



ORIGINAL RESEARCH

AI-driven user aesthetics preference prediction for UI layouts via deep convolutional neural networks

Baixi Xing¹ | Hanfei Cao¹ | Lei Shi² | Huahao Si³ | Lina Zhao³¹Zhejiang University of Technology, Hangzhou, China²Durham University, Durham, UK³Hangzhou Dianzi University, Hangzhou, China**Correspondence**Baixi Xing, Zhejiang University of Technology, Hangzhou, China.
Email: xingbaixi@zjut.edu.cn**Funding information**

Basic Scientific Research Project of Humanities and Social Sciences (Interdisciplinary Research Project) of Zhejiang University of Technology, Grant/Award Number: SKY-ZX-20200345; Natural Science Foundation of Zhejiang Province of China, Grant/Award Number: LY19F020047

Abstract

Leveraging the power of computational methods, AI can perform effective strategies in intelligent design. Researchers are pushing the boundaries of AI, developing computational systems to solve complex questions. The authors investigate the association of user preference for UI and deep image features, aiming to predict user preference level using deep convolutional neural networks (DCNNs) trained on a UI design image dataset. A total of 12,186 UI design images were collected from UI.cn and DOOOOR.com. Users' views and likes can help understand the implicit user preference level, which is set as the ground-truth annotation for the dataset. Six DCNNs, including VGG-19, InceptionNet-V3, MobileNet, EfficientNet, ResNet-50 and NASNetLarge were trained to learn the user preference of UI images. The experiment achieves an optimal result with a mean-squared error of 0.000214 and a mean absolute error of 0.0103 based on EfficientNet, which indicates that the proposed method provides the possibility in learning the pattern of user aesthetics preference for UI design. On the basis of the prediction model, a mobile application named 'HotUI' was developed for UI design recommendations.

KEYWORDS

computational aesthetics, deep convolutional neural networks, UI aesthetic assessment, user preference

1 | INTRODUCTION

In an era of AI-driven design, many researchers have delved deeply into how AI can influence and improve design. It is challenging to predict which GUI design may become the most popular one among Internet users. Despite the known controversial ethical and philosophical challenges, AI can be the catalyst for the evolution of design aesthetic evaluation.

The *views* and *likes* data of GUI design may help understand users' true preferences. Using these data, machine learning models can be built to make predictions of user preferences. Consequently, these predictive models can drive the innovation of design assessment techniques, and thus making AI smarter with design issues.

In UI design, aesthetics defines pleasing design qualities through human visual communication with user interfaces [1]. Lavie T. et al. Found that excellent designs have good

performance in both classical orderly and expressive creativity by the factor analysis [2]. Websites' credibility can be highly influenced by visual design aesthetics quality [3]. Similarly, website designs can greatly affect the first impression of users for a company after a very short exposure [4]. Harrison L. et al. analysed 1278 participants' ratings for infographic aesthetics and found that the user's first impression is largely influenced by colourfulness and visual complexity of design [5]. Moreover, UI design aesthetics follows certain design rules, which can be quantified as computational characters. In the study of Zheng X. S. et al. [6], symmetry, balance and equilibrium were evaluated for their correlation with user aesthetic and affective judgement. They revealed interesting patterns between image low-level features and design relevant dimensions, which indicate that it is feasible to develop aesthetics user preference judgement method based on computation algorithms for general evaluation of image design.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Cognitive Computation and Systems* published by John Wiley & Sons Ltd on behalf of The Institution of Engineering and Technology and Shenzhen University.

Consequently, we argue that the computer can recognise the best UI design from the visual aesthetic perspective using the UI image crowd analysis method. A meeting point, where the intelligent computational algorithm finds patterns, connects and inspires abstract UI graphics art. Here, we dedicate ourselves to the challenging task of assessing the user preference level of UI design and efforts put into the data mining of *likes/views* rating of UI image data. A promising UI preference model is explored using deep learning methods on a visual aesthetics basis.

In this study, we aim to build a user preference prediction model, which is trained on a UI design image dataset. There are four research contributions in this work: (1) A total of 12,187 UI images were collected to form the UI user preference database; (2) Implicit user preference level was explored by user data of *views* and *likes*. Specifically, the value of UI design *likes/views* proportion reflects the popularity of the design to some extent, which can be used as the user preference rating for the UI design image sample. (3) Six deep convolutional neural networks (DCNNs) were compared to build the optimal UI user preference prediction model in the experiment. Specifically, Inception-V3, MobileNet, Visual Geometry G (VGG-19), EfficientNet, ResNet-50 and NAS-NetLarge were applied to construct an effective user preference prediction model based on the image feature analysis. The optimal model result is achieved with a mean-squared error (MSE) of 0.000214 by EfficientNet in model validation. (4) A mobile application named '*HotUI*' was developed for UI design recommendations. The general pipeline of the study is illustrated in Figure 1.

The rest of this article is structured as follows: Section 2 reviews the related works of theoretical and practical studies in the literature; Section 3 provides the methodology of this study, including image processing, feature analysis and an introduction of the DCNNs network structure applied in the experiment; Section 4 introduces the experiment procedure, including the data collection and the model construction; Section 5 discusses and analyses the experimental results; Section 6 introduces the application *HotUI* developed based

on the prediction model; and Section 7 presents our conclusions and discusses some research opportunities in the future.

2 | RELATED WORK

Aesthetics is an important dimension in evaluating the attractiveness of images [7]. Data processing techniques can be applied to optimise the user experience and interaction procedure based on user behaviour features [8, 9]. UI design follows certain rules, including the spatial organization in grids layout [10, 11], spatial graph grammar (semantic grouping and interpretation of layout segments) [12], rule of three colours scheme, style trends [13] etc. Specifically, these design characters give us the opportunity to develop AI-driven methods for design evaluation. Colourfulness and complexity of a website are widely studied for aesthetic evaluation [14]. Reinecke K. [5] investigated the difference of visual preference for websites among people around the world using a dataset of 2.4 million ratings by 40 thousand of participants. A computational method based on a perceptual model was proposed in the study to present the estimation of the peak appeal of colourfulness and complexity, which can support website design and evaluation [15]. Symmetry is another factor that affects human aesthetic perception. It is interesting that males' judgement for website aesthetics can be influenced by symmetry, but it is not the case for females, according to the study conducted by Tuch A. N. et al. [16]. The conclusions of these studies can support designers and developers to know better about users' preference for interfaces.

There are great opportunities of AI-driven design in both academia and industry. Here, we reviewed the studies of UI aesthetics evaluation modelling using computational methods and list the related works in Table 1. In the existing studies, there are mainly two kinds of mechanisms for the user preference analysis:

- (1) One is using regression methods to predict the image ranking result based on image quality assessment.

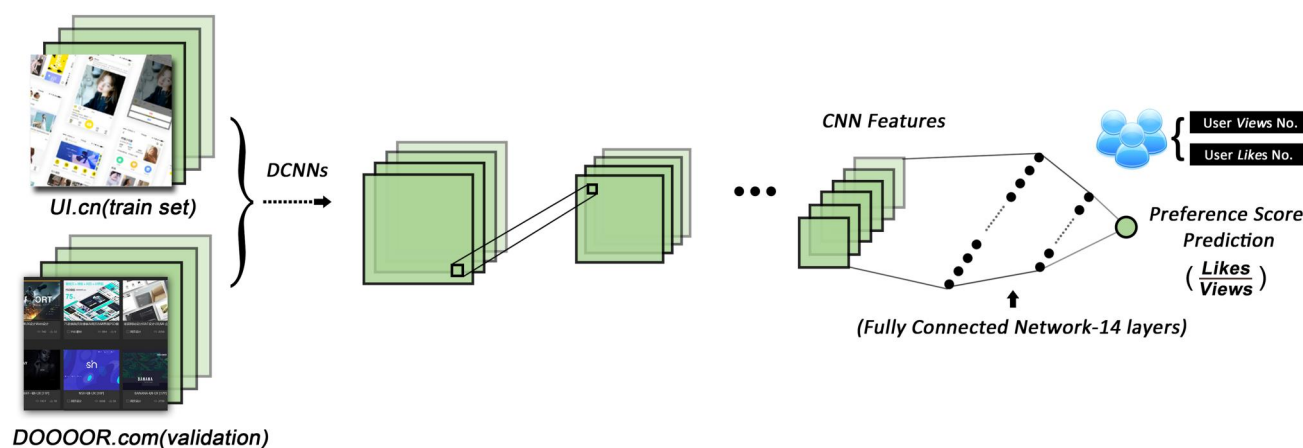


FIGURE 1 Research approach of UI user preference prediction modelling

TABLE 1 User preference prediction modelling for images and websites

Refer	Features	Classifier/Method	Descriptors	Dataset	Users	Results
Lerman K. 2007 [22]	Topic tags	Probabilistic model	User agreement of the search result	13,500 images	-	The image search precision was dramatically increased.
Fan J. 2008 [23]	Topic tags	Kernel principle component analysis	Personalised interest score	Flickr images 1.5 billion	-	JustClick system shows positive results in personalised image recommendation.
Lu D. 2011 [24]	Groups, tags	Rank SVM	Preference score	Flickr images	81,288	Improvement was achieved by the proposed approach compared to one-fit-all retrieval methods
Wu O. 2011 [25]	Visual and text features	LibSVM	High/Low visual quality	500 homepages	7	AMCC = 0.1846
Liu X. 2016 [26]	Image content and user profiles	A hybrid algorithm	-	0.3 million flickr photos	20,298 users	mAPS = 45.77
Cuntuku S. C. 2015 [29]	Visual and text-based features	a probabilistic approach	User like counts	Flickr 60,000 images	300 users	The proposed method improved the state-of-the-art performance by 12%.
S. Ohsawa, 2013 [30]	Log-sum-likes, avg-likes, log-max-likes etc.	SVR	User like counts	22,616,574 Pages in 190 categories in Facebook	-	A prediction root mean square of 0.810 was achieved.
Niu W. 2018 [21]	Tags, geographic features, visual factors	Neural personalised ranking	Preference ranking	Flickr YFCC 100M dataset	-	Outperform the baseline by 16.6% in precision.
Dou Q. 2019 [32]	Colour, complexity	Deep learning	Aesthetic rating	398 websites	40,000	Pearson $r = 0.85$, $q < 0.001$
Lu X. 2014 [33]	Global and local image features	RDCNN	High/Low aesthetic	AVA	-	MAP = 56.81%
Ren J. 2017 [34]	Image and contents	Active learning	Aesthetic score	Flickr 40000images/14 albums	-	Average correlation = 0.59
Suchecki M. 2017 [35]	CNN features	AlexNet, support vector machine (SVM), random Forest classifier	Aesthetic score	1.7 million photos	-	The trained model achieves accuracy of 70.9%
Li M. 2018 [36]	Visual balance	K-means	High/Low quality	65 images	118	$\alpha = 0.8$, $\beta = 0.2$, Poor (A) = excellent (R) = 1

TABLE 1 (Continued)

Refer	Features	Classifier/Method	Descriptors	Dataset	Users	Results
Reddy G. 2019 [37]	CNN features	EfficientNet	High/Low quality	1.2 million images	-	For almost all the attributes, the proposed method is superior to the existing model.
Bourguet M. 2018 [38]	Graphical user interface image features:	Segmentation methods, Ngo's aesthetics evaluation metrics	User agreement of the search result	Ten web designs	-	The segmentation problem is a bottleneck issue towards aesthetics evaluation.
Bakaev M. 2021 [40]	UI design elements	RLE, deflate, Hick's law, fitts' model, MHP	-	5000 VSWM-engaging tasks	164	IDM can be used to refine the layout of UI elements in order to lower VSWM load.

Abbreviations: AMCC, average miss classification cost; AVA, Aesthetic Visual Analysis; DCNN, deep convolutional neural networks; IDM, index of difficulty; KNN, k-nearest neighbour; mAPS, mean average precision; MAP, mean average precision; MHP, human Processor model; RDCNN, regularised deep convolutional neural networks; RLE, run-length encoding; SVM, support vector machine; SVR, support vector regression; VSWM, visual-spatial working memory; YFCC, Yahoo Flickr Creative Commons.

S. Singh et al. proposed a saliency-enforced convolutional neural network to predict image emotion and image aesthetics based on an image ranking framework. They achieved comparable results to the state-of-the-art methods [17]. Y. J. Liu et al. presented a learning-based Objective Quality Assessment method to predict the ranking of images. A general regression neural network was constructed to output the ranking result, which is demonstrated to be effective with better generalisation ability to various datasets [18, 19]. J. Zhou trained a fine-tuned generic model by convolutional neural network (CNN)-based regression for individual preference prediction. The experimental result shows that their method can predict personalised image aesthetic preference effectively [20]. W. Niu et al. proposed a neural personalised ranking model for Flickr image recommendations. They used the Flickr Yahoo Flickr Creative Commons 100 M dataset with implicit user feedback and proved that the result outperforms the baseline in precision. In their study, user preference of spatial, topics and visual was discussed as the evidence for ranking prediction [21].

- (2) Another approach is using DCNNs to predict user preference based on image features and user like counts in social media. Here, we mainly explored user aesthetic preference for GUIs with this approach. Many researchers have proposed different personalised image search models to assist Flickr users in getting their interested photos. K. Lerman et al. used the tag information on Flickr images to recommend personalised search results to users. They collected the top 4500 images for searching terms of tiger, newborn and beetle. A probabilistic model was applied to find the latent topics for personalising recommendations. The final search result has a significant improvement in search precision [22]. J. Fan et al. developed JustClick, a topic-driven personalised image recommendation system for image exploration on Flickr. The integrated method of the Kernel principal component analysis and hyperbolic

visualization were utilised to analyse the topic tags [23]. User profiles were investigated to extract the information of their implicit interests, including groups and tags information. Term frequency-inverse document frequency method was employed to set the edge weights of group-vector and tag-vector in this study. The result demonstrates that the improvement of the proposed approach was achieved compared to one-size-fit-all retrieval methods [24]. O. Wu et al. introduced a webpage layout extraction algorithm partitioning a page into blocks to evaluate the visual quality. Visual features and text features were investigated as the primary factors to learn the visual quality level. A total of 500 website homepages were rated by seven PhD students to form the dataset. They applied libSVM in the modelling and achieved an average miss-classification cost of 0.1846 as a result [19, 25]. X. Liu et al. collected 0.3 million Flickr images for user preference modelling and proposed a novel hybrid algorithm, which integrated the image content analysis with user sparse information. The method outperformed another latent factor-based method, with a mean Average percentile score of 45.77 [20, 26].

Besides, user '*likes*' is the direct evidence for user interest, which can reflect the popularity of entities. Research of user '*likes*' is significant to predict user preference and personalised application. Many researchers have addressed this topic in their study in various aspects. D. Lee et al. proposed the concept of 'Likeology' in 2015 with a comprehensive overview of user '*likes*' study based on social media data [27, 28]. S. C. Cuntuku et al. developed a computational model for user '*likes*' for images based on a collection of images crawled from Flickr. Both visual and text-based features were utilised to predict user '*likes*' with a probabilistic approach [29]. S. Ohsawa et al. proposed a model for *like* counts prediction. They crawled 22,616,574 Pages in 190 categories in Facebook, which have a

total of 30,685,646,909 *likes*. They found 10 most effective features for people, including log-sum-likes, avg-likes. A prediction root mean square of 0.810 was achieved in the study with a proposed *likes* predict function [30]. In 2015, they crawled a dataset of 2 billion *like* activities from 20 million users in Instagram. They compared the structural characteristics of Like Network and Follow Network. They also examined factors that influence *likes* counts [31].

Furthermore, we can refer to website aesthetics computing modelling for the image feature analysis and the experimental procedure. Q. Dou et al. presented 'Webthetics', a webpage aesthetic quantifying method based on deep learning networks. A total of 398 website screenshots rated by 40,000 users were used in the experiment. Hand-crafted image features of colourfulness and complexity were extracted for objective evaluation [32]. X. Lu et al. developed a regularised double-column convolutional neural network and achieved a mean average precision of 56.81% on the Aesthetic Visual Analysis dataset. The proposed method considers both the global views and local views of an image [33]. J. Ren et al. collected two image datasets for personalised image aesthetics study, including a dataset of 40,000 photos from Flickr and a dataset of real personal albums. A residual-based model adaptation scheme was introduced as an active learning algorithm in the experiment. The result indicates that the approach outperforms existing methods with an average correlation of 0.59 [34]. M. Suchecki et al. used a dataset of 1.7 million photos collected from Flickr for aesthetic value evaluation modelling. To suppress the 'time effect', they normalized each aesthetic value by calculating the number of *views* divided by the number of *days* since a photo was uploaded. Then, a fine-tuned AlexNet was proposed for the aesthetic value assessment [35]. M. Li et al. learnt the image aesthetic level based on visual balance scores. They assessed the visual balance score of each image according to the positional relationship between the gravity visual centre and the physical centre [36]. G. Reddy et al. proposed a multi-task deep CNN influenced by EfficientNet that collectively learns aesthetics score for images. They found that the scores of six aesthetics attributes (content, colour, harmony, depth of field, light, object emphasis, vivid colour) correlate considerably with image aesthetics scores [37]. In the field of image aesthetics computing, researchers also exhibit many promising directions and topics that can be further studied. M. Bourguet et al. pointed out that it is important to study the aesthetics aspect of user interface design and the segmentation problem could be further investigated in this area [38]. K. Srinivas et al. presented an efficient low light image enhancement framework to improve the image visual quality with lower distortions [39]. M. Bakaev et al. explored predictors of user engagement performance in their interaction with graphical user interfaces. Several derived independent UI design variables were considered, and they proposed the method to calculate the index of difficulty for the tasks engaging visual-spatial working memory in UI design [40].

Research of user preference is quite important in understanding user behaviour in social media and developing outstanding personalised services. Many existing studies have

addressed the importance from various aspects. Moreover, image visual features and user behaviour features were commonly utilised to evaluate user preference in the literature. According to existing studies listed above, visual aesthetic modelling for websites using computational methods was proved to be effective. However, most of the studies are centred on the investigation of visual appearance with subjective evaluation. Therefore, it is necessary to explore scientific approaches to quantify the level of preference among user crowd and build optimal models for user preference prediction based on the ground-truth evidence. In this work, user aesthetic preference study with DCNNs method on GUI images can present a promising way for this area. Successful user preference recognition of GUIs can be an effective way of GUI image/style retrieval and shed light on the 'Likeology' study in social media. Considering the massive scale of social data, it is significant to explore an automatic user preference prediction method.

3 | METHOD

3.1 | Research paradigm

Various machine learning methods, including support vector machine, AlexNet, EfficientNet, ResNet-50 and some other DCNNs have been proved to be feasible in assessing aesthetic quality and image quality. In particular, DCNNs achieve good performance in such tasks. Here, we used DCNNs to learn a user preference model using the UI design image dataset. The study focusses on building the prediction model based on the deep image features analysis. Firstly, a set of 10,000 UI image samples were crawled from UI.cn and 2186 samples from DOOOOR.com. Secondly, the value of *likes/views* was set as the label of user preference level for each image sample. Thirdly, the images were resized to be a standard size without cropping for different DCNNs as the network input. The detailed network configurations of VGG-19, InceptionNet-V3, MobileNet, EfficientNet, ResNet-50 and NASNetLarge applied in this study are presented in Table 2. The detailed procedure of the methodology is described as follows.

3.2 | UI design image data collection

UI design works exhibited in design websites are suitable image resources for UI user preference learning, since images displayed in these websites have ground-truth annotations of user *likes* and *views* data. In view of this, we selected two popular design websites 'UI.cn' and 'DOOOOR.com' to collect UI design images.

UI.cn (UI China), founded in 2008, has become a professional design platform in China after more than 11 years of development, and it has nearly 1 million registered members, 90% of which are professional user experience designers (including interaction designers, visual designers, user

TABLE 2 Convolutional network configuration of VGG-19, InceptionNet, MobileNet, EfficientNet, ResNet-50 and NASNetLarge

VGG-19 Input [224,224,3]	InceptionNet Input [299,299,3]	MobileNet Input [224,224,3]	EfficientNet Input [600,600,3]	ResNet-50 Input [224,224,3]	NASNetLarge Input [331,331,3]
Conv1	Conv1	Conv1	Conv1	Conv1	Conv_normalv1
[224,224,64]	[149,149,32]	[112,112,32]	[224,224,64]	[112,112,64]	[42,42,336]
[112,112,64]	[147,147,32]	[112,112,32]	[112,112,32]	[56,56,64]	
Conv2_x	Conv2_x	Conv2_x	MBConv2	Conv2_x	Conv_normalv2
[112,112,128]	[147,147,64]	[112,112,32]	[112,112,48]	[56,56,256]	[42,42,1008]
[56,56,128]	[73,73,64]	[112,112,64]	[56,56,80]		
		[56,56,64]			
Conv3_x	Conv3_x	Conv3_x	MBConv3	Conv3_x	Conv_reducev0
[56,56,256]	[71,71,80]	[56,56,128]	[56,56,160]	[28,28,512]	[21,21,1344]
[28,28,256]	[35,35,192]	[28,28,128]	[28,28,224]		
	[35,35,288]				
Conv4_x	3xInception	Conv3_x	MBConv4	Conv4_x	Conv_normalv3
[28,28,512]	[17,17,768]	[28,28,256]	[14,14,384]	[14,14,1024]	[21,21,2016]
[14,14,512]		[14,14,256]			
		[14,14,512]			
Conv5_x	5xInception	Conv4_x	MBConv5	Conv5_x	Conv_reducev1
[14,14,512]	[8,8,1280]	[14,14,512]	[7,7,640]	[7,7,2048]	[11,11,2688]
[7,7,512]			[7,7,2560]		
Dense3_x	2xInception	Conv5_x	Dense1	Dense3_x	Conv_normalv4
[25088]	[8,8,2048]	[7,7,512]	[1000]	[2048]	[11,11,4032]
[1000]		[7,7,1024]		[1000]	
Output	Dense1_x	Dense1_x	Output	Output	Dense
[1000]	[2048]	[1024]	[512]	[512]	[4032]
[512]	[1000]	[1000]			[1000]
	Output	Output			Output
	[1000]	[1000]			[512]
	[512]	[512]			

researchers etc.). The users of UI.cn come from all major cities in China, and it is also the most influential design platform in China. DOOOOR.com is a design platform to exhibit excellent design works abroad and provide designers with the brain food for design inspiration. It has over 198,000 registered members and a large number of works constantly enriched and updated.

They presented thousands of design works of UI and the number of user *views* and *likes* of each showcase. A total of 12,186 UI images were obtained to form the UI image database, including 10,000 images from UI.cn and 2186 images from DOOOOR.com.

We also collected *likes/views* ratings of the UI design images in the experiment, and the value of *likes/views* proportion was recorded as the ground-truth label for the images. The image samples are all in a resolution of 96 dpi and were set to the standard size according to the requirement of each applied DCNN. The images were used without cropping, in order to retain its complete visual information presented to users.

The major differences of the collected dataset compared to other datasets can be concluded in two aspects, including label setting method and image collection method.

(1) Label setting:

Here, we used the ratio of *likes/views* as the label, which can take the time scale into account. Considering that some older works may always receive more *likes* and *views* than very recent works despite their real aesthetics quality.

(2) Image collection:

Most existing aesthetics datasets consist of images of landscapes, plants and portraits. Many of them have diverse image content and the image content frameworks differ. Unlike GUI images, user-interface design usually follows certain rules of content arrangement, which has consistency in graphical presentation. A typical website interface has some basic design elements presented in a relatively standard design structure, including the header, banners, text part, and the footer. GUI images with consistent design content and framework can be suitable for aesthetics computing. Consequently, we constructed a new dataset with GUI images collected from UI.cn and DOOOOR.com with the ratio of *likes/views* as ground-truth annotation, revealing the aesthetics user preference level of each image, see Figures 2 and 3.

In the future study, social factors and user characters can be considered, such as image's shareability and user personality.

3.3 | Applied deep convolutional networks

VGG-19.

VGG-19 network structure.

For a learning image aesthetics level, which is highly related to an image user preference prediction task, VGG-19 achieves competitive performance in the literature. Consequently, a fine-

tuned VGG-19 network was applied in the prediction modelling in this work. Python was used to build the VGG-19 model for the UI image analysis [41]. A general introduction of the VGG network is presented below.

VGG-19 network uses 3×3 convolution filters in its construction. In the training procedure, the original images are resized to 224×224 pixels and be put into the convolutional network layers followed by three fully connected layers. A softmax layer is the final layer in the network. In this work, ReLU is used as the activation function.

Feature extraction by VGG-19.

In the initial stage, we extracted 25,088 image features in total, which were then reduced to be an output vector with 1000 features. Finally, a 512-dimensional feature vector was obtained by a fully connected network to form the feature set for training in the next step.

InceptionNet-V3

InceptionNet-V3 network structure.

For many multimedia processing tasks, the convolution network is the core of the latest solution. For most tasks, although the extension of model size and calculation cost can contribute to model accuracy improvement, computing efficiency is still the limiting factor in various scenarios. Inception-V3 network is a very deep convolution network developed by Google. Its goal is to utilise the increased computation effectively through proper deconvolution and active regularisation. The default input size of this model is 299×299 , with three channels [42].

Inception-V3 is developed with four characters: (1) the bottleneck of feature description should be prevented; (2) it can accelerate the speed of convergence by increasing the number of features; (3) the number of feature dimensions can be compressed to reduce the amount of computing; and (4) the depth and width of the entire network structure should be balanced.

The number of convolutions is reduced to be 3×3 . The module of Inception is a 17×17 grid with 768 filters. In the structure, there are five inception modules. After that, an $8 \times 8 \times 1280$ grid is obtained and a feature set of 2048 dimensions is got as the output.

InceptionNet-V3 feature extraction.

A total of 2048 dimensions of features were obtained from the network output. Then, it was reduced to be a set of 1000 features. Finally, a 512-dimensional output vector is got by a fully connected network to form the feature database.

MobileNet.

MobileNet network structure.

Presently, DCNNs have shown a great performance in multiple learning tasks. However, the demand of huge storage and the computing cost has seriously limited the application of DCNNs in many areas. MobileNet is a light weighted DCNN widely used for the image analysis. It can effectively compress storage and computation at the algorithm level. Its structure separates the convolution into several independent groups for computing. The core of MobileNet is to divide convolution into two parts, including depth-wise and point-wise separable convolutions, in order to compress the computation. This compressed structure causes some accuracy loss, but it uses

global hyperparameters to decide on the best size of the network for different scenarios. In the structure of MobileNet, there are 28 layers, with a BatchNorm (batch normalisation) and a rectified linear unit followed each layer. Before the final layer of the fully connected network, it uses average pooling to reduce the spatial resolution. And finally, the output was obtained by a softmax layer [43].

MobileNet feature extraction.

By applying MobileNet in CNN feature extraction, we first extracted a set of 1024-dimensional image features. Then, they were reduced to 1000 dimensions in the output layer. At last, 512-dimensional features were obtained by a fully connected network to form the feature set for training.

EfficientNet.

EfficientNet network structure.

M. Tan et al. proposed EfficientNet in 2019. It is designed to balance network specifications to achieve improvement in performance [44]. A new scaling method was proposed in the

network, which applies a simple and efficient composite coefficient to expand the network from depth, width and resolution. By using neural structure retrieval technology, the network can get the best set of parameters (coefficients). EfficientNet can achieve great model efficiency and model performance compared to the other DCNNs.

EfficientNet feature extraction.

A total of 1280 dimensions of features were obtained from the network output. Then, it was reduced to be a set of 1000 features. Finally, a 512-dimensional output vector is got by a fully connected network to form the feature database.

ResNet-50.

ResNet-50 network structure.

Because of its powerful representation ability, ResNet-50 is widely used in computer graphics and computer vision applications, and it has achieved a breakthrough in computing performance [45, 46]. The network structure is inspired by VGG-19. It introduces residual blocks into the VGG-19 network,

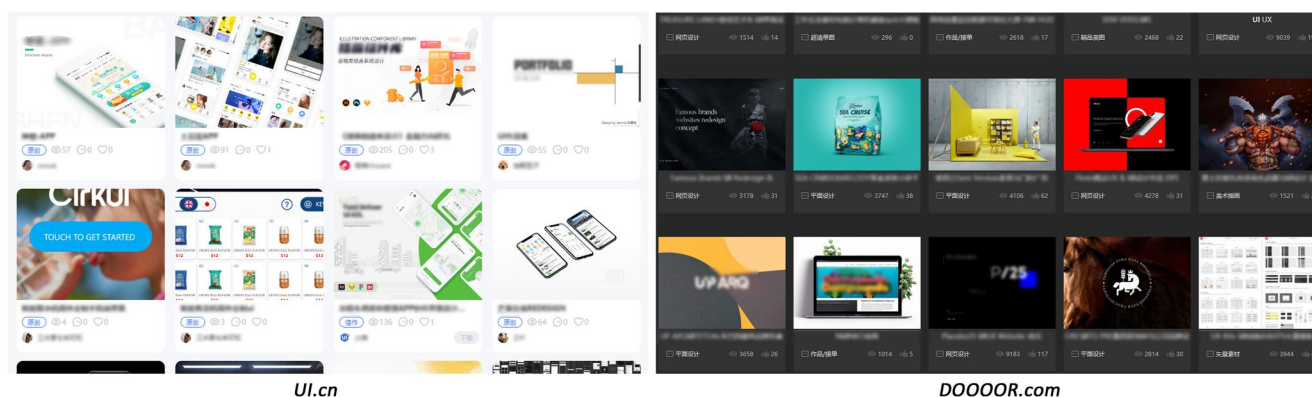


FIGURE 2 Samples of UI design images collected from UI.cn and D0000R.com



FIGURE 3 Labels can reflect implicit user preferences. Users can endorse the UI works with a *like* in UI.cn and D0000R.com. The user *views* data is also recorded

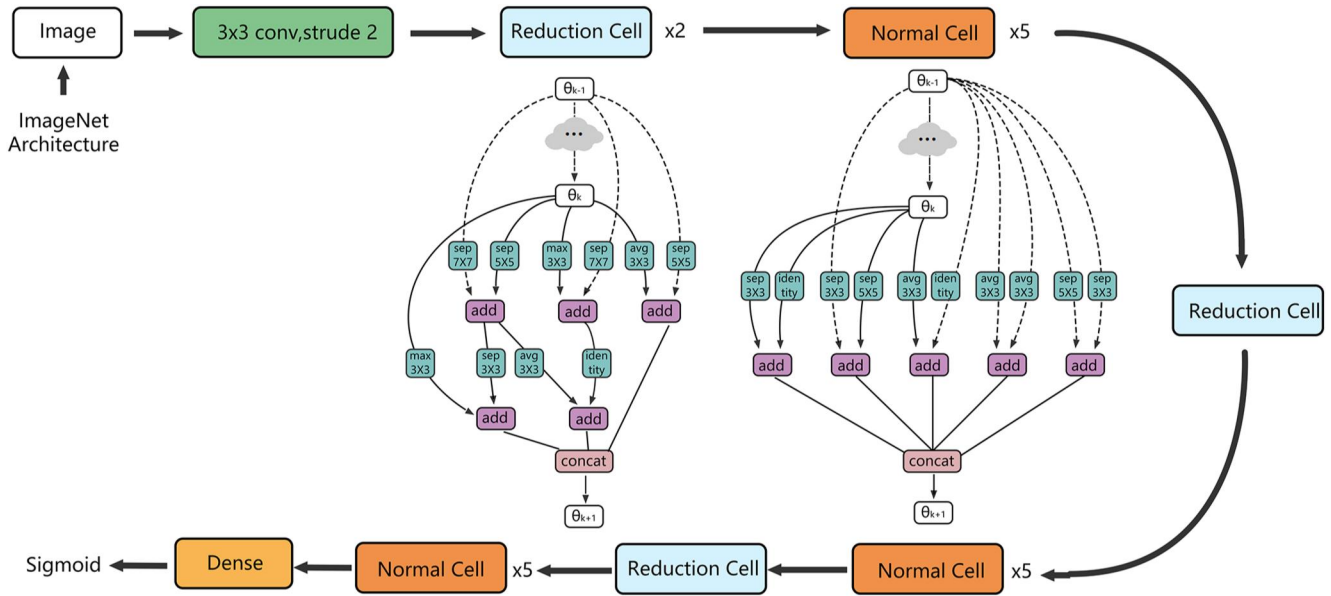


FIGURE 4 Network architecture of the improved NASNetLarge

which alleviates the problems faced in training very deep networks. The core of the residual block is to skip the direct connection of some layers of the model. These skip connection technology in ResNet can solve the problem of gradient disappearance in deep CNN. In this experiment, the input image is rescaled to 224×224 pixels to put into the network. ResNet-50 was used to extract the features. A fully connected network was used to get the output vector. We applied 2048 dimensional features to train the fully connected network for 200 epochs. In the training process, binary cross entropy was set as the loss function, and Adam was applied as the optimiser.

ResNet-50 feature extraction.

A total of 2048 dimensions of features were obtained from ResNet-50 extraction. Then, it was reduced to be a set of 1000 features. Finally, a 512-dimensional output vector is got by a fully connected network for the output vector.

NASNet.

NASNet network structure.

Zoph et al. designed significant network architecture for image classification [47]. They proposed a search for an architectural network block on a small dataset and then transfer the block to a larger dataset. The main contribution of this method is that they have designed a new search space, namely 'NASNet search space'. It can search for the best convolutional layer ('cell') on a small dataset and then apply this cell to the larger dataset. By stacking more of these network units, the network is migrated to more complex and larger datasets. This network convolutional architecture was named as 'NASNet architecture'. They also introduced a new regularisation technique called 'ScheduledDropPath', which can improve generalisation ability of NASNet. There are mainly two types of NASNet, including NASNetLarge and NASNetMobile. Here, we conducted cells reduction based on NASNetLarge structure, which can improve the computation efficiency. The detailed network architecture is presented in Figure 4.

NASNet feature extraction.

A total of 1000 dimensions of features were obtained from the network output. And then, a 512-dimensional output vector is got by a fully connected network to form the feature dataset.

Prediction model based on a fully connected network.

After extracting the features by different DCNNs, a fully connected network is used to build the prediction model. The prediction network has 14 layers. The output layer is sigmoid (mapping the result to 0–1) while the activation function in the other layers is Relu to prevent gradient disappearance or gradient explosion. In this experiment, the learning of UI user preference levels is a regression problem; therefore, there is no activation function Relu in the output layer, see Table 3.

4 | EXPERIMENT

In the modelling experiment, we built our prediction models on the UI.cn dataset and validated them with the dataset collected from DOOOOR.com, in order to manifest their generalisation ability. We compared the performance of VGG-19, InceptionNet-V3, MobileNet, EfficientNet, ResNet-50 and NASNetLarge to find the optimal model. Several model metrics were applied for result evaluation, including MSE, root-mean-squared error (RMSE), mean absolute error (MAE), and median absolute error (MDAE). We select the optimal model that performs the best with the lowest value of these metrics, which demonstrates its competitiveness in prediction. The detailed roadmap of the experiment is illustrated in Figure 5.

Many personalised search models constructed users' interest profiles to quantify the level of their preference, including user tagging and group information in social networks (see Table 1). Yet, researchers have not reached a consensus on how to evaluate user preference level for UI

design works. While in the web aesthetic exploration studies, they usually set the user aesthetic ratings of page screenshots as the prediction label. And user *likes* and *views* were regarded as the ground-truth implicit preference evidence [26]. There are various themes in design, including

TABLE 3 Prediction model using a fully connected network

Layer type	Layers	Output size
Dense1_x	fc, relu	[512]
	Batch normalisation	[512]
	fc, relu	[512]
Dense2_x	fc, relu	[256]
	Batch normalisation	[256]
	fc, relu	[256]
Dense3_x	fc, relu	[128]
	Batch normalisation	[128]
	fc, relu	[128]
Dense4_x	fc, relu	[64]
	Batch normalisation	[64]
	fc, relu	[64]
Dense5_x	fc, relu	[32]
	Batch normalisation	[32]
Dense6_x	fc, relu	[16]
	Batch normalisation	[16]
	fc, relu	[16]
Dense7_x	fc, relu	[8]
	Batch normalisation	[8]
	fc, relu	[8]
Output	fc, sigmoid	[1]

packaging design, graphics design and 3D design. UI design is selected in this study for its importance for Internet products success and its flexibility and charm for user perception. It is significant to study how users favour UI design and how to evaluate the UI design before the system or website is online. In order to explore the prediction model, we held the experiment in three steps: (1) user preference level computing and label normalisation; (2) image features normalisation; (3) UI user preference level regression modelling exploration.

(1) Label of UI user preference level.

There are two types of user preferences that are studied in the literature, which are explicit preference and implicit preference. Explicit user preference is also referred as explicit feedback, including user ratings and comments for items. Implicit preference is reflected by user *hits*, *views* and give *likes* history. The user-preference prediction method for UI design aims to establish an approach to forecast the popularity of the UI design work in its design concept stage. The objective of this experiment session is to seek for the approach to represent user preference knowledge for UI design. It is perceived from the number of users' *likes* and *views* of each image. To achieve this, we evaluated the majority users' UI preference level towards the design works by the analysis of user *likes* and *views* data. The rating data sparsity problem also exists as most of the cases in user preference prediction, since not all the users left their comments and rating after their viewing.

Due to the lack of comments in text in UI design websites, we applied implicit preference information of user *likes* and *views* as the label in this study.

- Firstly, P is set as the value of user preference level, N is the number of *likes* and M is the number of *views*, that $P = N/M$.
- Secondly, the value of P is normalized to the scale of 0–1 to be the preference label for the next experiment session.

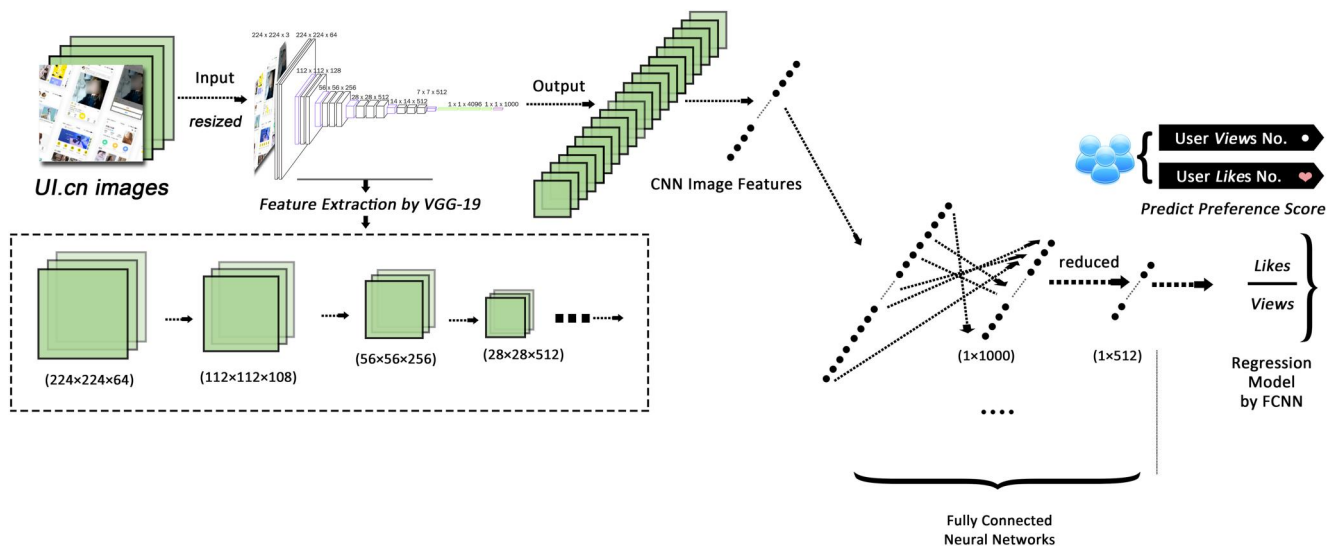


FIGURE 5 Detailed experiment scheme for UI user preference prediction

(2) CNN features normalisation.

We conducted normalisation on the image features extracted by DCNNs. Consequently, the data centralisation can be realized. It conforms to the data distribution mechanism and can improve the generalisation ability of the model. Firstly, the image pixel was normalized from 0–255 to 0–1. Then, the normalized pixel value was used to extract features. Finally, the extracted features were normalized to the scale of 0–1 to form the feature dataset. Normalized label distribution of the UI.cn dataset and the [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) dataset is presented in Figure 6.

(3) UI user preference level regression modelling.

Six types of DCNNs methods were implemented in the model exploration to find the optimal model, including VGG-19, Inception-V3, MobileNet, EfficientNet, ResNet-50 and NASNetLarge. The regression model was established to predict the user preference level indicated by the value of *likes/views*. The proportion of the model training, testing and validation in the UI.cn dataset is about 3:1:1, in which 6000 samples were used for training, 2000 samples were applied in testing and the other 2000 samples were employed for validation. After that, the dataset of 2186 samples collected from [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) was used for cross-dataset model validation.

5 | RESULTS AND DISCUSSION

The metrics of MSE and MAE were measured to reveal the model performance. The results of the validation of the model are presented in Tables 4 and 5. The best model prediction accuracy of user preference level is achieved by EfficientNet with an MSE of 0.000214 and an MAE of 0.0103 in validation of the UI.cn dataset. And the best result is obtained by ResNet-50 with an MSE of 0.00383 and an MAE of 0.0344 in cross-dataset model validation.

Table 2 and Table 3 show the information of the applied DCNNs architecture for feature extraction and the regression model. MobileNet, Inception-V3, VGG-19, EfficientNet,

ResNet-50 and NASNetLarge were applied to extract image features from the collected UI design images. The 14 layers fully connected network was used to flatten the extracted data for preference score prediction. In the last layer, we used one single neuron and Relu was the activation function to output the predicted preference rating.

Adam was utilised as the optimiser in this experiment. Since we conducted the preference prediction task as a regression problem, MSE, RMSE, MAE and MDAE were selected as the metrics to evaluate the regression performance. The details of model parameters setting are provided in Table 4 and Table 5. And MSE was used as the loss function in the comparison in the model validation experiment. The prediction model validation result using the [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) dataset is proved to be steady, which shows that the model has good generalisation across different datasets.

The convergence curve of loss in model validation is shown in Figure 7. As we can see in the model loss curves, with the increase in network training epochs, the loss value is decreasing, which indicates the effectiveness in predicting a preference score. After a training process of multiple epochs, the loss value of VGG-19 achieves 0.000233, the loss of MobileNet is 0.000266, the loss of InceptionNet is 0.000326, the loss of EfficientNet is 0.000214, the loss of ResNet-50 is 0.000292 and the loss of NASNetLarge is 0.000239, when it no longer has the potential for loss decrease. Consequently, EfficientNet achieves the lowest MSE in the UI.cn dataset. In the validation experiment on the [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) dataset, the loss value of VGG-19 is 0.00462, the loss of MobileNet is 0.00386, the loss of InceptionNet is 0.00446, the loss of EfficientNet is 0.00414, the loss of ResNet-50 is 0.00383 and the loss of NASNetLarge is 0.00471. It can be seen from the comparison result that ResNet-50 achieves a lowest MSE in the [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) dataset. To be mentioned, ResNet-50 and VGG-19 are converging asymptotically in a more stable and smooth way in both UI.cn and [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) datasets validation experiments.

The results show that the prediction accuracy of the five networks is very close. Therefore, the time cost of the model

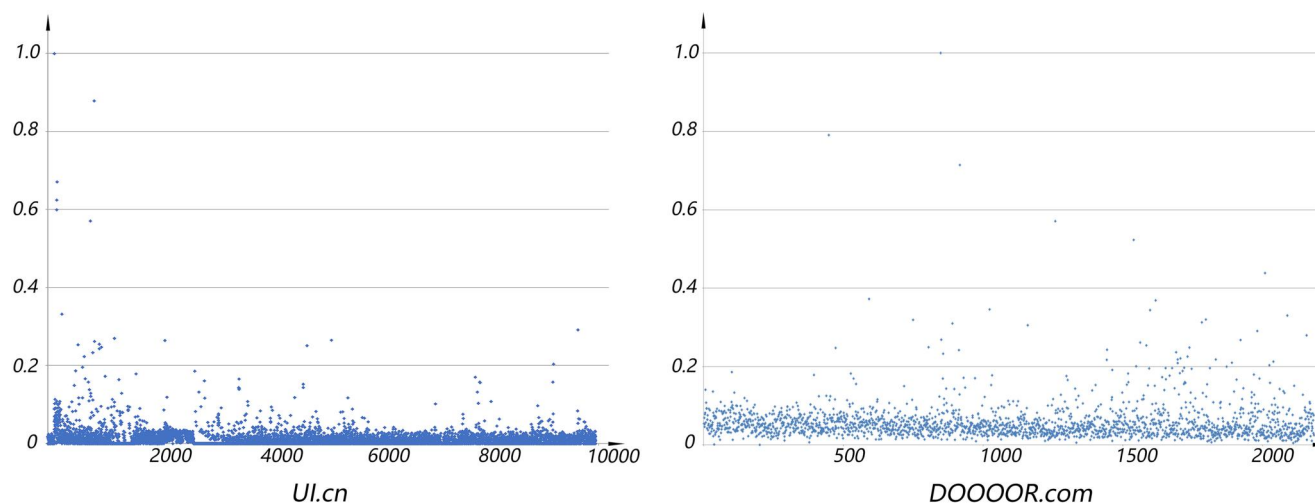


FIGURE 6 Normalized label distribution of the UI.cn dataset and the [DOOOOR.com](https://researchonlinelibrary.wiley.com/doi/10.1049/csc2.12055) dataset

TABLE 4 Validation result of user *likes/* *views* prediction using the UI.cn image dataset

Algor.	Results				<i>Parameters setting</i>
	MSE	RMSE	MAE	MDAE	
InceptionNet-V3	0.000326	0.0181	0.0137	0.0148	Input: 299×299 extraction_iter: 38716, Activation: Sigmoid Epoch:258
MobileNet	0.000266	0.0163	0.0125	0.0143	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:400
VGG-19	0.000233	0.0153	0.0112	0.0128	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:323
EfficientNet	0.000214	0.0146	0.0103	0.0105	Input: 600×600 extraction_iter: 38716, Activation: Sigmoid Epoch:200
ResNet-50	0.000292	0.0171	0.0126	0.0111	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:200
NASNetLarge	0.000239	0.0154	0.00732	0.0113	Input: 331×331 extraction_iter: 38716, Activation: Sigmoid Epoch:200

TABLE 5 Validation result of user *likes/* *views* prediction using the DOOOOR.com image dataset

Algor.	Results				<i>Parameters setting</i>
	MSE	RMSE	MAE	MDAE	
InceptionNet-V3	0.00446	0.0668	0.0421	0.0318	Input: 299×299 extraction_iter: 38716, Activation: Sigmoid Epoch:258
MobileNet	0.00386	0.0621	0.0348	0.0226	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:400
VGG-19	0.00462	0.0680	0.0436	0.0328	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:323
EfficientNet	0.00414	0.0643	0.0381	0.0269	Input: 600×600 extraction_iter: 38716, Activation: Sigmoid Epoch:200
ResNet-50	0.00383	0.0619	0.0344	0.0222	Input: 224×224 extraction_iter: 38716, Activation: Sigmoid Epoch:200
NASNetLarge	0.00471	0.0686	0.0446	0.0338	Input: 331×331 extraction_iter: 38716, Activation: Sigmoid Epoch:200

should be considered as well, to find the best model. The DCNNs network is relatively shallow using a small UI image dataset, it does not require a large amount of computational resources. Table 6 illustrates the time cost, number of parameters and FLOPs of feature extraction models. It shows that EfficientNet has the most time consuming, whereas MobileNet has the best efficiency in modelling. After image feature extraction, the feature dimensions were fed into the prediction model. The total number of parameters is 795,353 and the FLOPs number is 803,367 in the prediction model.

According to the model evaluation results, the model performs well on both UI.cn and DOOOOR.com datasets validation experiments. Therefore, the experimental result indicates the feasibility of our proposed method.

6 | APPLICATION DEVELOPMENT

On the basis of user aesthetics preference prediction modelling, we pursued the industrialisation of UI design recommendation modelling. Considering that the mobile application is becoming a new trend of multimedia retrieval, we delved into developing a mobile application *HotUI*, which can provide user aesthetic

preference prediction service and exhibit aesthetic UI design works based on image aesthetic computing. The user interface of the application is presented in Figure 8.

Users can upload the UI design work image in *HotUI* and the aesthetics score of the image will be computed and presented. The application will predict the *likes/views* ratio of this image to indicate its user preference level. Additionally, aesthetic UI design works with a high score of *likes/views* ratio in the database will be exhibited in the recommendation list, presented in the order of the user aesthetics preference score.

7 | CONCLUSIONS AND FUTURE WORK

AI affects the design process as assistant tools and creativity inspiration. We should be aware of how important the data of user behaviour and rating is when delivering meaningful information for AI-driven design development. In this study, we have investigated how to train a deep learning model on UI design image data, to make a trial in user preference prediction modelling. The prediction of user preference of UI design images is challenging for its subjective and abstract

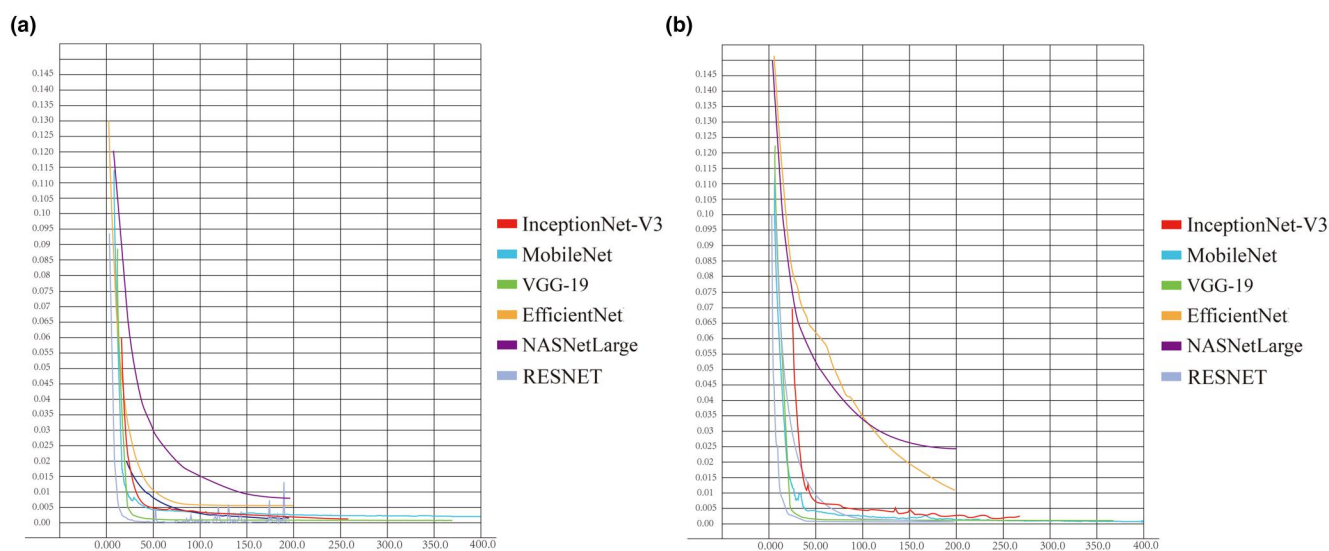


FIGURE 7 (a) Model loss curve during deep convolutional neural networks (DCNNs) validation process using the UI.cn dataset; (b) Model loss curve during DCNNs validation process using the DOOOOR.com dataset

Algor.	Time cost (s)		Trainable parameters	FLOPs
	UIChina(10001)	DOOOOR(2186)		
InceptionNet-V3	1339.19	244.71	22.83 M	5730.28 M
MobileNet	1100.37	172.19	5.48 M	226.04 M
VGG-19	1879.47	328.26	143.67 M	19646.96 M
EfficientNet	2200.48	403.26	66.35 M	38449.58 M
ResNet-50	1227.53	205.33	24.56 M	4110.52 M
NASNetLarge	1620.23	332.15	6.28 M	1253.32 M

TABLE 6 Model time cost, number of model parameters and FLOPs in the feature extraction process via deep convolutional neural networks (DCNNs)



FIGURE 8 Demo interface of *HotUI* mobile application

nature. Previous studies have shed light on how to use computational algorithms to make image preference prediction in practice [16, 48].

In this work, triggered by the interest of interdisciplinary research of design and AI, we presented research on UI user preference prediction modelling with DCNNs, and discussed the possibility of evaluating design schemes in the design industry. UI image data were collected from two popular UI design websites, including UI.cn and DOOOOR.com. A total of 12,186 UI design images were obtained to build the database. We extracted CNN image features to build the model based on the collection, in view of the good performance of DCNNs in existing studies. The value of *likes/views* was set as the ground-truth annotation, which expresses the user preference for the design work exhibition. We built the prediction model with the dataset collected from UI.cn and made validation with the dataset gathered from DOOOOR.com to test the generalisation of the model. The experimental result shows that the proposed method achieves an optimal result with an MSE of 0.000214 and MAE of 0.0103 based on EfficientNet, which suggests that it is an effective way in user preference prediction. Moreover, *HotUI* application was developed on the basis of user aesthetic preference prediction modelling.

In further study, a deep analysis of UI design layout elements with computational methods can be conducted in the experiment. The quantification of design characters, including colour, style, text, and figures schemes can be used in feature extraction and processing procedure. The weights for

assessment of different parts in a UI layout can be studied with user study and affects the computational model construction. Additionally, UI design assessment applications based on user preference prediction model can be further improved according to the user feedback. Besides, the scale of UI user preference database should be further extended to improve the model performance based on a larger number of design images collected over time.

ACKNOWLEDGEMENTS

Conceptualization: Baixi Xing; data curation: Baixi Xing, Hanfei Cao, Lina Zhao; methodology: Baixi Xing, Huahao Si; writing—original draft: Baixi Xing; application development: Huahao Si; writing—review and editing: Baixi Xing, Lei Shi, Hanfei Cao. This study is partly supported by the Basic Scientific Research Project of Humanities and Social Sciences (Interdisciplinary Research Project) of Zhejiang University of Technology (SKY-ZX-20200345) and the Natural Science Foundation of Zhejiang Province of China (LY19F020047).

CONFLICT OF INTEREST

There is no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available on request due to privacy/ethical restrictions.

ORCID

Baixi Xing  <https://orcid.org/0000-0001-6778-3650>

REFERENCES

1. Lisbeth, T.: The role of aesthetics in web design. *Nord. Rev.* 28(1), 63–76 (2008)
2. Lavie, T., Tractinsky, N.: Assessing dimensions of perceived visual aesthetics of web sites. *Int. J. Hum. Comput. Stud.* 60(3), 269–298 (2004)
3. Robins, D., Holmes, J.: Aesthetics and credibility in web site design. *Inf. Process. Manag.* 44(1), 386–399 (2008)
4. Maria, D., Rafael, J., Thielsch, M.T.: Effects of different website designs on first impressions, aesthetic judgements and memory performance after short presentation. *Interact. Comput.* 28(45), 52–56 (2015)
5. Harrison, L., Reinecke, K., Chang, R.: Infographic aesthetics: designing for the first impression. In: *Proc. Of Acm Conference on Human Factors in Computing Systems*, pp. 1–4 (2015)
6. Zheng, X.S., et al.: Correlating low-level image statistics with users' rapid aesthetic and affective judgments of web pages. In: *Proc. International Conference on Human Factors in Computing Systems*, pp. 1–11 (2009)
7. Wu, Z., Gu, Z.Y., Dong, Z.: Multi-dimensional aesthetics mining for social photo recommendation. In: *Proc. Of International Conference on Human-Computer Interaction*, pp. 407–412 (2013)
8. Gu, Z., et al.: Visualizing group user behaviors for social network interaction design iteration. *Proc. of International Conference of Design*, 36–45 (2015)
9. Liu, X., Yin, J.: Stacked residual blocks based encoder–decoder framework for human motion prediction. *Cognitive Computation and Systems*. 2(4), 242–246 (2020)
10. Dayama, N.R., et al.: GRIDS: interactive layout design with integer programming. *Proc. of CHI*. 20, 1–13 (2020)
11. Bauerly, M., Liu, Y.: Computational modeling and experimental investigation of effects of compositional elements on interface and design aesthetics. *Int. J. Hum. Comput. Stud.* 64(8), 670–682 (2006)
12. Jun, K., Kang, Z., Zeng, X.: Spatial graph grammars for graphical user interfaces. *ACM Trans. Comput. Hum. Interact.* 13, 268–307 (2006)
13. Dong, R.: Minimalist style of UI interface design in the age of self-media. In: *Proc. Of International Conference on Information and Social Science*, pp. 217–221 (2019)
14. Michailidou, E., Harper, S.: Visual complexity and aesthetic perception of web pages. In: *Proc. Of the 26th Annual International Conference on Design of Communication*, pp. 215–225 (2008)
15. Katharina, R., Krzysztof, G.: Quantifying visual preferences around the world. In: *Proc. Of Conference on Human Factors in Computing Systems*, pp. 11–21 (2014)
16. Tuch, A.N., Bargasavila, J.A., Opwis, K.: Symmetry and aesthetics in website design: it's a man's business. *Comput. Hum. Behav.* 26, 1831–1837 (2010)
17. Singh, S., Sanchez, V., Guha, T.: Ensemble network for ranking images based on visual appeal. *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing*, 4507–4511 (2020)
18. Chen, Y., Liu, Y.J., Lai, Y.K.: Learning to rank retargeted images. In: *Proc. Of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4743–4751 (2017)
19. Liu, Y.J., et al.: Ranking-preserving cross-source learning for image retargeting quality assessment. *IEEE Trans. Pattern Anal. Mach. Intell.* 42(7), 1798–1805 (2020)
20. Zhou, J., et al.: Joint regression and learning from pairwise rankings for personalized image aesthetic assessment. *Computational Visual Media*. 7(2), 241–252 (2021)
21. Niu, W., Caverlee, J., Lu, H.: Neural personalized ranking for image recommendation. In: *Proc. Of the 11th ACM International Conference on Web Search and Data Mining*, pp. 423–431 (2018)
22. Lerman, K., Plangprasopchok, A., Wong, C.: Personalizing image search results on Flickr. In: *Proc. Of AAAI workshop on Intelligent Techniques for Information Personalization*, pp. 1–11 (2007)
23. Fan, J., et al.: JustClick: personalized image recommendation via exploratory search from large-scale Flickr images. *IEEE Trans. Circ. Syst. Video Technol.* 19(2), 273–288 (2009)
24. Lu, D., Li, Q.: Personalize search on Flickr based on searchers preference prediction. In: *Proc of WWW*, pp. 81–82. (2011)
25. Ou, W., et al.: Evaluating the visual quality of web pages using a computational aesthetic approach. *Proc. of WSDM*. 11, 337–346 (2011)
26. Liu, X., Tsai, M.H., Huang, T.: Analyzing User Preference for Social Image Recommendation, pp. 1–10. *Comput. Res. Repository* (2016)
27. Lee, D., Liu, H.: Like and recommendation in social media. In: *Proc. Of the World Wide Web Conference*, pp. 1539–1540 (2015)
28. Lee, D.: Likeology: modeling, predicting, and aggregating likes in social media. In: *Proc. Of the 8th ACM Conference*, pp. 13–13 (2016)
29. Guntuku, S.C., Roy, S., Lin, W.: Evaluating visual and textual features for predicting user likes. *Proc. of IEEE International Conference on Multimedia and Expo*, 1–6 (2015)
30. Ohsawa, S., Matsuo, Y.: Like prediction: modeling like counts by bridging facebook pages with linked data. In: *Proc. Of the World Wide Web Conference*, pp. 541–548 (2013)
31. Jang, J.Y., et al.: Generation like: comparative characteristics in instagram. In: *Proc. Of ACM Conf. on Human factors in computing systems*, pp. 4039–4042 (2015)
32. Dou, Q., et al.: Webhetics: quantifying webpage aesthetics with deep learning. *Int. J. Hum. Comput. Stud.* 124, 56–66 (2018)
33. Lu, X., et al.: RAPID: rating pictorial aesthetics using deep learning. In: *Proc. Of ACM International Conference on Multimedia*, pp. 457–466 (2014)
34. Ren, J., et al.: Personalized image aesthetics. *Proc. of ICCV*, 638–647 (2017)
35. Suchecki, M., Trzcinski, T.: Understanding aesthetics in photography using deep convolutional neural networks. In: *Proc. Of Signal Processing algorithms, architectures, arrangements, and applications*, pp. 149–154 (2017)
36. Li, M., Lv, J., Tang, C.: Aesthetic assessment of paintings based on visual balance. *IET Image Process.* 13(14), 2821–2828 (2019)
37. Viswanatha, R., Mukherjee, S., Thakur, M.: Measuring photography aesthetics with deep CNNs. *IET Image Process.* 14(8), 1561–1570 (2020)
38. Bourguet, M.: Metrics-based evaluation of graphical user interface aesthetics: the segmentation problem. *Proc. of ACM International Conference*, 31–39 (2018)
39. Srinivas, K., Bhandari, A.: Low light image enhancement with adaptive sigmoid transfer function. *IET Image Process.* 14(4), 668–678 (2020)
40. Bakaev, M., Razumnikova, O.: What makes a UI simple? Difficulty and complexity in tasks engaging visual-spatial working memory. *Future Internet*. 13(1), 21 (2021)
41. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: *Proc. Of International Conference on Learning Representations*, pp. 1–14 (2014)
42. Xia, X., Xu, C., Nan, B.: Inception-v3 for flower classification. In: *Proc. Of the 2nd International Conference on Image*, pp. 783–787. *Vision and Computing* (2017)
43. Howard, A.G., et al.: MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, arXiv:1704.04861v1. pp. 1–9 (2017)
44. Tan, M., Le, Q.V.: EfficientNet: rethinking model scaling for convolutional neural networks. *Proc. of Int. Conf. Mach. Learn.*, 1–11 (2019)
45. Akiba, T., Suzuki, S., Fukuda, K.: Extremely large minibatch SGD: training ResNet-50 on ImageNet in 15 Minutes. arXiv:1711.04325v1 [cs.DC]. (2017)
46. Rezende, E., et al.: Malicious software classification using transfer learning of ResNet-50 Deep neural network. In: *Proc. Of The 16th*, pp. 1011–1014. *IEEE International Conference On Machine Learning And Applications* (2017)
47. Zoph, B., et al.: Learning transferable architectures for scalable image recognition, arXiv:1707.07012v4. pp. 1–14 (2018)
48. Zheng, X.S., et al.: Correlating low-level image statistics with users' rapid aesthetic and affective judgments of web pages. In: *Proc. Of International Conference on Human Factors in Computing Systems*, pp. 1–10 (2009)

How to cite this article: Xing, B., et al.: AI-driven user aesthetics preference prediction for UI layouts via deep convolutional neural networks. *Cogn. Comput. Syst.* 4(3), 250–264 (2022). <https://doi.org/10.1049/ccs2.12055>