PROCEEDINGS OF SPIE

SPIEDigitalLibrary.org/conference-proceedings-of-spie

Deep learning for the extraction of sketches from spectral images of historical paintings

Zhang, Qunxi, Cui, Shan, Liu, Lu, Wang, Jiaxin, Wang, Jun, et al.

Qunxi Zhang, Shan Cui, Lu Liu, Jiaxin Wang, Jun Wang, Erlei Zhang, Jinye Peng, Sotiria Kogou, Florence Liggins, Haida Liang, "Deep learning for the extraction of sketches from spectral images of historical paintings," Proc. SPIE 11784, Optics for Arts, Architecture, and Archaeology VIII, 1178407 (8 July 2021); doi: 10.1117/12.2593680

SPIE.

Event: SPIE Optical Metrology, 2021, Online Only

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021 Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Deep learning for the extraction of sketches from spectral images of historical paintings

Qunxi Zhang[a,d], Shan Cui[b], Lu Liu[b*,d], Jiaxin Wang[b], Jun Wang[b*,d], Erlei Zhang[e], Jinye Peng[b,d], Sotiria Kogou[c], Florence Liggins[c],Haida Liang[c]
[a]Shaanxi History Museum, no.91, Xiaozhai East Road, Yanta District, Xi'an, Shaanxi Province, P.R. China; [b]School of Information Science and Technology, Northwest university, no.1, Xuefu Avenue, Chang'an District, Xi'an, Shaanxi Province, P.R. China; [c]School of Science and Technology, Nottingham Trent University, Nottingham NG11 8NS, UK; [d]Shaanxi Province Silk Road Digital Protection and Inheritance of Cultural Heritage Collaborative Innovation Center, no.1, Xuefu Avenue, Chang'an District, Xi'an, Shaanxi Province, P.R. China; [e]Northwest A&F University, No.3, Taicheng Road, Yangling, Shaanxi Province, P.R. China

ABSTRACT

Sketch extraction is of great value for historians to copy and study historical painting styles. However, most of the existing sketch extraction methods can successfully perform extraction only if the sketches are well preserved, but for paintings with severe conservation issues, the extraction methods need to be improved. Therefore, we propose a sketch extraction method using spectral imaging and deep learning. Firstly, the spectral image data is collected and the bands sensitive to the sketches are extracted by using the prior knowledge of the sketches (e.g. near infrared bands will be chosen if the sketches are made of carbon ink). A publicly available image dataset of natural scenes is used to pre-train the bi-directional cascade network (BDCN). The network parameters in the model are then fine-tuned by using the sketches drawn by experts based on images of painted cultural objects, so as to solve the problem of insufficient sketch dataset of painted cultural objects and enhance the generalization ability of the model. Finally, the U-Net is used to further suppress unwanted information, to make the sketch clearer. Experimental results show that the proposed method can extract clear sketches even with faded paintings and the presence of unwanted information or instrumental noise. It is superior to the other six advanced extraction methods in visual and objective comparison. The proposed deep learning method is also compared with an unsupervised clustering method using Self-Organising Map (SOM) which is a 'shallow learning' method where pixels of similar spectra are grouped into clusters without the need for data labeling by experts.

Keywords: Sketch extraction; Relics digital protection and inheritance; Spectral images; Spectral imaging; Edge detection; Deep learning; Mural

1. INTRODUCTION

As an ancient civilization in the world, China has a large number of painted cultural relics with high artistic value, which are of great significance for our study and inheritance of history and culture. However, due to environmental changes and man-made destruction, many existing painted cultural relics have been damaged. Painted cultural objects are some of the most fragile amongst various other cultural relics. In order to ensure the long-term preservation of painted cultural relics, obtaining sketches is a key step in protecting and restoring them and it is one of the key information captured in a traditional conservation and archaeological survey report.

The traditional methods of sketch extraction are mostly manually drawn. These methods are not only time-consuming and labor-intensive, but also affected by external factors such as the drawing skills of copyists. Its accuracy and efficiency cannot be guaranteed. Moreover, manual hand-drawing can only be seen with naked eyes, which ignores the information of many painted cultural relics outside the visible light range. Therefore, they are unreliable methods to obtain the sketches of painted cultural relics.

*liulu@nwu.edu.cn; phone +86 17792220909; fax 88308995
*jwang@nwu.edu.cn; phone +86 15591807509; fax 88308995

Optics for Arts, Architecture, and Archaeology VIII, edited by Haida Liang, Roger Groves, Proc. of SPIE
Vol. 11784, 1178407 · © 2021 SPIE · CCC code: 0277-786X/21/$21 · doi: 10.1117/12.2593680

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

In recent years, a series of sketch extraction methods have been developed based on edge detection[1-5]. These methods can be grossly divided into two categories: traditional approaches[7-12] and deep learning-based approaches[13-20]. Traditional edge operators often make use of the underlying features to detect the edge. In 1963, Lawrence Roberts proposed the first edge detection operator—Robert. The Robert edge operator is a 2 * 2 template that takes the difference between two adjacent pixels in the diagonal direction. Later, the Sobel operator[3], the Prewitt operator6 and the widely used Canny operator[10] have been proposed successively. Since they are calculated only based on gradient and susceptible to noise, the lines without obvious gradient change will be lost, making the extracted sketch incoherent. After that, P. Arbelaez et al. proposed a method of artificially designed features based on information theory - gPbowl-ucm algorithm[11], and Dollar P et al. proposed Fast Edge Detection Using Structured Forests (SE) algorithm[12]. Although the edge detection methods using low-level features have made great progress, they still have limitations. Semantic information is difficult to capture using low-level cues.

With the development of deep learning technology, especially the emergence of Convolutional Neural Network (CNN), a large number of edge detection methods based on deep learning have emerged. CNN is constructed by imitating the biological visual perception mechanism. Its convolutional kernel parameter sharing in hidden layers and the sparsity of interlayer connections enable the network to have a stable effect on lattice features with a small amount of computation. In 2015, Xie et al.[13] proposed Holistically-Nested Edge Detection (HED) to detect and extract edges of natural images in a nested way. In 2017, Liu et al.[14] proposed Richer Convolutional Features for Edge Detection (RCF), which improved the edge detection effect on the basis of HED method by using a multi-scale strategy. In 2019, Fu et al.[15] proposed a segmentation network for salient target detection. He et al.[17] proposed a Bi-Directional Cascade Network (BDCN) and realized the fusion of multi-scale features. Therefore, the accuracy of the research method for edge detection of natural images has been significantly improved. However, for the painted cultural relics, due to the limited amount of data and the imperfect condition of the paintings due to deterioration and other issues, the CNN-based algorithms cannot achieve the ideal result.

In this work, we have proposed a relic sketch extraction framework based on spectral images and deep learning. Firstly, the spectral data cubes are preprocessed, and BDCN is used to detect from different scales to perform preliminary sketch extraction. Then the preliminary extraction results are put into the U-Net for refinement and noise reduction. Finally, a clear and accurate sketch of painted cultural relics is obtained.

Contributions of this paper can be summarized as follows:

- We propose a new framework for relic sketch extraction based on spectral imaging and deep learning. For spectral images with damaged, degraded and complex backgrounds, ideal sketches can also be extracted.
- We use a large number of natural images to pre-train the BDCN network, and a small amount of cultural relic data to fine-tune the model, which solves the problem of insufficient image data of painted cultural relics.
- We use U-Net to refine the sketches extracted, reduce the ambiguity and noise, so that the structural integrity and accuracy of the extracted sketches are greatly improved.

The structure of this paper is organized as follows. The related works are introduced in Section 2. The structures of our proposed method are described in Section 3. The experimental results are shown in Section 4. The conclusion of the full text is summarized in Section 5.

2. RELATED WORKS

2.1 Spectral imaging

Spectral imaging typically collects reflectance spectra from thousands to millions of spatial pixels within a given field of view. Compared with ordinary images, spectral images not only contain spectral information but also combine spectral with spatial information. In recent years, the rapid development of spectral imaging in terrestrial applications has also influenced the field of cultural heritage restoration[32-37].

The analysis and collection of the mural information was obtained by a commercial grating-based hyperspectral imaging system, which has 128 spectral channels spanning the wavelength range from 377nm to 1037nm. The scanning was achieved by non-contact ‘push-broom’ scheme. It was non-invasive and non-contact while allowing the extraction of

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

information that cannot be observed by the naked eye. Moreover, most paints are more transparent to near-infrared bands, which means that the hyperspectral imager can reveal the original sketches of the mural.

2.2 Bi-Directional cascade network (BDCN)

BDCN[17] designs a new lightweight CNN for edge detection and a Scale Enhancement Module (SEM)[21] composed of multiple parallel convolutions with different dilation rates for multi-scale learning. Compared with the previous deep-level networks like ResNet50[23], or image pyramid methods, BDCN does not significantly increase the parameters, nor does it cause redundant computing, which makes good use of shallow CNN to mine multi-scale clues. BDCN also designs a bi-directional cascade structure so that each layer can learn the characteristics of a specific scale, and the specific layer supervision of each layer is inferred by the bi-directional cascade structure itself. In this way, each layer is focused on a specific scale, so that training can be carried out more effectively. However, BDCN needs a lot of data support, which is a great challenge for the limited cultural relic data. BDCN is also vulnerable to unwanted information or instrumental effects. Therefore, it is difficult to achieve the ideal result for damaged or blurred cultural relic images.

2.3 U-Net

U-Net is a deep learning network for semantic segmentation based on full convolution network, which is widely used in medical image segmentation[22]. The network is composed of a group of symmetrical encoders and decoders and does not contain the full connection layer. It is named U-Net because its structure model is like the letter U, which is an improvement of fully convolutional network. The encoders use the convolution layers and pooling layers for down-sampling and feature extraction, while the decoders use the convolution layers and pooling layers for up-sampling. The down sampled and up sampled feature maps of the same dimension are spliced through skip connections, so that the high-level and low-level features can be better fused to make multi-scale prediction. Finally, the predictions with the same size as the original images are obtained. U-Net can be applied to the application scenarios with a small number of samples. In addition, U-Net is trained fast and has a low cost, so it can be easily applied to more fields, such as image denoising[24 25] or segmentation[26-28].

3. METHODOLOGY

The proposed sketch extraction framework is shown in Fig. 1. Firstly, we calculate the average of the far red to near infrared bands of the spectral images that has the best signal to noise (650nm to 900nm), and pre-train the BDCN with the publicly available image dataset BSDS500[29] of natural scenes. Next, the images of the paintings are used for fine-tuning of the model. The preliminary results are then fed into the U-Net as input for further refinement and denoising. The ideal sketches of the painted cultural relics are finally obtained.

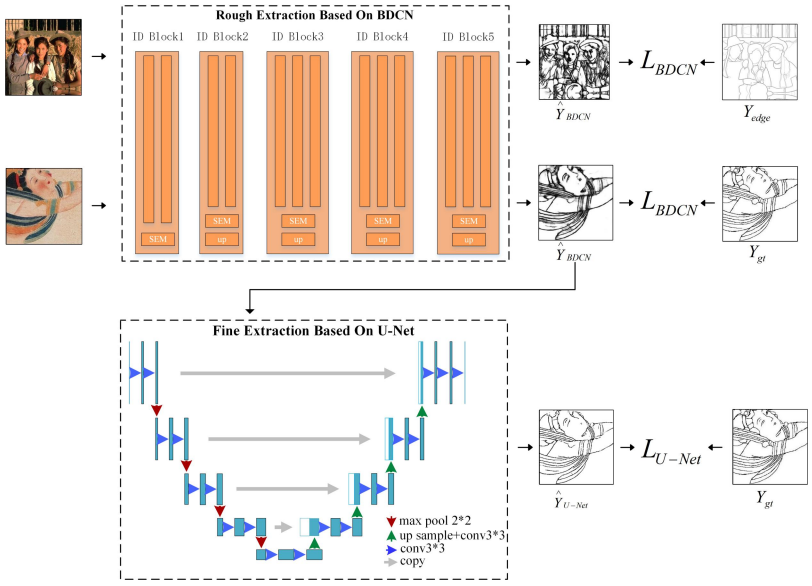

Figure 1. The proposed sketch extraction framework.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

3.1 Data preprocessing

The ancient murals are mostly affected by the environmental conditions and man-made destructions, so there are many problems such as surface mud deposits left over during excavation, natural degradation, conservation intervention and so on. We first use the prior knowledge of the sketch to select the spectral bands that shows the sketches best. For the murals painted with carbon-based material, we can choose the far red to near infrared bands (650-900nm) to produce an image that is the average of these bands, which reveals the sketches with the highest contrast.

3.2 Rough extraction based on BDCN

BDCN often needs a large amount of data as support, which is a big problem for the painted cultural relics with scarce data. Therefore, we first use the natural image dataset to pre-train the model, and then use the painted cultural relic data to fine-tune the model, which can not only effectively use the features in the natural images, such as texture and shape, but also solve the problem of insufficient painted cultural relic dataset.

The network consists of five Incremental Detection Blocks (ID Block). Each ID Block is connected with a pooling layer to gradually expand the receptive field of the next block. A group of multiple parallel convolutions with different dilation rates are inserted to complete the multi-scale representation of image features. For an input feature map $x \in R^{H \times W}$ with a convolution filter $w \in R^{h \times w}$, the output of dilated convolution can be calculated as:

$$y_{ij} = \sum_{m,n}^{h,w} x_{[i+r \cdot m, j+r \cdot n]} \cdot w_{[m,n]} \tag{1}$$

where r is the dilation rate and represents the step size of the sampled input feature map. The formula shows that the dilated convolution can expand the receptive field of neurons, but it does not increase the parameters or reduce the resolution. It completes the multi-scale representation of the image well, and avoids the problems of too many parameters, high training cost and long training time.

In our work, in order to accelerate the convergence, we use labeled public natural image dataset to pre-train BDCN. First, the ground truth of objects in natural image dataset is used to guide learning. The loss function of BDCN is designed as follows:

$$L_{BDCN} = L\left(\hat{Y}_{BDCN}, Y_{edge} \right) \tag{2}$$

where $\hat{Y}_{BDCN}$ is the predicted value of BDCN, Y_{edge} means the ground truth of natural images. We predict each pixel value. For ground truth, we define the positive samples and negative samples. Positive samples are edge pixels and negative samples are non-edge pixels. In order to distinguish them better, we also introduce a threshold γ, that is $Y_+ = \{y_i, y_j > \gamma\}$, $Y_- = \{y_i, y_j = 0\}$. Because the distribution of edge and non-edge pixels is very different, we adopt a class balanced cross-entropy loss $L(\cdot)$, and only consider the pixels corresponding to Y_+ and Y_- when computing the loss. We define $L(\cdot)$ as follows:

$$L\left(\hat{Y}, Y \right) = -\alpha \sum_{y_j \in Y_-} \log\left(1 - \hat{y}_j \right) - (1-\alpha) \sum_{y_j \in Y_+} \log \hat{y}_j \tag{3}$$

where $\hat{y}_j = P(X_j; W)$, X_j represents the activation value of the pixel j, W represents all the learning parameters in our architecture. $\alpha = |Y_+| / (|Y_+| + |Y_-|)$ and $1-\alpha = |Y_-| / (|Y_+| + |Y_-|)$ represent the number of edge pixels and non-edge pixels respectively, which are used to control the weight of positive and negative samples.

After pre-training, the painted cultural relic images are used to fine-tune BDCN. We use the ground truth of painted cultural relics Y_{gt} instead of the real edge of natural images Y_{edge} to guide training. Y_{gt} contains more cultural relic

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

image features, which can help us to extract sketches of painted cultural relics. The ground truth is obtained by manually tracing the sketches from the averaged far-red to near infrared bands (650-900nm).

3.3 Fine extraction based on U-Net

Due to various degradation and damage in the painted cultural relics, the sketches extracted by BDCN may appear blurred and incorrect in some places. In order to solve these problems, the U-Net-based method is used to further refine and denoise the sketches. U-Net is a full convolution network, which is a symmetric U-shaped structure including compression path and expansion path. The images can be denoised and reconstructed by U-Net.

The structure of the proposed method based on U-Net is shown in Fig. 2. The encoder is displayed on the left side, which is composed of four blocks. Each block uses three effective convolutions and a maximum pooling layer to down sample the images. After down sampling, the number of feature maps becomes twice of the original, and finally a feature map with the size of 32 * 32 is obtained. The decoder is shown on the right side of the network, which is also composed of four blocks. Each block firstly multiplies the size of the feature map by 2 through deconvolution, and at the same time reduces its number by half. The left symmetric down-sampling layer is connected with the up-sampling layer by using skip connection, so that the features obtained by the down-sampling layer can be directly transferred to the up-sampling layer. The fused feature map contains the information of different receptive fields in the encoder, which makes the network more accurate. The loss function of U-Net is defined as follows:

$$L_{U-Net} = L\left(\hat{Y}_{U-Net}, Y_{gt}\right) \tag{4}$$

where $\hat{Y}_{U-Net}$ is the prediction result of U-Net and Y_{gt} is the ground truth of the painted cultural relics.

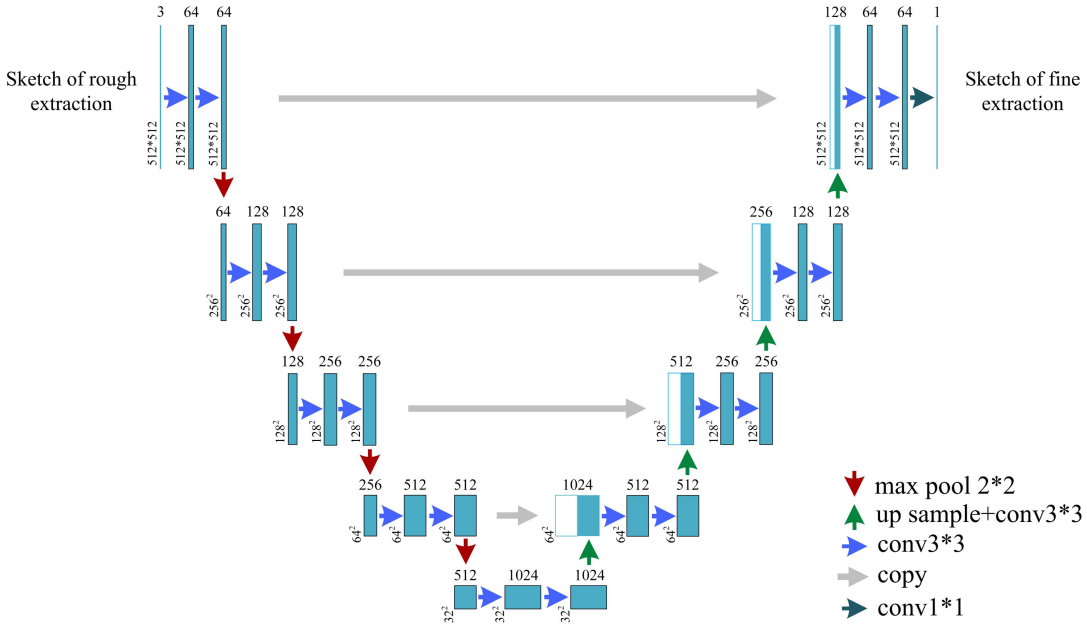

Figure 2. The framework of fine extraction based on U-Net.

We can use the multi-scale receptive field in the encoder path to suppress the influence of damage in the cultural relic images by using U-Net, and finally achieve the effects of denoising, deblurring and thinning sketches.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Dataset

In our experiments, two datasets are used. One is the public natural image dataset BSDS500, the other is the painted cultural relics dataset.

The public natural image dataset BSDS500 is used to pre-train the BDCN model. The BSDS dataset contains 500 natural images and reference edge labels. The painted cultural relics dataset contains 41 images collected by the Northwest university team. All the reference edge labels are drawn by experienced experts. We randomly cropped 41 images into 82 sub-images and used the data enhancement strategy to make the training more effective. In order to evaluate the performance of the model, we collected six spectral image cubes of paintings from the Eighteen Arhats of Fengguo temple. Some of them are images with clean background and less degradation, while others are images with more complex background.

4.2 Experimental setups

We implemented the network by using PyTorch. The initial learning rate and weight decay are set to 1e-6 and 2e-4 respectively in the rough extraction based on BDCN, and 1e-4 and 2e-4 respectively in the fine extraction based on U-Net.

The proposed method is compared with six state-of-the-art edge detection methods, including Canny[10], Flow-Based Difference-of-Gaussians (FDoG)[38], SE[12], RCF[14], BDCN[17], U-Net[22] and unsupervised self-organizing map clustering (SOM)[6]. The root mean square error (RMSE)[30], structural similarity index (SSIM)[30], and average precision (AP)[31] are used to evaluate the sketches extracted by different methods. RMSE is used to measure the deviation between the predicted values and the true values. SSIM mainly measures the differences of brightness, contrast and structure between the extraction results and the reference edge labels. The smaller value of RMSE and the larger value of SSIM indicate better performance of the model. We use AP to evaluate the accuracy of the model, expecting a higher value.

4.3 Experimental results

4.3.1 Comparison with other image-based methods

The quantitative results of comparison between our method and comparison methods are shown in Table 1, where the best results are shown in bold. In terms of indicators, the proposed method is superior to the other six methods in RMSE, SSIM and AP, which are 1~6%, 7~27% and 0.5~38% respectively. The experimental results show that the proposed method has better performance in structural integrity, accuracy and noise suppression. Qualitative visual inspection also shows that the sketches produced by the proposed method are more satisfactory. According to the evaluation results of SSIM, CNN-based methods (RCF, BDCN, Ours) are higher than traditional methods (Canny, FDoG, SE), which confirms the advantages of deep learning-based methods in sketch extracting. Compared with BDCN, the proposed method is about 19% higher on SSIM, 6% lower on RMSE and 24% higher on AP, which indicates the effectiveness of the proposed method and the necessity of the extraction method based on U-Net. Compared with U-Net, the proposed method achieves 24% improvement on SSIM, 1% improvement on RMSE and 24% improvement on AP, respectively, confirming the importance of rough extraction step.

Figure 3 shows the sketches extracted by different methods and their reference edge labels. As shown in Fig. 3, Canny, a method based on gradient calculation (Fig. 3c) failed in cases where the gradient change is not obvious, as broken or undetectable lines appeared, especially at the ears in the second and fifth images. FDoG (Fig. 3d) is sensitive to 'noise' or unwanted information, so there are many unexpected lines in the extraction results. The sketches extracted by SE (Fig. 3e) are blurred, especially when extracting an image with a complex background like the fourth image. Although RCF (Fig. 3f) can suppress unwanted information due to degradation, the extracted sketches are not clear and many details are lost. The extraction results of BDCN (Fig. 3g) are relatively complete, but its 'noise' suppression is insufficient. For example, the background part of the first image and the head of the second image are greatly affected by the unwanted information. Moreover, the sketches are too thick, which makes it easy to have the sketches merged when extracting dense areas, such as the fourth image. The sketches extracted by U-Net (Fig. 3h) are incomplete, and the results are poor for the unclear parts of the original images. For example, almost no useful sketch can be extracted in the body part of the figure in the sixth image. Compared with these edge detection methods, the results of our method are not only more complete and more detailed, but also contain less unwanted information. These experimental results show our advantages in detail extraction from real historic paintings with various damages and deterioration. In other words, our

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

method achieves better results both in objective indicators and subjective visual inspection than other state-of-the-art methods.

Table 1. Comparison with other methods.

Method	RMSE	SSIM	AP
Canny[10]	0.3126	0.6683	0.4138
FDoG[38]	0.2575	0.7672	0.6362
SE[12]	0.2648	0.6461	0.2614
RCF[14]	0.2948	0.8647	0.2923
BDCN[17]	0.3068	0.7486	0.4030
U-Net[22]	0.2535	0.6869	0.6186
Ours	**0.2451**	**0.9344**	**0.6409**

Figure 3. Comparison of sketch extraction results using spectral imaging of painted cultural relics, (a) the NIR images (average of 650-900 nm); (b) 'ground truth' (manual sketch by experts based on (a)); (c) Canny; (d) FDoG; (e) SE; (f) RCF; (g) BDCN; (h) U-Net; (i) Ours;

4.3.2 Comparison with a spectra-based method

The methods discussed above all relied on just one image, that is the averaged image of spectral bands in the far-red and near infrared wavelength range (650-900 nm) where the sketches appear with the best contrast. The reflectance spectra collected by the hyperspectral image system has not been fully utilized. In this section, we cluster those pixels with similar spectral reflectance as a means of segmenting the spatial image into different regions representing different

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

painted regions. For example, carbon-based sketches will be clustered together while sketches drawn with other inks or paints would form different clusters. Here we use an unsupervised clustering algorithm based on artificial neural networks, self-organising map (SOM), to cluster the pixels[6 39]. To improve the signal-to-noise ratio of the individual input spectral bands for clustering, the 128 band image cubes were reduced to 16 bands by averaging every 8 bands. The 9 bands (bands 2-10 in the 16 bands) covering the spectral range from 410 nm to 780 nm were then selected for clustering analysis (band 1 had very low signal to noise ratio and bands 11-16 were blurred due to chromatic aberration of the hyperspectral imaging lens and were thus discarded). This reduced spectral image cube has a spectral resolution of ~40nm which is sufficient for discriminating between different painting materials, especially since reflectance spectra of pigments are smooth and do not have sharp features (only a handful of pigments have absorption features smaller than 10nm in this spectral range). The SOM based algorithm is unsupervised and does not require labeling by experts. It is a 'shallow learning' method, where it learns from the data itself and it uses the spectral information for segmentation in contrast to the methods described above. The results of two of the above cases are shown in Fig. 4 for detailed comparison. The SOM results using spectra for segmentation of the sketches found not only sketches in carbon-based ink but also in red paint demonstrating the advantage of using the spectral information (Fig. 4g). On the other hand, SOM clustering sometimes groups regions affected by surface mud with the black sketches and fails in distinguishing the hair from the outlines as they were both painted by carbon ink and therefore had the same spectra and segmented together.

As it was shown in Kogou et al.[39], Principal Component Analysis (PCA) and Independent Component Analysis (ICA) can sometimes reveal more than the individual spectral band images, PCA and ICA were performed on the spectral image cubes, and in some cases, they did show more sketches than the averaged near infrared images (Fig. 3a and Fig. 4b) or the 'ground truth' shown in Fig. 3b and Fig. 4c. Fig. 4 first row shows that the PCA and ICA images revealed more sketches than the 'ground truth', which is not surprising because the 'ground truth' was based on the visual inspection of the averaged 650-900nm image (Fig. 4b) which did not show some of the sketches found in the PCA and ICA images. This also means any technique that revealed more sketches than the so called 'ground truth' will be penalized in the quality indicators, despite revealing the hidden true sketches. This highlights the difficulty of knowing what the 'ground truth' is in the study of real painted cultural objects rather than mockups. It also shows that quality indicators need to be used with caution.

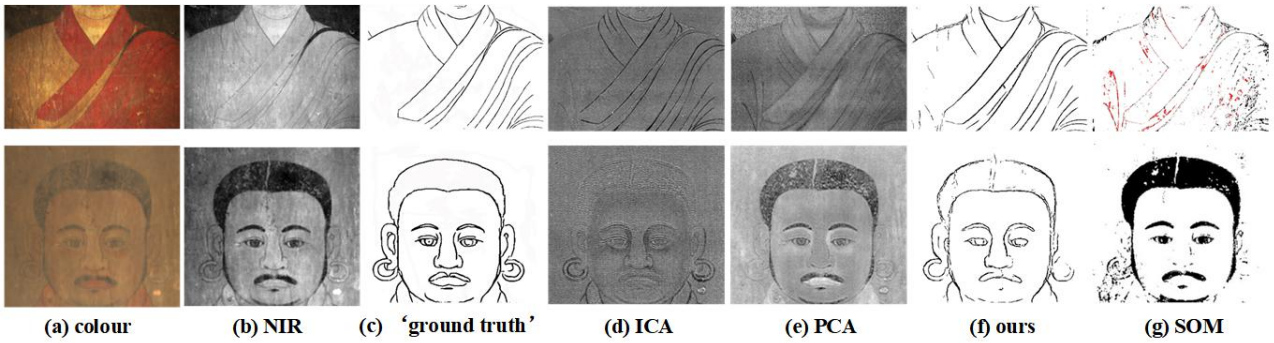

Figure 4. Two examples corresponding to the second and third row of Fig. 3 showing the colour images, the NIR images (average of 650-900 nm, same as Fig. 3a), the 'ground truth' sketches (same as in Fig. 3b), the sketches extracted using our proposed method 'ours', the ICA and PCA component images from the reduced 9 band spectral image cube that best reveals the sketches and the SOM cluster maps that extracts the areas of black sketches (black) and red sketches (red).

5. CONCLUSIONS

In this paper, we proposed a new framework for sketch extracting based on spectral imaging and deep learning. The framework is composed of BDCN and U-Net. For the data with fading or serious protection problems, we use the prior knowledge of sketch to preprocess them. In the extraction based on BDCN, we adopt the method of pre-training and fine-tuning to solve the problem of insufficient data of painted cultural relics. Then the results are further optimized based on U-Net to extract more details and suppress noise. The proposed method can be used not only for natural images, but also for spectral images of cultural relics with various conservation issues.

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

Compared with other methods, we verify the effectiveness and practicability of the proposed method. To sum up, the proposed method can extract ideal sketches from murals with serious conservation issues and complex background, which shows that it is an effective and promising method for sketch extracting.

The paper has also shown that a potential future improvement can be made by utilizing the full spectral image cube to reveal the more of the hidden sketches first (e.g. using PCA or ICA) before applying the proposed method. In addition, it has also shown the difficulties of knowing what the 'ground truth' is in a real painted cultural object and the potential issues of using quality indices based on the so called 'ground truth' to compare algorithms.

ACKNOWLEDGEMENT

Sincere thanks to all the experts and staff of Fengguo Temple in Yi County, Jinzhou, for their assistance in data collection. The Project Supported by the Key Research and Development Program of Shaanxi (No. 2021ZDLGY15-06); the Xi'an Key Laboratory of Intelligent Perception and Cultural Inheritance (No. 2019219614SYS011CG033); National Social Science Foundation of China (Grant No. 20BKG031); Open Research Fund of CAS Key Laboratory of Spectral Imaging Technology (Grant No. LSIT201920W); Changjiang Scholars and Innovative Research Team in University (No. IRT_17R87). Financial support from UK Natural Environment Research Council (NE/R014868/1) and Arts and Humanities Research Council (AH/T013184/1) are gratefully acknowledged.

REFERENCES

[1] Liu, J., Lu, D., and Shi, X., "Interactive sketch generation for Dunhuang frescoes," Lecture Notes in Computer Science 3942(1), 943-946 (2006).
[2] He, J., Wang, S., Zhang, Y., and Zhang, J., "A computational fresco sketch generation framework," Proc. International Conference on Multimedia and Expo Workshops, 1-6 (2013).
[3] Pan, G., Sun, D., Zhan, R., and Zhang, J., "Mural sketch generation via style-aware convolutional neural network," Proc. Computer Graphics International 2018, 239-245 (2018).
[4] Xu, D., Liu, Y., and Wang, X., "Line drawing generation method for ancient chinese murals," Computer Engineering 42(5), 244-248 (2016).
[5] Sun, D., Zhang, J., Pan, G., and Zhan, R., "Mural2sketch: A combined line drawing generation method for ancient mural painting," Proc. International Conference on Multimedia and Expo 2018, 1-6 (2018).
[6] Kogou, S., Lee, L., Shahtahmassebi, G., and Liang, H., "A new approach to the interpretation of XRF spectral imaging data using neural networks," X-Ray Spectrometry 50, 310-319 (2021).
[7] Kittler, J., "On the accuracy of the Sobel edge detector," Image and Vision Computing 1(1), 37-42 (1983).
[8] Kang, H., Lee, S., and Chui, C. K., "Coherent line drawing," Proc. International symposium on Nonphotorealistic animation and rendering 5, 43-50 (2007).
[9] Martin, D. R., Fowlkes, C. C., and Malik, J., "Learning to detect natural image boundaries using local brightness, color, and texture cues," IEEE transactions on Pattern Analysis and Machine Intelligence 26(5), 530-549 (2004).
[10] Canny, J., "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence 8(6), 679-698 (1986).
[11] Arbeláez, P., Maire, M., Fowlkes, C., And Malik, J., "Contour detection and hierarchical image segmentation," IEEE Transactions on Pattern Analysis and Machine Intelligence 33 (5), 898-916 (2011).
[12] Dollár, P., Zitnick, C. L., "Fast edge detection using structured forests," IEEE Transactions on Pattern Analysis and Machine Intelligence 37 (8), 1558-1570 (2015).
[13] Xie, S., and Tu, Z., "Holistically-nested edge detection," International Journal of Computer Vision 125(1-3), 3-18 (2017).
[14] Liu, Y., Cheng, M., Hu, X., Wang, K., and Bai, X., "Richer convolutional features for edge detection," Proc. IEEE conference on computer vision and pattern recognition, 3000-3009 (2017).
[15] Fu, K., Zhao, Q., And Gu, I. Y., "Refinet:a deep segmentation assisted refinement network for salient object detection," IEEE Transactions on Multimedia 21(2), 457-469 (2019).
[16] Bertasius, G., Shi, J., and Torresani, L., "Deepedge: A multi-scale bifurcated deep network for top-down contour detection," Proc. Conference on Computer Vision and Pattern Recognition 7, 4380-4389 (2015).

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use

[17] He, J., Zhang, S., Yang, M., Shan, Y., and Huang, T., "Bi-directional cascade network for perceptual edge detection," Proc. Computer Vision and Pattern Recognition, 3828-3837 (2019).
[18] Shen, W., Wang, X., Wang, Y., Bai, X., and Zhang, Z., "Deepcontour: A deep convolutional feature learned by positive-sharing loss for contour detection," Proc. Computer Vision and Pattern Recognition, 3982-3991 (2015).
[19] Liu, Y., and Michael, S. L.., "Learning relaxed deep supervision for better edge detection," Proc. Computer Vision and Pattern Recognition, 231-240 (2016).
[20] Deng, R., Shen, C, Liu, S., Wang, H., and Liu, X., "Learning to predict crisp boundaries," Lecture Notes in Computer Science 11210, 570-586 (2018).
[21] Liang, C. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L., "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," IEEE transactions on pattern analysis and machine intelligence 40(4), 834-848 (2018).
[22] Ronneberger, O., Fischer, P., and Brox, T., "U-net: Convolutional networks for biomedical image segmentation," Lecture Notes in Computer Science 9351, 234-241 (2015).
[23] He, K., Zhang, X., Ren, S., and Sun, J., "Deep residual learning for image recognition," IEEE Conference on computer vision and pattern recognition, 770-778 (2016).
[24] Mao, X., Shen, C., and Yang, Y., "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," Proc. Conference on Neural Information Processing Systems, (2016).
[25] Zhang, K., Wang, M., Zuo, L., Zhang, S., and Yan., Z., "Toward convolutional blind denoising of real photographs," Proc. Conference on Computer Vision and Pattern Recognition, 1712-1722 (2019).
[26] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., and Liang, J., "Unet++: A nested u-net architecture for medical image segmentation," Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support 11045, 3-11 (2018).
[27] Weng, Y., Zhou, T., Li, Y., and Qiu, X., "Nas-unet: Neural architecture search for medical image segmentation," IEEE Access 7, 44247-44257 (2019).
[28] Farahani, A., and Mohseni, H., "Medical image segmentation using customized u-net with adaptive activation functions," Neural Computing and Applications (2020).
[29] Arbelaez, P., Maire, M., Fowlkes, C., and Malik, J., "Contour detection and hierarchical image segmentation," IEEE Trans Pattern Anal Mach Intel 33(5), 898-916 (2011).
[30] Wang, Z., Bovik, A. C., and Simoncelli, E. P., "Image quality assessment: from error visibility to structural similarity," IEEE transactions on image processing 13(4), 600-612 (2004).
[31] Martin, D. R., Fowlkes, C. C., and Malik, J., "Learning to detect natural image boundaries using local brightness, color, and texture cues," IEEE Transactions on Pattern Analysis and Machine Intelligence 26(5), 530-549 (2004).
[32] Liang, H., Advances in multispectral and hyperspectral imaging for archaeology and art conservation. Appl. Phys. A 106, 309-323 (2012).
[33] Wu, T, Li, G., Yang, Z., Zhang, H., Lei, Y., Wang, N., and Zhang, L., "Shortwave infrared imaging spectroscopy for analysis of ancient paintings," Applied spectroscopy 71(5), 977-987 (2017).
[34] Daniel, F., Mounier, A., Pérez-Arantegui, J., Pardos, C., Prieto-Taboada, N., Vallejuelo, S., and Castro, K., "Hyperspectral imaging applied to the analysis of Goya paintings in the museum of Zaragoza (spain)," Microchemical Journal 126, 113-120 (2016).
[35] Bai, D., Messinger, D. W., and Howell, D., "Hyperspectral analysis of cultural heritage artifacts: pigment material diversity in the Gough map of Britain," Optical Engineering 56(8), 0091-3286 (2017).
[36] Pan, N., and Hou, M., "The extraction and fusion of faint mural based on feature transform of hyperspectral images," In 2016 4th International Workshop on Earth Observation and Remote Sensing Applications (EORSA), 161-164 (2017).
[37] Peng, J., Yu, K., Wang, J., Zhang, Q., Wang, L., and Fan, P., "Mining painted cultural relic patterns based on principal component images selection and image fusion of hyperspectral images," Journal of Cultural Heritage 36, 32-39 (2019).
[38] Kang, H., Lee, S., and Chui, C. K., "Coherent line drawing," Proc. International symposium on Nonphotorealistic animation and rendering, 43-50 (2007).
[39] Kogou, S., Shahtahmassebi, G., Lucian, A., Liang, H., Shui, B., Zhang, W., Su, B., Van Schaik, S., From remote sensing and machine learning to the history of the Silk Road: large scale material identification on wall paintings. Scientific Report 10, 19312 (2020)

Downloaded From: https://www.spiedigitallibrary.org/conference-proceedings-of-spie on 28 Jul 2021
Terms of Use: https://www.spiedigitallibrary.org/terms-of-use