Mural Sketch Generation via Style-aware Convolutional Neural Network

Gang Pan
School of Computer Software
Tianjin University
pangang@tju.edu.cn

Di Sun*
School of Computer Science and Information Engineering
Tianjin University of Science and Technology
dsun@tust.edu.cn

Rui Zhan
School of Computer Software
Tianjin University
1573111291@qq.com

Jiawan Zhang
School of Computer Software
Tianjin University
jwzhang@tju.edu.cn

ABSTRACT

Sketch is one of the most important art expression forms for traditional Chinese painting. This paper presents a complete sketch generation framework for ancient mural paintings. First, we propose a deep learning network to perform mural-to-sketch prediction by combining meaningful convolutional features in a holistic manner. A dedicated mural database with fine-grained ground truth is built for network training and testing. Then we design a style-aware image fusion approach by detecting the specific feature region in a mural, from which the artistic style can be maximally preserved. Experimental results have demonstrated its validity in extracting style mural sketch. This work has the potential to provide a computer aided tool for artists and restorers to imitate and restore time-honored paintings.

CCS CONCEPTS

• **Computing methodologies** → **Image processing**; • **Human-centered computing** → *HCI theory, concepts and models*; *User interface programming*;

KEYWORDS

Line drawing, sketch, edge extraction, mural digital protection

ACM Reference Format:
Gang Pan, Di Sun, Rui Zhan, and Jiawan Zhang. 2018. Mural Sketch Generation via Style-aware Convolutional Neural Network. In *CGI 2018: Computer Graphics International 2018, June 11–14, 2018, Bintan Island, Indonesia*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3208159.3208160

*Corresponding author.

CGI 2018, June 11–14, 2018, Bintan Island, Indonesia

ACM ISBN 978-1-4503-6401-0/18/06. . . $15.00
https://doi.org/10.1145/3208159.3208160

1 INTRDUCTION

Dunhuang murals are famous cultural treasure in the world heritage as well as buddhist art. Unfortunately these invaluable murals have been severely damaged by both human and nature factors through thousands of years. One of the most important traditional preservation methods is to imitate those precious murals. Chinese traditional paintings belong to claborate-style painting. Generally speaking, the production process can be divided into three steps: first to draw a sketch, then to fill the color, and at last to finalize with a dark ink line. Among these, the most important step of imitation procedure is to get the mural sketch. Sketch mainly uses lines to outline objects, which is both the basic element of visual image and the necessary means for art modeling.

Traditional manual imitation procedure for mural sketch involves laborious work. For purpose of eliminating subjectivity, photography is regarded as a convenient option, painters work on sketch copying according to the digital mural picture or slide show. The whole process is still time-consuming. Moreover, it is prone to errors due to draftsmens painting skill.

Advanced image analysis and computer vision techniques enable the computers to improve the efficiency of imitation work greatly. An interactive sketch generation method is presented in[20]. The algorithm first obtain the contour of a mural image which is composed of many connected curves, and then rendered the strokes by learning styles from examples. Another work[12] is about a computational framework to generate sketch by extracting hierarchical structure blocks and replacing missing contents with interaction. These two work are inspiring and supportive, but the results can be further improved.

Since most ancient paintings involve a lot of noise due to the destruction of human or natural factors, its difficult to directly extracting sketch from original mural images with low-level image processing methods. moreover, existing edge detection or image segmentation techniques usually fail to produce a good artistic illustration, as they do not deal with aesthetic aspects. Hence extracting a style sketch from a mural image is very useful and challenging.

In the past few years, convolutional neural networks (C-NNs) have become popular in the computer vision community by substantially advancing the state-of-the-art of various tasks, including image classification[18, 31, 33], object detection[10, 11, 28] and semantic segmentation[5, 22] etc. Since CNNs have a strong capability to learn high-level representations of natural images automatically, there is a recent trend of using convolutional networks to perform edge detection. Some well-known CNN-based methods have pushed forward this field significantly[3, 9, 13, 21, 30, 35].

In this paper, we address mural sketch generation problem by developing a new CNN-based sketch extraction framework, which aims to convert a mural image into a sketch. We treat sketch generation as a task of learning image features and a filtering task. Essentially, our underlying hypothesis is that Convolutional Neural Network is able to find some distinctive information from a clustered background, and thus there remains only important signals that correspond to mural sketches.Especially, we incorporate artistic style information by using specific feature region detection for mural artwork.

We aim at quickly generating high-quality mural lines which can (i) help record and research our precious cultural heritage; (ii) better guide people to work on mural copying and creation; (iii) use for mural disease labeling so as to protect the ancient mural paintings.

Our contributions are as follows:

(1) We formulate a deep model that fully exploits staged convolutional features of objects to perform the mural-to-sketch prediction in a holistic manner.

(2) We design a collaborative representation method for sketch optimization in the principle of mural style retain. Specifically, we detect a specific "style region" and enhance the information of this region so as to preserve artistic style to the maximum extent possible.

2 RELATED WORK

2.1 Line extraction

Sketch uses simple lines to convey information. The key technique of sketch generation involves line extraction, simply called edge detection. As an important low level image analysis technique, edge detection has been extensively analyzed in the fields of computer vision and pattern recognition over the last 20 years.

Early classic methods such as Sobel detector[17], Canny detector[4, 24] mainly consider locate sharp changes in the intensity function. Due to the high efficiency of this method, there are still some uses, but its accuracy is hard to meet the requirements of modern applications. Especially, the fresco strokes usually have a certain thickness so as to present different style visually, while feature lines extracted from these operators have two edges for one stroke.

With the vigorous development of deep learning, there have been a recent wave of CNN based methods including N4-Fields[9], DeepEdge[3], DeepContour[30], CSCNN[13], HED[35] and RCF[21]. The above CNN-based models have advanced the state-of-the-art significantly, but these methods usually only adopt CNN features from the last layer of each conv stage. Specifically, When directly applying these methods to fresco images, the results usually have no obvious coherence from the perspective of drawing, unlike human painting.

What's more, many image-based non-photorealistic rendering (NPR) techniques use lines to help create illustrations. A variety of methods[6, 8, 16, 25, 26] have been reported on A-mong these, Kang et al.[15] proposed a flow-based anisotropic filtering framework named FDoG (Flow-based Difference-of-Gaussians) to extract lines, this work has a preferable accuracy and less noise. We have attempted this method to process the mural images. It is inevitable to do image denoising or smoothing for input data and manually adjust the parameters when applying this kind of method to mural sketch extraction. This process is complicated.

To sum up, we need to specialize in sketch generation technique for ancient murals.

2.2 Specific Object Detection for artwork

When it comes to art, artistic style plays a significant role in identifying the origins of an artwork. In order to generate a sketch that can well preserve the mural style, we consider introducing face details such as eyes, eyebrows or mouths.

Facial point detection is critical for face recognition and analysis, and has been studied extensively in recent years. Existing approaches can be generally divided into two categories: classifying search windows or directly predicting keypoint positions (or shape parameters). Many used Adaboost[19], SVM[2, 27], or random forest[1] classifiers as component detectors and detection was based on local image features. Convolutional networks and other deep models have been successfully used in vision tasks such as face detection and pose estimation, face parsing, image classification, and scene parsing.

To the best of our knowledge, current datasets available in this field are mainly for natural persons, while our research object is artistic figure. When these existing methods are applied directly to the artwork, the results usually fail, as it lacks essential artistic data analysis.

There are also art-related research work, but the main focus is on artistic data analysis[32] or artistic style transfer, seldom involves specific object detection. Westlake et al.[34] pose an work of detecting People in Artwork with CNNs. A People-Art dataset consisting of people from photos, cartoons and different artwork movements is contributed in this work. Our focus area is even smaller, that is, face detection in mural artwork.

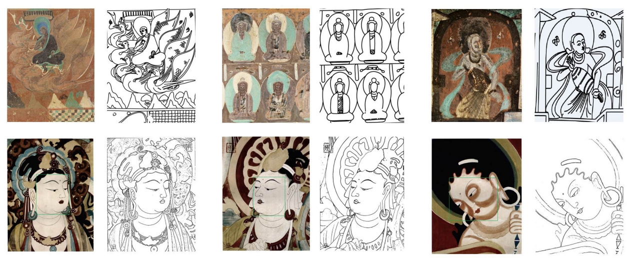

Figure 1: Examples of our mural dataset. The second row with figures are marked with green box which can be further used for mural feature region detection.

3 DATA SET

Our research is based on a field study at Dunhuang Mogao Grottoes, so we conduct a focus-group interview with the staff at Archaeological Institute in Dunhuang Research Academy, observe their existing working pattern, and collect relevant data including high quality photographic images and the sketches manually drawn by archaeological draftsmen which have been open published in the form of archaeological report. We obtain 624 mural-sketch images in total based on the above two parts. Fig. 1 shows some examples.

4 METHODOLOGY

An overview of our approach is shown in Fig.2. We start by introducing the architecture of our network (Sec. 4.1) from which a mural-to-sketch prediction is performed based on our dataset. To further obtain an artistic style sketch, we detect the facial region in a mural painting, and generate the corresponding sketch for this specific region(Sec. 4.2). Essentially, our underlying hypothesis is that the style characteristics can be well expressed by information of these feature regions. Then, we introduce a new principle followed by collaborative representation to fuse the mural sketch(Sec. 4.3).

4.1 Network Architecture of Sketch Extraction

Recent work has shown that a fine-tuning convolutional neural network is an efficient way to obtain a good performance on new tasks. We explicitly learn sketch information from mural images via our network. Note that this type of Convolutional Neural Network architecture is first proposed by [35], which is a deep learning model used for image edge detection.

The original network structure is initialized based on an ImageNet pretrained VGG16 layer model[31] by adding additional side output layers and replacing fully connected layers by fully convolutional layers with 1 × 1 kernel size. Each fully convolutional layer is then connected to an up-sampling layer to ensure that the outputs of all the stages are the same size.

Given sketch characteristics, we design an improved version, as illustrated in Fig.**??**. We design the network to address two important issues: (1) we train and predict the entire image end-to-end (i.e., it is holistical); (2) we incorporate multi-scale and multi-level learning for deep image features. The architecture does not have to be as deep as the previous version. The reason is that the features gradually become ambiguous and generate a poor sketch along with the deeper layer. Meanwhile, as described in[21], even though the convolutional neural network is designed with multiple inner features, the use of outer multi-scale input improves the final sketch generation significantly.

To obtain a more accurate prediction, we run the single network on multi-scale input images, as illustrated in Figure 2(A). This strategy occurs at both the training stage, such as data augmentation, and at the testing stage such as ensemble testing. This approach is particularly common in non-deep-learning based methods[7].

Compared with HED[35], our modifications can be described as following:

(1) We provide the network with six-scale images and the corresponding ground truths as the training set. We give the training images six-scales (0.8, 1.0, 1.2, 1.4, 1.6, and 2.0) ,rotate the images to 4 different angles (0°, 90°, 180°, and 270°) and flip them with different axes (up-down, left-right, and no flip) as input for training examples.

(2) We remove the layer of conv 5_3 and reserve the layers of conv1_1, conv2_2, conv3_3 and conv4_3.

(3) HED only considers the last conv layer in each stage of VGG16, in which lots of helpful information to edge detection is missed. We use a fully convolutional network to combine features from each CNN layer efficiently. Each conv layer is connected to a conv layer with kernel size 1 × 1 and channel depth 21. Thus it can capture more object or object part boundaries accurately across a larger range of scales.

4.2 Feature Region Detection for Mural Style Retain

The sketch results obtained from the above multi-scale deep network model have better precision, but the style characteristics have not yet been taken into consideration.

Ancient Chinese murals are composed of lines. These line strokes have a specific artistic style of "straightforward skill", which is sketch in traditional ink and brush style. The strokes used for the drafting and finalization focus more on the line shape and thickness of the strokes. Taking figure painting as an example, this painting style is mainly reflected by facial details such as eyes, eyebrows or mouths. Without taking these factors into consideration, the results are alike only in appearance but spirit.

To this end, we consider to design a specific feature region based sketch extraction method, from which the style information of the feature regions can be maximally retained. The key of this issue is twofold: one is how to detect these feature regions, and the other is what this so-called style should be.

It is worth mentioning that we have obtained some deep features with multi-scale in the process of deep learning. According to the staff of Dunhuang Research Institute, the

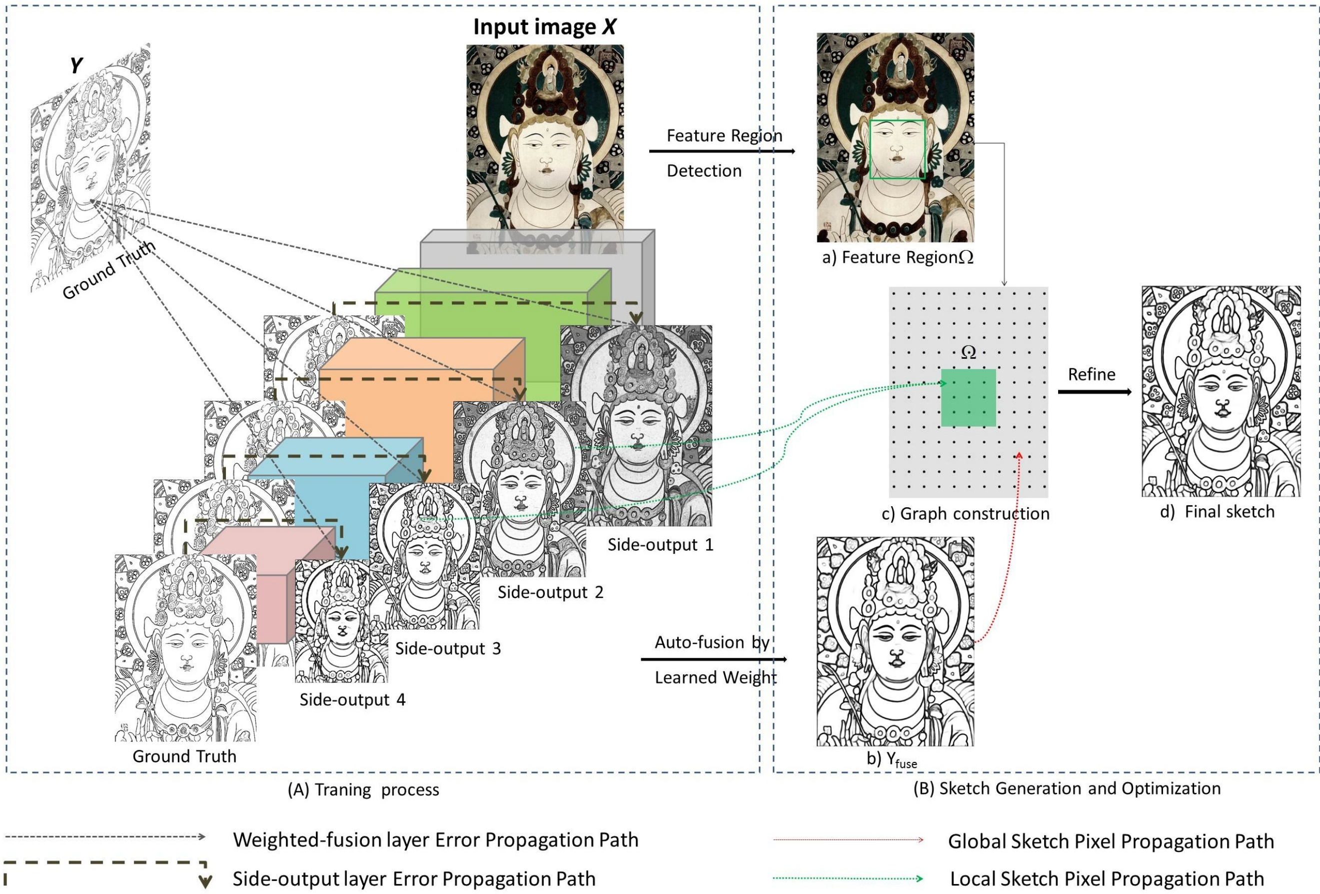

Figure 2: The flowchart of Our work.

side outputs from the 2nd and 3rd layers show more style features which should be retained. We will introduce details in next subsection.

For feature region detection, we utilize TensorBox to detect the face area of mural image. TensorBox supports several data formats for bounding boxes description, so we can train on our own data that the face areas are marked with bounding boxes, as illustrated in the 2nd row of Fig. 1.

The automatically detected result for face area can be seen as Fig. 2a). Note that the face detection for artwork is a challenging problem. We do not solve this general problem in our work. Instead, we design an application-specific solution. We apply interactive refinement to adjust candidate region with a low false positive rate. This step is necessary, because some automatic detected images may contain incorrect object items due to errors in object detection sometimes. Furthermore, the bounding-box based segmentation also needs improvement. The user can interactively refine the segmentation with methods like those in [29].

4.3 Style-Aware Sketch Optimization

As described in [35], the fused/averaged result is produced by individual side outputs at different scales. It has been statistically validated that the learned weighted-fusion achieves best F-score, while directly averaging all of layers produces better average precision. We discover that convolutional features become coarser gradually and intermediate layers contain lots of useful fine details for style illustration. Directly averaging all of layers produces better average precision, but also losts the style characteristics. In principle, we might choose existing image fusion method and optimize the combination, but a simple merging method might not result in an artistic combination. Our novel fusion method suited to specific style-retain contains Three steps.

Firstly, we fuse a global sketch from the multi-scale side outputs using the learned weighted-fusion. A weighted-fusion layer is added to the network which simultaneously learn the fusion weight during training. The loss function at the fusion layer L_{fuse} becomes

$$L_{fuse}(W, w, h) = Dist(Y, \hat{Y}_{fuse})$$

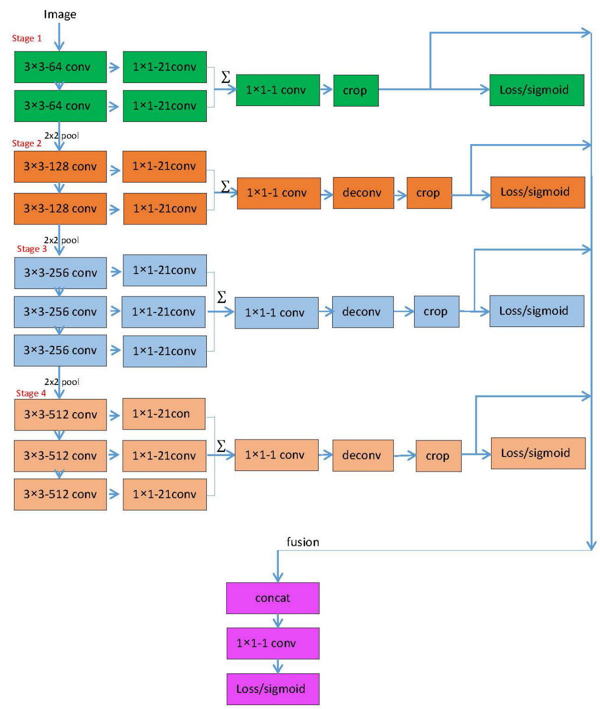

Figure 3: Our network architecture. The input is an image with arbitrary sizes, and our network outputs an edge possibility map in the same size.

where $\hat{Y}_{fuse}$ represents the fusion result. Dist(.,.) is a defined distance measure between the ground truth label map and the fused predictions. Our paper uses cross-entropy loss for this purpose. Thus, the following objective function is minimized via a traditional standard stochastic gradient, and we update the parameter by back propagation:

$$(W, w, h)^* = argmin(L_{side}(W, w) + L_{fuse}(W, w, h))$$

where W denotes the collection of all standard network layer parameters. Suppose in the network we have M side-output layers. Each side-output layer is also associated with a classifier, in which the corresponding weights are denoted as $w = (w^1, ..., w^M)$. h is the fusion weight. L_{side} denotes the image-level loss function for sideoutputs. We emphasize here that all the side-output predictions are obtained in one pass, this enables us to fully investigate different configurations of combining the outputs at no extra cost.

Secondly, we compute a local sketch for the detected style region by averaging the corresponding results from 2nd and 3rd side outputs. The reason for partial average is because we discover that convolutional features become coarser gradually, and intermediate layers contain lots of useful fine details for art style illustration. We also try to use weighted sum, but we find the simple average works very well.

Finally, we optimize the combination of these images into a final sketch. Merging those two readily available outputs further boost the performance. The final merging f is then computed as

$$f(p) = \begin{cases} Average(\hat{Y}^2_{side}(p) + \hat{Y}^3_{side}(p)) & if\ p \in \Omega \\ \hat{Y}_{fuse}(p) & \text{other} \end{cases}$$

where Ω is the style region obtained from face detection in last subsection.

5 EXPERIMENTS AND RESULTS

5.1 Experiment settings

Dataset Splits and Pre-processing: There are 312 mural-sketch pairs in the introduced dataset. Of these, we use 180 pairs for training, 50 pairs for validation, and the rest for testing. Before we conduct the experiments, we use the edge strength to crop object of both photos and sketches for coarse alignment. Then we resize all cropped murals/sketches to the same size of 430 × 550.

Implementation Details: We implement our network using the publicly available Caffe[14] which is well-known in this community. The VGG16 model is used to initialize our network. In training, the weights of 1 × 1 conv layer in stage 1-4 are initialized from zero-mean Gaussian distributions with standard deviation 0.01 and the biases are initialized to 0. All experiments in this paper are finished using a NVIDIA TITAN X GPU.

5.2 Results

Fig.4 shows mural sketch results obtained from the test images. Our method performs consistently well on a variety of images with different noise or broken content. Note that the resulting strokes are clean, smooth, and coherent, with little or no dispersement. Even in tough noisy images, our work has been shown to produce satisfied results.

Fig. 4 (b-d) show the comparison of our method with other popular line extraction techniques, including Canny, gPb, and HED. We use a deep-learning model with multi-scale and crop layer to obtain abundant information and the result is better than traditional hand-crafted features. From the mural line-drawing perspective, our method outperforms others in that it is not only capable of generating smooth and coherent lines but also capable of depicting perceptually interesting painting style.

Generally speaking, it is easy to fade or local damage for mural paintings due to the destruction of human or natural factors, directly extracting sketches from original mural images would involve a lot of noise. Traditional line extraction methods for this issue usually need to do image preprocessing, while mural image denoising is tough and complicated with popular smoothing detectors, such as L0-smoothing[36], Bilateral Filtering[23]. In addition, due to the addition of pretreatment, the contents of the mural itself has been modified, in fact, the result is not precise enough. It is worth mentioning that our method does not require image denoising

(a) (b) (c) (d) (e)

Figure 4: Results and comparison with other techniques. (a) Original image. (b-d) show edge responses from the Canny, gPb, and HED operators respectively. (e) displays our results.

or smoothing operations. This shows that our method has better robustness and practicability.

6 CONCLUSION

In this paper we have posed a style-aware mural sketch generating method based on convolutional neural network. We built a dedicated mural database with fine-grained ground truth for network training and testing, and utilized a fully convolutional network to accumulate the features from each CNN layer. To illustrate the specific artistic style of mural paintings, we designed a style-retained image fusion approach based on the detected feature region in mural images. A collaborative representation solution were proposed to integrate the global sketch and the local sketch. We demonstrated the performance by extensive mural images. Our work is not only capable of generating smooth and coherent sketches but also is capable of showing artistic styles. Sketch is generally considered the cornerstone of nonphotorealistic rendering, we believe that it will be very suitable for realistic mural copying, archaeological drawing and related culture applications.

Acknowledgements: The authors would like to thank Dunhuang Academy for providing the research data. This work was supported by Tianjin Philosophy and Social Science Planning Program under grant TJSR15-008, National Social Science Foundation under grant 15XMZ057 and MSRA CCNP 2016.

REFERENCES

[1] Brian Amberg and Thomas Vetter. 2012. Optimal landmark detection using shape models and branch and bound. In *IEEE International Conference on Computer Vision*. 455–462.

[2] P. N. Belhumeur, D. W. Jacobs, D. J. Kriegman, and N Kumar. 2013. Localizing parts of faces using a consensus of exemplars. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 12 (2013), 2930–2940.

[3] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. 2014. DeepEdge: A multi-scale bifurcated deep network for top-down contour detection. 52, 3 (2014), 4380–4389.

[4] J Canny. 1986. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8, 6 (June 1986), 679–698. https://doi.org/10.1109/TPAMI.1986.4767851

[5] Liang Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. 2014. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs. *Computer Science* 4 (2014), 357–361.

[6] Doug Decarlo and Anthony Santella. 2002. Stylization and abstraction of photographs. *Acm Transactions on Graphics* 21, 3 (2002), 769–776.

[7] P Dollr and C. L. Zitnick. 2015. Fast Edge Detection Using Structured Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 8 (2015), 1558–1570.

[8] Jan Fischer, Dirk Bartz, and Wolfgang Straer. 2005. Stylized Augmented Reality for Improved Immersion. 2005 (2005), 195–325.

[9] Yaroslav Ganin and Victor Lempitsky. 2014. *N^4 -Fields: Neural Network Nearest Neighbor Fields for Image Transforms.* Springer International Publishing. 536–551 pages.

[10] Ross B. Girshick. 2015. Fast R-CNN. *CoRR* abs/1504.08083 (2015). arXiv:1504.08083 http://arxiv.org/abs/1504.08083

[11] Ross B. Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2013. Rich feature hierarchies for accurate object detection and semantic segmentation. *CoRR* abs/1311.2524 (2013). arXiv:1311.2524 http://arxiv.org/abs/1311.2524

[12] Jianing He, Shan Wang, Yi Zhang, and Jiawan Zhang. 2013. A computational fresco sketch generation framework. In *2013 IEEE International Conference on Multimedia and Expo (ICME)*. 1–6. https://doi.org/10.1109/ICME.2013.6607509

[13] Jyh Jing Hwang and Tyng Luh Liu. 2015. Pixel-wise Deep Learning for Contour Detection. *Computer Science* (2015).

[14] Jia, Yangqing, Shelhamer, Evan, Donahue, Jeff, Karayev, Sergey, Long, and Jonathan. 2014. Caffe: Convolutional Architecture for Fast Feature Embedding. (2014), 675–678.

[15] Henry Kang, Seungyong Lee, and Charles K. Chui. 2007. Coherent Line Drawing. In *Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering (NPAR '07)*. ACM, New York, NY, USA, 43–50. https://doi.org/10.1145/1274871.1274878

[16] Hyung W. Kang, Charles K. Chui, and Uday K. Chakraborty. 2006. A unified scheme for adaptive stroke-based rendering. *The Visual Computer* 22, 9 (2006), 814–824.

[17] J Kittler. 1983. On the accuracy of the Sobel edge detector. *Image and Vision Computing* 1, 1 (1983), 37 – 42.

[18] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. 2012. ImageNet classification with deep convolutional neural networks. In *International Conference on Neural Information Processing Systems*. 1097–1105.

[19] Lin Liang, Rong Xiao, Fang Wen, and Jian Sun. 2008. Face Alignment Via Component-Based Discriminative Search. In *European Conference on Computer Vision*. 72–85.

[20] Jianming Liu, Dongming Lu, and Xifan Shi. 2006. Interactive sketch generation for dunhuang frescoes. In *International Conference on Technologies for E-Learning and Digital Entertainment*. 943–946.

[21] Yun Liu, Ming Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. 2017. Richer Convolutional Features for Edge Detection. In *Computer Vision and Pattern Recognition*.

[22] Jonathan Long, Evan Shelhamer, and Trevor Darrell. 2015. Fully convolutional networks for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*. 3431–3440.

[23] R. Manduchi and C. Tomasi. 1998. Bilateral Filtering for Gray and Color Images. In *Computer Vision, IEEE International Conference on(ICCV)*, Vol. 00. 839. https://doi.org/10.1109/ICCV.1998.710815

[24] P. Meer and B. Georgescu. 2001. Edge Detection with Embedded Confidence. *IEEE Trans. Pattern Anal. Mach. Intell.* 23, 12 (Dec. 2001), 1351–1365. https://doi.org/10.1109/34.977560

[25] Orzan, Alexandrina, Bousseau, Adrien, Barla, Thollot, and Jo. 2007. Structure-preserving manipulation of photographs. (2007).

[26] F. A. Pellegrino, W. Vanzella, and V. Torre. 2004. Edge detection revisited. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 34, 3 (June 2004), 1500–1518. https://doi.org/10.1109/TSMCB.2004.824147

[27] Deva Ramanan. 2012. Face detection, pose estimation, and landmark localization in the wild. In *IEEE Conference on Computer Vision and Pattern Recognition*. 2879–2886.

[28] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. 2015. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *CoRR* abs/1506.01497 (2015). arXiv:1506.01497 http://arxiv.org/abs/1506.01497

[29] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. 2004. "GrabCut": interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH*. 309–314.

[30] Wei Shen, Xinggang Wang, Yan Wang, Xiang Bai, and Zhijiang Zhang. 2015. DeepContour: A deep convolutional feature learned by positive-sharing loss for contour detection. (2015), 3982–3991.

[31] Karen Simonyan and Andrew Zisserman. 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition. *Computer Science* (2014).

[32] Gjorgji Strezoski and Marcel Worring. 2017. OmniArt: Multi-task Deep Learning for Artistic Data Analysis. (2017).

[33] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. 2014. Going Deeper with Convolutions. *CoRR* abs/1409.4842 (2014). arXiv:1409.4842 http://arxiv.org/abs/1409.4842

[34] Nicholas Westlake, Hongping Cai, and Peter Hall. 2016. Detecting People in Artwork with CNNs. In *European Conference on Computer Vision*. 825–841.

[35] Saining Xie and Zhuowen Tu. 2015. Holistically-Nested Edge Detection. In *IEEE International Conference on Computer Vision*. 1395–1403.

[36] Li Xu, Cewu Lu, Yi Xu, and Jiaya Jia. 2011. Image smoothing via L0 gradient minimization. *Acm Transactions on Graphics* 30, 6 (2011), 174.