Painting Style Transfer for Head Portraits using Convolutional Neural Networks

Ahmed Selim*
CONNECT center, Trinity College Dublin

Mohamed Elgharib*
Qatar Computing Research Institute, HBKU

Linda Doyle
CONNECT center, Trinity College Dublin

Figure 1: *Our painting transfer for different examples. The input photograph is shown at the top left and the example paintings are shown in the insets. Our approach transfers the example paintings and maintains the input identity. In addition it maintains the integrity of the facial structures. We use a convolutional neural network approach and impose novel spatial constraints to avoid facial deformations. Example paintings by (in clock-wise direction); Michael D. Edens, Patrick Earle, Aaron Mitchell, Linden Holman, Paul Wright, Gwenn Seemel and Ian Fleming.*

Abstract

Head portraits are popular in traditional painting. Automating portrait painting is challenging as the human visual system is sensitive to the slightest irregularities in human faces. Applying generic painting techniques often deforms facial structures. On the other hand 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 limits their domain of applicability. We present a new technique for transferring the painting from a head portrait onto another. Unlike previous work our technique only requires the example painting and is not restricted to a specific style. We impose novel spatial constraints by locally transferring the color distributions of the example painting. This better captures the painting texture and maintains the integrity of facial structures. We generate a solution through Convolutional Neural Networks and we present an extension to video. Here motion is exploited in a way to reduce temporal inconsistencies and the shower-door effect. Our approach transfers the painting style while maintaining the input photograph identity. In addition it significantly reduces facial deformations over state of the art.

Keywords: NPAR, painting transfer, VGG, spatial constraints, Gatys, video, Gain maps, portrait, deformations

Concepts: •**Computing methodologies** → **Non-photorealistic rendering;**

*Joint first author, sorted alphabetically

SIGGRAPH '16 Technical Paper, July 24-28, 2016, Anaheim, CA,
ISBN: 978-1-4503-4279-7/16/07
DOI: http://dx.doi.org/10.1145/2897824.2925968

ACM Reference Format
Selim, A., Elgharib, M., Doyle, L. 2016. Painting Style Transfer for Head Portraits using Convolutional Neural Networks. ACM Trans. Graph. 35, 4, Article 129 (July 2016), 18 pages. DOI = 10.1145/2897824.2925968 http://doi.acm.org/10.1145/2897824.2925968.

1 Introduction

Portrait painting is an important genre in the art world. Its popularity spans a wide space from paintings by street artists to priceless pieces such as Van Gogh self portraits and the Mona Liza. With the rise of non photo-realistic rendering [Kyprianidis et al. 2013], several techniques for automated painting were proposed [Gatys et al. 2015; Zhao and Zhu 2011; Wang et al. 2013; Hertzmann et al. 2001; Hertzmann 1998; Collomosse and Hall 2005; Gooch et al. 2004; Kyprianidis et al. 2013]. However most of these techniques are designed for generic images and hence when applied to head portraits often deform facial structures [Gatys et al. 2015; Zeng et al. 2009; Ashikhmin 2003; Wang et al. 2004; Wang and Tang 2009]. As the human visual system is very sensitive to facial irregularities [Sinha et al. 2006; McKone et al. 2007], such deformations are unacceptable [Zhao and Zhu 2011; Wang et al. 2013].

[Gatys et al. 2015] presented the most recent generic painting transfer technique. Painting texture is transferred to the input image through the VGG convolutional neural network [Simonyan and Zisserman 2014]. The rendered painting is a mixture between two terms; the first maintains identity of the input image while the second transfers the painting texture. The absence of spatial constraints in their second term however leads to poor capturing of the painting texture and generates irregularities. Such deformations are more problematic in head portraits (see Fig. 3, top row).

Few authors attempted to reduce portrait deformations by exploiting the human facial geometry [Gooch et al. 2004; Zhao and Zhu 2011; Wang et al. 2013; Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; DiPaola 2007; Meng et al. 2010]. However most of these techniques are mainly designed for graphite/sketch painting [Gooch et al. 2004; Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Meng et al. 2010]. Limited work is developed for style-independent transfer [Zhao and Zhu 2011; Wang et al. 2013]. However they are based on image analogies where a training image pairs are required; the original unfiltered image and its artistic deception. This limits their applicability.

In the special case of weak painting texture one may consider applying color transfer techniques [Pitie et al. 2005; Reinhard et al. 2001; Bae et al. 2006]. Such techniques transfer the color palette of an example image to an input image. To render visually pleasant results for head portraits different regions require different treatment e.g. forehead, cheeks, chin and so on. [Shih et al. 2014] presented a solution for this by transferring the local statistics through a multi-scale image representation. Here the notion of power maps is used akin to [Bae et al. 2006; Li et al. 2005; Su et al. 2005]. This captures the local spatial distribution of light sources on the face. Their technique is tailored to color transfer for head portraits and generates impressive results. However it fails with painting transfer due to the strong texture commonly available in painting styles.

We present a technique for head portrait painting transfer. Given an input photograph and an example portrait painting, a new painting is generated that follows the example style (see Fig. 1). Unlike previous work our technique is not restricted to a specific style and is single example driven, not requiring the original unpainted version of the example image. We use [Gatys et al. 2015] formulation to transfer the painting texture. We impose novel spatial constraints during transfer using the notion of gains maps [Bae et al. 2006; Li et al. 2005; Su et al. 2005]. This is motivated by [Shih et al. 2014] recent success on transferring the local statistics during color transfer. Our gain maps are applied on VGG features [Simonyan and Zisserman 2014] and captures the local spatial color distribution of the example painting. This better transfers the texture of the painting styles while maintaining the facial structures. In addition it reduces facial deformations over state of the art while maintaining input photograph identity (see Fig. 3, bottom row). We extend our technique to image sequences by incorporating motion information. This generates temporally coherent results and reduces the shower-door effect. We examined our approach on a wide variety of painting styles. We also examined different forms of motion.

Aspects of novelty in this work include:

1. The first approach for single-example head portrait painting not constrained by a specific style.
2. Novel spatial constraints for portrait painting through gain maps.
3. Handling videos for head portrait painting by exploiting motion information.

The next section reviews state of the art followed by a detailed discussion of [Gatys et al. 2015]. We then present our painting algorithm and show how it significantly reduces deformations over state of the art. Results are presented, extension to video are proposed followed by conclusion.

2 Related Work

2.1 Painting Transfer

Painting transfer is a subset of a computer graphics topic known as non photo-realistic rendering (NPR) [Kyprianidis et al. 2013]. Current painting techniques can be classified into three main categories: stroke based, texture transfer based and parts transfer based. Stroke based techniques [Zeng et al. 2009; Hertzmann 1998; Hertzmann 2001; Litwinowicz 1997; Lu et al. 2010; Collomosse and Hall 2002; Collomosse and Hall 2005; Lin 2010; Hays and Essa 2004; Meier 1996; Haeberli 1990; Gooch et al. 2002; Zhao and Zhu 2011; Wang et al. 2013] render the painting by simulating the brush-stroke placement process. Here different brush attributes such as its scale, orientation, color and opacity often guide the placement process. Texture transfer based techniques [Hertzmann et al. 2001; Lee et al. 2010; Ashikhmin 2003; Kim et al. 2009; Wang et al. 2004; Gatys et al. 2015] modifies the input image in a way to follow sample textures. Here ideas from textures synthesis are commonly used [Ashikhmin 2001; Efros and Leung 1999; Efros and Freeman 2001]. Parts transfer based techniques [Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Chen et al. 2001; Meng et al. 2010] are based on parsing the input image into different parts. A database of painted parts is then queried and used to reconstruct the final painting. Most painting transfer techniques are designed for generic images and when applied to head portraits generate facial deformations. An example of such techniques is [Gatys et al. 2015], which is the most recent in literature. Sec. 3 discusses this technique in details.

Few authors attempted to address painting deformations by exploiting the shape of the examined objects. For head portrait facial landmarks guide the painting process [Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Meng et al. 2010; Zhao and Zhu 2011; Wang et al. 2013]. Facial landmarks are extracted through either active appearance model (AAM) [Cootes et al. 1998] or active shape model (ASM) [Cootes et al. 1995], and they consist of eyes, eyebrows, nose, mouth and the facial outline. In [Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Meng et al. 2010] parts based techniques for painting transfer are proposed. Here non parametric sampling reconstructs the painting from a database of painted landmarks. Furthermore in [Chen et al. 2002b; Chen et al. 2002a] pels are modeled as MRFs and spatial

smoothness is imposed on the reconstructed painting. Despite interesting results being generated, most portrait techniques are designed for graphite/sketch painting [Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Meng et al. 2010].

Limited work exist for head portrait painting that is not constrained by specific styles [Zhao and Zhu 2011; Wang et al. 2013]. Developed techniques are based on image analogies [Hertzmann et al. 2001] where a set of original images and their corresponding artistic deception is required. [Zhao and Zhu 2011] renders a painting by transferring the strokes from an example painting to an input photograph. The input photograph is first matched with its nearest neighbor in a database of image analogies. The strokes of the corresponding painting are then warped over the input photograph using thin-plate spline (TPS) transformation [Barrodale et al. 1993]. [Wang et al. 2013] approach also transfers the strokes from an example painting. However they learn a relation between strokes attributes and the unpainted version of the example image. The learning phase uses dense sift features [Lowe 1999] for BoVW representation of the facial features. Pels of the reconstructed painting are modeled as MRFs and spatial smoothness is imposed on the transferred strokes. Interesting results are generated for both [Zhao and Zhu 2011; Wang et al. 2013]. However the image analogies requirement limits their domain of applicability since we often do not have the original unpainted version of the the example image.

2.2 Color Transfer

Color transfer techniques transfer the color palette of an example image to an input image [Pitie et al. 2005; Reinhard et al. 2001; Bae et al. 2006]. They do not require the unfiltered version of the example image and hence unlike painting techniques [Zhao and Zhu 2011; Wang et al. 2013] they are not based on image analogies. Recently [Shih et al. 2014] presented a solution tailored for head portraits. Their approach borrows the notion of power maps from [Bae et al. 2006; Li et al. 2005; Su et al. 2005] to robustly transfer local statistics. This captures the local spatial distribution of light over the face and hence generates a visually pleasing and artistic effect. Spatial constraints are imposed by warping the example image over the input. Similar to [Chen et al. 2001; Chen et al. 2002b; Chen et al. 2002a; Chen et al. 2004; Meng et al. 2010; Zhao and Zhu 2011; Wang et al. 2013] facial landmarks are extracted to guide the warping process. Impressive results are generated. However attempting to transfer painting through [Shih et al. 2014] often fails as it does not capture the strong texture commonly present in painting styles.

2.3 Extension to video

Applying painting transfer frame by frame independently often generate two main artifacts; flickering and shower-door. Flickering occur due to temporal inconsistencies in the transferred texture. Shower door is an effect where the transferred texture drifts away from the underlying object. In such effect the video appears as if it is viewed through a painted glass window. Several authors attempted to address those issues by exploiting the motion of the original input sequence [Hertzmann 2001; Litwinowicz 1997; Lu et al. 2010; Lin 2010; Hays and Essa 2004; Meier 1996; Hashimoto et al. 2003; O'Donovan and Hertzmann 2012; Wang et al. 2010]. In [Hertzmann 2001; Litwinowicz 1997; Lu et al. 2010; Lin 2010; Hays and Essa 2004] strokes are propagated along the estimated motions. In [Lin 2010] flickering is further reduced by gradually fading out the strokes that are deviating from the input sequence. None of these techniques however are designed for head portraits.

[Shih et al. 2014] proposed a video extension for their color transfer technique. A database of several painting examples is searched to find the best match to the first frame. Style is transferred to the first frame and the resulting style representations are propagated to the remaining frames through optical flow. Transfer is performed on the remaining frames. This generates temporally coherent results.

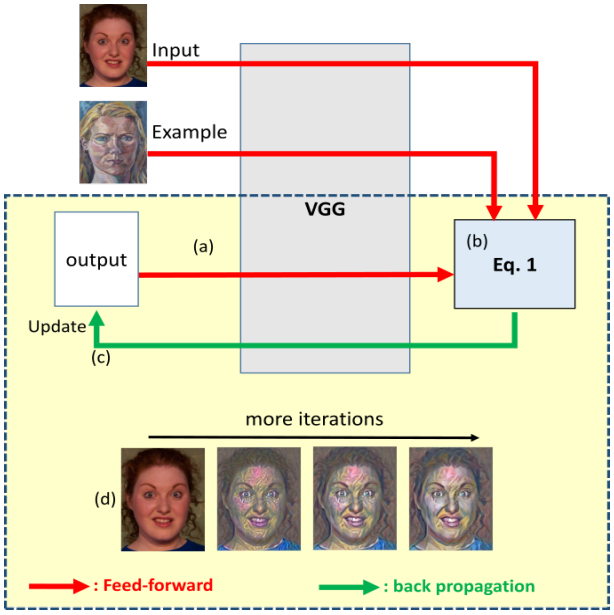


Figure 2: *Overview of Gatys et al. An input photograph and an example painting are fed to a VGG network. Eq. 1 is then minimized using gradient descent. The solution (output) is initialized with the input photograph and updated iteratively. For each iteration: (a) the feature maps are extracted in a feed-forward step (b) Eq. 1 is minimized and (c) error gradients are back-propagated through the network to update the input. (d) shows the output at (from left) iteration* 0, 30, 150 *and* 300 *respectively. Example painting by Patrick Earle.*

3 Painting Through Convolutional Neural Networks (CNN)

Deep Neural Networks attracted the attention of the research community for some time now [Krizhevsky et al. 2012; Taigman et al. 2014]. For image processing applications Convolutional Neural Networks (CNN) is the most commonly used class of Deep Neural Networks. They consist of a set of convolutional layers for extracting features from the input images. In the training stage CNNs learn image representations that become more abstract along the processing hierarchy. Recently [Gatys et al. 2015] introduced an approach for painting transfer by using the CNN representations (see Fig. 2). Here they used the VGG network architecture (see Fig. 4). Their approach estimates an output painting O by transferring textures of the example painting E while being constrained by the input image I. This is achieved by minimizing the following loss function

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \alpha_l \mathcal{L}_1^l + \Gamma \sum_{l=1}^{L} \beta_l \mathcal{L}_2^l, \tag{1}$$

$$\mathcal{L}_1^l = \frac{1}{2N_l D_l} \sum_{ij} (F_l[O] - F_l[I])_{ij}^2, \tag{2}$$

$$\mathcal{L}_2^l = \frac{1}{2N_l^2} \sum_{ij} (F_l[O]F_l[O]^T - F_l[E]F_l[E]^T)_{ij}^2. \tag{3}$$

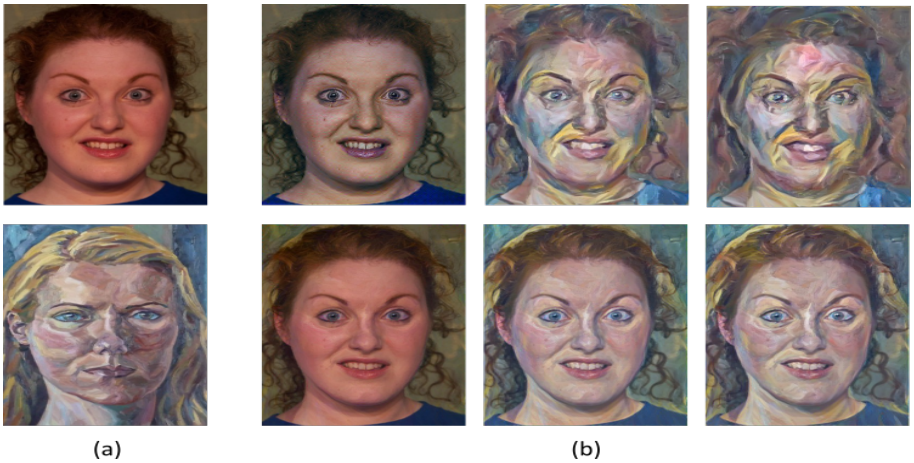


Figure 3: *(a) shows an input photograph (top) and an example painting (bottom). (b) shows the corresponding painting transfer using Gatys et al. (top row) and our approach (bottom row). Each column shows the result at different* Γ *(from left* $\Gamma = 1, 10, 100$ *respectively). The lack of spatial constraints in Gatys et al. (Eq. 1 second term) often leads to facial deformations. Our novel spatial constraints avoids such deformations. This transfers the painting style while maintaining the input identity and the integrity of the facial structures. Example painting by Patrick Earle.*

At the lth convolutional layer, there are N_l filters each with a vectorized feature map of size D_l (D_l is the number of elements in the filter response). This creates the feature matrix $F_l[.] \in R^{N_l \times D_l}$. Here (i, j) indexes the feature matrix. $(F_l[I], F_l[E], F_l[O])$ are the feature matrices of the input photograph I, the example painting E and the desired output painting O at the lth convolutional layer. L is the total number of layers in the examined network, where (α_l, β_l) are weights to configure layer preferences. Γ is a weight that compromises between the integrity of the input image (Eq. 1, first term) and the amount of painting texture transfer (Eq. 1, second term).

To estimate the output painting, image O is initialized with an arbitrary image, which we fix to the input photograph in all our experiments. The feature maps are extracted in a feed-forward step and gradient descent is used to minimize Eq. 1. Error gradients are then back-propagated through the network to update the input (see Fig. 2). This process is applied iteratively until a maximum number of iterations is reached. In all our experiments we fix this to 300 iterations. More detail on feed-forward and back-propagation optimization can be found in [Mahendran and Vedaldi 2015].

Fig. 3 (top row) shows typical painting result of [Gatys et al. 2015] at different values of Γ. Larger Γ represents more painting due to Eq. 1, second term. The absence of spatial constraints in this term however often leads to structural deformations. Such deformations are problematic in head portraits as our human visual system is sensitive to the slightest facial irregularities [Sinha et al. 2006; McKone et al. 2007].

4 Example-Driven Spatial Constraints for Portrait Painting

Fig. 5-6 show an overview of our algorithm. We seek to avoid facial deformations during painting transfer while mainlining the identity of the input photograph. First we align the example painting on the input photograph. Using the notion of gain maps ([Shih et al. 2014; Bae et al. 2006; Li et al. 2005; Su et al. 2005]) we modify the photograph feature maps to account for the spatial color variations in the painting example. This imposes spatial constraints during painting transfer and prevents facial deformation (see Fig. 6, a-d). Optionally we composite the face over a new background to avoid ghosting from the example painting (see Fig. 5, Stage B). Here [Levin et al. 2008] matting is used. The rest of this section discusses our modified input photograph features and propose our solution for the final painting.

4.1 Modified feature maps

Without loss of generality, in the rest of this paper we will consider the VGG network [Simonyan and Zisserman 2014] (see Fig. 4). Our approach focuses on the features generated at the convolutional layers only[1] (denoted by l). We propose modified feature maps to transfer the local color distributions of the example painting E onto the input photograph I. For this we modify input features $F_l[I]$ using the gain maps G_l as follows

$$F_l[M] = F_l[I] \times G_l, \tag{4}$$

$$G_l = \frac{F_l[E]}{F_l[I] + \epsilon}, \tag{5}$$

ϵ avoids division by 0 and is fixed to 10^{-4}. Note that Eq. 4 generates $F_l[M] \approx F_l[E]$ which kills the input identity. On the other hand enforcing the gain maps G_l to be close to 1 maintains the input image while killing the painting style. Hence to reach a balance between the painting style and input identity we clamp G_l through

$$G_{l\,\text{clamped}} = \max(\min(G_l, g_{\text{max}}), g_{\text{min}}) \tag{6}$$

Here $(g_{\text{min}}, g_{\text{max}})$ are the minimum and the maximum gains to control the amount of feature transfer from the example painting. $(g_{\text{min}}, g_{\text{max}}) = (0.7, 5)$ in all experiments (unless stated otherwise).

4.2 Solving for the painting

We modify Eq. 1 to account for our new input features maps $F_l[M]$ (Eq. 4) as follows

[1] When referred to feature maps at a certain convolutional layer, we mean the features generated after the linear filtering followed by the rectified linear units (ReLU).

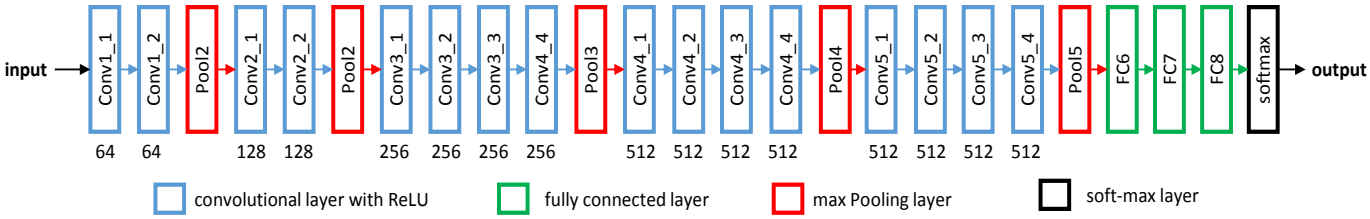


Figure 4: *Architecture of the VGG network. The number of filters is written under each convolutional layer.*

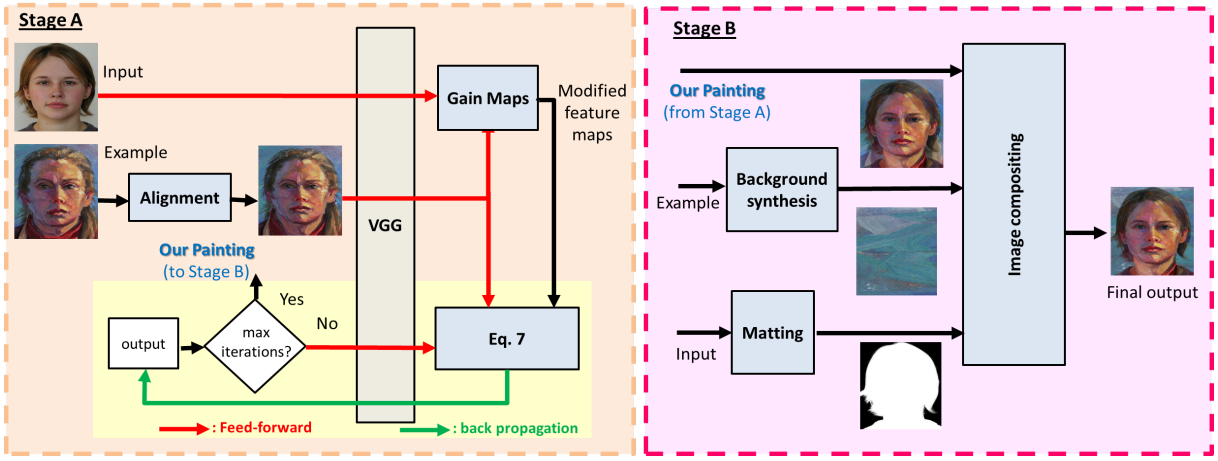


Figure 5: *Overview of our painting transfer. An example painting is first aligned over the input photograph. VGG features are extracted and gain maps are estimated (Eq. 5). This generates modified feature maps (Eq. 4). Eq. 7 is minimized iteratively using feed-forward and back-propagation as in Sec. 3 (third paragraph). This process is terminated when a maximum number of iterations is reached (set to 300 in all experiments). The generated painting resembles the style of the example image and maintains the integrity of the facial structures. We optionally remove possible ghosting from the example painting by compositing the face over an arbitrary background. Laplacian matting of Levin et al. is used here. Example painting by Lesley Spanos.*

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \left(\alpha_l \frac{1}{2N_l D_l} \sum_{ij} (F_l[O] - F_l[M])_{ij}^2 \right)$$

$$+ \Gamma \sum_{l=1}^{L} \left(\beta_l \frac{1}{2N_l^2} \sum_{ij} (F_l[O]F_l[O]^T - F_l[E]F_l[E]^T)_{ij}^2 \right) \quad (7)$$

$F[O]$ is the VGG features of the desired output painting O. Γ configures the level of texture transfer (second term) and is fixed to 100 in all experiments. α_l and β_l configure the layer preference and are switched ON for layers 3 and 4 only (conv3_1 and conv4_1). $\alpha_l = \beta_l = 0.5$ for $l \in \{3, 4\}$. We found this values empirically and they are fixed in all experiments.

Our technique performs three main tasks: 1. preserves the input identity 2. transfers the local spatial color distribution of the painting style and 3. reduces deformations during texture transfer (from Eq. 7, second term). Those tasks are achieved through our gain maps (Eq. 7, first term). However since the gain G estimates local statistics between the input photograph I and the example painting E, the examined sites need to be aligned. We detect facial landmarks in both images using [Saragih et al. 2009]. This generates 66 points describing the lower facial contour, eyes, eyebrows, nose and mouth. Using the facial landmarks we align the example E over the input I with image morphing [Beier and Neely 1992] followed by sift-flow [Liu et al. 2008]. The former aligns the landmarks such as the eyes and mouth. Sift-flow often aligns the facial contour.

We minimize Eq. 7 using a combination of feed-forward and back-propagation, similar to Sec. 3 (third paragraph). We minimize Eq. 7 iteratively until we reach a maximum number of iterations. This is set to 300 iterations in all experiments.

Fig. 6 shows the process for generating the modified features maps. Fig. 6 (a) shows the impact of the gain maps on the input photograph. Fig. 6 (a) is equivalent to minimizing Eq. 7 with $\Gamma = 0$. (a) resembles the local color distribution of the example painting; the brightness of one half of the face is different from the other half. This modified input biases the final painting in a way that maintains the integrity of the facial structures (see Fig. 6-b). For this we minimize Eq. 7 with the default non-zero Γ value. Our results compare favorably with [Gatys et al. 2015] which generates significant deformations (Fig. 6-d).

5 Results

We performed experiments on a wide variety of input photographs and example paintings. The input photographs are collected from Shih et al. [Shih et al. 2014]. The example paintings is a collection of various styles generated by different artists. We have generated 108 different paintings. This manuscript contains 56 of them. The remaining paintings are in the supplementary material.

We compared our algorithm against [Gatys et al. 2015] which is the most recent approach for painting transfer. We compared against two variations: Gatys et al. and Gatys et al. +. In the former we use the parameters and network design recommended by the original

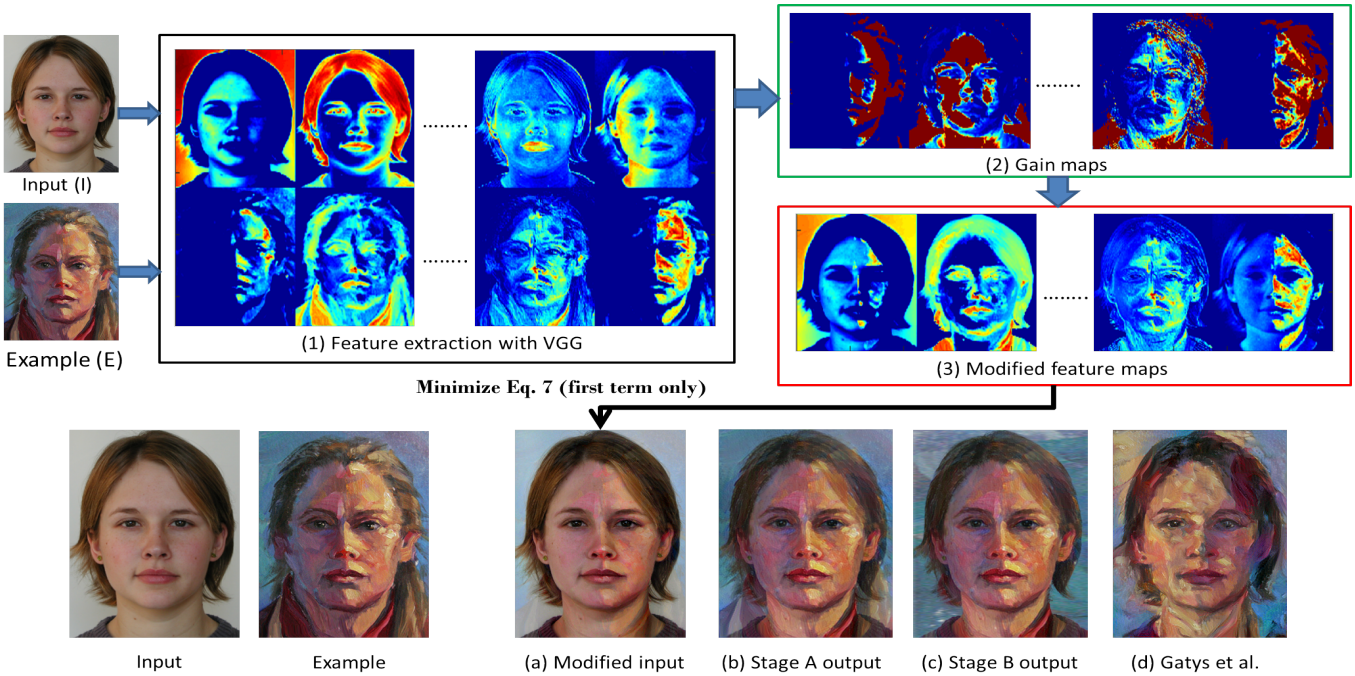


Figure 6: *Estimating our gain maps (Eq. 5) and its impact on painting transfer. (1) Feature maps are first extracted for the input photograph and the example painting. The VGG network is examined at different scales (i.e. convolutional layers). For simplicity we only show one layer. (2) Gain maps (Eq. 5) are estimated and (3) the input photograph features are modified (Eq. 4). The modified features transfer the local color distribution of the painting example onto the input photograph. For instance here half of the example face is brighter than the other, while all the input face is of similar brightness. Applying the gain maps on the input photograph makes half of its face brighter than the other, similar to the example painting (see (a) Modified input). This biases our final painting to maintain the integrity of the facial landmarks (see (b)). We use matting to composite the face on a new background (c). This removes example painting ghosting. Our approach transfers the painting style and reduces facial deformations significantly over state of the art Gatys et al. (d). Example painting by Lesley Spanos.*

work of [Gatys et al. 2015]. The latter version however has the following modifications:

- The original, un-normalized VGG network with max pooling is used [Simonyan and Zisserman 2014]
- The image size of the style and content is around 450^2 pixel in total
- The content weight = 1 and style weight = $1e3$
- The optimization is run for 1000 iterations
- The publicly available implementation of [Johnson 2015] is used

In both comparisons our techniques better captures the texture of the painting styles while maintaining the identity of the input photograph (Fig. 9-11). In addition our modified feature maps reduces the facial deformations over both approaches.

We extended our technique to image sequences by modifying the alignment step (Sec. 5.2). We performed experiments on different variations of head movements including large translation, tilt and rotation (see Fig. 19-20). We also processed a wide variety of facial expressions that are temporally impulsive. We compared against applying the still image algorithm and results show our temporal modification generates more temporally coherent results and reduces the shower door effect. For video we examined a total of 5 styles on 4 different people (2 male, 2 female). In total we processed around 3000 frames. All videos are available in the supplementary material.

5.1 Still Images

Fig. 7-8 show the result of applying our algorithm on 6 input photographs; 3 male and 3 female. For each input we transfer the style of the example painting (left-most column). Fig. 7-8 show that our algorithm transfers the example painting while maintaining the integrity of the facial structures. In addition it keeps the identity of the input photograph.

Fig. 9-11 compare our approach against Gatys et al. and Gatys et al. +. Our modified input features (Sec. 4.1) imposes spatial constraints during painting transfer. This better captures the texture of the painting styles and maintains the integrity of the facial structures over state of the art. Furthermore the lack of spatial constraints in Eq. 1 (second term) leads to facial deformations in Gatys et al. and Gatys et al. +. For instance in Fig. 9 (rows 1,2) and Fig. 11 (rows 2,3) Gatys et al. generates irregularities around the eyes. In addition it generates irregularities around the cheeks/eyes in Fig. 10 (rows 3,5). Gatys et al. + generates irregularities around the forehead in Fig. 9 (row 2) and Fig. 10 (rows 1,3), around the cheeks/eyes in Fig. 10 (row 5) and on the left eye in Fig. 11 (row 2). In addition it generates blue beard and blue hair in Fig. 11 (row 4).

Fig. 12 shows the modified input for Fig. 8 (third column and last column). Here different clamping factors are examined where $(g_{\text{min}}, g_{\text{max}}) = (0.7, 5)$ generates the best balance between input integrity and painting texture.

Handling Ghosting from the Example Paintings: Fig. 13 shows cases where the example image bleeds into the final painting. This

Figure 7: *Our painting transfer on different input photographs (first row). We examine multiple example paintings (first column). For each example our results are shown in the same row. Example paintings are by (from top): Paul Wright, Aaron Mitchell, Aaron Nagel and Bill Sharp (two rows).*

Figure 8: *Our painting transfer on different input photographs (first row). We examine multiple example paintings (first column). For each example our results are shown in the same row. Example paintings are by (from top): Patrick Earle, Andrew Salgado, Elizabeth Hristova - Lisa, Lesley Spanos and Gwenn Seemel.*

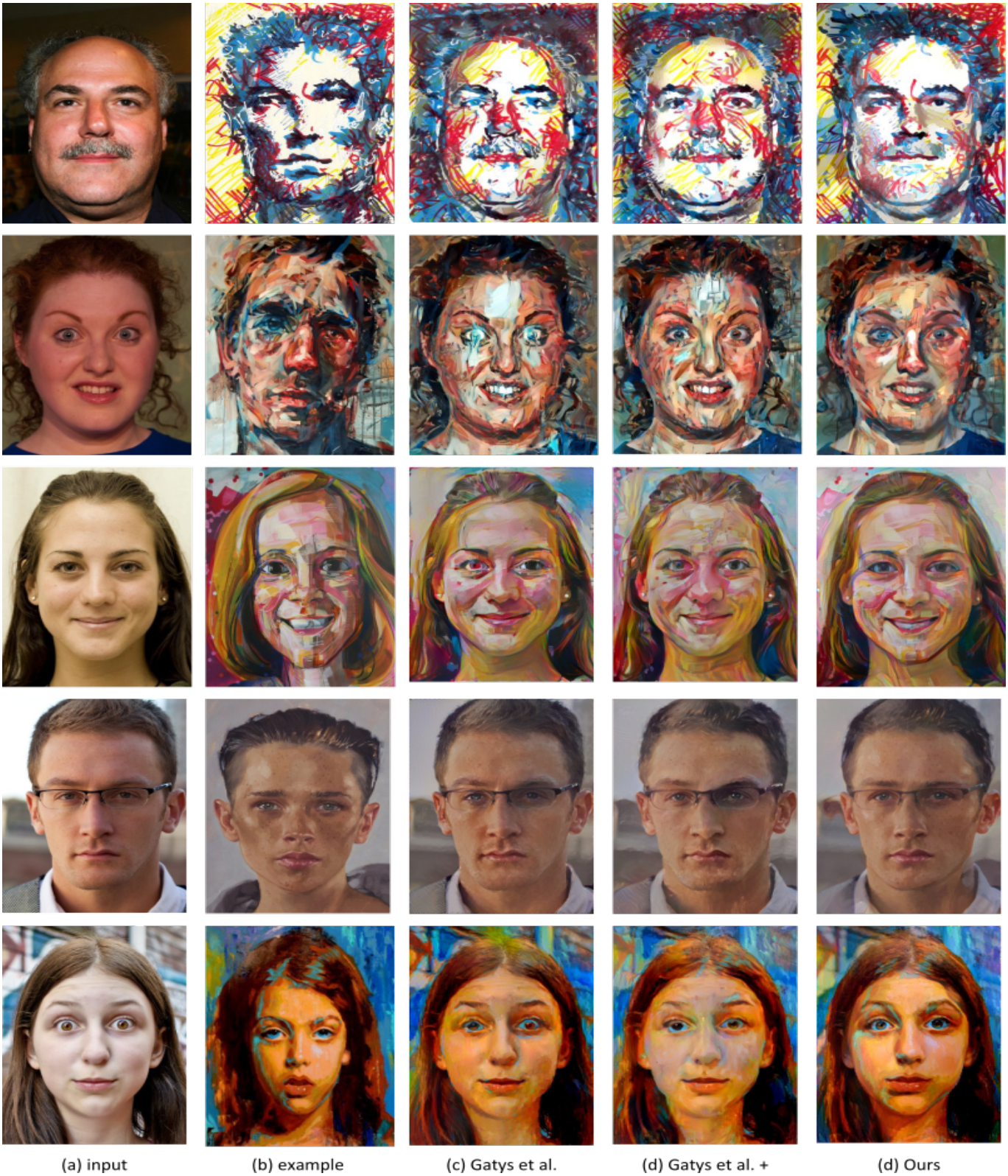


Figure 9: *Painting transfer for different input photographs (first column). Here we examine multiple example paintings (second column). For each input-example pair we show Gatys et al., Gatys et al. + and our results. Our approach better captures the texture of the painting styles than Gatys et al. and Gatys et al. +. In addition it reduces facial deformations over both approaches. For instance Gatys et al. generates irregularities around the eyes in rows 1,2. Gatys et al. + generates irregularities around the forehead in row 2 and around the cheeks in row 5. In addition it generates a red nose in row 3. Example paintings are by (from top): Michael D. Edens, Andrew Salgado, Gwenn Seemel, Aaron Nagel and Elizabeth Hristova - Lisa.*

Figure 10: *Painting transfer for different input photographs (first column). Here we examine multiple example paintings (second column). For each input-example pair we show Gatys et al., Gatys et al. + and our results. Our approach better captures the texture of the painting styles than Gatys et al. and Gatys et al. +. In addition it reduces facial deformations over both approaches. For instance Gatys et al. generates irregularities around the forehead in row 2 and around the cheeks/eyes in rows 3,5. Gatys et al. + generates irregularities on the forehead in rows 1,3, around the cheeks/eyes in row 5 and on the mouth corner in row 4. Example paintings are by (from top): Patrick Earle, Bill Sharp, Henry Lamb, Aaron Nagel and Lesley Spanos.*

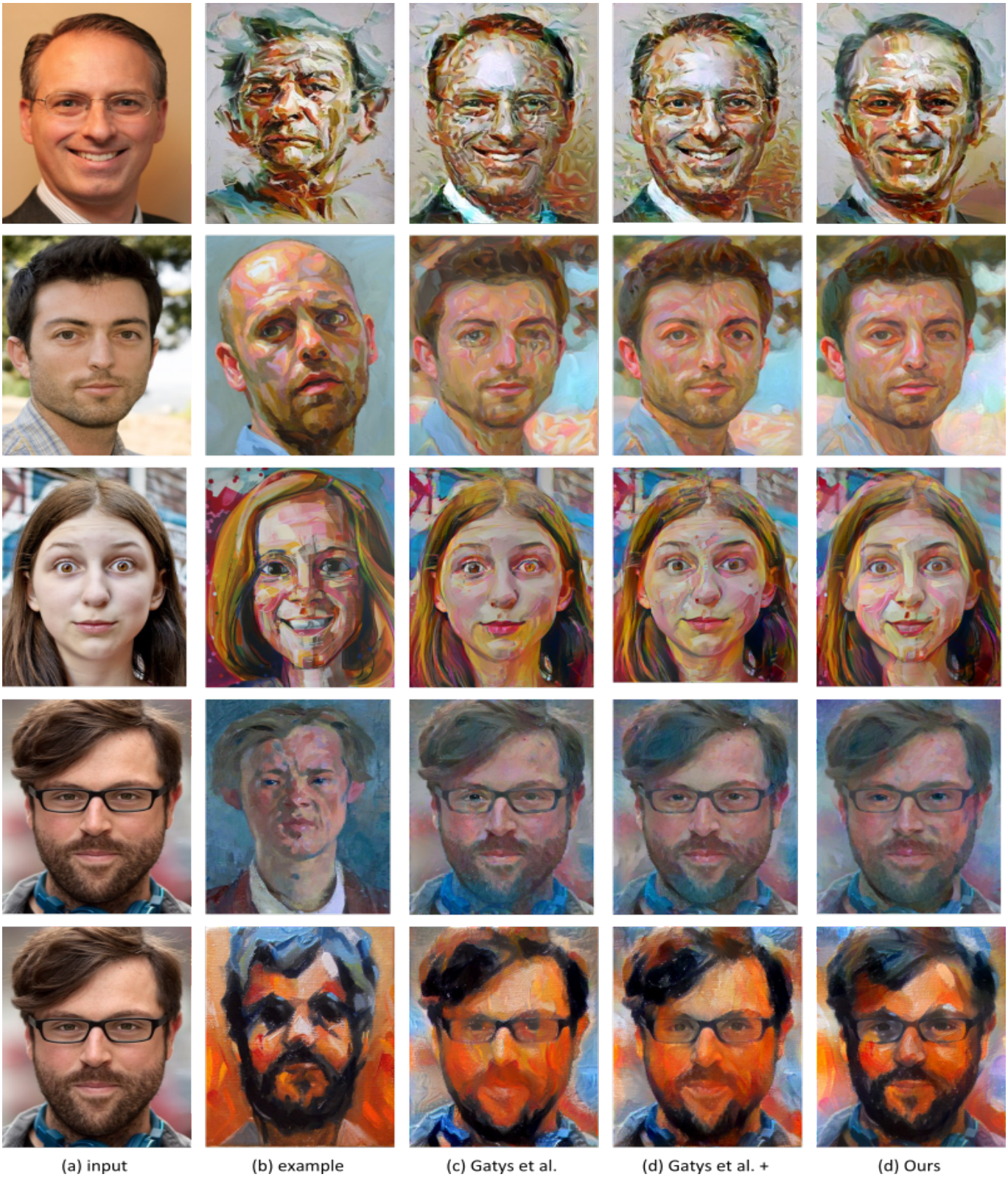


Figure 11: *Painting transfer for different input photographs (first column). Here we examine multiple example paintings (second column). For each input-example pair we show Gatys et al., Gatys et al. + and our results. Our approach better captures the texture of the painting styles than Gatys et al. and Gatys et al. +. In addition it reduces facial deformations over both approaches. For instance in rows 2,3 Gatys et al. generates irregularities around the eyes. Gatys et al. + generates irregularities in row 4 (blue hair and blue beard) and around the left eye in row 2. Example paintings are by (from top): Paul Wright (two rows), Gwenn Seemel, Ian Fleming and Aaron Mitchell.*

Figure 12: *Modified input with different clamping factors (from left)* $(g_{\min}, g_{\max}) = (0, 1)$, $(1, 100)$ *and* $(0.7, 5)$. *The original input and example are shown in Fig. 8 (third column and last row). (c) generates the best balance between painting texture and the input integrity.*

generates a ghosting effect. We remove such ghosting by compositing the face over a new background. We use [Levin et al. 2008] to extracts the matts and we synthesize a new background with [Efros and Freeman 2001]. This eliminates the ghosting.

Impact of alignment and gain maps: Fig. 14 analysis the effect of alignment and gain maps in our method. The absence of gain maps lead to the loss of the painting style texture (see Fig. 14, c). The removal of the alignment on the other hand generates a ghosting effect on the face (see around chin in Fig. 14, d). Our full algorithm however avoids both problems (see Fig. 14, e).

Profile portrait Fig. 15 shows an example of applying our algorithm on a profile portrait. We use [Levin et al. 2008] to extract a matte and composited our painting over a newly synthesized background to generate Fig. 15 (c).

Photographic stylization Fig. 16 shows two results from [Shih et al. 2014]. We used [Shih et al. 2014] publicly available code. We both handle two different problems and hence results are different. Future work can address extending our work to photographic stylization.

Running Time The core of our algorithm is optimizing Eq. 7 using a combination of feed-forward and back-propagation. This takes around 100 seconds per image (300 iterations) on a GeForce GTX 780. Here an image is resized to 500×378 pels. The preceding alignment step takes about 5.3 seconds per image with an unoptimized MATLAB code. An i7-3740QM CPU @2.7 GHz is used with 8 GB of RAM.

Limitations The alignment step of Sec. 4.2 is error-prone mainly due to sift-flow. Clamping the gain through Eq. 6 accommodates for such inaccuracies in many cases (Fig. 17). However it fails when there is strong misalignment between the input photograph and example painting. This can be generated from different head poses (Fig. 18, first row) or from hair mismatch (Fig. 18, last row). Here [Gatys et al. 2015] can perform better (see d and e).

5.2 Extension to Video

Recall in Sec. 4.2 we align the example painting over the input image using a combination of morphing and sift-flow. We then estimate the final painting by minimizing Eq. 7. This gives good results for still images (Fig 7-11). However applying the same process frame by frame independently for an image sequence generate temporal inconsistencies and shower-door effect. This is due to inaccuracies in sift-flow estimation. Hence for video we resort to a different alignment procedure. We choose one key frame and align the painting example over that frame using sift-flow and morphing. We then estimate the optical flow between the key frame and remaining frames. The generated vectors align the key frame example painting over the remaining frames. This way we estimate sift-flow on only one frame and hence avoids potential temporal incoherency in its estimation. We then minimize Eq. 7 on each frame independently. Finally we optionally use a motion-compensated moving average filter to further smooth the output. Similarly, to composite the video over a new background we manually provide the matting strokes for only the key frame. The estimated motion vectors propagate the strokes to the remaining frames. Matting is then applied on each frame independently.

Fig. 19-20 show few frames from the examined videos. We processed around 3000 frames in total with 5 painting styles and 4 different people (2 male and 2 female). The videos contain a wide variety of facial expressions aswell as different variations of head movements. This includes large translation, tilt and rotation. In all cases we transfer the painting style and maintain temporal coherency. All videos are available in the supplementary material.

Fig. 19 (first three rows) show few images from one of the examined sequences. The subject is moving her head freely and performing various facial expressions (mouth opening, laughing, sad, happy, neutral and so on). Our algorithm transfers the example painting while maintaining the input identity. In addition it maintains the integrity of the facial structures even in the mouth region where there is strong covering/uncovering. This generates a final painting that does not flicker in time (see supplementary material for full video). Note that for this sequence we matched the painting background to the input background through matting. This simplifies the painting transfer and makes the result more temporally coherent. For this sequence we set $(g_{\min}, g_{\max}) = (0.8, 2)$ to better handle the covering/uncovering of the mouth. We also composited the painted face on a new background. The background is synthesized manually.

Fig. 20 (last two rows) shows the impact of our video modifications as opposed to applying the still image algorithm. Applying the still image algorithm frame by frame generates inaccurate alignments. Such misalignments generate a clear shower-door/wrap-paper effect in the final painting (see Fig. 20, last row). Our video modification however significantly reduces such artifacts (Fig. 20, Our video handling). We reduce sift-flow errors by estimating it on the key frame only. Optical flow then generates good enough alignment for the remaining frames by propagating the key-frame alignment. This can handle a good variety of head movements. Our approach however struggles when there is a strong change in head pose. For this the wrap-paper effect can be more noticeable. To reduce this problem future work can explore introducing more key frames and imposing temporal consistency explicitly during style transfer.

6 Discussion

We presented a technique for head portrait painting. Current techniques poorly capture the painting texture, deform facial structures and are mainly based on image analogies. This limits their domain of applicability. Our technique uses a convolutional neural network approach with novel spatial constraints. This transfers the painting style while maintaining the integrity of the facial structures. Unlike current approaches our technique is not based on image analogies and hence requires only the example image. In addition it is not constrained to a specific style. We examined a wide variety of painting styles and input photographs, including different variations in facial outline, expressions, hair, gender, lightening and skin color. We presented an extension to image sequences and examined a variety of global (head) and local (expressions) movements. We generated temporally coherent results by exploiting the motion

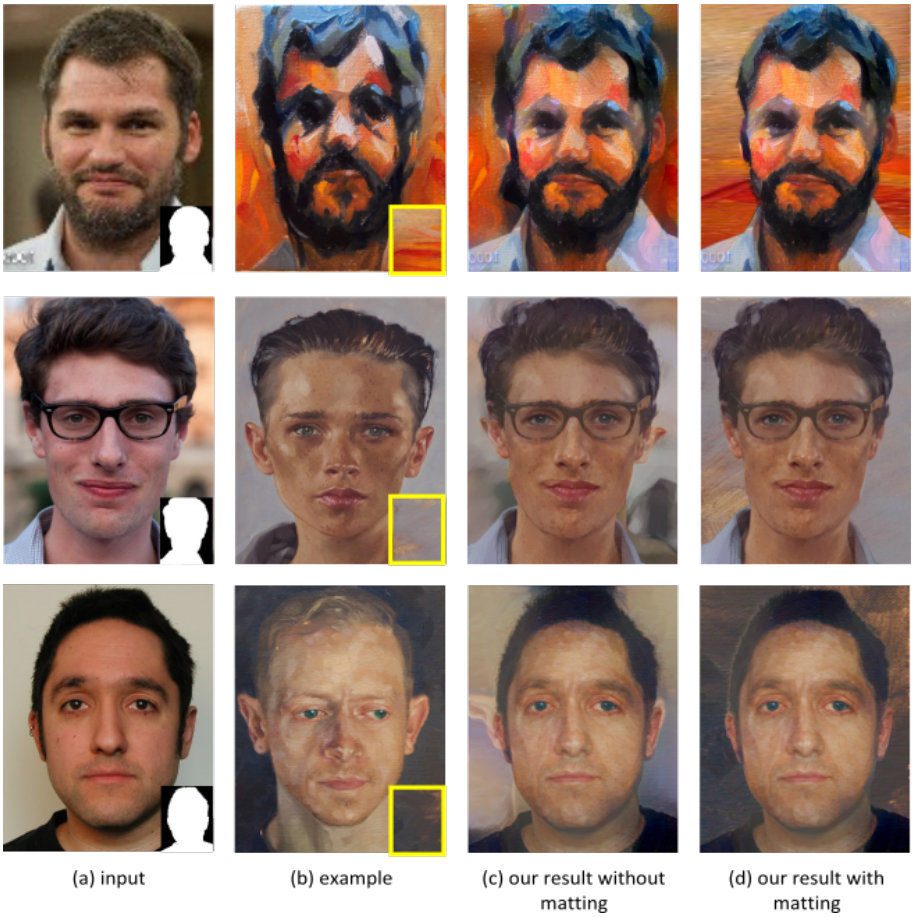


Figure 13: *In some cases the example painting generates ghosting artifacts in our transfer. We remove such ghosting by compositing the face over a new background. Levin et al. matting is used here and shown in first column (insets). The new backgrounds are shown in the second column (insets). Example paintings are by (from top): Aaron Mitchell and Aaron Nagel (two rows).*

information between the input frames. Future work can examine painting and color transfer for a wider variety of objects.

Acknowledgements

We thank Professor Fredo Durand for his valuable discussions and the anonymous SIGGRAPH reviewers for their comments. We also thank the artists for allowing us to use their paintings. We thank (in paper display order): Michael D. Edens, Patrick Earle, Aaron Mitchell ('Self Portrait', 2011, oil on canvas, www.mitchellaaron.com), Linden Holman, Paul Wright, Gwenn Seemel, Ian Fleming, Lesley Spanos ('Me', 2011, oil, spanosart.com), Aaron Nagel, Bill Sharp ('self-portrait', 2014, oil on linen and 'portrait of sara', 2014, oil on linen panel), Andrew Salgado, Elizabeth Hristova - Lisa, Henry Lamb, Sandra Lo, Nick Lepard and Lu Cong (lucong.info). Experiments were performed on the machines maintained by the Trinity Center for High Performance Computing and by Qatar Computing Research Institute. Thanks to Rachel O' Dwyer, Elma Avdic, Omar Abou-Selo and Ahmed Musleh for being our video models. This work was partially funded by Science Foundation Ireland (SFI) grant number 10/CE/i853.

References

ASHIKHMIN, M. 2001. Synthesizing natural textures. In *ACM Symposium on Interactive 3D Graphics*, 217–226.

ASHIKHMIN, N. 2003. Fast texture transfer. *Computer Graphics and Applications 23*, 4, 38–43.

BAE, S., PARIS, S., AND DURAND, F. 2006. Two-scale tone management for photographic look. In *ACM SIGGRAPH*, 637–645.

BARRODALE, I., SKEA, D., BERKLEY, M., KUWAHARA, R., AND POECKERT, R. 1993. Warping digital images using thin plate splines. *Pattern Recognition 26*, 2, 375–376.

BEIER, T., AND NEELY, S. 1992. Feature-based image metamorphosis. *ACM SIGGRAPH 26*, 2, 35–42.

CHEN, H., XU, Y.-Q., SHUM, H.-Y., ZHU, S.-C., AND ZHENG, N.-N. 2001. Example-based facial sketch generation with non-parametric sampling. In *International Conference on Computer Vision*, 433–438.

CHEN, H., LIANG, L., QING XU, Y., YEUNG SHUM, H., AND NING ZHENG, N. 2002. Example-based automatic portraiture. In *Asian Conference on Computer Vision*.

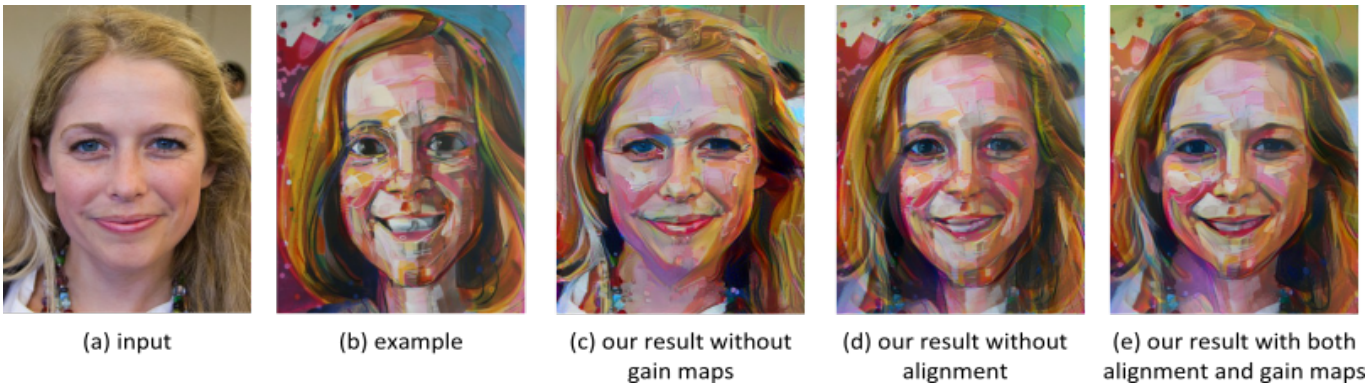


Figure 14: *Analysis of our algorithm components. The absence of gain maps lead to the loss of the painting style texture (c). The removal of the alignment generates a ghosting effect on the face (see the chin, d). Our full algorithm avoids both problems (e). Example painting by Gwenn Seemel.*

Figure 15: *Applying our algorithm on a profile portrait. Original target painting of Colin by Sandra Lo.*

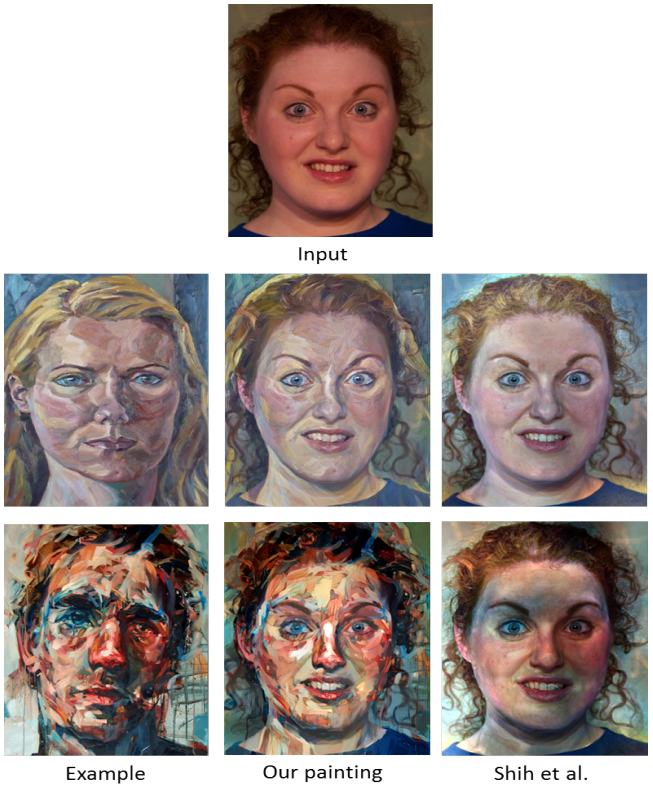


Figure 16: *Comparison against Shih et al. photographic stylization. Shih et al. does not capture the often strong painting texture. Example painting by (from top) Patrick Earle and Andrew Salgado.*

CHEN, H., ZHENG, N., LIANG, L., LI, Y., XU, Y.-Q., AND SHUM, H.-Y. 2002. Pictoon: a personalized image-based cartoon system. In *ACM Multimedia*, 171–178.

CHEN, H., LIU, Z., ROSE, C., XU, Y., SHUM, H.-Y., AND SALESIN, D. 2004. Example-based composite sketching of human portraits. In *International Symposium on Non-photorealistic Animation and Rendering*, 95–153.

COLLOMOSSE, J., AND HALL, P. 2002. Painterly rendering using image salience. In *Eurographics*, 122–128.

COLLOMOSSE, J., AND HALL, P. 2005. Genetic paint: A search for salient paintings. In *EvoMUSART*, 437–447.

COOTES, T. F., TAYLOR, C. J., COOPER, D. H., AND GRAHAM, J. 1995. Active shape model: Their training and application. *Computer Vision and Image Understanding 61*, 1, 38–59.

COOTES, T. F., EDWARDS, G. J., AND TAYLOR, C. J. 1998. Active appearance models. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Springer, 484–498.

DIPAOLA, S. 2007. Painterly rendered portraits from photographs using a knowledge-based approach. In *SPIE: Human Vision and Imaging*.

EFROS, A. A., AND FREEMAN, W. T. 2001. Image quilting for texture synthesis and transfer. In *ACM SIGGRAPH*, 341–346.

Figure 17: *Sift-flow can generate alignment artifacts (c). Our approach however handles many of such artifacts (d). Example paintings are by (from top): Patrick Earle and Elizabeth Hristova - Lisa.*

EFROS, A., AND LEUNG, T. 1999. Texture synthesis by non-parametric sampling. In *International Conference on Computer Vision*, 1033–1038.

GATYS, L. A., ECKER, A. S., AND BETHGE, M. 2015. A neural algorithm of artistic style. *CoRR abs/1508.06576*.

GOOCH, B., COOMBE, G., AND SHIRLEY, P. 2002. Artistic vision: Painterly rendering using computer vision techniques. In *International Symposium on Non-photorealistic Animation and Rendering*, 83–88.

GOOCH, B., REINHARD, E., AND GOOCH, A. 2004. Human facial illustrations: Creation and psychophysical evaluation. *SIGGRAPH 23*, 1, 27–44.

HAEBERLI, P. 1990. Paint by numbers: Abstract image representations. *SIGGRAPH Computer Graphics and Interactive Techniques 24*, 4, 207–214.

HASHIMOTO, R., JOHAN, H., AND NISHITA, T. 2003. Creating various styles of animations using example-based filtering. In *Computer Graphics International*, 312–317.

HAYS, J., AND ESSA, I. 2004. Image and video based painterly animation. In *International Symposium on Non-photorealistic Animation and Rendering (NPAR)*, 113–120.

HERTZMANN, A., JACOBS, C. E., OLIVER, N., CURLESS, B., AND SALESIN, D. H. 2001. Image analogies. In *SIGGRAPH*, 327–340.

HERTZMANN, A. 1998. Painterly rendering with curved brush strokes of multiple sizes. In *SIGGRAPH*, 453–460.

HERTZMANN, A. 2001. Paint by relaxation. In *Computer Graphics International*, 47–54.

JOHNSON, J., 2015. Torch implementation of neural style algorithm. https://github.com/jcjohnson/neural-style.

KIM, S. Y., MACIEJEWSKI, R., ISENBERG, T., ANDREWS, W. M., CHEN, W., SOUSA, M. C., AND EBERT, D. S. 2009. Stippling by example. In *International Symposium on Non-Photorealistic Animation and Rendering*, 41–50.

KRIZHEVSKY, A., SUTSKEVER, I., AND HINTON, G. E. 2012. Imagenet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*, 1097–1105.

KYPRIANIDIS, J. E., COLLOMOSSE, J., WANG, T., AND ISENBERG, T. 2013. State of the art: A taxonomy of artistic stylization techniques for images and video. *IEEE Transactions on Visualization and Computer Graphics 19*, 5, 866–885.

LEE, H., SEO, S., RYOO, S., AND YOON, K. 2010. Directional texture transfer. In *International Symposium on Non-Photorealistic Animation and Rendering*, 43–48.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2008. A closed-form solution to natural image matting. *IEEE Transactions on PAMI 30*, 2, 228–242.

LI, Y., SHARAN, L., AND ADELSON, E. H. 2005. Compressing and companding high dynamic range images with subband architectures. *SIGGRAPH 24*, 3, 836–844.

LIN, L. 2010. Painterly animation using video semantics and feature correspondence. In *NPAR*, 73–80.

LITWINOWICZ, P. 1997. Processing images and video for an impressionist effect. In *SIGGRAH Computer Graphics and Interactive Techniques*, 407–414.

LIU, C., YUEN, J., TORRALBA, A., SIVIC, J., AND FREEMAN, W. T. 2008. Sift flow: Dense correspondence across different scenes. In *European Conference on Computer Vision*, 28–42.

LOWE, D. G. 1999. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, 1150–1157.

LU, J., SANDER, P. V., AND FINKELSTEIN, A. 2010. Interactive painterly stylization of images, videos and 3d animations. In *ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, 127–134.

MAHENDRAN, A., AND VEDALDI, A. 2015. Understanding deep image representations by inverting them. In *Computer Vision and Pattern Recognition (CVPR)*, 5188–5196.

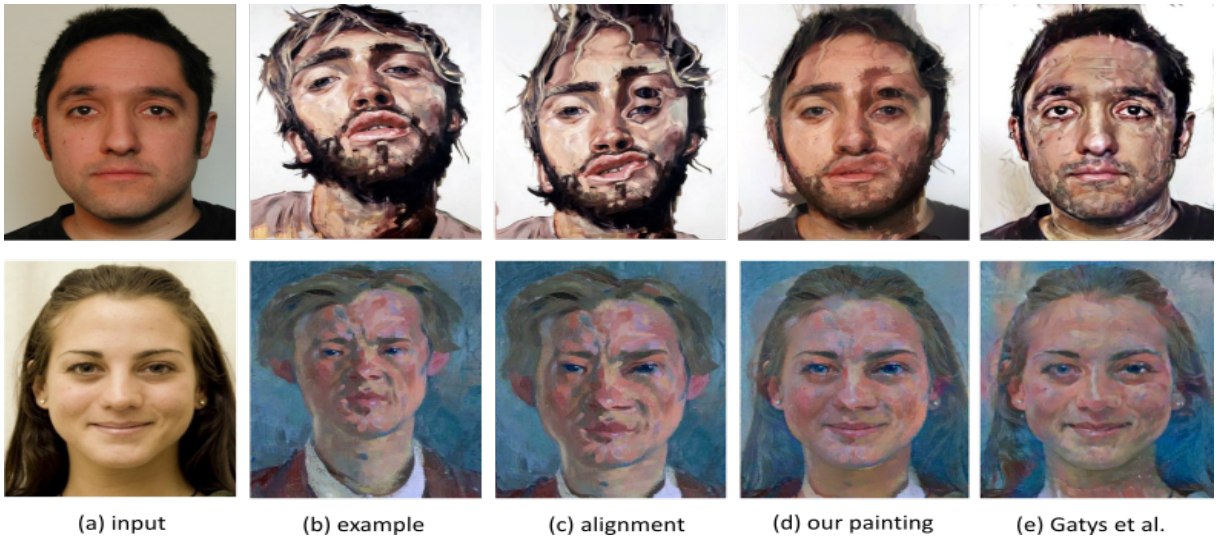


Figure 18: *Our algorithm can struggle in case of strong misalignments between the input photograph and example painting. This can be generated from different head poses (top) or from hair mismatch (bottom). For such cases Gatys et al. can perform better. Example paintings are by (from top): Nick Lepard and Ian Fleming.*

McKone, E., Kanwisher, B., and Duchaine, B. 2007. Can generic expertise explain special processing for faces? *Trends in Cognitive Science 1*, 8–15.

Meier, B. J. 1996. Painterly rendering for animation. In *Computer Graphics and Interactive Techniques*, 477–484.

Meng, M., Zhao, M., and Zhu, S.-C. 2010. Artistic papercut of human portraits. In *ACM International Conference on Multimedia (Short Paper)*, 931–934.

O'Donovan, P., and Hertzmann, A. 2012. Anipaint: Interactive painterly animation from video. *IEEE Transactions on Visualization and Computer Graphics 18*, 3, 475–487.

Pitie, F., Kokaram, A., and Dahyot, R. 2005. N-dimensional probability density function transfer and its application to color transfer. In *International Conference on Computer Vision*, 1434–1439.

Reinhard, E., Adhikhmin, M., Gooch, B., and Shirley, P. 2001. Color transfer between images. *Computer Graphics and Applications 21*, 5, 34–41.

Saragih, J. M., Lucey, S., and Cohn, J. 2009. Face alignment through subspace constrained mean-shifts. In *International Conference of Computer Vision*, 1034 – 1041.

Shih, Y., Paris, S., Barnes, C., Freeman, W. T., and Durand, F. 2014. Style transfer for headshot portraits. *SIGGRAPH 33*, 4, 148:1–148:14.

Simonyan, K., and Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556*.

Sinha, P., Balas, B., Ostrovsky, Y., and Russell, R. 2006. Face recognition by humans: Nineteen results all computer vision researchers should know about. *Proceedings of the IEEE 94*, 11, 1948–1962.

Su, S. L., Durand, F., and Agrawala, M. 2005. De-emphasis of distracting image regions using texture power maps. In *International Workshop on Texture Analysis and Synthesis in conjunction with ICCV'05*, 119–124.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. 2014. Deepface: Closing the gap to human-level performance in face verification. In *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, 1701–1708.

Wang, X., and Tang, X. 2009. Face photo-sketch synthesis and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 31*, 11, 1955–1967.

Wang, B., Wang, W., Yang, H., and Sun, J. 2004. Efficient example-based painting and synthesis of 2d directional texture. *IEEE Transactions on Visualization and Computer Graphics 10*, 3, 266–277.

Wang, T., Collomosse, J., Slatter, D., Cheatle, P., and Greig, D. 2010. Video stylization for digital ambient displays of home movies. In *International Symposium on Non-Photorealistic Animation and Rendering*, 137–146.

Wang, T., Collomosse, J., Hunter, A., and Greig, D. 2013. Learnable stroke models for example-based portrait painting. In *British Machine Vision Conference*, 36.1–36.11.

Zeng, K., Zhao, M., Xiong, C., and Zhu, S.-C. 2009. From image parsing to painterly rendering. *SIGGRAPH 29*, 1 (Dec.), 2:1–2:11.

Zhao, M., and Zhu, S.-C. 2011. Portrait painting using active templates. In *Non-Photorealistic Animation and Rendering (NPAR)*, 117–124.

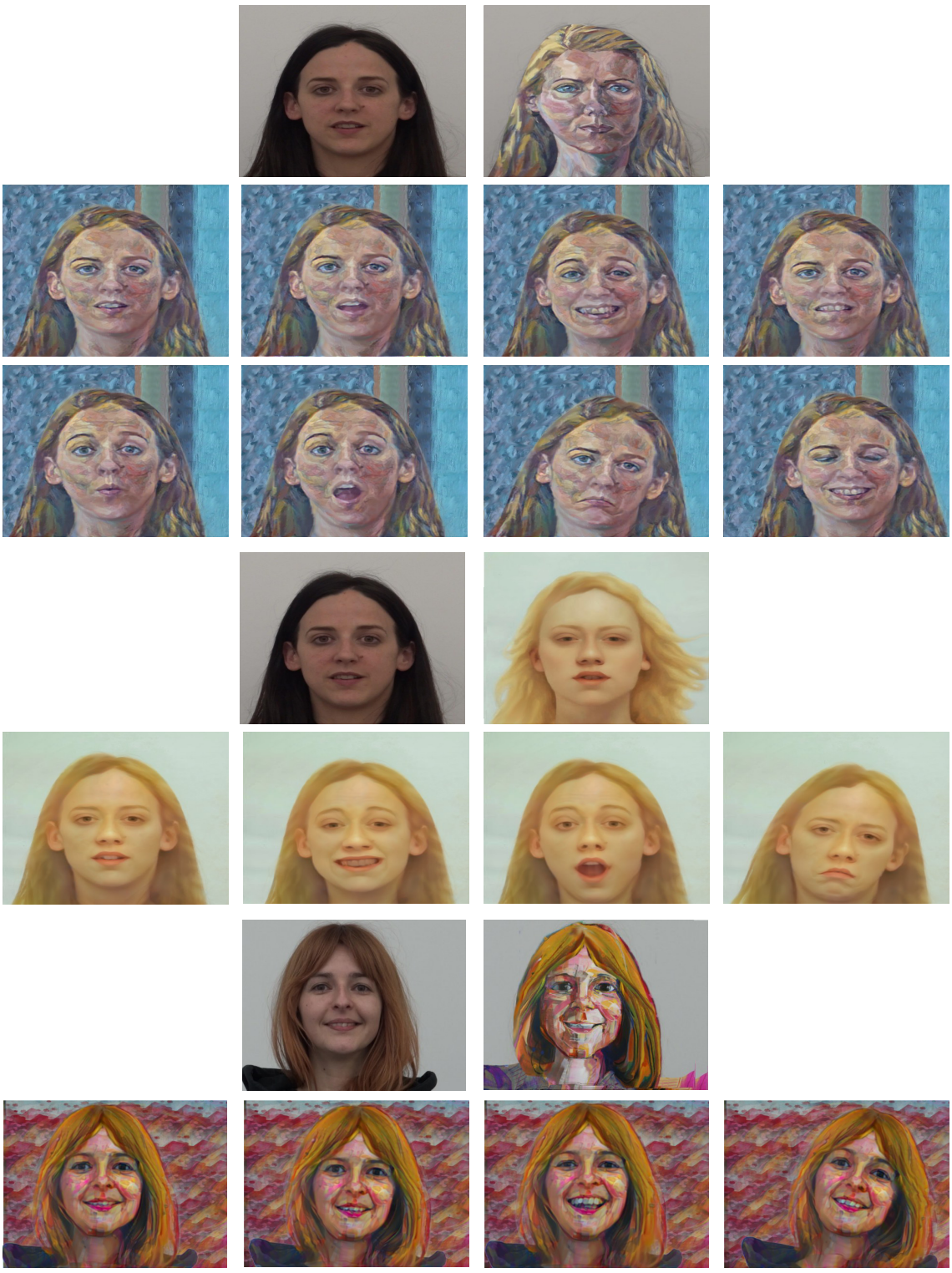
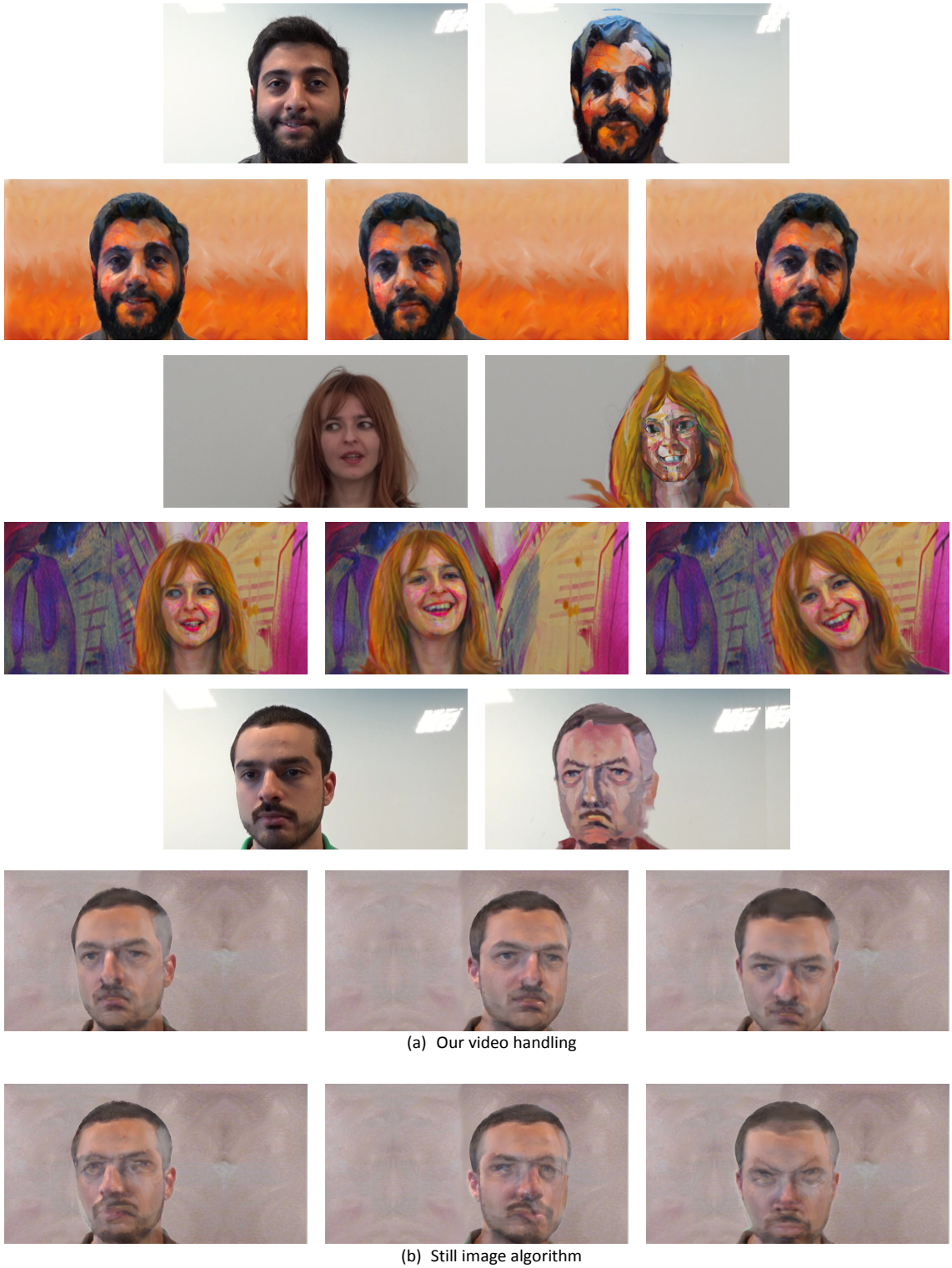


Figure 19: *Handling videos with our approach. For each sequence the first row shows one frame with its input and example painting. The second row shows paintings generated by our approach. The sequences undergo moderate head movement with strong temporally impulsive facial expressions. Please see supplementary material for full videos. Original example paintings by (from top): Patrick Earle, Lu Cong and Gwenn Seemel.*



(a) Our video handling

(b) Still image algorithm

Figure 20: *Handling videos with our approach. For each sequence the first row shows one frame with its input and example painting. The second row shows paintings generated by our approach. The sequences undergo strong head movement including translation, rotation and tilt. Last row: The result of applying our still image algorithm without any temporal modifications. This however generates strong shower-door effect. Please see supplementary material for full videos. Original example paintings by (from top): Aaron Mitchell, Gwenn Seemel and Bill Sharp.*