Communication. 2016; 47: 511-518.

52. Chao Z, Ce Z, Xun X, Liu Y, Xiao J, Tillo T. Visual aesthetic understanding: Sample-specific aesthetic classification and deep activation map visualization, Signal Processing: Image Communication. 2018; 67:12-21.

53. Ma S, Liu J, Chen CW. A-Lamp: Adaptive Layout-Aware Multi-patch Deep Convolutional Neural Network for Photo Aesthetic Assessment, IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017; 722-731.

54. Yu W, Zhang H, He X, Chen X, Xiong L, Qin Z. Aesthetic-based Clothing Recommendation, Proceedings of the 2018 World Wide Web Conference. 2018.

55. Zhang L, Gao Y, Zimmermann R, Tian Q, Li X, Fusion of Multichannel Local and Global Structural Cues for Photo Aesthetics Evaluation. IEEE TRANSACTIONS ON IMAGE PROCESSING. 2014; 23(3): 1419-1430.

56. Murray N, Marchesotti L, Perronnin F. AVA: A Large-Scale Database for Aesthetic Visual Analysis, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2012; 2408-2416.

57. Kucer M, Loui AC, Messinger DW. Leveraging Expert Feature Knowledge for Predicting Image Aesthetics, IEEE TRANSACTIONS ON IMAGE PROCESSING. 2018; 27(10): 5100-5113.

58. Lee D, Liu H. Like and recommendation in social media. Proceedings of the World Wide Web Conference. 2015; 1539-1540.

59. Lee D. Likeology: modeling, predicting, and aggregating likes in social media, Proceedings of the 8th ACM Conference. 2016; 13.

60. Guntuku S C, Roy S, Weisi L. Evaluating visual and textual features for predicting user 'likes', Proceedings of IEEE International Conference on Multimedia and Expo (ICME), Turin, Italy. 2015; 1-6.

61. Khosla A, Das Sarma A, Hamid R. What makes an image popular? Proceedings of the World Wide Web Conference. 2014; 867–876.

62. Ohsawa S, Matsuo Y. Like prediction: modeling like counts by bridging facebook pages with linked data, Proceedings of the World Wide Web Conference. 2013; 541–548.

63. Jang J Y, Han K, Shih P C, Lee D. Generation like: Comparative characteristics in instagram, Proceedings of ACM Conf. on Human factors in computing systems. 2015; 4039-4042.

64. Jang J, Han K, Lee D. No Reciprocity in "Liking" Photos: Analyzing Like Activities in Instagram, Proceedings of the ACM Conf. on Hypertext and Social Media. 2015; 273-282.

65. Youyou W, Kosinski M, Stillwell D. Computer-based personality judgments are more accurate than those made by humans, Proceedings of the National Academy of Sciences (PNAS). 2015; 1036-1040.

66. He R, McAuley J. VBPR: Visual Bayesian Personalized Ranking from Implicit Feedback. 2016; arXiv:1510.01784v1 [cs.IR].

67. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition, Proceedings of International Conference on Learning Representations. San Diego, USA. 2015.

68. Fu X, Yan J, Fan C. IMAGE AESTHETICS ASSESSMENT USING COMPOSITE FEATURES FROM OFF-THE-SHELF DEEP MODELS, Proceedings of the International Conference on Image Processing. 2018; 3528-3533.

69. Meng X, Gao F, Shi S, Zhu S, Zhu J. MLANs: Image Aesthetic Assessment via Multi-Layer Aggregation Networks, Proceedings of International Conference on Image Processing Theory, Tools and Applications. 2018; doi: 10.1109/IPTA.2018.8608132.

70. Akiba T, Suzuki S, Fukuda K. Extremely Large Minibatch SGD: Training ResNet-50 on ImageNet in 15 Minutes. 2017; arXiv:1711.04325v1 [cs.DC].

71. Rezende E, Ruppert G, Carvalho T, Ramos F, Geus P. Malicious Software Classification Using Transfer Learning of ResNet-50 Deep Neural Network, Proceedings of The 16th IEEE International Conference On Machine Learning And Applications. 2017: 1011-1014.

72. Hu J, Shen L, Albanie S, Sun G, Wu E. Squeeze-and-Excitation Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence. 2020; 42(8): 2011-2023.

73. Cheng S, Wang J, Zhang L, Wei Q. Motion Imagery-BCI Based on EEG and Eye Movement Data Fusion, IEEE Transactions on Neural Systems and Rehabilitation Engineering. 2020; 28(12): 2783-2793.

74. Cheng S, Fan J, Dey A. Smooth Gaze: a framework for recovering tasks across devices using eye tracking, Personal and Ubiquitous Computing. 2018; 22(3):489-501.