Frontiers of Information Technology & Electronic Engineering
www.jzus.zju.edu.cn; engineering.cae.cn; www.springerlink.com
ISSN 2095-9184 (print); ISSN 2095-9230 (online)
E-mail: jzus@zju.edu.cn

Cantonese porcelain classification and image synthesis by ensemble learning and generative adversarial network*

Steven Szu-Chi CHEN[1], Hui CUI[1], Ming-han DU[1], Tie-ming FU[2], Xiao-hong SUN[2], Yi JI[†‡2], Henry DUH[1]

[1]*Department of Computer Science and Information Technology, La Trobe University, Melbourne 3086, Australia*

[2]*School of Art and Design, Guangdong University of Technology, Guangzhou 510006, China*

[†]E-mail: jiyi001@hotmail.com

Received Aug. 6, 2019; Revision accepted Nov. 18, 2019; Crosschecked Dec. 12, 2019

Abstract: Accurate recognition of modern and traditional porcelain styles is a challenging issue in Cantonese porcelain management due to the large variety and complex elements and patterns. We propose a hybrid system with porcelain style identification and image recreation modules. In the identification module, prediction of an unknown porcelain sample is obtained by logistic regression of ensembled neural networks of top-ranked design signatures, which are obtained by discriminative analysis and transformed features in principal components. The synthesis module is developed based on a conditional generative adversarial network, which enables users to provide a designed mask with porcelain elements to generate synthesized images of Cantonese porcelain. Experimental results of 603 Cantonese porcelain images demonstrate that the proposed model outperforms other methods relative to precision, recall, area under curve of receiver operating characteristic, and confusion matrix. Case studies on image creation indicate that the proposed system has the potential to engage the community in understanding Cantonese porcelain and promote this intangible cultural heritage.

Key words: Cantonese porcelain; Classification; Generative adversarial network; Creative arts
https://doi.org/10.1631/FITEE.1900399 **CLC number:** TP751

1 Introduction

Cultural heritage represents the way ancestors lived, and is considered a precious and irreplaceable resource that people and governments put forth great effort and expense (Lowenthal, 2005). Cultural heritage can be classified as intangible and tangible cultural heritages (Smith and Akagawa, 2008). Intangible cultural heritage represents knowledge and skills that are preserved and passed on through generations, including social practices, performing arts, and traditional craftsmanship (Kurin, 2004). Porcelain is one of the most valuable forms of Chinese intangible cultural heritage (Ji et al., 2019). Differing from other types of Chinese porcelain, Cantonese porcelain is glazed at low temperature and decorated by woven gold (China Intangible Cultural Heritage Network, 2008). However, the inheritance of Cantonese porcelain techniques is challenging. Younger generations are becoming less interested in such craftsmanship. Therefore, it is necessary to promote Cantonese porcelain craft to prevent its disappearance.

To facilitate the inheritance of Cantonese porcelain, the identification of porcelain categories is an essential and important step. The identification of porcelain styles is a challenging issue because recognition requires specific domain knowledge. Due to the large variety of wares, elements, and patterns in Cantonese porcelain, even for experienced craftsman, there can be inter- and intra-observer variations. For

‡ Corresponding author

* Project supported by the Guangzhou Science and Technology Project, China (No. 2018GZMZYB17)

ORCID: Steven Szu-Chi CHEN, http://orcid.org/0000-0001-6019-7034

example, there may be a vase with a modern pattern in the window, line, background, and picture elements, but the style of this vase is categorized as traditional. Even with the same patterns in the line element, a vase and a plate can be classified as different styles.

With the development of computer science techniques, computing models have attracted significant interest in classification and image synthesis. Computing models are efficient and effective for complex and non-linear relationships among features, which have been explored for the identification of creative work in the visual art domain. Feature extraction is a common method in the classification of paintings. Zujovic et al. (2009) proposed a simple and efficient feature extraction method for the classification of different painting styles. Neural networks play an important role in computational classification of artwork. For example, Lecoutre et al. (2017) evaluated various neural networks in the identification of artistic styles, and proposed a residual neural network with a retraining process for art style classification. Compared with western paintings, traditional Chinese paintings involve many diverse styles. Meng et al. (2018) trained several classical convolutional neural networks in three styles of traditional Chinese paintings and achieved a high prediction accuracy by editing the specific network. Another classifier for traditional Chinese paintings based on support vector machine (SVM) was proposed by Jiang et al. (2006). This classifier employed low-level features to achieve a high-level classification task. Bao et al. (2010) proposed an algorithm that can extract scripts from Chinese paintings to reveal the artistic conception of a painting. Note that this algorithm can also be used for identification and authentication purposes. However, the automated classification of Cantonese porcelain remains a challenging task due to the large class variations and immersive design elements between modern and traditional porcelains.

Generating synthesized images via image-to-image translations can be achieved by conventional purpose-specific machinery (Efros and Freeman, 2001; Hertzmann et al., 2001; Buades et al., 2005) and convolutional neural networks (CNNs). Although the learning process is automated, significant effort is required to design effective losses for a CNN to identify what must be minimized (Isola et al., 2017). It has been demonstrated that the generative adversarial networks (GANs) proposed by Goodfellow et al. (2014) can automatically learn a loss function that does not require manual adjustment, thus attracting interest in research and industry communities. For examples, GAN-based models have been used to colorize (Iizuka et al., 2016; Larsson et al., 2016; Zhang et al., 2016) and restyle (Zhu et al. 2017) images. For medical research and applications, given magnetic resonance images, a GAN-based model that uses ResNet as a generator and regular CNN as a discriminator was developed by Emami et al. (2018) to generate synthesized CT images. The HEMIGEN model proposed by Dirvanauskas et al. (2019) can generate generic image of cells from human embryo images during cell development. In addition, conditional GAN (cGAN) models can generate new images based on ambiguous and incomplete information (Li et al., 2019), and can remove the background from images (El Hattami et al., 2019) or videos (Chen, 2019). SIGAN, a GAN-based model proposed by Mao et al. (2018), can generate visually similar and semantically consistent images. Sketch-to-image models have been proposed to translate sketches to portraits, Pokémon (Isola et al., 2017), or colorful cartoon images (Liu et al., 2018). For example, the auto-painter model adds Wasserstein loss to assist supervised training to overcome model collapse and enable better convergence. With either a sketch or a sketch with color blocks, auto-painter can transform a sketch into colorful cartoon images (Liu et al., 2018).

To assist the management of Cantonese porcelain and engage the community in understanding and enjoying this cultural heritage, we propose a system with two modules for porcelain style identification and synthesis. Our first contribution is the classification module for extraction, representation, and explanation of underlying porcelain design signatures during data-driven identification. This is achieved using an ensemble model with principal component analysis (PCA) and a discriminative and redundancy quantization strategy. Another contribution is the Cantonese porcelain recreation module to engage the community in interaction and appreciation of this intangible cultural heritage. The cGAN-based module enables users to create an abstract image with desired objects of interest and design elements.

2 Proposed method

2.1 Data collection

We collected 603 images of Cantonese porcelain from the Lingnan Exhibition Hall, the Guangzhou Thirteen Hongs Museum, and the Guangzhou Liwan Museum at the Guangdong University of Technology. Each image is associated with a set of design elements and patterns (Fig. 1).

There are seven elements, i.e., style, ware, topic, line, background, window, and picture. Style is considered a target in the classification task, where there are 279 modern and 324 traditional style samples. The attributes of each element and the corresponding number of observed images are given in Table 1. The resolution of the images is 350×350 dpi, and the original size of the image is 6000×4000 pixels.

2.2 Framework of the proposed system

Framework of the proposed system with the Cantonese porcelain style identification and image recreation modules is shown in Fig. 2. Given images of Cantonese porcelain, the first module involves feature extraction and an ensemble model for classification. The second module takes a mask of porcelain elements designed by users as an input and generates a synthesized image of Cantonese porcelain style based on a cGAN model.

2.2.1 Porcelain style classification module

The Cantonese porcelain style classification and prediction module is illustrated in Fig. 3. Note that the elements of window, picture, line, and background have multiple attributes; thus, one hot encoding is

Table 1 Attributes of each element and the number of observed images of Cantonese porcelain

Element	Attribute								
Style	Modern (279)	Traditional (324)							
Ware	Bowl (20)	Cup (21)	Jar (40)	Vase (284)	Plate (78)	Tub (6)	Pot (44)	Painting (36)	Others (74)
Topic	Animal (160)	Flower and bird (8)	Flower (181)	Plant (64)	Contemporary (11)	Landscape (119)	Portrait (60)		
Line	Dogtooth (8)	Draghand (84)	Modern (217)	Pig nose (94)	Wishful-plantain (45)	Fisheye (4)	Topped (15)	Twisted-rope (8)	Reverse-emblem (4)
Background	Water-wave (3)	Triple-lined-coin (8)	Swastika (10)	Square (3)	Sharkskin (91)	Triple-lined (5)	Modern (67)	Pepper (14)	Crack-emblem (8)
Window	Begonia (45)	Modern (95)	Wishful-begonia (14)	Butterfly (7)	Historical (60)	Nail (19)	Wishful (3)	Pillow (46)	Round-corner (4)
Picture	Flower (8)	Modern (218)							

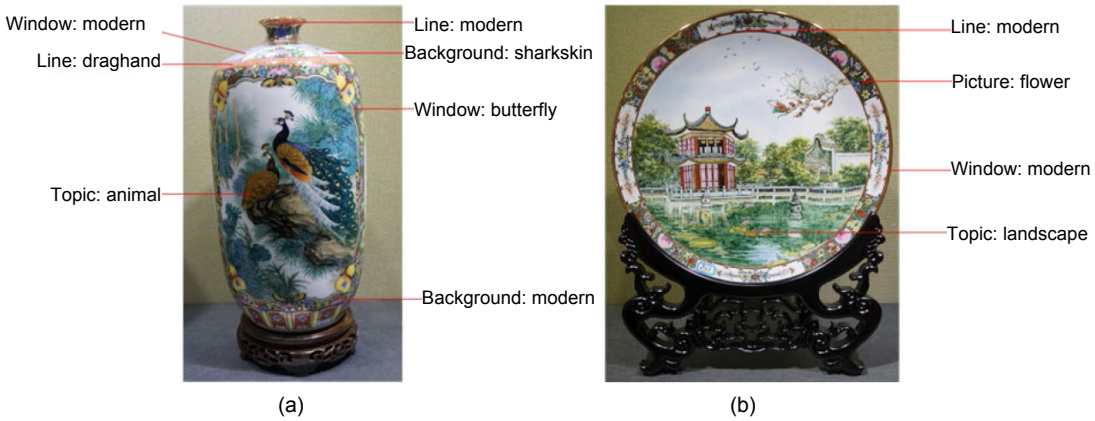

Fig. 1 Images of Cantonese porcelain with labeled design elements and patterns: (a) a vase; (b) a plate

performed to transform those elements to numeric features. In total, for each image, there are M=35 features. To reduce the dimensionality of features, we employ PCA and a discriminative and redundancy quantization strategy (feature selection (FS)) to transform the original feature space and select significant features to represent the characteristics of the image samples.

PCA for feature representation and dimensionality reduction is performed as follows: Let $\overline{\boldsymbol{F}}_{N\times M}$ be mean-subtracted features, where N is the number of samples and M is the number of features. Eigenvectors $\{\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \ldots, \boldsymbol{\lambda}_M\}$ are obtained and sorted in descending order by singular value decomposition of the covariance matrix $\overline{\boldsymbol{F}}_{N\times M}$. The corresponding eigenvectors $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \ldots, \boldsymbol{X}_M\}$ are obtained. Given p% information to be preserved, we obtain K principal components. Here, feature dimensionality is reduced from M to K by projecting features into a lower K-dimensional space constructed by the first K eigenvectors. The transformed features are obtained as projected weights $W=\{w_1, w_2, \ldots, w_K\}$ of the samples on each of the K principal components.

The FS strategy ranks the feature significance by calculating the information gain ratio (Quinlan, 1986), ReliefF (Kira and Rendell, 1992), χ^2 (Cochran, 1954),

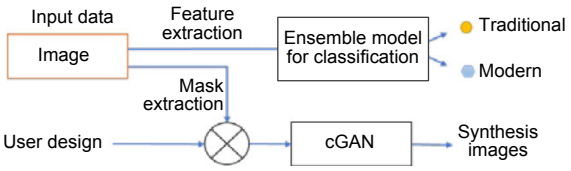

Fig. 2 Framework of the proposed system for Cantonese porcelain identification and image synthesis

and the fast correlation-based filter (FCBF) (Yu and Liu, 2003). The information gain ratio is the ratio of the information gain to the intrinsic value, which is used to reduce bias toward multi-valued features by considering the number and size of all the samples when selected a feature. The information gain ratio is defined as follows:

$$\text{gain.ratio}(\boldsymbol{F}, \boldsymbol{f}_{nm}) = \text{IG}/\text{IV}, \tag{1}$$

where n=1, 2, …, N, m=1, 2, …, M, $\boldsymbol{F}$ represents the feature matrix, IG the information gain for the m^{th} feature, and IV the intrinsic value. IG and IV are expressed as

$$\begin{aligned}\text{IG}(\boldsymbol{f}_{n^*}, \boldsymbol{f}_{*m}) = H(\boldsymbol{f}_{n^*}) - \sum_{v\in \boldsymbol{f}_{*m}} \Bigg(& \frac{|\{\boldsymbol{f}_{n^*} \in \boldsymbol{F} \mid \boldsymbol{f}_{nm} = v\}|}{|\boldsymbol{f}_{n^*}|} \\ & \cdot H(\{\boldsymbol{f}_{n^*} \in \boldsymbol{F} \mid \boldsymbol{f}_{nm} = v\}) \Bigg),\end{aligned} \tag{2}$$

$$\begin{aligned}\text{IV}(\boldsymbol{f}_{n^*}, \boldsymbol{f}_{*m}) = -\sum_{v\in \boldsymbol{f}_{*m}} \Bigg\{ & \frac{|\{\boldsymbol{f}_{n^*} \in \boldsymbol{F} \mid \boldsymbol{f}_{nm} = v\}|}{|\boldsymbol{f}_{n^*}|} \\ & \cdot \log_2 \left(\frac{|\{\boldsymbol{f}_{n^*} \in \boldsymbol{F} \mid \boldsymbol{f}_{nm} = v\}|}{|\boldsymbol{f}_{n^*}|} \right) \Bigg\},\end{aligned} \tag{3}$$

where $\boldsymbol{f}_{n^*}$ denotes all the features of the n^{th} sample, $\boldsymbol{f}_{*m}$ the m^{th} feature vector in $\boldsymbol{F}$, and H the entropy value.

ReliefF can differentiate data samples that are near to each other, and is defined as follows:

$$d(f_{nm}, f_{ym}) < d(f_{nm}, f_{zm}), \tag{4}$$

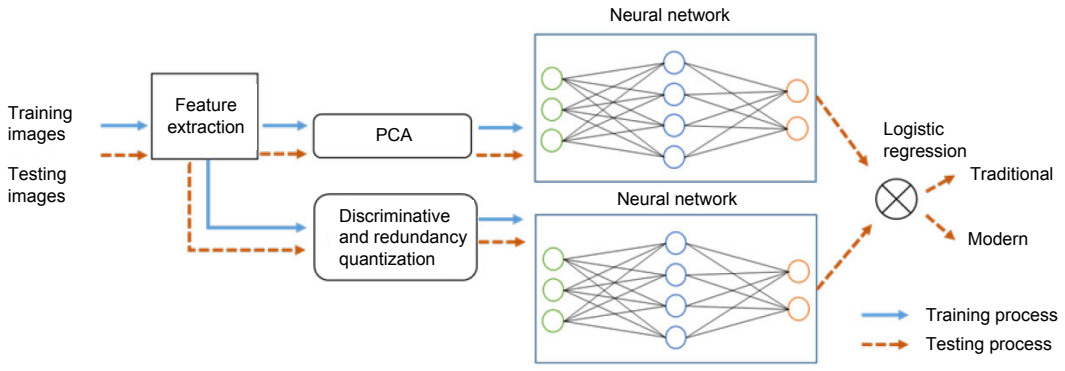

Fig. 3 The proposed ensemble model for classification of modern and traditional Cantonese porcelain

where y is the nearest neighbor of the n^{th} random sample within the same class, z a sample from another class, and d the Euclidean distance between two samples. Note that the feature weight of $\boldsymbol{f}_m$ increases if inequality (4) is satisfied (vice versa).

The χ^2 test is employed to determine whether there are significant differences between the expected and observed features in one class or more classes. Assuming the null hypothesis is correct as $r\to\infty$, the limiting distribution of χ^2 is defined as follows:

$$\chi^2=\sum_{i=1}^{C}\frac{(f_i-q_i)^2}{q_i}=\sum_{i=1}^{C}\frac{f_i^2}{q_i}-r, \tag{5}$$

where C denotes mutually exclusive classes with respective observed features f_i (i=1, 2, …, C), r the observation in a random sample from a population classified into c mutually exclusive classes, and p_i the probability that an observation will fall into the i^{th} class. Therefore, we have the expected numbers $q_i=rp_i$ for all i, where p_i and q_i are defined as follows:

$$\sum_{i=1}^{C}p_i=1, \tag{6}$$

$$\sum_{i=1}^{C}q_i=r\sum_{i=1}^{C}p_i=\sum_{i=1}^{C}f_i. \tag{7}$$

FCBF is a multivariate FS method that begins with a full set of features and uses a symmetrical uncertainty (SU). SU is normalized information theoretic measure that uses entropy H and conditional entropy values to calculate the dependencies of features and finds the best subset using a backward selection technique with a sequential search strategy. SU=1 indicates that one feature's value can be completely predicted by another feature, while SU=0 indicates that two features are completely independent. SU is defined as

$$\mathrm{SU}(f_{nm},f_{tm})=2\frac{H(f_{nm})-H(f_{nm}\mid f_{tm})}{H(f_{nm})+H(f_{tm})}, \tag{8}$$

where t, n, m indicate features, f_{nm} and f_{tm} are features of random samples, $H(f_{nm}|f_{tm})$ the conditional entropy, and $H(f_{nm})$ the entropy of f_{nm}. $H(f_{nm}|f_{tm})$ and $H(f_{nm})$ are defined as follows:

$$H(f_{nm}\mid f_{tm})=-\sum P(\boldsymbol{f}_t)\sum P(\boldsymbol{f}_n\mid\boldsymbol{f}_t)\log_2 P(\boldsymbol{f}_n\boldsymbol{f}_t). \tag{9}$$

$$H(f_{nm})=-\sum P(\boldsymbol{f}_n)\log_2 P(\boldsymbol{f}_n), \tag{10}$$

where $P(\boldsymbol{f}_n)$ is the probability of $\boldsymbol{f}_n$.

The importance and discriminative of each feature are obtained by the fusion of normalized ranking scores of the information gain.ratio, ReliefF, FCBF, and χ^2.

Given W and top-ranked features in the training process, two neural networks are trained. To better associate the classification results of the two models, the prediction of an unknown porcelain sample is obtained by an ensemble model, in which logistic regression is employed to integrate the results.

2.2.2 Cantonese porcelain recreation module

The second module for porcelain image recreation is developed based on a cGAN (Isola et al., 2017). In the proposed method, we extract masks of porcelain elements (Fig. 4). In total, there are 196 images and 23 types of objects, such as crane, fruit tree, frame, boat, building, stone, bamboo, birds, chicken, dog, cat, butterfly, and fisherman.

This training procedure is shown in Fig. 5. Given image I and corresponding mask S, discriminator D learns to classify a fake image created by generator G and an input $\{I, S\}$, while G attempts to generate fake images to fool discriminator D. Unlike traditional GANs, in the cGAN, both G and D observe the input mask. The objective of a cGAN is defined as follows:

$$G^*=\arg\min_G\max_D L_{\mathrm{cGAN}}(G,D)+\lambda L_{L_1}(G), \tag{11}$$

where $\min_G$ and $\max_D$ indicate that generator G attempts to minimize the objective function and discriminator D attempts to maximize the objective function, respectively. L_{L_1} is the loss calculated by distance L_1. λ is a parameter to adjust the weight of L_{L_1}, indicating that G attempts to fool D (in a conventional GAN), and at the same time, needs to be as similar with the ground truth as possible by distance L_1. L_{cGAN} is the loss of the cGAN and is defined as follows:

$$\begin{aligned}L_{\mathrm{cGAN}}(G,D)=&E_{I,O}[\log D(I,O)]\\&+E_{I,U}\{\log(1-D[I,G(I,\boldsymbol{U})])\},\end{aligned} \tag{12}$$

where E is the expected value in probability theory, O is the output image and $\boldsymbol{U}$ is a random noise vector. A trained cGAN model can be obtained when G^* is converged or the pre-defined number of epochs is reached.

Given a trained cGAN model, a synthesized image can be obtained by providing a creative mask or an abstract image designed by the user as an input.

2.3 Methods for comparison

To evaluate the proposed model and investigate the contributions of PCA, the proposed feature ranking strategy, and the ensemble model, we compare the proposed model (Stack-PCA-FS-NN) with a neural network with only PCA (PCA-NN), a neural network with only FS (FS-NN), and an ensemble model of two neural networks without PCA and FS (Stack-NN) for feature dimension reduction.

2.4 Evaluation methods and parameter settings

2.4.1 Evaluation metrics

All classification models are evaluated by 10-fold cross-validation. The average results over the testing dataset are obtained for performance comparison. The prediction performance is evaluated according to the area under curve of receiver operating characteristic (AUC ROC), precision, recall, and confusion matrix. Here, precision represents the ratio of the number of correctly predicted positive samples to the number of all the samples predicted as positive, expressed as

$$\text{Precison}=\frac{\text{TP}}{\text{TP+FP}}, \tag{13}$$

where TP is true positive (the number of correctly

Fig. 4 Cantonese porcelain images and masks with labeled objects

Tree, house, bridge, boat, birds, fisherman, and bamboo are shown in dark green, light blue, brown, blue, orange, pink, and bright green, respectively. References to color refer to the online version of this figure

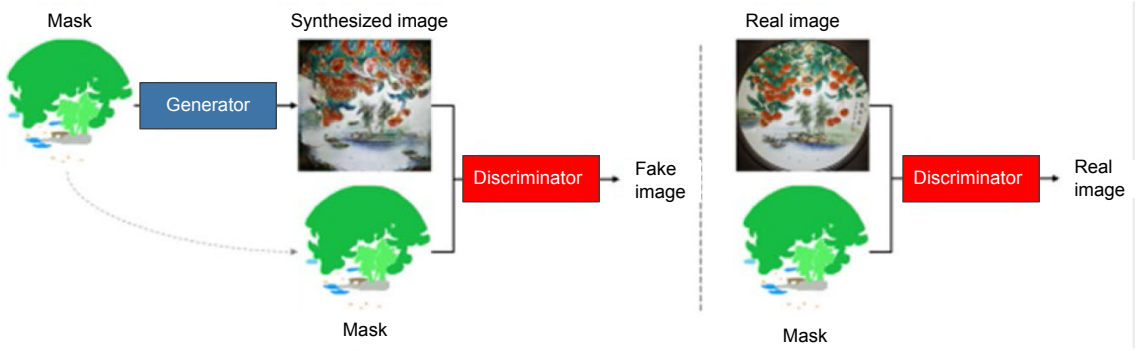

Fig. 5 Training process of the cGAN-based model for mask-to-image generation

Given mask S, discriminator D learns to distinguish fake set {synthesized image, mask} and real set {real image, mask}. Generator G learns to synthesize images to deceive discriminator D. cGAN learns a loss according to the results of discriminator D, while simultaneously training generator G to minimize the loss

identified positive samples), and FP is false positive (the number of falsely identified negative samples). Recall is the ratio of the number of correctly predicted positive samples to the number of all the positive samples, expressed as

$$\text{Recall}=\frac{\text{TP}}{\text{TP+FN}}, \tag{14}$$

where FN is false negative (the number of falsely identified positive samples). F_1=2×Precision×Recall/(Precision+Recall), compromising that precision and recall are calculated.

The AUC ROC is defined by the false positive rate (FPR) and true positive rate (TPR) as the x and y axes (illustrating relative tradeoffs between TP and FP). TPR and FPR are defined as follows:

$$\text{TPR}=\text{Recall}=\frac{\text{TP}}{\text{TP+FN}}, \tag{15}$$

$$\text{FPR}=\frac{\text{FP}}{\text{FP+TN}}, \tag{16}$$

where TN is true negative (the number of correctly predicted negative samples). Note that the higher the AUC ROC values, the better the classification performance.

2.4.2 Parameter settings

In the first classification module, ReLu is used for activation in the neural network, where the number of neurons is set to 100. Here, the maximum number of iterations is 200. Note that the neural network settings are the same for all the methods compared in the ablation study, to evaluate the contributions of the ensemble structure, PCA, and the feature ranking strategy to classification performance. Thus, we do not conduct automated selection of the best parameters for the networks in the original design of the system. Note that an automated parameter selection mechanism (Połap et al., 2018) can be used for a large dataset with complicated design elements and features.

K is set to 10 in PCA, and the top-12 ranked features are selected in the feature ranking strategy. The effects of the parameter settings are discussed in Section 3.

In the second cGAN-based module, the parameters are the same as those in Isola et al. (2017), in which the epoch was set to 200 and λ was set to 100.

3 Results and discussion

3.1 Classification results

Classification results obtained by the proposed Stack-PCA-FS-NN model and all other models are shown in Table 2. The AUC ROC results are shown in Fig. 6. Overall, the ensemble models outperformed other models in terms of AUC, precision, and recall. PCA-NN improved the performance in terms of AUC when compared with FS-NN with slightly increased precision and recall. The ensemble model with reduced and transformed features by FS and PCA outperformed the Stack-NN, and the proposed model achieved the best performance with AUC ROC of 0.944, precision of 0.87, and recall of 0.872.

Table 2 Classification performances of the proposed model and other methods with and without PCA, discriminate feature selection (FS), and ensemble strategy

Model	AUC ROC	Precision (mean±std)	Recall (mean±std)	F_1
FS-NN	0.920	0.832±0.006	0.828±0.007	0.830
PCA-NN	0.936	0.840±0.010	0.833±0.007	0.837
Stack-NN	0.906	0.851±0.010	0.839±0.008	0.845
Stack-PCA-FS-NN	0.944	0.87±0.005	0.872±0.006	0.872

Confusion matrices are given in Fig. 7. The NN with features transformed by PCA outperformed the NN with top-ranked features transformed by FS when predicting samples of the traditional class, while FS-NN obtained better classification results over samples from the modern class than the PCA-NN method. The ensemble model without feature dimensionality reduction achieved comparable performance over two classes. The proposed Stack-PCA-FS-NN significantly improved classification performance against the traditional class.

The effects of the number of principal components in PCA are investigated by varying K from 7 to 20. The AUC ROC, recall, and the corresponding variance covered by PCA are shown in Fig. 8. As can be seen, both AUC and recall reached an initial peak value when the number of principal

components is 10. Thus, K is set to 10, when the variance covered is 54%.

We investigate the original distributions of the elements, patterns, and features in traditional and modern classes using a tree structure. As shown in Fig. 9, among all 603 cases, 324 cases are from the traditional class. If the target is a bowl, painting, or pot, 92% of the porcelains are from the modern class. For a cup, jar, plate, tub, or vase, and other objects, 62.8% of the porcelains are from the traditional class. Among those traditional porcelains, if modern patterns are present in the image, 82.5% of them are traditional porcelain (the rest are modern). If the picture pattern is not modern, it is difficult to identify the type of the class, because the probability that the item is traditional is 51.5%, which is almost the same as that of modern. Among the modern samples, if the target is a cup, jar, plate, or tub, 86.8% of them are traditional. As a result, it is challenging to model the relationships between features and the target. A quantitative evaluation of the discrimination and redundancy of features would contribute to identifying significant features.

The top-35 ranked features identified by the proposed discriminative and redundancy quantification strategy are shown in Fig. 10. As can be seen, the most informative features include a modern pattern in the picture element, a draghand pattern in the line element, the number of different patterns in the window element, sharkskin in background, etc. Among the top-12 ranked features, the number of different patterns in various elements is

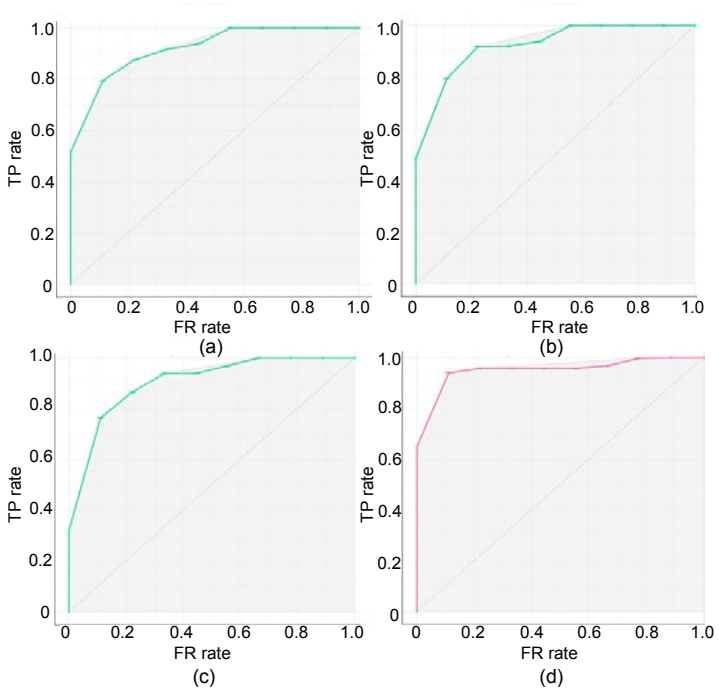

Fig. 6 AUC ROC results of FS-NN (a), PCA-NN (b), Stack-NN (c), and Stack-PCA-FS-NN (d)

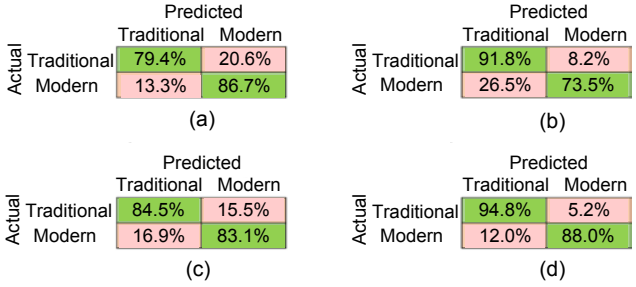

Fig. 7 Confusion matrices of FS-NN (a), PCA-NN (b), Stack-NN (c), and Stack-PCA-FS-NN (d)

four, and the number of patterns in the line element is three. As demonstrated by the confusion matrices shown in Fig. 7, the transformed feature obtained by PCA contributes to the classification of traditional class, while the proposed feature ranking strategy works better with modern class data. We also investigate the effects of the top-ranked features on classification. The AUC and recall are shown in Fig. 11 when the top-k ranked features (k=[7,19]) are used in the experiments. As shown in Fig. 11, both AUC and recall reached the local minimum when the number of selected top features is 10. In addition, performance improves, and reaches the local maximum when top-12 ranked features are used. The classification results in terms of AUC and recall change slightly when the number of top features is larger than 14.

3.2 Synthesis results

The loss over time in the training process is shown in Fig. 12. It take 22 h to train a conditional GAN model using only a CPU.

Three examples of synthesized images and the

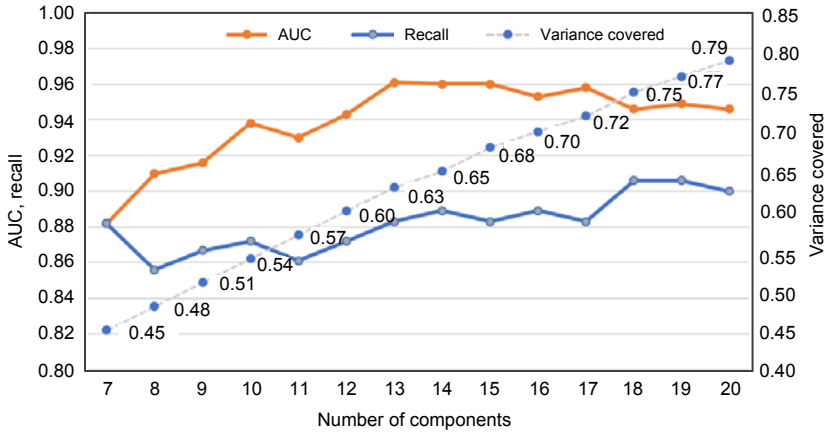

Fig. 8 Effects of the number of components and corresponding variance covered in PCA on AUC and recall of the proposed classification model

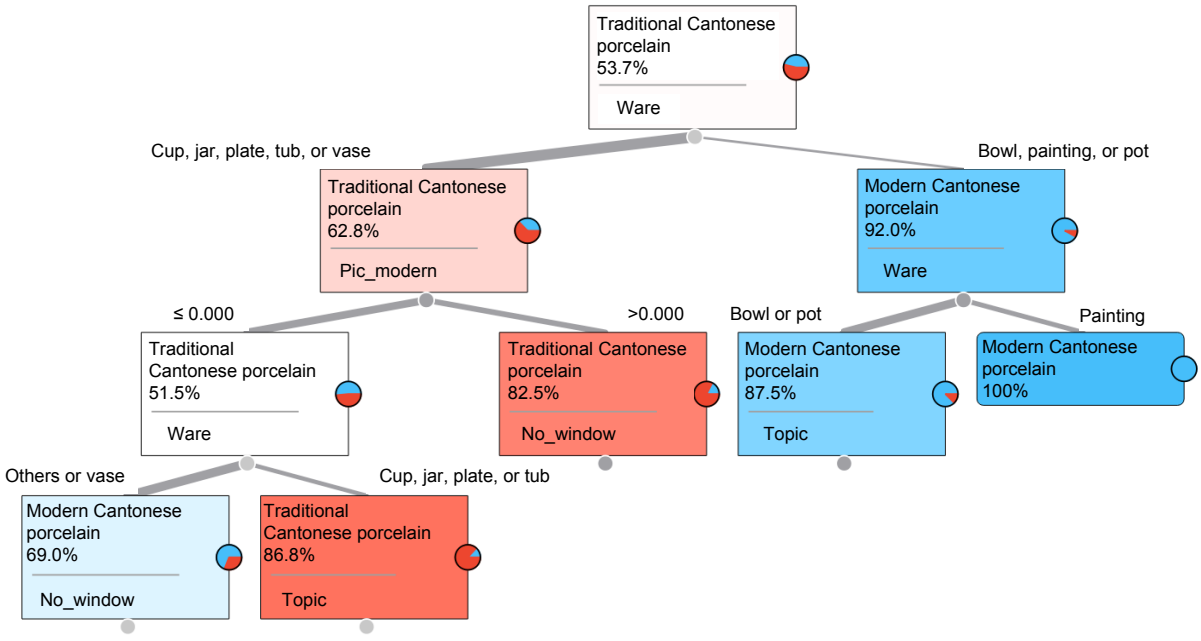

Fig. 9 Visualization of decision-making schema with feature importance in a tree structure

Traditional class is in red, and the modern class is in blue. References to color refer to the online version of this figure

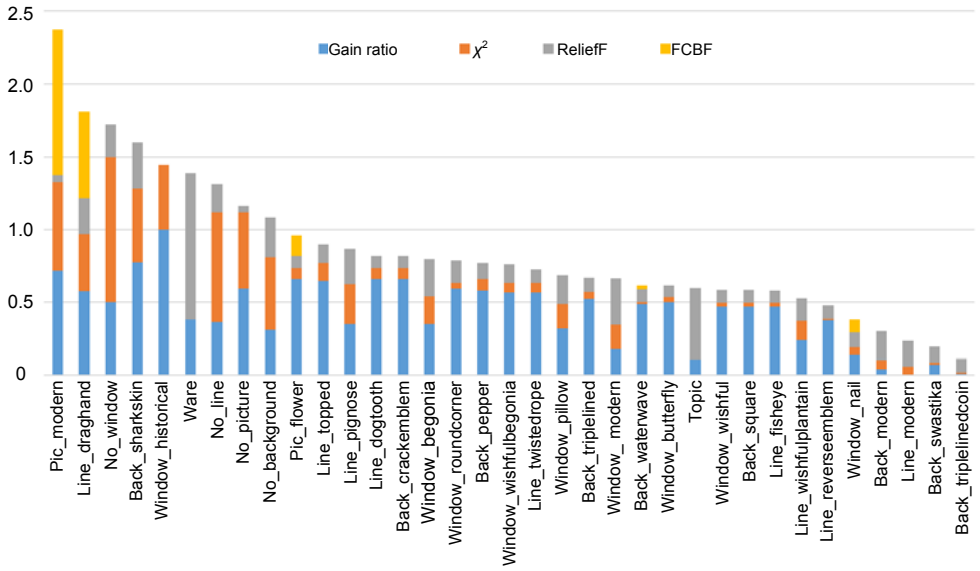

Fig. 10 Feature ranking of four metrics for the proposed discriminative and redundancy quantification strategy
Among the top-35 features, seven are line features and eight are background features. References to color refer to the online version of this figure

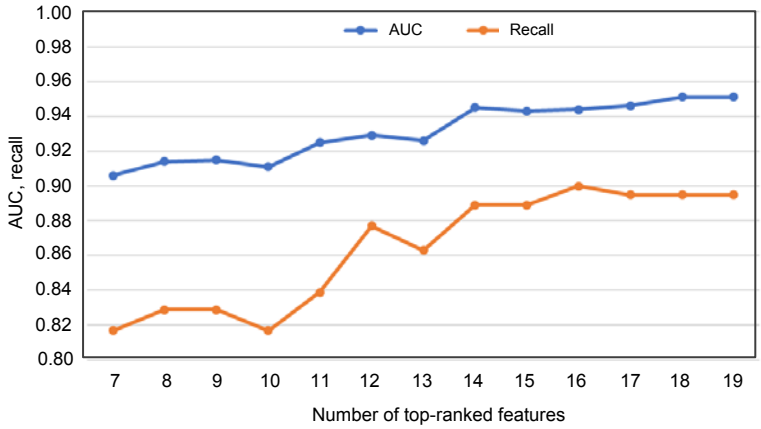

Fig. 11 Effects of the number of top-ranked features on classification results

corresponding user input are shown in Fig. 13. Given a simple abstract image with only a fruit tree painting, the synthesized image is shown in Fig. 13a. After adding drawings of cranes and a picture frame, a synthesized image with cranes is produced (Fig. 13b). A scenery synthesis (Fig. 13c) was obtained when the user provides a more complicated drawing with mixed fruit trees, boats, bamboo, stones, birds, and bridges. As demonstrated by Fig. 13, even with simple abstract drawings in the images, the proposed system can generate Cantonese porcelain style images and fill in the background based on the learning from real porcelain images in the training dataset.

In the current system, there are 23 different types of objects, e.g., crane, fruit tree, frame, boat, building, stone, bamboo, birds, chicken, dog, cat, butterfly, and fisherman. The user has the flexibility to select one or more desired object masks for drawing; however, this is limited to those that exist in system. Thus, the validity of the proposed system for recreating Cantonese style images is proved by 23 different objects, 7 design elements, and 35 features. One of the reasons that might cause ambiguity of the synthesis images is that objects in Cantonese porcelain are relatively limited compared with those in natural images. In the future, we plan to implement transfer learning from a larger

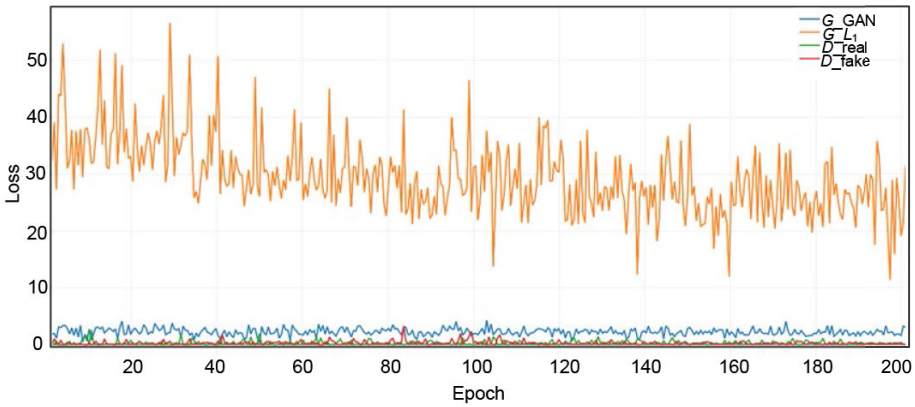

Fig. 12 Loss over time in the training process of cGAN model (*G* is the generator and *D* is the discriminator). References to color refer to the online version of this figure)

Fig. 13 Synthesized images of Cantonese porcelain produced by the proposed system given simple (a) and mixed (b and c) abstract drawings

public dataset, e.g., COCO (Lin et al., 2014), with a broader range of objects. In addition, we plan to employ automated parameter searching methods, e.g., a multi-threaded learning control mechanism (Połap et al., 2018), for larger dataset with more complicated design elements and features.

4 Conclusions

In this stufy, we have proposed a system for interpretable Cantonese porcelain style classification. Results of an ablation study demonstrated that ensembled neural networks of PCA and discriminative features outperformed other methods regarding the classification performance. The proposed system provided a module for synthesizing Cantonese porcelain style images from user designed abstract images of Cantonese porcelain elements. Case studies on synthesized images of Cantonese porcelain using 23 objects demonstrated that the proposed system had the potential to engage the community in enjoying and protecting such intangible cultural heritage.

Compliance with ethics guidelines

Steven Szu-Chi CHEN, Hui CUI, Ming-han DU, Tie-ming FU, Xiao-hong SUN, Yi JI, and Henry DUH declare that they have no conflict of interest.

References

Bao H, Liang Y, Liu HZ, et al., 2010. A novel algorithm for extraction of the scripts part in traditional Chinese painting images. Proc 2nd Int Conf on Software Technology and Engineering, p.V2-26-V2-30. https://doi.org/10.1109/ICSTE.2010.5608756

Buades A, Coll B, Morel JM, 2005. A non-local algorithm for image denoising. IEEE Computer Society Conf on Computer Vision and Pattern Recognition, p.60-65. https://doi.org/10.1109/CVPR.2005.38

Chen KH, 2019. Image Operations with cGAN. http://www.k4ai.com/imageops/index.html

China Intangible Cultural Heritage Network, 2008. Cantonese Porcelain Inheritance Project. http://www.ihchina.cn/project_details/14453/ [Accessed on July 16, 2019] (in Chinese).

Cochran WG, 1954. Some methods for strengthening the common χ^2 tests. *Int Biom Soc*, 10(4):417-451. https://doi.org/10.2307/3001616

Dirvanauskas D, Maskeliūnas R, Raudonis V, et al., 2019. HEMIGEN: human embryo image generator based on generative adversarial networks. *Sensors*, 19(16):3578. https://doi.org/10.3390/s19163578

Efros AA, Freeman WT, 2001. Image quilting for texture synthesis and transfer. Proc 28th Annual Conf on Computer Graphics and Interactive Techniques, p.341-346. https://doi.org/10.1145/383259.383296

El Hattami A, Pierre-Doray É, Barsalou Y, 2019. Background removal using U-net, GAN and image matting. https://github.com/eti-p-doray/unet-gan-matting [Accessed on July 14, 2019].

Emami H, Dong M, Nejad-Davarani SP, et al., 2018. Generating synthetic CTs from magnetic resonance images using generative adversarial networks. *Med Phys*, 45(8): 3627-3636. https://doi.org/10.1002/mp.13047

Goodfellow IJ, Pouget-Abadie J, Mirza M, et al., 2014. Generative adversarial nets. Proc 27th Int Conf on Neural Information Processing Systems, p.2672-2680.

Hertzmann A, Jacobs CE, Oliver N, et al., 2001. Image analogies. Proc 28th Annual Conf on Computer Graphics and Interactive Techniques, p.327-340. https://doi.org/10.1145/383259.383295

Iizuka S, Simo-Serra E, Ishikawa H, 2016. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Trans Graph*, 35(4), Article 110. https://doi.org/10.1145/2897824.2925974

Isola P, Zhu JY, Zhou TH, et al., 2017. Image-to-image translation with conditional adversarial networks. IEEE Conf on Computer Vision and Pattern Recognition, p.1125-1134. https://doi.org/10.1109/CVPR.2017.632

Ji Y, Tan P, Chen SC, et al., 2019. Kansei engineering for E-commerce Cantonese porcelain selection in China. 21st Int Conf on Human-Computer Interaction, p.463-474. https://doi.org/10.1007/978-3-030-22646-6_34

Jiang SQ, Huang QM, Ye QX, et al., 2006. An effective method to detect and categorize digitized traditional Chinese paintings. *Patt Recogn Lett*, 27(7):734-746. https://doi.org/10.1016/j.patrec.2005.10.017

Kira K, Rendell LA, 1992. A practical approach to feature selection. Machine Learning Proc, p.249-256. https://doi.org/10.1016/B978-1-55860-247-2.50037-1

Kurin R, 2004. Safeguarding intangible cultural heritage in the 2003 UNESCO convention: a critical appraisal. *Museum Int*, 56(1-2):66-77. https://doi.org/10.1111/j.1350-0775.2004.00459.x

Larsson G, Maire M, Shakhnarovich G, 2016. Learning representations for automatic colorization. Proc 14th European Conf on Computer Vision, p.577-593. https://doi.org/10.1007/978-3-319-46493-0_35

Lecoutre A, Négrevergne B, Yger F, 2017. Recognizing art style automatically in painting with deep learning. Proc 9th Asian Conf on Machine Learning, p.327-342.

Li WB, Zhang PC, Zhang L, et al., 2019. Object-driven text-to-image synthesis via adversarial training. https://arxiv.org/abs/1902.10740

Lin TY, Maire M, Belongie S, et al., 2014. Microsoft COCO: common objects in context. Proc 13th European Conf on Computer Vision, p.740-755. https://doi.org/10.1007/978-3-319-10602-1_48

Liu YF, Qin ZC, Wan T, et al., 2018. Auto-painter: cartoon image generation from sketch by using conditional Wasserstein generative adversarial networks. *Neurocomputing*, 311:78-87. https://doi.org/10.1016/j.neucom.2018.05.045

Lowenthal D, 2005. Natural and cultural heritage. *Int J Herit Stud*, 11(1):81-92. https://doi.org/10.1080/13527250500037088

Mao XF, Wang SH, Zheng LY, et al., 2018. Semantic invariant cross-domain image generation with generative adversarial networks. *Neurocomputing*, 293:55-63. https://doi.org/10.1016/j.neucom.2018.02.092

Meng QY, Zhang HH, Zhou MQ, et al., 2018. The classification of traditional Chinese painting based on CNN. Proc 4th Int Conf on Cloud Computing and Security, p.232-241. https://doi.org/10.1007/978-3-030-00009-7_22

Połap D, Woźniak M, Wei W, et al., 2018. Multi-threaded learning control mechanism for neural networks. *Fut Gener Comput Syst*, 87:16-34. https://doi.org/10.1016/j.future.2018.04.050

Quinlan JR, 1986. Induction of decision trees. *Mach Learn*, 1(1):81-106. https://doi.org/10.1007/BF00116251

Smith L, Akagawa N, 2008. Intangible Heritage. Routledge, London, UK.

Yu L, Liu H, 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. Proc 20th Int Conf on Machine Learning, p.856-863.

Zhang R, Isola P, Efros AA, 2016. Colorful image colorization. Proc 14th European Conf on Computer Vision, p.649-666. https://doi.org/10.1007/978-3-319-46487-9_40

Zhu JY, Park T, Isola P, et al., 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. IEEE Int Conf on Computer Vision, p.2223-2232. https://doi.org/10.1109/ICCV.2017.244

Zujovic J, Gandy L, Friedman S, et al., 2009. Classifying paintings by artistic genre: an analysis of features & classifiers. IEEE Int Workshop on Multimedia Signal Processing, p.1-5. https://doi.org/10.1109/MMSP.2009.5293271