Multimedia Tools and Applications (2019) 78:19565–19586
https://doi.org/10.1007/s11042-019-7270-8

Photographic painting style transfer using convolutional neural networks

Asad Khan[1] · Muhammad Ahmad[2] · Nuzhat Naqvi[3] · Faisal Yousafzai[4] · Jing Xiao[1]

Received: 26 April 2018 / Revised: 20 December 2018 / Accepted: 24 January 2019 /
Published online: 15 February 2019

Abstract
We propose a novel automatic photographic painting style technique with a single example image by using Convolutional Neural Networks (CNN). The photographic painting style is a challenging problem in the research community. Even though, researchers have been trying to obtain good results on painting style, but not much has been done on photographic stylization. Portrait painting techniques are mainly designed for the graphite style and/or are based on image analogies; an example painting as well as its original unpainted version are required. This preceding issue is a motivation of our proposed methods. As a result, our method extends the limits of their domain of applicability. We present a novel multi-convolutional-learning technique that is developed for both images (NPR/PR) labeling, style transmission and elevating a particular unified CNN model per weight sharing. A new painting technique is generated that follows the example style in the example image and maintains the integrity of facial structures. We believe this novel interpretation connects these two important research fields and could enlighten future researches. Moreover, our proposed technique is not restricted to headshot images or specific styles as our method can also change the photographic painting style in the wild.

Keywords Photographics Painting Style (PPS) · Non-Photorealistic Rendering (NPR) · Photorealistic Rendering (PR) · Digital painting · Example-based painting

✉ Jing Xiao
xiaojing@scnu.edu.cn

Asad Khan
asadustc@gmail.com

1 School of Computer Science, South China Normal University, Guangzhou, People's Republic of China

2 Institute of Robotics, Computer Science Department, Innopolis University, Innopolis, Russia

3 School of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, People's Republic of China

4 Military College of Engineering, National University of Sciences and Technology (NUST), Islamabad, Pakistan

1 Introduction

Artwork image processing in non-photorealistic rendering (NPR) field remains a dynamic part of the research. To address the issue of style transfer using deep neural networks, currently, many techniques have been contributed accordingly to enhance the algorithmic capacity of proposed systems. Style transformation of an image to another is an interesting fact, but complex to be implemented as per existing research [15, 21, 44]. The major issue which is a challenge since its start for computer graphic researchers is that how to accumulate human faces along their complexities and the vast number of degrees freedom. Faces are the most common object in artwork and painting, so digital image operations also focus on this fundamental object attractively. As compared to manual painting or drawing of faces, the computer is a fast and inexpensive medium of artistic style transformer which does not use massive laborious operations. The available solutions have a difference between, an assumption of input data, priority and their techniques which are used by them.

For tackling the style transfer of an image to another, diverse methods have been proposed. There are some ways to handle the intrinsic ambiguities in transferring the Photographics Painting Style (PPS). Through the increase of non-photo-realistic rendering [25], numerous practices were proposed for automated painting [9, 15, 18, 19, 25, 46, 51]. Although, the existing methods are designed to deal with generic images, their implications, however, fail for the photographic style portraits. This failure can be observed as they instantly deform facial structures [4, 15, 45, 49, 50]. A human visual system is very sensitive to make over, and due to this reasons such deformations are unacceptable [46, 51], as described by Sinha et al. [39] and McKone et al. [32]. Many efforts have been made for the development of efficient methods for transformation of automatic style [10, 11, 19, 24, 37]. Gatys et al. [15], in his latest study, offered a seminal work of painting texture that moved to the input picture through the VGG convolutional neural network [38]. To address this issue, the proposed method takes an image that matches content and style statistics, which are based on the neural activations of each layer in CNN. This innovative approach gained good quantitative results, and the number of follow-up work is impartial from its improved version [21, 26, 35, 44].

Gatys et al. [15] and Selim et al. [36] presented latest generic painting transfer methods. They transferred it to real-world photographs utilizing the neural networks that captured the style of artistic images nicely. To distinct and reassemble content and style proceeding approach they used advanced level feature illustrations of the images from concealed layers of the VGG-CNNs [38]. This has been done by formulating an optimized problem, that starts with white noise and search new visual, and further shows related neural stimulations while content and similar characteristic associations are the style image. Starts of white noise generate the formulating of an optimized problem desirable hunt new image which demonstrates alike neural activations. However, content and similar feature relationships are the style visuals. Through continual examination, our visual scheme has developed sensitively to an assembly of the human face where slight abnormality will be noticed rapidly. Furthermore, Selim et al. [36], propound a headshot portrait technique for the transfer of the painting style to another image. In order to resolve the portrait painting (through gain maps) in proceeding technique, spatial constraints have been used. This solution stimulates texture to contest against the imperfect capturing of the painting texture and irregularities. On another hand, this approach provides focus on headshot portraits and it also produces results with artifacts only if the face poses are not similar. Additionally, it does not fulfill the photographic style because they use alignment.

Photographic style painting is a challenging problem for researchers these days. Even though they succeeded to generate few worthy results, however, they ignore to maintain the focus on stylization of photographs. Stylization is pretty challenging because a human optical organism is sensitive anti smallest variability in human faces or body structures. Numerous types of distortions are also alternative issues which are also associated with stylization. In [7, 46, 51], researchers attempted to decrease the portrayal of the distorted issue. For this matter, the author exploited the facial geometry of the human face. Partial areas of non-photorealistic rendering are the illustration of human faces which studied for graphite or color sketch and artistic binarization separately [5, 6, 8, 16, 33, 49]. Style-independent methods depend on picture equivalences where an exercise pair picture require original unfiltered along with the creative deception.

In this paper, we propose an automatic photographic painting style technique for a single example image by using the CNN. Current painting system generates to follow the example style in the example image. The proposed system does not restrict headshot images which could be changed the painting style in the wild. Additionally, presented algorithm does not require choosing the same pose and image size between source and target images. For taking the result, we use Gatys et al. [15] construction to transfer the painting texture. Moreover, the algorithm imposes the labeling map for both images and trains it on a single network. It is a novel learning method which simultaneously performs labeling and style transfer. Labeling of the whole image assists to automatic transfer of style to the content image. This motivation has been taken from Selim et al. [36], upon recent success in transferring the head portrait painting.

Currently, on social media such as Facebook, Whatsapp, and Wechat etc., painting style profile/display picture is much popular among the users. Most of the applications simply focus on image painting style. They are not able to focus on example style if a user tries to imply a photographic painting style; these applications deform the facial information. To fill this gap, our proposed technique preserves the semantic information to avoid the facial deformation, and obtains the better result compared to existing techniques. Applying the painting style to the whole image instead of the headshot is another additive characteristic of our proposed technique. In addition, texture transformation is another characteristic to the painting styles while maintaining the facial structures.

Following are the main contributions of proposed work:

- we propose a first approach for single-example photographic painting style in the wild.
- a novel multi-convolutional-learning technique is developed for both image (NPR/PR) labeling, style transfer and optimizing a single interconnected CNN model with weight sharing.
- an efficient and fast full-size labeling style transfer testing method is introduced.

2 Related work

Nowadays, style transfer is an active topic in both academia and industry. Researchers have paid much concentration in creating photo-realistic images of synthetic objects and computer graphics. Modern time NPR (non-photorealistic rendering) methods are classified into four main categories: texture transfer based, part-transfer based, stoke-based rendering, and artistic style transfer via deep learning.

Texture transfer based techniques Texture transfer algorithms receipt two images the source texture and the target image as an input [4, 14, 19, 22, 27, 45], and alters the input image to admiring the follow sample textures. This algorithm modifies the target image and substitutes some high-frequency evidence with the source texture. The ideas from textures synthesis are commonly used [3, 10, 11]. Texture-based systems share common properties with patch-based texture in-filling methods [10, 11]. These methods seek to fill holes in images by searching visually similar patches. In example-based rendering, the patches are unmatched within the source image to be condensed as a substitute within the exemplar source image. Texture in-filling algorithm is sustained a careful balance between fidelity of the patch identical and the spatial coherence in the rendering.

Part-transfer based techniques Following methods parse the input image into diverse portions [5–8, 33]. A painted portions database is queried and used to recreate the ultimate painting. Chen et al. [7, 8] obtained PicToon cartoon system that generates a custom-made cartoon from an input image. An image-based cartoon generator, interactive Cartoon Editor for exaggeration and a speech-driven cartoon animator are the major components of the presented system. Chen et al. [5] recommended example-based approach which could be generated by various styles of portraits. Generation of human portrait sketches from interactive computer system is also suggested by Chen et al. [6]. Meng et al. [33] proposed a technique to render artistic paper-cut of human portraits. To accomplish this, the technique integrates bottom-up and top-down cues to better determine the binary values.

Artistic style transfer via deep learning Recently, neural style transfer [15] has demonstrated remarkable results for image stylization. It takes the advantage of the powerful representation of deep Convolutional Neural Networks (CNN). The goal of this technique is to contest simultaneously the graphic style of first image. In this way, the style of an image can be substituted with one or another without altering the inclusive semantic content of the image. This novel technique attracts many follow-up works for different aspects of improvements and applications. Speedup the iterative optimization procedure in [15], Johnson et al. [21] and Ulyanov et al. [44] trained a feed-forward generative network for fast neural style transmission. To improve the transfer results in [15], different complementary schemes are proposed, including spatial constraints [36]. There are also some extension works to apply neural style transfer to other applications. Ruder et al. [35] incorporated temporal consistence terms by penalizing deviations between frames for video style transfer. Selim et al. [36] proposed novel spatial constraints through gain map for portrait painting transfer.

More recently, to further allow arbitrary style transfer in feed forward networks [20, 29]. The back end idea is to match the statistics of content features at intermediate layers to that of the style features, and then train a decoder to turn features to the image. Non-parametric neural style transfer method is firstly proposed by Li et al. [28]. Another representative work is called Deep Analogy [30], which proposes accurate semantic-level patch match algorithm by considering bidirectional constraint and pyramids refinement. Wang et al. [47, 48] developed an algorithm for human activity recognition in videos with CNN features which has received increasing attention in multimedia understanding. Song et al. [40–42] optimized graph-learning-method which is a general framework, and theoretically, it can be incorporated with other graph-based-learning technique. Zhu et al. [53] produced an algorithm in image analysis; the images are often represented by multiple visual features (also known as multiview features), that aim to better interpret them for achieving a remarkable performance of the learning. Gao et al. [13] developed an algorithm end-to-end framework named aLSTMs, an attention-based LSTM model with semantic consistency, to transfer videos to natural sentences.

3 Overview of proposed approach

We propose an automatic photographic painting style technique for a single example image by using CNN. Generated new painting technique follows the example style in the input image. The proposed technique does not restrict headshot images and has the ability to change the painting style in the wild. Gatys et al. [15] formulation to transfer the painting texture has been used to transfer the style. Our algorithm imposes the labeling map for both images and trains through a single network. It is a novel learning method which simultaneously performs labeling and style transfer. Labeling of the whole image assists to automatic transfer of style to the content image. Experimental result demonstrates that our algorithm achieves satisfactorily, compared to state-of-the-art algorithms. Our experiment also includes the results of hair parsing which are produced consecutively from the unified framework. Existing schemes rarely address hair parsing because it is a challenging area of research.

The proposed method has twofold benefits; firstly, its training process with non-structured losses is as likely competent as the procedure of existing CNNs. Lastly, to learn for compelling labeling consequent advantage of acclaimed strategy, we use transformation under pairwise expression into a logistic loss and semantic image boundaries. Style can be transmitted from different parts of the style image to corresponding fragments in the content image.

3.1 Face labeling convolutional learning

We express the problem of face labeling image X as a conditional random fields (CRFs) model $P(Y|X) = \frac{1}{Z}\exp(-E(Y,X))$; the place Z is the segment function and Y is set of irregular variables; $y_i \in Y$ well-defined on each pixel i. Every variable y_i takes a label from an situated for labels $\ell = 1, 2, .., K$. To study the label-dependency, we present 4-connected-chart (ν, ε) where the place of each node denotes one pixel $i \in \nu$ and edges denote the associations between two contiguous pixels $i, j \in \varepsilon$. Therefore, the CRF model can be communicated as energy function $E(Y, X)$ for two information reliant terms:

$$E(\mathbf{Y}, \mathbf{X}) = \sum_{i \in \mathcal{V}} E_u(y_i, x_i) + \lambda \sum_{(i,j) \in \mathcal{V}} E_b(y_i, y_j, \mathbf{x}_{ij}). \tag{1}$$

The unartistic term $E_u(y_i, x_i)$ measures those effort cost from claiming variable y_i dependent upon those picture patch x_i, medium at the pixel i and the pairwise term $V(y_i, y_j, x_(ij))$ converts those consistency function for contiguous variables y_i, y_j provided for their overlap patch $x_(ij)$. On addition, λ is the mingling constant. We present a multi-class classifier $P_u(y_i = \ell|x_i, \omega_u)$ to express those label cosset to the unartistic term,

$$E_u(y_i, \mathbf{x}_i, \omega_u) = -\log P_u(y_i = \ell|\mathbf{x}_i, \omega_u). \tag{2}$$

On measuring the consistency about two contiguous pixels i, j in the pairwise-term, we present another label $z_(ij) = 1$, if $y_i \neq y_j$ What's more $z_(ij) = 0$, generally. Consequently, those pairwise-term, will be characterized toward the yield of a binary classifier $P_b(z_(ij) = 1|x_(ij), \omega_u)$,

$$E_b(y_i, y_j, \mathbf{x}_{ij}, \omega_b) = -\log P_b(z_{ij} = 1|\mathbf{x}_{ij}, \omega_b). \tag{3}$$

This research, we utilize 9 layers CNN neural system pointed towards unartistic and pairwise classifiers as they give end-to-end predictions denied of handcrafted characteristics. Taking in from claiming CNN parameters ω_u and ω_b mutually with the CRF model will be challenging, since this transform necessities will investigate combinatorial labeling space and expansive parameter space separately. On grip, this problem requires CNNs plan that ought to prepare unartistic and pairwise-terms independently. It has been recognized that both CNNs are based on local image patches that might offer fundamental share of the same features in the lower level layers [38]. Conceivably extensive situated of parameters starting with two CNNs might cause in fitting issues. In this investigation, we recommend a single unified CNN approach which is advantageous for both pairwise and unartistic classifiers. This approach permits to share all the features inside an absolute CNN, two classifiers have the ability will delight in preferred generalization capacity along higher computational effectiveness.

We define a loss functions for a pairwise-classifiers. We represent the parameters of the joint CNN system toward ω, and the characteristic reaction concentrated from those topmost intermediate layer from claiming CNN by $h_i = h(x_i, \omega)$. Thus, output of the pairwise classifier is provided for by a logistic function,

$$P_b(z_{ij} = 1|h_i, \omega_b) = \frac{1}{1 + \exp(-\omega_b^{\top} h_i)}, \tag{4}$$

and accordingly, the logistic loss for pairwise term is

$$L_b(z_{ij}, x_{ij}, \omega, \omega_b) = -log P_b(z_{ij} = 1|h_i, \omega_b). \tag{5}$$

Based on this loss functions (4), we train the unified CNN,

$$\min_{\omega}\{O_b(\omega, \omega_b)\},$$

$$O_b(\omega, \omega_b) = \mathbb{E}\left(\sum_{i,j\in\varepsilon} L_b(z_{i,j}, x_{i,j}, \omega, \omega_b)\right) + \phi(\omega, \omega_b). \tag{6}$$

The place $O_b(\omega, \omega_b)$ will be the required loss to those binary-classifiers through every last one of training samples. In addition, $\phi(\omega, \omega_b)$ would be the regularization expression. That system is updated through gradients of the logistic loss work for back propagation. The span of the output maps may be littler similarly as compared to the original picture, which may be a limit amongst constraints for image testing approach. This debate rose because of the down-sampling strides in the max pooling layers. Mostly existing methodologies make up-sample those low-resolution map of the picture size [12]. A handful of input images are propagated in a forward manner to obtain the output maps possessing the original sizes. These inputs images are produced by moving the original input images with single or multiple pixels, depending on the zooming factor, on both the x and y axes, as portrayed to [34]. A zooming factor for 2, will be an approach how should produce an up-sampled output map.

Our algorithmic fill in build for two pooling layers along down-sampling stride of (2). Likewise 4×4 times of ahead propagations starting with moving input images produce an up-sampled output map for a zooming variable for 4 for a comparable way. Those impacts for convolutional operation might settle on the extent from claiming the inconsistence of original image. To avoid such issues, we get a last output map toward re-scaling of the accurate enter picture size. It will be paramount that 16 passes for forward propagation are substantially quicker over applying the convolutional network to patches during each pixel from claiming image.

3.2 CNN architecture

For the most part, CNNs operates on a patch level focused during each pixel, and those labeling pipeline is based on sliding window information, and the place patches need to be overlapped [43]. We develop a comparative architecture similar but deeper [38] than that of [23], with 7 convolutional and 2 fully connected layers. The inputs are 512×512 absolute scale patches which would be passed through two top sequential convolutional units. The filter about 5×5, is placed in every convolutional layer accompanied toward the one max pooling layer with a down-sampling stride of 2. An alternate stack for little convolutional units for a receptive field of 3×3 without pooling layer is also a part of introducing technique. Each layer is prepared with rectification (ReLU) nonlinearity and local response normalization (LRN) layer separately.

Our network concatenates on extra labeling channels m^l for size M toward the same resolution, computed by down-sampling where static labeling-map is specified as input. The effect is another output for $N + M$ channels, indicating s^l and c^l labels as a need to be every layer individually. In front of concatenation, the semantic channels are weighted toward parameter η to provide an extra user control point:

$$s^\ell = x^\ell \| \eta m^\ell. \tag{7}$$

For style images, the activations to those input images and its labeling-map are concatenated together as s_s^ℓ. For output image, the current activations x^ℓ and the input content's labeling map are concatenated as s^ℓ.

3.3 Photographic painting style transfer

3.3.1 Model

In image processing applications Convolutional Neural Networks (CNN) is a more common use class of Deep Neural Networks. For features extractions, a set of convolutional layers has been used from the input images. Our loss function is derived from Gatys et al. [15] and compares image statistics extracted from a fixed pre-trained descriptor CNN. We use VGG network [38], which are pre-trained for image classification on the ImageNet ILSVRC 2012 data. Our contribution builds on a semantic style transfer approach for photographic painting style transfer. Here, we use optimization to minimize content reconstruction error E_c (weighted by α) and style re-mapping error E_s (weight by β). See Gatys et al. [15] for details about E.

$$E = \alpha E_c + \beta E_s \tag{8}$$

First, we introduce an augmented CNN (see Fig. 2) that incorporates label information, then we define the input label map and its representation, and finally show how the algorithm is able to exploit this additional information.

3.4 Solving for the photographic painting

Patches stand for $k \times k$ are extracted from the labeling layers and denoted by the function ϕ. $\phi(s_s^\ell)$ and $\phi(c_c^\ell)$ stand for the style image patches and content image patches respectively. $\phi(s^\ell)$ and $\phi(c^\ell)$ stand for the current image patches. For any patch i in the current image and layer ℓ, its nearest neighbor $NN(i)$ is computed using normalize cross correlation taking into account weighted labeling map:

$$NN(s_i) := \arg \min_j \frac{\phi_i(s).\phi_j(s_s)}{|\phi_i(s)|.|\phi_j(s_s)|}, \tag{9}$$

and

$$NN(c_i) := \arg\min_j \frac{\phi_i(c).\phi_j(c_c)}{|\phi_i(c)|.|\phi_j(c_c)|}. \tag{10}$$

The style error E_s between all the patches i of layer ℓ in the current image to the closest style patch and content error E_c between all the patches i of layer ℓ in the current image to the closest content patch are defined as the sum of the Euclidean distances respectively:

$$E_s(s, s_s) = \sum_i \|\phi_i(s) - \phi_{NN_{(s_i)}}(s_s)\|^2, \tag{11}$$

$$E_c(c, c_c) = \sum_i \|\phi_i(c) - \phi_{NN_{(c_i)}}(c_c)\|^2. \tag{12}$$

Note that the information from the labeling-map in m^ℓ is used to compute the best matching patches and contribute to the loss value, but it is not part of the derivative of the loss relative to the current pixels; the differences in activation x^1 compared to the style patches cause an adjustment of the image itself via the L-BFGS algorithm [52]. By using an augmented CNN that's compatible with the original, existing patch-based implementations can use the additional labeling information without changes. In fact, the introduction of the η parameter from (7) provides a convenient way to introduce semantic style transfer incrementally.

We modify (8) to account for our new input labeling maps (12) and (11) as follows,

$$E = \alpha \sum_i \|\phi_i(c) - \phi_{NN_{(c_i)}}(c_c)\|^2 + \beta \sum_i \|\phi_i(s) - \phi_{NN_{(s_i)}}(s_s)\|^2. \tag{13}$$

α and β configure the layer preference and, $\alpha = \beta = 0.5$.

We found these values empirically and they are fixed in all experiments.

4 Experiments and results

To get the wanted results, numerous examinations have been attained on a wide assortment about input photograph and sample painting. Input photographs are assembled starting with helen [1] dataset which is publicly accessible. Additionally, Printerest mobile application is used to style era upon accumulation from claiming sample painting [2]. The experiments were performed on a desktop PC with a Intel(R) Core(TM)i7 − 2820QM CPU @ 2.30GHz 2.30 GHz, 8 GB RAM, and NVIDIA GTX 1070 GPU. We also compare the time cost of all these methods. Table 1 gives the average running time of each method on 512 × 512 image pairs. Our method is faster as compared to all other techniques.

Figure 3, we have performed a comparison exclusively with Selim et al. [36]. It can be seen that a misalignment between the example and focus pictures makes awful visual impacts. However, our system effectively transfers the painting style, therefore, there is no need for any arrangement and this solves a constraint of the technique proposed by Selim et al. [36]. Moreover, the recommended strategy for the most part meets expectations for any photographic work of painting style exchange similarly as it is not confined with head

Table 1 Comparison with existing painting style transfer techniques

Method	Gatys et al. [15]	Selim et al. [36]	Liao et al. [30]	Li et al. [28]	Gu et al. [17]	Lu et al. [31]	Ours
Time (s)	390	110	121	191	121	11.5	100

portraits. Consequently, our strategy likewise solves proficiently the opposite constraint of the strategy which perused from Selim et al. [36]. In some cases, the method presented by Selim et al. [36] generate ghosting artifacts for the example painting transfer. As a solution, it is used composting of face over a new background. To achieve this purpose they extracted the matts and synthesized a new background. Whereas, our algorithm, automatically handles this problem by using labeling map. It does not need to transfer the target image into a new background.

Figures 4 and 5 analyze our proposed method with Gatys et al. [15] and Selim et al. [36] respectively. Note that our system does not confine to head portraits and alignment between sample and the target image. We performed a comparison on head portraits for the purpose to evaluate the approach of Selim et al. [36] that is not worth of effort for random photographic styles. Our examination outcomes indicate that our methodology supports the texture by proficiently catching it from painting style superior to the systems by Gatys et al. [15] and Selim et al. [36]. Moreover, it keeps the personality of painting style and exceptionally supports those integument and visual impacts about example pictures. An inconvenience about the shortcoming in composition to a methodology of Selim et al. [36] is observed at rows 1 & 5 of Fig. 4. Furthermore, another disfigurement and unpredictability in the face on row 5 of Fig. 5 is observed from the result for Selim et al. [36]. In Fig. 4, row 1, 3, 4, show the irregularities and shortcoming in texture of the result of Gatys et al. [15] scheme. Shortcoming in composition, texture and integrity of the face is also reflected in row 1, 3, 4 of Fig. 5 . Thus, in the last row for Fig. 5, there are irregularities and deformities are due to powerless upkeep about composition.

Fig. 1 Requisitions of our photographic painting exchange strategies around diverse illustration pictures. The input photos are demonstrated at top left and the instance painting are shown in the insets. A convolutional neural system methodology is also a labeling map on dodge facial deformations. Our approach maintains the identity of transfer images and integrity of a facial structure. To evade facial deformations the proposed approach uses convolutional neural network along with labeling map. Example style paintings by (in a clockwise direction); Francoise Nielly, Paul Wright, Andrew Salgado, Patrick Earle, Gwenn Seemel, Aaron Nagel, and Keith Haring

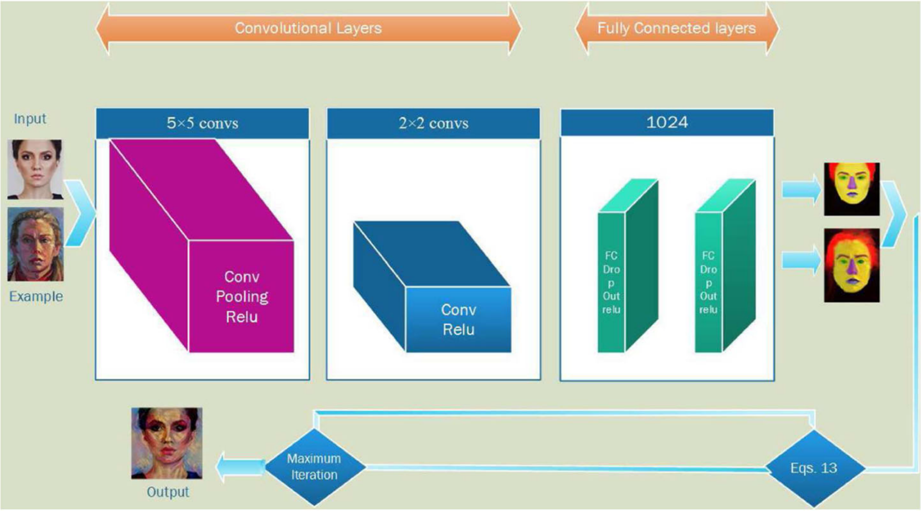

Fig. 2 The algorithm pipeline of our photographic painting style transfer. Face labeling of both photo-realistic and non-phot-realistic images then transfer the photographic painting style to the Input image. This process is terminated at 500 iterations, where results have no visible changes with further iterations. The generated painting resembles the style of the example image and maintains the integrity of the facial structures. Example painting style by Lesley Spanos

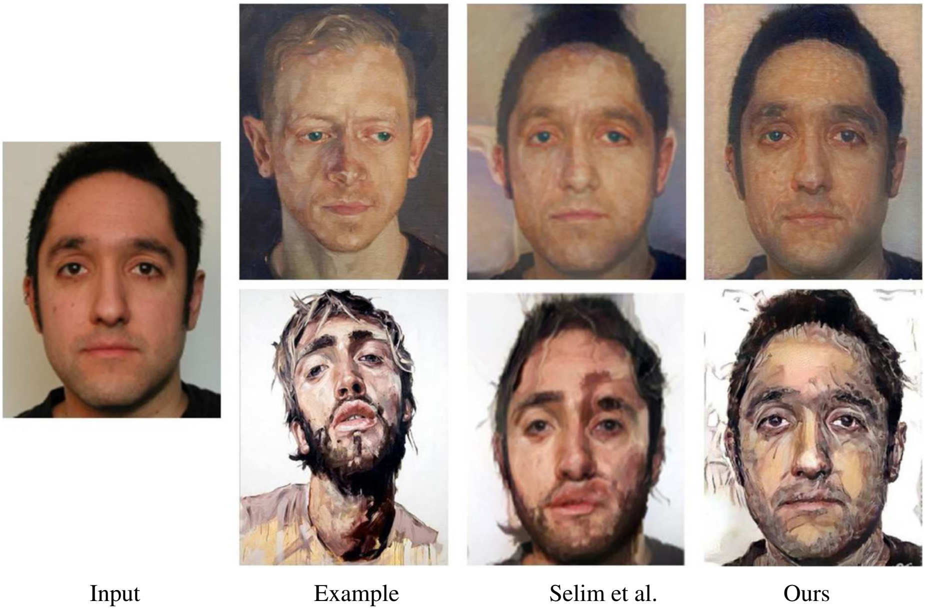

Fig. 3 Photographic painting style transfer in which the example and target images have different alignments. The results in second column is produced by the method of Selim et al. [36] Our results are shown in third column. Paintings style are by (from top): Aaron Nagel and Nick Lepard

Figure 6, a comparison is performed against the result of Gatys et al. [15] scheme. It can be seen that the system of Gatys et al. [15] is unable to transfer painting style where it mainly transfers those color between pictures without preserving the boundary and

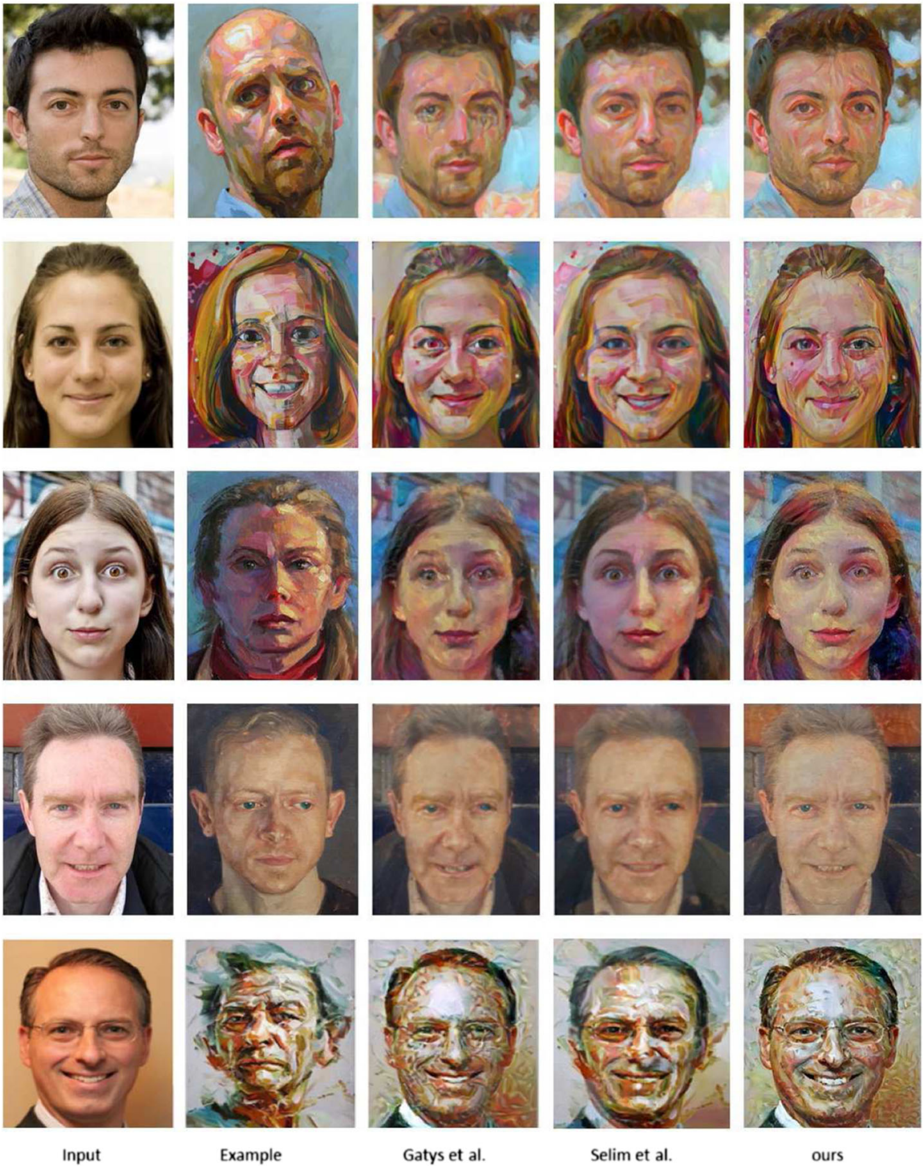

Fig. 4 photographic painting exchange to distinctive input pictures (first column). Here we inspect various cases painting (second column). For every input example pair we show Gatys et al. [15] , Selim el In. [36], along our results. Our approach better captures the composition of the painting style over Gatys et al. [15], Selim et al. [36]. Furthermore it lessen facial deformations as compare of both methodologies. Sample painting need aid by (from top): Paul Wright, Gwenn Seemel, Lesley Spanos, Aaron Nagel, and Paul Wright

background. Irregularities alongside color exchange and shortcoming in texture recommend that our technique is most fine transferring in shade and painting style, at the same time preserving those foregrounds, experiences, and limit as compared with existing techniques.

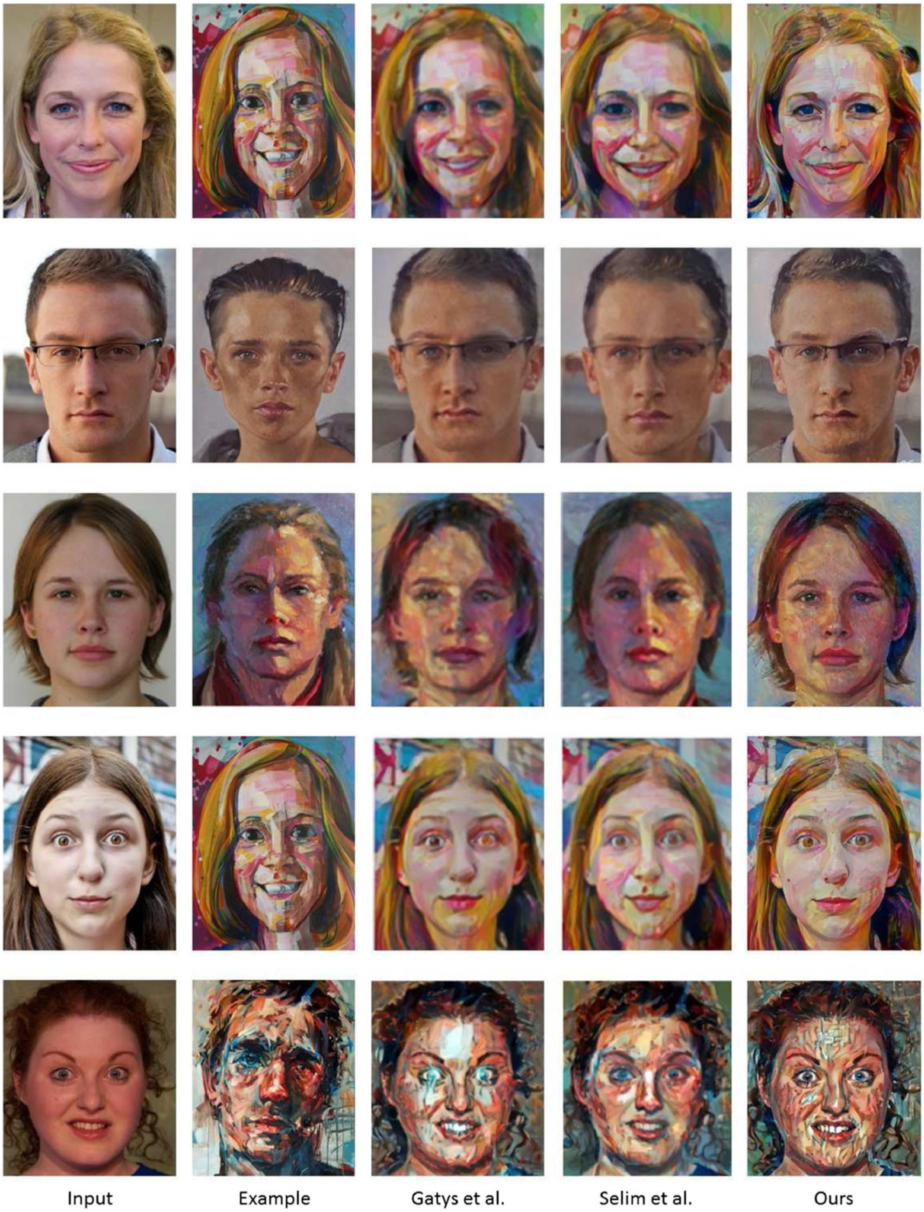

Fig. 5 Photographic painting exchange for different input pictures (First column). Here we analyze various multiple examples painting (2nd column). To every input sample pair we hint at Gatys et al. [15], Salim et al. [36] and our effect. Our methodology superior captures those texture of the painting style over Gatys et al. [15] also Salim et al. [36] additionally, reduces facial deformations through both methodologies. Example painting (starting with top): Gwenn Seemel, Aaron Nagel, Lesley Spanos, Gwenn Seemel and Andrew Salgado

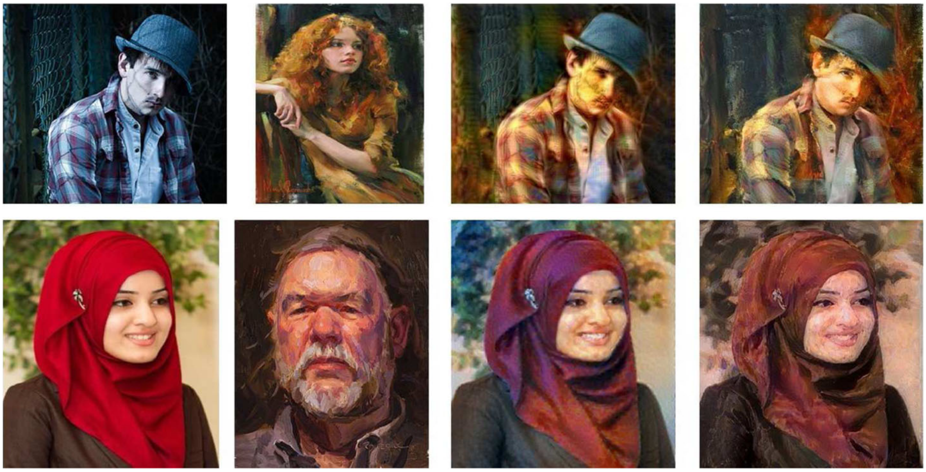

Fig. 6 An examination between our technique against the perusing method of Gatys et al. [15]. Input pictures (Initially column), example painting style (second column) comes by Gatys et al. [15] (third column) and our effects (final one column) would demonstrated. Sample painting style by (from top) Patrick Earle and Andrew Salgado

Fig. 7 Photographic painting exchange for different Input pictures (first column), example painting style (second column) comes by Lu et al. [31] (third column) and our effects (final column) are demonstrated. Our methodology superiorly captures the texture of the painting style over Lu et al. [31]. It additionally reduces facial deformations through the proposed method

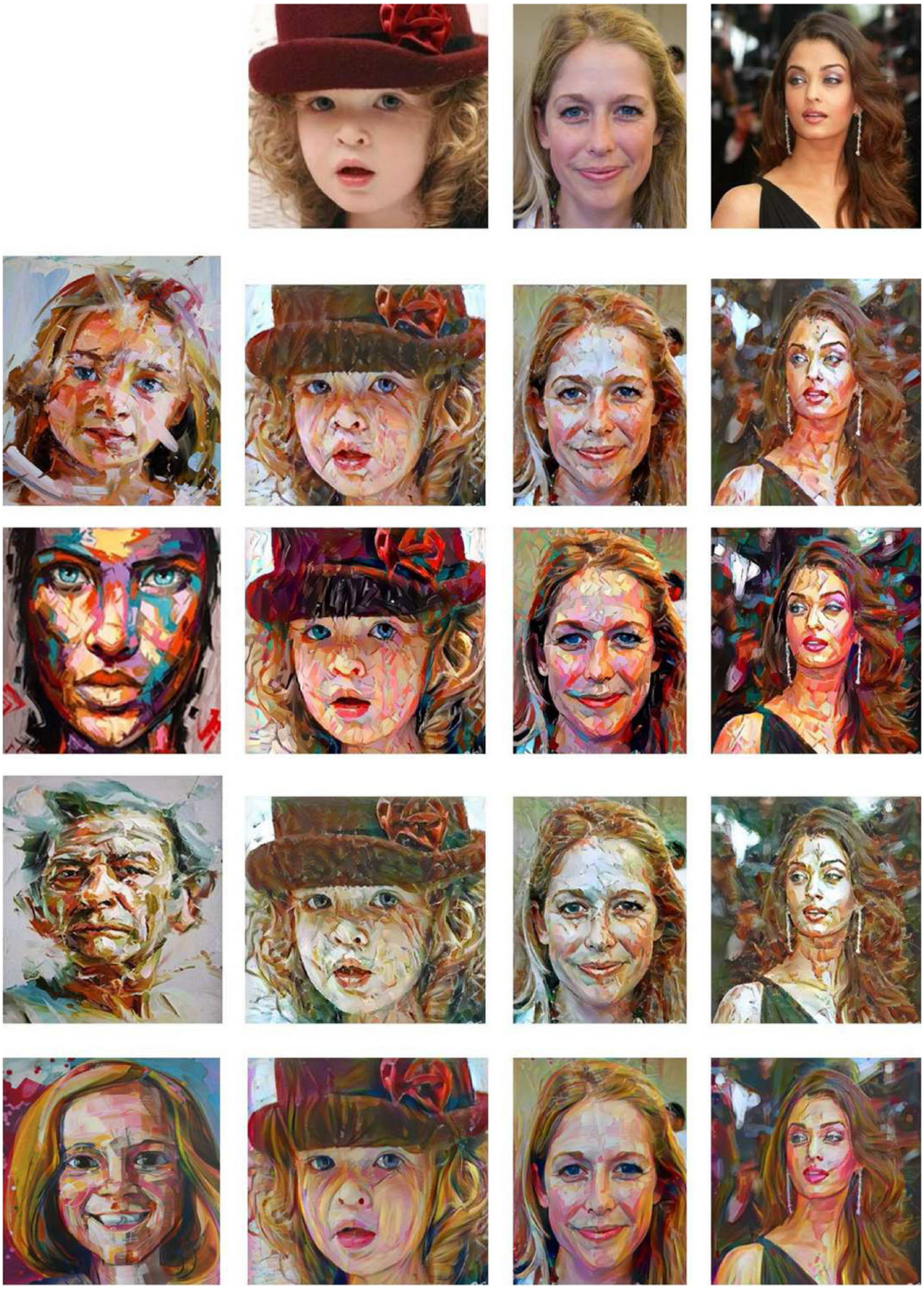

Fig. 8 Painting style transfers around separate information pictures (first row). We inspect various illustration sketches (first column). For every sample our results demonstrated in the same row. Example painting (from top): Patrick Earle, Francoise Nielly, Paul Wright and Gwenn Seemel

In Fig. 7, a comparison is performed against the results of Lu et al. [31] technique. We can see that Lu et al. [31] algorithm has produced some artifacts, row 1, 2, 3 artifacts on face, eyes, and face and neck respectively. On another hand, our algorithm superiorly captures those texture of the painting style over Lu et al. [31] as shown in Fig. 7. We further

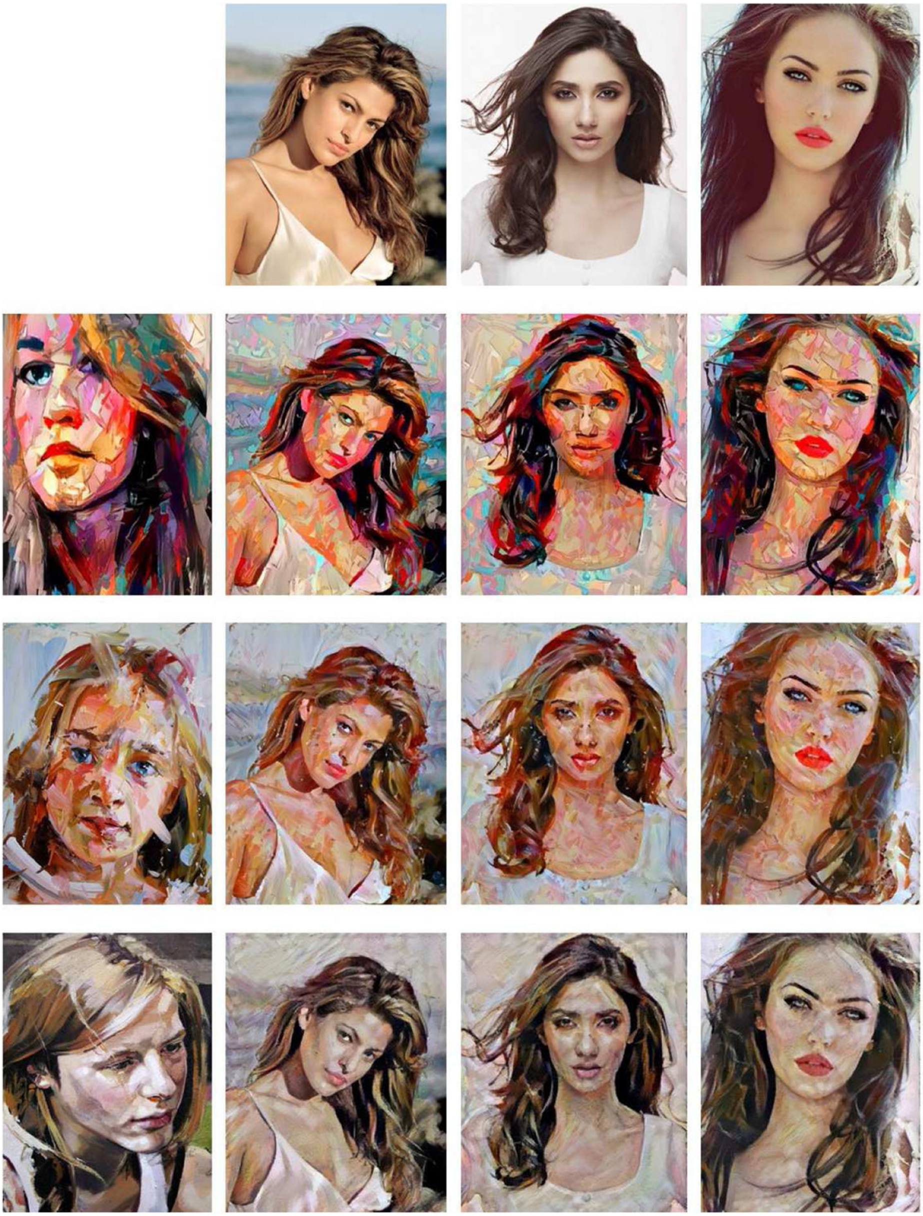

Fig. 9 Painting style transfers on diverse inputs images (1st row). We inspect numerous examples painting (1st column). For example our outcomes are shown in the identical row. Example painting is by (from top): Francoise Nielly, Patrick Earle, and Jenny Saville

can reduce the facial deformations through this methodology. Our findings predict that this approach is widely applicable to various inputs and produces better quality than existing methods.

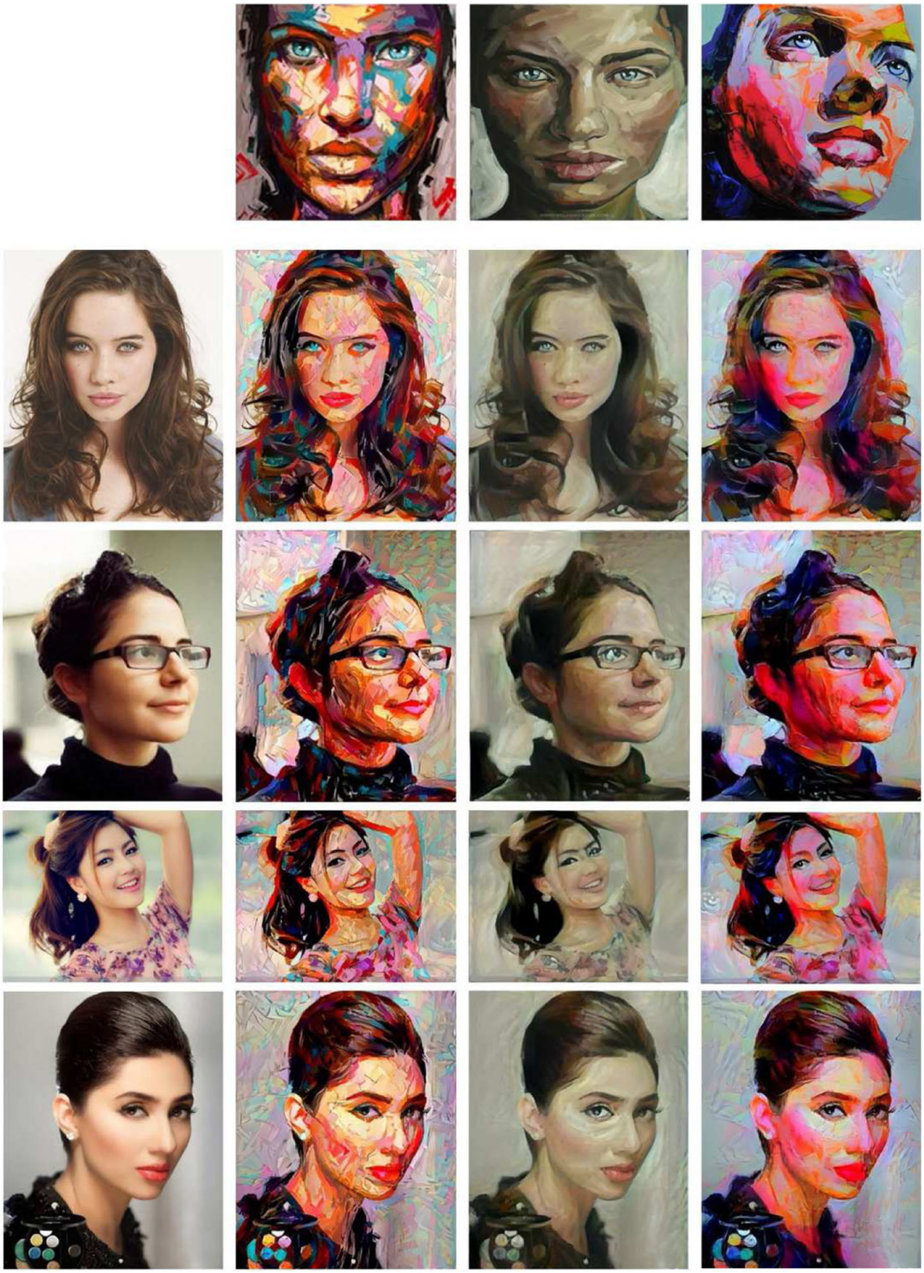

Fig. 10 Painting style transfer on dissimilar example painting (first row) . We inspect numerous inputs images sample (first column). For each example our results are shown in the same column. Example painting are by (from left) : Francoise Nielly, Jenny Saville and Francoise Nielly

Our exhibited approach is much better than as compared with results produced by Gatys et al. [15] and Selim et al. [36], as these are weak methodologies to art exchange or painting transfer. Quantitative values demonstrated that our propound methodology has superior capacity to catch those compositions of the photographic painting styles. Additionally, it offers great administer to maintain the identity of an input photo concerning with existing methodologies [15]. Our approach decreases facial deformations eventually utilizing changed labeling maps, which is an alternate dissimilar trademark for displaying algorithm. Figures 1, 2, 3, 4, 5, 6, 7, 8, 9 and 10, indicates the effects from 28 input photographic style images; in which 7 male and 21 female pictures are placed separately. Additionally, Fig. 1, delineated that our method transfers photographic painting style and support the integrity of the facial structures including input photo preserves its personality victoriously. In Figures 7–10, we applied our algorithm on pictures with side pose portrayal and perceive that our system is suitable for side pose pictures and it produces the outcomes from claiming the same quality with respect to different pictures.

5 Discussion and summary

The results from existing techniques for painting style transfer are failed to maintain the visual effects of the resulting images, and further poorly capture the painting texture. It is equally deformed facial structures which are mainly based on image analogies. In a recent study, researchers get the attention of deep learning in vision, which provides the chance to apply deep learning to artistic style transfer. These methods produce better results on headshot portraits with confinement that the face posed must be similar. Nevertheless, these methods have superiority over previous approaches by means of performance. The proceeding research tried to reduce the portrait deformation issues by exploiting the facial geometry of human. In this research, we study the major drawback of photographic painting style and proposed a new automatic photographic painting style technique through a single example image by using CNN. Our methodology is based on the multi-convolutional-learning technique that involves both images (NPR/PR), labeling, style transfer and optimizing a single interconnected CNN model. We obtained good stylization quality and it is not restricted to headshot images or specific styles, but also able to change the painting style in the wild. Photographic painting style transfer of human face between two images can be further studied for multiple images in future work.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant No. 61202296, 61872153) and the National Science Foundation of Guangdong province No. 2018A030313318.

References

1. Helen (2016) http://www.ifp.illinois.edu/vuongle2/helen/
2. Printerest (2017) https://www.pinterest.com/
3. Ashikhmin M (2001) Synthesizing natural textures. In: Proceedings of the 2001 symposium on interactive 3D graphics. ACM, pp 217–226

4. Ashikhmin N (2003) Fast texture transfer. IEEE Comput Graph Appl 23(4):38–43
5. Chen H, Liang L, Xu Y-Q, Shum H-Y, Zheng N-N (2003) Example-based automatic portraiture. Chinese journal of computers-chinese edition- 26 2:147–152
6. Chen H, Liu Z, Rose C, Xu Y, Shum H-Y, Salesin D (2004) Example-based composite sketching of human portraits. In: Proceedings of the 3rd international symposium on Non-photorealistic animation and rendering. ACM, pp 95–153
7. Chen H, Xu Y-Q, Shum H-Y, Zhu S-C, Zheng N-N (2001) Example-based facial sketch generation with non-parametric sampling. In: Proceedings 8th IEEE international conference on computer vision, ICCV 2001, vol 2. IEEE, pp 433–438
8. Chen H, Zheng N-N, Liang L, Li Y, Xu Y-Q, Shum H-Y (2002) Pictoon: a personalized image-based cartoon system. In: Proceedings of the tenth ACM international conference on multimedia. ACM, pp 171–178
9. Collomosse JP, Hall PM (2005) Genetic paint: A search for salient paintings. In: Workshops on applications of evolutionary computation. Springer, pp 437–447
10. Efros AA, Freeman WT (2001) Image quilting for texture synthesis and transfer. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, pp 341–346
11. Efros AA, Leung TK (1999) Texture synthesis by non-parametric sampling. In: The proceedings of the 7th IEEE international conference on computer vision, 1999, vol 2. IEEE, pp 1033–1038
12. Farabet C, Couprie C, Najman L, LeCun Y (2013) Learning hierarchical features for scene labeling. IEEE Trans Pattern Anal Mach Intell 35(8):1915–1929
13. Gao L, Guo Z, Zhang H, Xu X, Shen HT (2017) Video captioning with attention-based lstm and semantic consistency. IEEE Trans Multimed 19(9):2045–2055
14. Gatys L, Ecker AS, Bethge M (2015) Texture synthesis using convolutional neural networks. In: Advances in neural information processing systems, pp 262–270
15. Gatys LA, Ecker AS, Bethge M (2016) Image style transfer using convolutional neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2414–2423
16. Gooch B, Reinhard E, Gooch A (2004) Human facial illustrations creation and psychophysical evaluation. ACM Trans Graph 23(1):27–44
17. Gu S, Chen C, Liao J, Yuan L (2018) Arbitrary style transfer with deep feature reshuffle. CoRR: 1805.04103
18. Hertzmann A (1998) Painterly rendering with curved brush strokes of multiple sizes. In: Proceedings of the 25th annual conference on Computer graphics and interactive techniques. ACM, pp 453–460
19. Hertzmann A, Jacobs CE, Oliver N, Curless B, Salesin DH (2001) Image analogies. In: Proceedings of the 28th annual conference on Computer graphics and interactive techniques. ACM, pp 327–340
20. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. In: IEEE international conference on computer vision (ICCV)
21. Johnson J, Alahi A, Fei-Fei L (2016) Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. Springer, pp 694–711
22. Kim SY, Maciejewski R, Isenberg T, Andrews WM, Chen W, Sousa MC, Ebert DS (2009) Stippling by example. In: Proceedings of the 7th international symposium on non-photorealistic animation and rendering. ACM, pp 41–50
23. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems 25. Curran Associates, Inc., pp 1097-1105
24. Kwatra V, Essa I, Bobick A, Kwatra N (2005) Texture optimization for example-based synthesis. ACM Trans Graph 24(3):795–802
25. Kyprianidis JE, Collomosse J, Wang T, Isenberg T (2013) State of the art: a taxonomy of artistic stylization techniques for images and video. IEEE Trans Vis Comput Graph 19(5):866–885
26. Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z et al (2016) Photo-realistic single image super-resolution using a generative adversarial network. arXiv: 1609.04802
27. Lee H, Seo S, Ryoo S, Yoon K (2010) Directional texture transfer. In: Proceedings of the 8th international symposium on non-photorealistic animation and rendering. ACM, pp 43–48

28. Li C, Wand M (2016) Combining markov random fields and convolutional neural networks for image synthesis. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 2479–2486
29. Li Y, Fang C, Yang J, Wang Z, Lu X, Yang M-H (2017) Universal style transfer via feature transforms. In: Advances in neural information processing systems
30. Liao J, Yao Y, Yuan L, Hua G, Kang SB (2017) Visual attribute transfer through deep image analogy. ACM Trans Graph 36(4):120:1–120:15
31. Lu M, Zhao H, Yao A, Xu F, Chen Y, Zhang L (2017) Decoder network over lightweight reconstructed feature for fast semantic style transfer. In: 2017 IEEE international conference on computer vision (ICCV), pp 2488–2496
32. McKone E, Kanwisher N, Duchaine BC (2007) Can generic expertise explain special processing for faces? Trends Cogn Sci 11(1):8–15
33. Meng M, Zhao M, Zhu S-C (2010) Artistic paper-cut of human portraits. In: Proceedings of the 18th ACM international conference on multimedia. ACM, pp 931–934
34. Pinheiro PHO, Collobert R (2013) Recurrent convolutional neural networks for scene parsing. CoRR 1306.2795
35. Ruder M, Dosovitskiy A, Brox T (2016) Artistic style transfer for videos. In: German conference on pattern recognition. Springer, pp 26–36
36. Selim A, Elgharib M, Doyle L (2016) Painting style transfer for head portraits using convolutional neural networks. ACM Trans Graph 35(4):129
37. Shih Y, Paris S, Barnes C, Freeman WT, Durand F (2014) Style transfer for headshot portraits. ACM Trans Graph 33(4):148:1–148:14
38. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556
39. Sinha P, Balas B, Ostrovsky Y, Russell R (2006) Face recognition by humans Nineteen results all computer vision researchers should know about. Proc IEEE 94(11):1948–1962
40. Song J, Gao L, Nie F, Shen HT, Yan Y, Sebe N (2016) Optimized graph learning using partial tags and multiple features for image and video annotation. IEEE Trans Image Process 25(11):4999–5011
41. Song J, Guo Y, Gao L, Li X, Hanjalic A, Shen HT (2018) From deterministic to generative: multi-modal stochastic RNNs for video captioning. IEEE Trans Neural Netw Learn Syst. https://doi.org/10.1109/TNNLS.2018.2851077
42. Song J, Zhang H, Li X, Gao L, Wang M, Hong R (2018) Self-supervised video hashing with hierarchical binary auto-encoder. IEEE Trans Image Process 27(7):3210–3221
43. Tompson J, Jain A, LeCun Y, Bregler C (2014) Joint training of a convolutional network and a graphical model for human pose estimation. In: Proceedings of the 27th international conference on neural information processing systems, NIPS'14. MIT Press, Cambridge, pp 1799–1807
44. Ulyanov D, Lebedev V, Vedaldi A, Lempitsky V (2016) Texture networks Feed-forward synthesis of textures and stylized images. In: Interational conference on machine learning (ICML)
45. Wang B, Wang W, Yang H, Sun J (2004) Efficient example-based painting and synthesis of 2d directional texture. IEEE Trans Vis Comput Graph 10(3):266–277
46. Wang T, Collomosse JP, Hunter A, Greig D (2013) Learnable stroke models for example-based portrait painting. In: British machine vision conference
47. Wang X, Gao L, Song J, Shen H (2017) Beyond frame-level cnn Saliency-aware 3-d cnn with lstm for video action recognition. IEEE Signal Process Lett 24(4):510–514
48. Wang X, Gao L, Wang P, Sun X, Liu X Two-stream 3-d convnet fusion for action recognition in videos with arbitrary size and length. 1–1
49. Wang X, Tang X (2009) Face photo-sketch synthesis and recognition. IEEE Trans Pattern Anal Mach Intell 31(11):1955–1967
50. Zeng K, Zhao M, Xiong C, Zhu S-C (2009) From image parsing to painterly rendering. ACM Trans Graph 29(1):2
51. Zhao M, Zhu S-C (2011) Portrait painting using active templates. In: Proceedings of the ACM SIGGRAPH/Eurographics symposium on non-photorealistic animation and rendering. ACM, pp 117–124
52. Zhu C, Byrd RH, Lu P, Nocedal J (1997) Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. ACM Trans Math Softw 23(4):550–560
53. Zhu X, Li X, Zhang S (2016) Block-row sparse multiview multilabel learning for image classification. IEEE Trans Cybern 46:450–461

Asad Khan is currently working as a Post-Doc Fellow and Teaching Instructor at South China Normal University, Guangzhou, China. His research interests include image processing, computer vision, Deep learning, computational photography, Hyperspectral imaging, and Wearable Computing. He received his Bachelor degree in Applied Mathematics from Government College University Faisalabad (GCUF), Pakistan, in 2010, and a Master degree in Mathematical Modeling and Scientific Computing from Air University Islamabad (AIU), Pakistan, in 2012. He joined University of Science and Technology (USTC) in 2014 and obtained a Ph.D. degree in Image and Video Processing in 2017.

Muhammad Ahmad is currently working as Research Fellow and Teaching Instructor at Innopolis University, Innopolis, Russia. His research interests include Image and Signal Processing, Machine learning, Hyperspectral imaging and Wearable Computing. He is an active reviewer for several top tier Journals and IEEE Transactions including but not limited to Springer Nature, IEEE Transactions on Industrial Electronics, IEEE Transactions on Neural Network and Learning Systems, IEEE Geoscience and Remote Sensing Letters, IEEE Geoscience and Remote Sensing Magazine, IEEE Transaction on Multimedia, IEEE Computers, IEEE Sensors, IEEE Access, Neural Computing and Applications, Measurement Science and Technology, Transactions on Internet and Information Systems, and IET Image processing.

Nuzhat Naqvi received her MS degree in Computer Science from Iqra University Karachi in 2017. She is currently doing her PhD degree from Dept. of Electronic Engineering and Information Science; University of Science and Technology, China. She is formerly working in a multinational company named Marubeni-Itochu Tubulars Asia Pte Ltd, Singapore as an IT Manager, which provides services to Oil and Gas Industry. She is presently working on machine learning, object detection and recognition individually. Her main research interest includes Information Security, Steganography, Cryptography, Watermarking, Deep learning, and also have experienced of generation of descriptive image paragraphs.

Faisal Yousafzai is currently the Assistant Professor of Mathematics and Director Students Affairs Office at Military College of Engineering, National University of Sciences and Technology, Islamabad, Pakistan. He holds a PhD Degree in Mathematics from University of Science and Technology of China, China under CAS-TWAS President's Fellowship. He did his MS in Mathematics from COMSATS, Abbottabad, Pakistan under CIIT Merit Scholarships. His research areas are Fuzzy Logics/Algebra and Non-associative semigroups. Dr. Faisal Yousafzai has over 45 international research publications to his credit.

Jing Xiao received PhD degree from the School of Computing, National University of Singapore, Singapore in 2005 and the B. S. and M.S. degrees in computer science from Wuhan University, Wuhan, China, in 1997 and 2000 respectively. She is now a professor at the School of Computer Science, South China Normal University, Guangzhou, China. Her research interests include information retrieval and data mining.