Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Arbitrary Style Transfer Using Neurally-Guided Patch-Based Synthesis

Ondřej Texler[a,*], David Futschik[a], Jakub Fišer[b], Michal Lukáč[b], Jingwan Lu[b], Eli Shechtman[b], Daniel Sýkora[a]

[a] *Faculty of Electrical Engineering, Czech Technical University in Prague, Czechia*
[b] *Adobe Research, USA*

ARTICLE INFO

Article history:
Received January 9, 2020

Keywords: style transfer, neural networks, example-based, patch-based synthesis

ABSTRACT

We present a new approach to example-based style transfer combining neural methods with patch-based synthesis to achieve compelling stylization quality even for high-resolution imagery. We take advantage of neural techniques to provide adequate stylization at the global level and use their output as a prior for subsequent patch-based synthesis at the detail level. Thanks to this combination, our method keeps the high frequencies of the original artistic media better, thereby dramatically increases the fidelity of the resulting stylized imagery. We show how to stylize extremely large images (e.g., 340 Mpix) without the need to run the synthesis at the pixel level, yet retaining the original high-frequency details. We demonstrate the power and generality of this approach on a novel stylization algorithm that delivers comparable visual quality to state-of-art neural style transfer while completely eschewing any purpose-trained stylization blocks and only using the response of a feature extractor as guidance for patch-based synthesis.

1. Introduction

In recent years, advances in neural style transfer and guided patch-based synthesis made the field of computer-assisted stylization very popular. Various publicly available software solutions (see, e.g., Prisma [4], DeepArt [1], StyLit [2], FaceStyle [5]) successfully brought the style transfer concepts to consumers. These applications enjoy popularity among casual users due to their novelty factors. However, they are not addressing the needs of professional users who demand high-resolution, high-quality output accurately preserving the textural details of the original artistic exemplar.

Though guided patch-based synthesis approaches [2, 5] can meticulously preserve fine-grained details, they require preparation of guidance channels. These guidance channels are important for establishing meaningful correspondences between the target image and the source style exemplar. Previous work designed guidance channels for specific use cases such as faces [5], but designing meaningful guidance automatically in general case remains a difficult problem. On the other hand, neural-based style transfer [1, 6] does not require explicit guidance to produce good stylization effects at a global level. Nevertheless, due to its convolutional nature, it usually fails to preserve low-level details such as brush strokes or canvas structure that are important to retain the fidelity of the underlying artistic media.

Neural techniques are also limited to work at lower resolutions (typically below 1K), which does not suit the need for FullHD, 4K or higher resolution used in real production settings. A similar limitation also holds for guided patch-based synthesis where the processing time grows significantly with increasing output resolution. Neural style transfer algorithms also have the problem of exhausting GPU memories where going beyond 4K resolution becomes impossible under current hardware constraints.

In this paper, we propose a straightforward approach which overcomes the aforementioned limitations by combining neural

[*] Corresponding author: Tel.: +420-22435-7502
e-mail: texleond@fel.cvut.cz (Ondřej Texler)

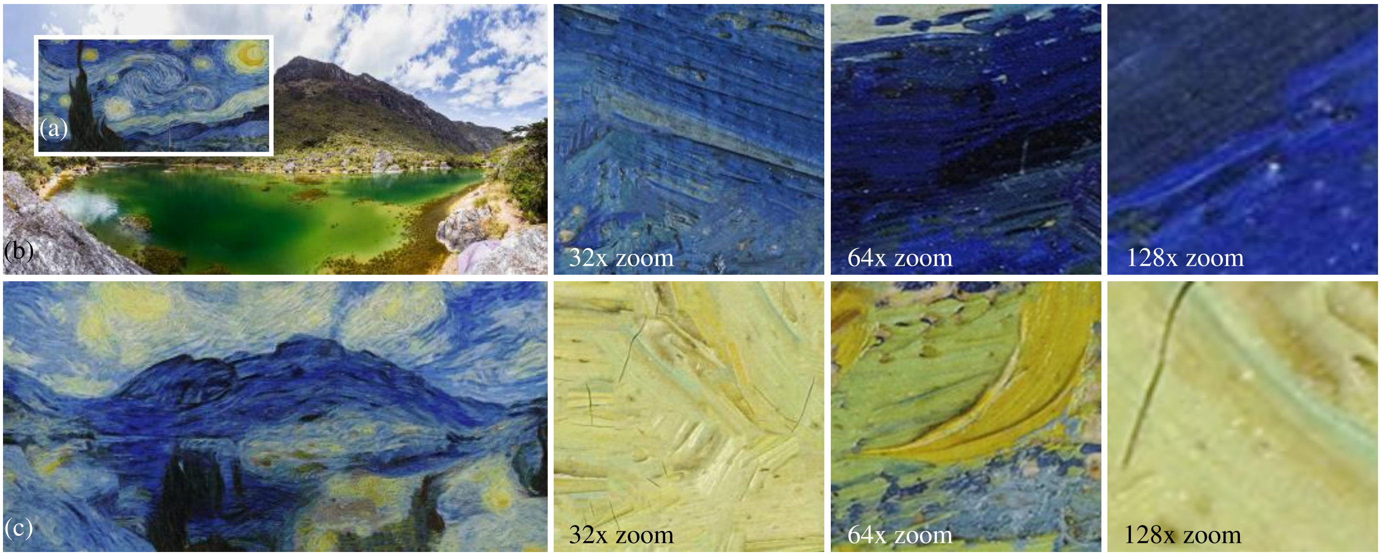

Fig. 1: An example of stylizing an extremely high-resolution image using our proposed method: (a) style exemplar of 26400 × 13100 px, (b) content image of the same resolution, (c) low resolution result of [1] enhanced and enlarged by our method to the mentioned resolution. To the right, zoom-in patches of different parts of (c) up to zoom of 128× are shown; see all the individual brush strokes and its sharp boundaries. Also, notice how well the structure of the original canvas and little cracks of the painting are preserved.

style transfer, patch-based synthesis, and dense correspondence field upscale. We first apply neural style transfer to obtain semantically meaningful stylization at a global level without the need of user intervention, and then use patch-based synthesis to remove low-level artifacts and restore the color and fine details to retain the fidelity of the original style, see Fig. 2. To significantly reduce computational overhead instead of running patch-based synthesis on the full resolution, we only upscale the dense correspondence field computed at a lower resolution level. We demonstrate that such a simple upscaling step can be performed quickly while still providing comparable visual quality as the full-fledged synthesis. This enables us to achieve high-quality stylization of extremely large images (see Fig. 1 where an image of 346Mpix is stylized). Our approach is generalized and can utilize any existing neural stylization method. We demonstrate this generality on a variant of our style transfer algorithm that directly uses the response of a neural network as a guide for patch-based synthesis. We developed a prototype of our method in the form of a Photoshop plug-in and put it into the hands of professional artists.

2. Related Work

One of the key tasks of non-photorealistic rendering [8] is to deliver stylized depictions of photos or synthetic scenes which preserve high-level information captured in the scene while on a detail level the resulting image resembles the artistic look.

Stroke-based approaches were one of the first techniques that enabled generation of stylized imagery. Rotated and translated brush strokes from a predefined set are placed according to some guiding information (e.g., the direction of image gradients). This technique is applicable both in 2D [9] and 3D [10] environment producing quite compelling results. Nevertheless, the main drawback here is the restriction to a predefined set of strokes, which limit the variety and fidelity of the stylized output. Such a limitation can partly be alleviated by introducing example-based brushes [11, 12]; nevertheless, the final look is still limited to a subset of styles that can be simulated by a composition of brush strokes.

To address this issue a more robust and general example-based approach called *Image Analogies* was pioneered by Hertzmann et al. [3]. Given an arbitrary style exemplar and a set of guidance channels, the stylized image can be produced using guided patch-based synthesis [7, 13, 2]. This approach has been successfully applied to various stylization scenarios including fluid animations [14], 3D renders [2, 15], facial animations [5] or video clips [16]. Nevertheless, a common drawback of this method is that it requires the preparation of custom-tailored guidance to deliver compelling stylization quality. Furthermore, an extensive computational overhead at higher resolutions makes those techniques difficult to use in production.

Neural-based style transfer approaches recently became popular due to advances made by Gatys et al. [1], they successfully applied the pre-trained convolutional neural network VGG [17] to the problem of style transfer. The core idea of their method is to match statistics in the domain of VGG [17] features of both the content and style images. They further extended this idea in [18] to introduce control over spatial location, color information, and scale of features. While these techniques produce impressive results for some particular style exemplars, they usually suffer from loss of high-frequency details of the style exemplar which is inevitably caused by the convolutional nature of the underlying neural network. Moreover, mentioned neural techniques usually have considerable computational overhead and memory footprint.

Although a feed-forward network can be pre-trained to speed up the stylization [4, 19, 20, 21], every new style requires additional costly training. Recently, adoption of encoder–decoder

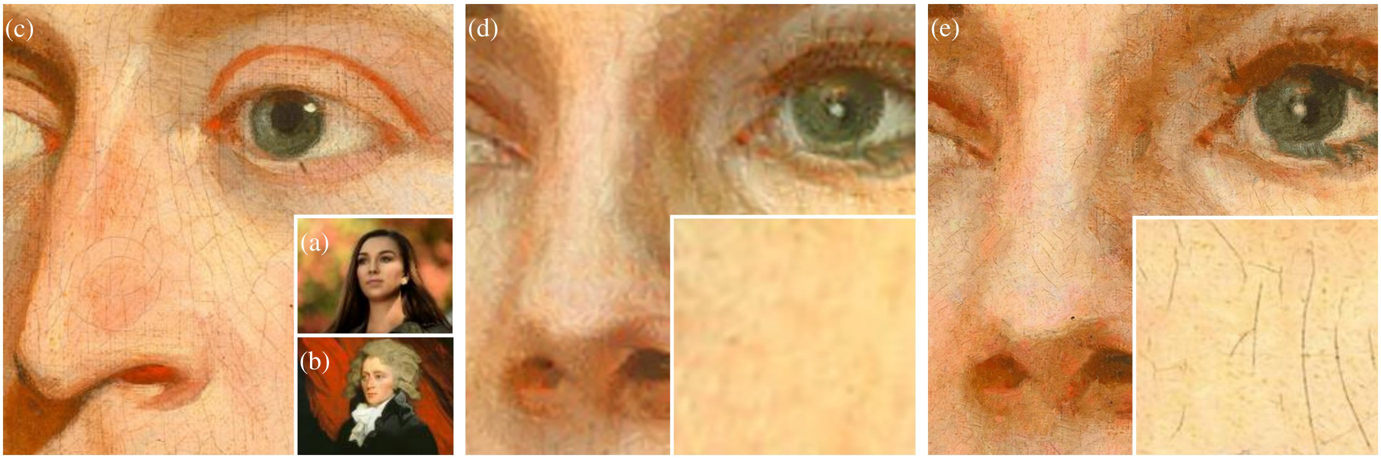

Fig. 2: An example of enhancing the result of neural-based approach using our method: (a) target photograph, (b) style exemplar of the same size, (c) 6× zoom-in to the style exemplar, (d) the output of neural-based method DeepArt [1] is capable to perform convincing stylization; nevertheless, the image contains artifacts caused by the parametric nature of the used neural network. High-frequency details like the structure of strokes and canvas are largely lost, sacrificing the visual quality of the original artistic medium. In contrast, our method (e) brings significant quality improvement, it restores the individual brush strokes and boundaries between them faithfully, the result better reproduces the used artistic medium as well as canvas' structure. Note how the cracks of the original artwork are preserved; although zoom-in patches are shown, we encourage the reader to zoom-in even further.

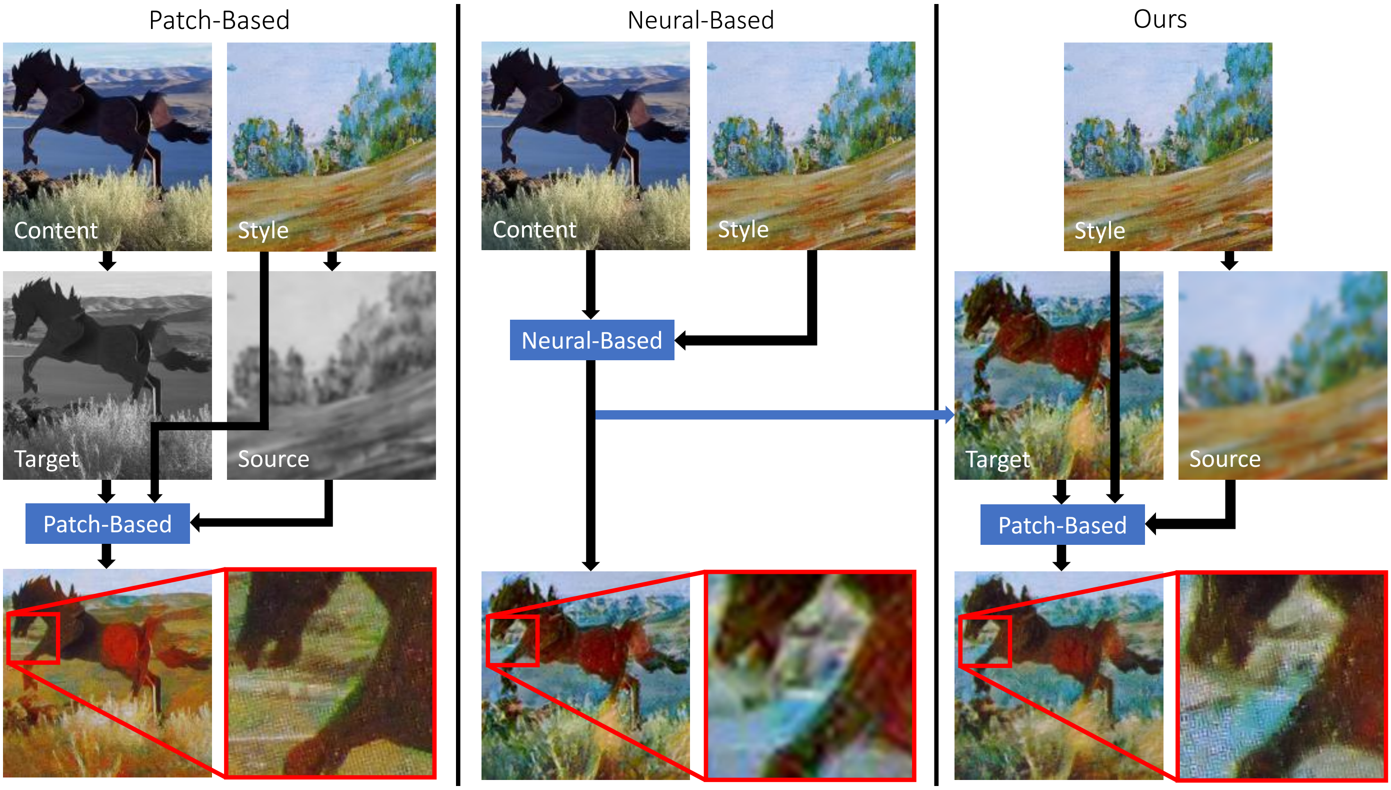

Fig. 3: Simplified scheme of a patch-based, neural-based, and our hybrid style transfer method: The left column shows a patch-based approach [2] with guidance based on blurred grayscale images as proposed in the original Image Analogies method [3]. The resulting image has high texture quality and preserves artistic attributes and canvas structure well; however, the result does not properly respect the content semantics, causing water to become brown. The middle column shows a neural-based approach [1], no guidance channels are needed and global style properties and image semantic are preserved well. However, the resulting image lacks high-frequency details of the original style exemplar, contains artifacts, and colors that are not present in the original style. The right column represents our method where low-resolution neural transfer result is used as a guidance channel for patch-based style transfer. Our result attenuates the neural artifacts and restores the original color and texture of the style exemplar.

scheme was proposed [22, 23, 24] to enable arbitrary style transfer in a feed-forward fashion. Here the encoder, usually convolution layers of the VGG, is used to get the feature representations (statistics) of the content and style, which are then combined, and a pre-trained decoder is used to turn the latent features back into the image. Nevertheless, all these techniques still suffer from convolutional artifacts leading to a lower quality of the synthesized imagery at a pixel level.

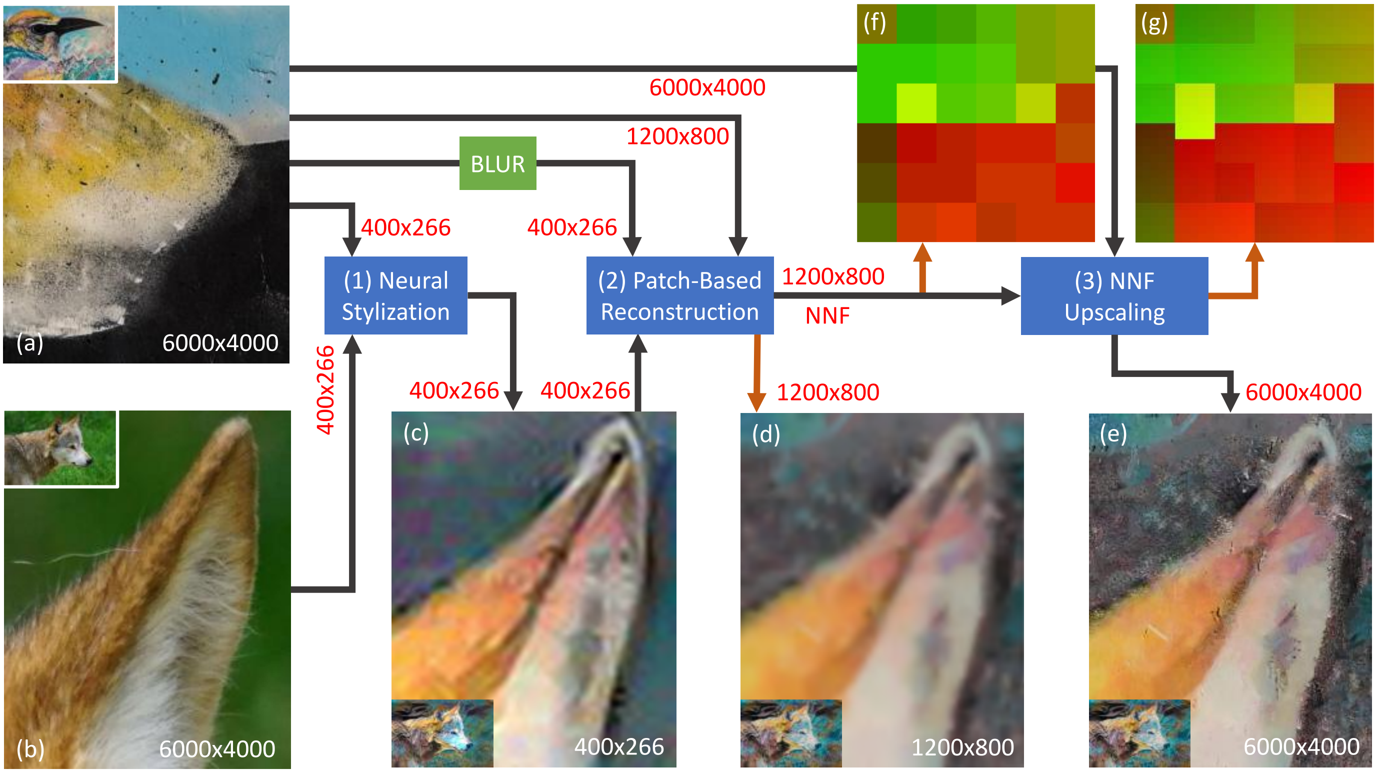

Fig. 4: Proposed pipeline: (a) style exemplar and (b) content image are both subsampled α–times and processed by a neural-based style transfer method (Sec. 3.1) which results in low resolution image (c) where fine details are missing and artifacts are apparent (see green and purple checkerboard artifacts). Next, low resolution result (c) from the previous step, style image (a) in the same resolution as (c), and β–times subsampled style image (a) are used as an input to a patch-based synthesis algorithm (Sec. 3.2) which outputs dense nearest neighbor field (NNF) (f) from which the corresponding image (d) can be produced using voting step [7]. Finally, in NNF upscaling step (Sec. 3.3) the low-resolution NNF (f) is upscaled β–times to the original resolution (g). Patch coordinates in NNF (f) and (g) are encoded as red and green color levels. Note subtle color gradients in (f), which indicate the presence of fine patch coordinates in upscaled NNF that points to the patches in the original high-resolution style exemplar (a). Given the upscaled NNF (g) and the style exemplar in its original resolution (a), high-resolution, and a perfectly sharp final result is created using voting step (e).

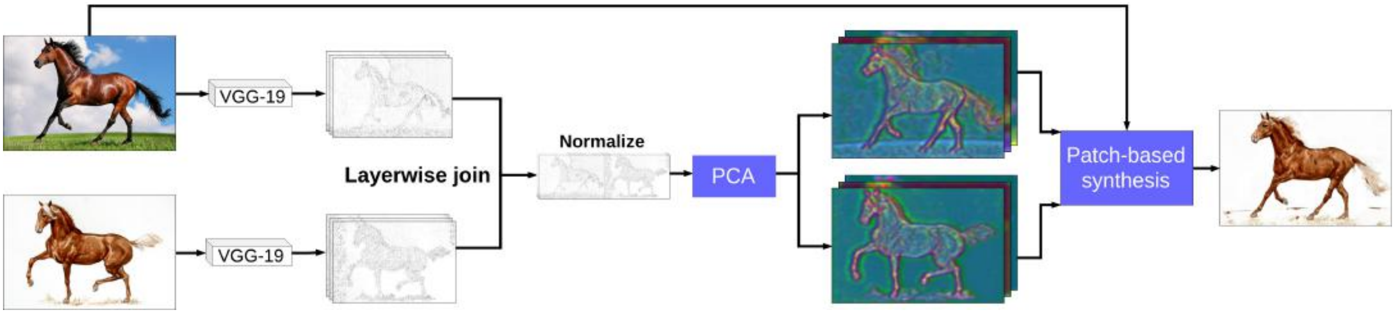

Fig. 5: An overview of our VGG-guided style transfer pipeline: we start with a target image and a style exemplar, extract their VGG-19 features, normalize them, reduce their dimensionality using PCA, and use these as guidance for subsequent patch-based synthesis. Even though the proposed pipeline is straightforward, it yields convincing output.

Recently, attempts to combine patch-based and neural-based techniques were proposed. Li et al. [25] search local neural patches from the style image concerning the structure of a content image, which leads to better reproduction of local textures. Liao et al. [26] later extended this idea in their *Deep Image Analogy* framework which adapts the concept of *Image Analogies* [3] in the domain of VGG features. Gu et al. [6] recently proposed to perform reshuffle in spirit of [13] to reduce the overuse of particular features. Futschik et al. [27] use patch-based method [5] to generate a larger dataset of stylized portraits which is then used to train a generative adversarial network capable of reproducing similar quality results as those in the underlying dataset. Although these techniques can notably improve the stylization quality and better preserve high-frequency details, they still heavily rely on the space of VGG features and do not explicitly enforce textural coherence on a pixel level in color domain [7] which is essential to retain the fidelity of the original style exemplar.

3. Our Approach

We propose an approach to combine patch-based synthesis with neural style transfer methods. The proposed pipeline overcomes three crucial obstacles which prevent existing stylization approaches from being used in real production: first, lower texture quality of neural-based techniques; second, the necessity of specific guidance for patch-based methods; and third, the resolution limitation which affects the usability of both approaches. Our framework allows easy switching to the newest future inventions in either neural-based or patch-based techniques.

As our first step, given the exemplar *Style* and the target image *Content*, we use an arbitrary neural-based style transfer method to synthesize an initial result (see Fig. 3 middle column). The resulting image on its own lacks high-frequency details of the style exemplar, and contains artifacts such as geometric distortions and colors that are not present in the original style. Also, the original contrast is usually artificially exaggerated, and edges are not sharp. However, on the other hand, it nicely preserves global style properties such as color distribution and respects the image semantics in general.

Our key idea is to use the low-resolution neural style transfer result as a guiding channel for patch-based synthesis. This enables us to combine the advantages of both techniques and to address the aforementioned limitations (see Fig. 3 right column). In particular, a pair of guidance channels *Source* and *Target* is needed for guided patch-based synthesis. We use blurred style exemplar as the *Source* guide and the low-resolution neural style transfer result as the *Target* guide. After running the guided patch-based synthesis, our result (Fig. 3 right column, bottom) effectively attenuates the neural artifacts and restores the color and texture of the original style exemplar.

Fig. 4 illustrates our entire pipeline which consists of three main parts: neural-based style transfer method, guided patch-based synthesis, and nearest neighbor field (NNF) upscaling method. Those individual steps are described in more detail in the following sections.

3.1. Neural-Based Style Transfer

Both *Style* (Fig. 4a) and *Content* (Fig. 4b) images are first subsampled by a coefficient α. This step is necessary not only to overcome the resolution restrictions but, more importantly, to suppress various high-frequency artifacts caused by neural-based techniques (α essentially defines the *working resolution* of a neural-based method). The α–times subsampled neural-based result (Fig. 4c) is then used as a guide for the patch-based synthesis method. Its resolution will be improved later in our pipeline.

3.2. Guided Patch-Based Synthesis

The output from the neural method (Fig. 4c) is used as a *Target* guide image in the patch-based method. Our pipeline does not assume any particular patch-based method; we used StyLit [2] algorithm for synthesis, however, we adapt its original error metric for measuring patch similarity to our needs. Let $\mathcal{S}$ be a style exemplar, O an output image, and $G^{\mathcal{S}}$ and $G^{\mathcal{T}}$ source and target guides, for matching two patches $p \in G^{\mathcal{S}}$ and $q \in G^{\mathcal{T}}$; we use the following error metric:

$$E(\mathcal{S}, O, G^{\mathcal{S}}, G^{\mathcal{T}}, p, q) = \|\mathcal{S}(p) - O(q)\|^2 + \lambda_g \|G^{\mathcal{S}}(p) - G^{\mathcal{T}}(q)\|^2 \quad (1)$$

where λ_g is a weighting factor for guiding channel and the first term helps to preserve *texture coherence* by directly matching colors in patches of *Style* to those in the output image O. Of all the images, only O is iteratively updated during the optimization process described in StyLit [2].

To obtain *Source* guide image, we use the already subsampled style image used in the previous step (Sec. 3.1), and upsample it back to its original resolution. To encourage the patch-based synthesis to find good correspondences for the style transfer, equivalent subsampling followed by upsampling needs to be done for both the *Source* and *Target* images. In spirit of *Color Me Noisy* [28], an additional low-pass filter can be applied on the *Source* image to let the synthesis algorithm deviate more from the initial solution, thus making the final result more abstract.

In Fig. 4d the result of patch-based synthesis is depicted in color for clarity, nevertheless, internally in our processing pipeline we use only the resulting nearest neighbor field (Fig. 4f) which is subsequently upsampled (Fig. 4g) and turned into a high-resolution image in the next step.

3.3. NNF Upscaling

Given the computed NNF–nearest neighbor field (Fig. 4f) and the style exemplar in its original resolution (Fig. 4a), a *voting step* (c.f. [7]) needs to be performed in order to reconstruct the final image. To reduce the computational overhead, we perform the patch-based synthesis (Sec. 3.2) at β–times lower resolution than the original target resolution (thus β essentially defines the *working resolution* of a patch-based method). Next, the resulting **nnf** (Fig. 4f) is upscaled by a factor of β to obtain the **NNF** (Fig. 4g) of the same resolution as the target image as follows:

$$\mathbf{NNF}(x, y) = \mathbf{nnf}(x/\beta, y/\beta) \cdot \beta + (x \bmod \beta, y \bmod \beta) \quad (2)$$

Finally, we perform a voting step using **NNF** to produce the final high-resolution result precisely preserving the characteristics of the canvas and the original artistic medium (Fig. 4e).

4. VGG-Based Guidance

One of the limitations of the proposed base algorithm introduced in the previous section is that it relies on color information to establish correspondences between style exemplar and the target image. This drawback could lead to an ambiguity that may introduce visible stylization artifacts (see Fig. 6).

In this section, we introduce a variant of our style transfer pipeline that uses features extracted by the convolutional layers of a classification network for guidance directly rather than relying on a neural style transfer algorithm to produce initial color domain stylization. The aforementioned neural responses provide more discriminative guidance than colors and thus can

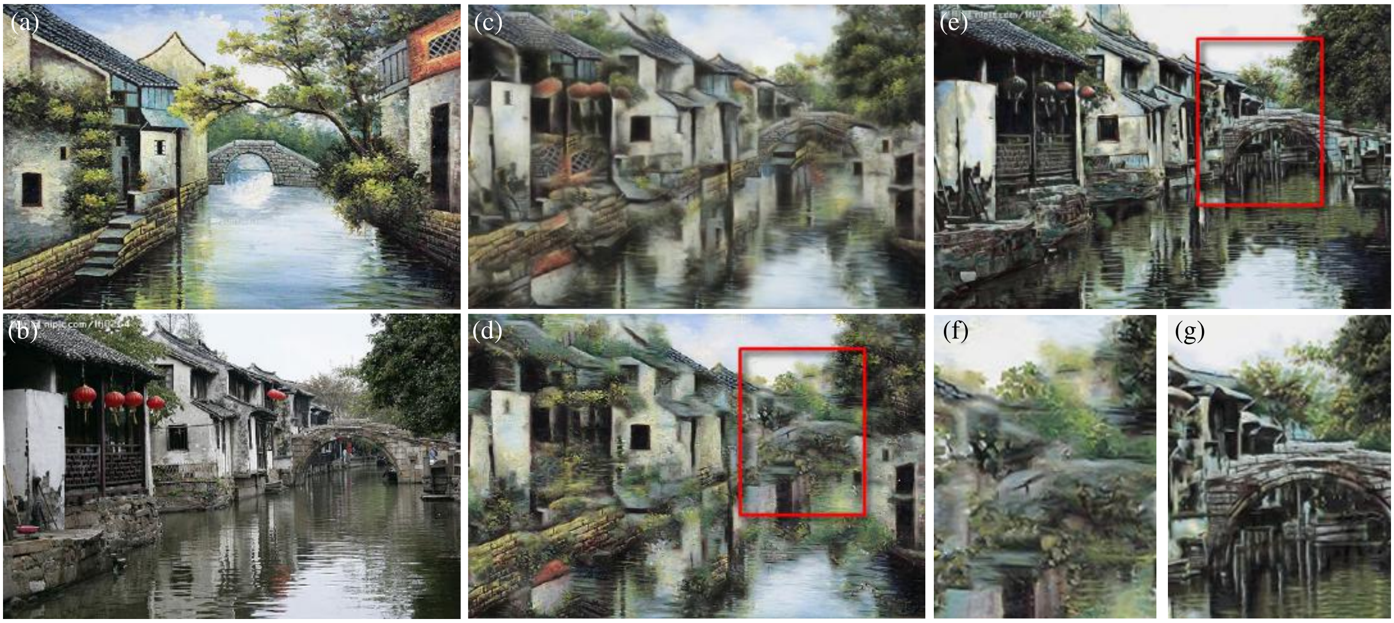

Fig. 6: Demonstration of the problem when patch-based synthesis has to rely on ambiguous color guidance: (a) style exemplar, (b) target image, (c) output of Gu et al. [6], (d) output of our basic algorithm with color-based guidance, (e) output of our style transfer algorithm with neural guidance. Note how our VGG-guided algorithm better preserves the semantics of the target photo, cf. details in (f) and (g).

preserve global semantics of the target while still keeping the benefits of patch-based optimization.

Our approach is inspired by modern optimization-based neural style transfer techniques of Liao et al. [26] and Gu et al. [6] that rely on computationally demanding global descent through a complicated loss function using an optimizer like L-BFGS. Although this approach is conceptually similar to the patch-based optimization framework, in our case expensive global descent is approximated by a highly efficient approximate nearest-neighbor matching.

The algorithm first extracts neural features for both the source and target image in multiple scales (see Fig. 5). Specifically, we run the input images through the neural network on four resolutions: 1344×1344, 896×896, 448×448 and 224×224. This set was chosen to capture a broader range of neural features.

For this purpose, we use VGG-19 network architecture trained on the ImageNet dataset [17]. After running a feed-forward pass on the input image, features are extracted from 6 different layers of the network. The layers used are *conv*2_2, *conv*3_1, *conv*3_4, *conv*4_1, *conv*4_4, and *conv*5_1. Features are extracted after applying the ReLU activation.

These neural features capture localized semantic similarities found in both images and can be used to guide the patch-based synthesis. However, the high dimensionality of these per-pixel features might significantly compromise both the performance and the quality of the patch-matching step. To avoid this, we reduce the feature dimension using PCA [29]. In particular, we treat each feature vector as an independent point and process feature maps in groups of the same resolution. The number of principal components we extract varies by feature map resolution. We use top 3 components at 1344×1344, top 6 components at 896×896, and finally top 12 components for the two remaining resolutions. We normalize the resulting values to $[0, 255]$ interval and resample them to the required resolution using bicubic upsampling. This can either be lower resolution, typically used in neural techniques, or full resolution of the target image. Lastly, we run the patch-based synthesis algorithm of Fišer et al. [2] to produce the final stylized image. The output is visually comparable to the state-of-the-art [26, 6] (see Fig. 7).

5. Results

We implemented our method both for CPU and GPU, using C++ and CUDA, respectively.

The parameter α is set to make the input images to the neural-based method approximately 400–500 pixels wide. In the case when the input images are already of low-resolution, we set α to be at least 2—to ensure the patch-based synthesis will have enough freedom to fix some of the artifacts caused by the neural-based approach. The α—sub-sampling allows us to get the result from a neural-based approach much faster or use a method that does not support high-resolution input. Moreover, it allows us to significantly suppress some of the artifacts of neural approaches. The parameter β allows us to stylize images of size 346Mpix or even larger, and to get the final result much faster (see an extreme-resolution result in Fig. 1 and our supplementary material). We observed that if the parameter β is in range 1–4, the perceived loss in the quality is almost negligible. If the parameter β is in range 6–10, when zooming closely, one can observe some repetition artifacts, however, the image is sharp and the overall quality is still satisfactory.

We measured run-time and memory performance. For detailed run-time measurement on mid-range laptop see graph in Fig. 8. On a desktop PC, the computational overhead is even lower, e.g., on NVIDIA Quadro M2000, stylizing the image of

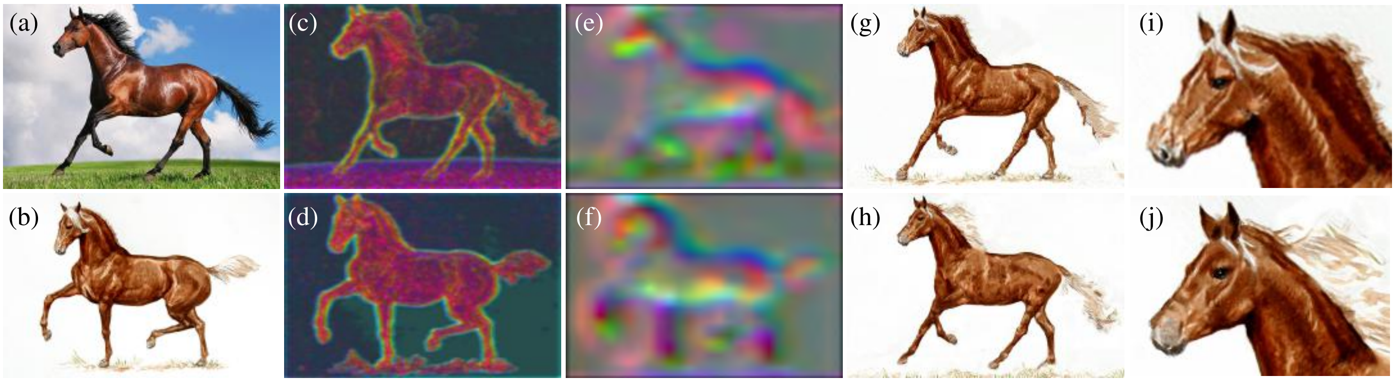

Fig. 7: An example result from our VGG-guided style transfer algorithm: (a) target image, (b) style exemplar, corresponding compressed VGG-responses of low- (c, d) and high-level (e, f) features used as a guide for patch-based synthesis, (g) output of Liao et al. [26], (h) output of our style transfer framework with neural guidance, note how our method can deliver comparable visual quality, cf. details in (i) and (j).

size 160Mpix takes between 3–30 seconds depending on the selection of the parameter β. Increasing the parameter β causes a linear increase in the computational time, while the number of pixels grows exponentially. Our method requires a few hundred MBs of RAM/GPU memory. The exact amount depends on the resolution of the input images and the value of the parameter β.

The performance of the neural-based step depends on a particular method. However, because the input is of very low resolution, 400–500 px wide, the run-time typically ranges between hundreds of milliseconds and several seconds. Most neural-based approaches cannot stylize images larger than 4K-by-4K due to GPU memory constraints. Although there is a possibility to decompose the synthesis into a set of tiles that are processed separately and stitched together, the resulting image would still suffer from the convolutional nature of used neural network introducing disturbing high-frequency artifacts and colors not present in the original style exemplar.

We plugged several different state-of-the-art neural-based style transfer techniques into our framework (see Fig. 9 and 10). In all cases, applying patch-based synthesis with neural transfer output as guidance produces better results than using the neural-based approach alone. The most noticeable differences are visible in (1) the original colors (e.g., saturated pixels that do not appear in the original style exemplar are removed), (2) suppression of checkerboard artifacts caused by deconvolution [30], and (3) results are sharper containing important high-frequency details of the original brush strokes and underlying canvas structure. Fig. 1 demonstrates stylization of a 346Mpix image. Despite the huge resolution, the result is still perfectly sharp and preserves well characteristics of the original artistic media.

To demonstrate the benefit of using the output of the neural approach to guide the patch-based synthesis, we compared our method to the guidance based only on blurred grayscale images (Fig. 3 left column) as proposed in the original Image Analogies method [3], the result does not properly respect the content semantics, causing trees to become pink.

In Fig. 11, we present additional results of our VGG-guided style transfer algorithm. These demonstrate the proposed method can produce convincing stylization without the need to use existing neural techniques as a preprocess.

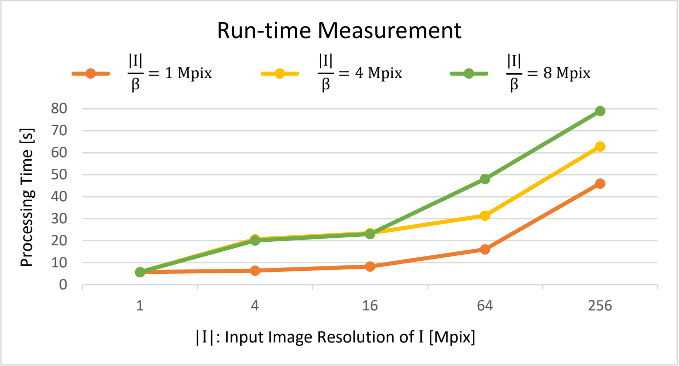

Fig. 8: Performance of our method (full pipeline–Fig. 4, excluding the neural part) on images ranging from resolution of 1Mpx, (i.e. 1000 × 1000 px) to extremely large resolution of 256Mpix (i.e., 16000 × 16000 px). Orange, yellow, and green lines show a case where the parameter β was set such that the patch-based method was run on a resolution of 1Mpix, 4Mpix, and 8Mpix respectively. The measurement was done on a mid-range laptop with NVIDIA GTX 1050 graphics card.

Finally, in Fig. 12, we demonstrate a UI prototype of our method running in Photoshop.

6. Limitations and Future Work

Although in most cases, our approach is capable of delivering significantly better and visually more pleasing results than the underlying neural technique itself, it still relies on the neural result as the initial solution. Due to this reason, we cannot fix large-scale artifacts produced by the neural-based method (see Fig. 13). In the current pipeline, only high-frequency artifacts can be suppressed. When zooming in, the improvement in the texture quality is immediately visible, nevertheless, looking from a distance, high-resolution image obtained by our method may appear almost identical as the result of the underlying neural approach.

As future work, we would like to tackle the issue commonly seen in neural techniques, i.e., many different colors are mixed together within a single coherent region or when the same mixture of colors is used to stylize semantically different regions (see an example in Fig. 14). To address this problem, we see

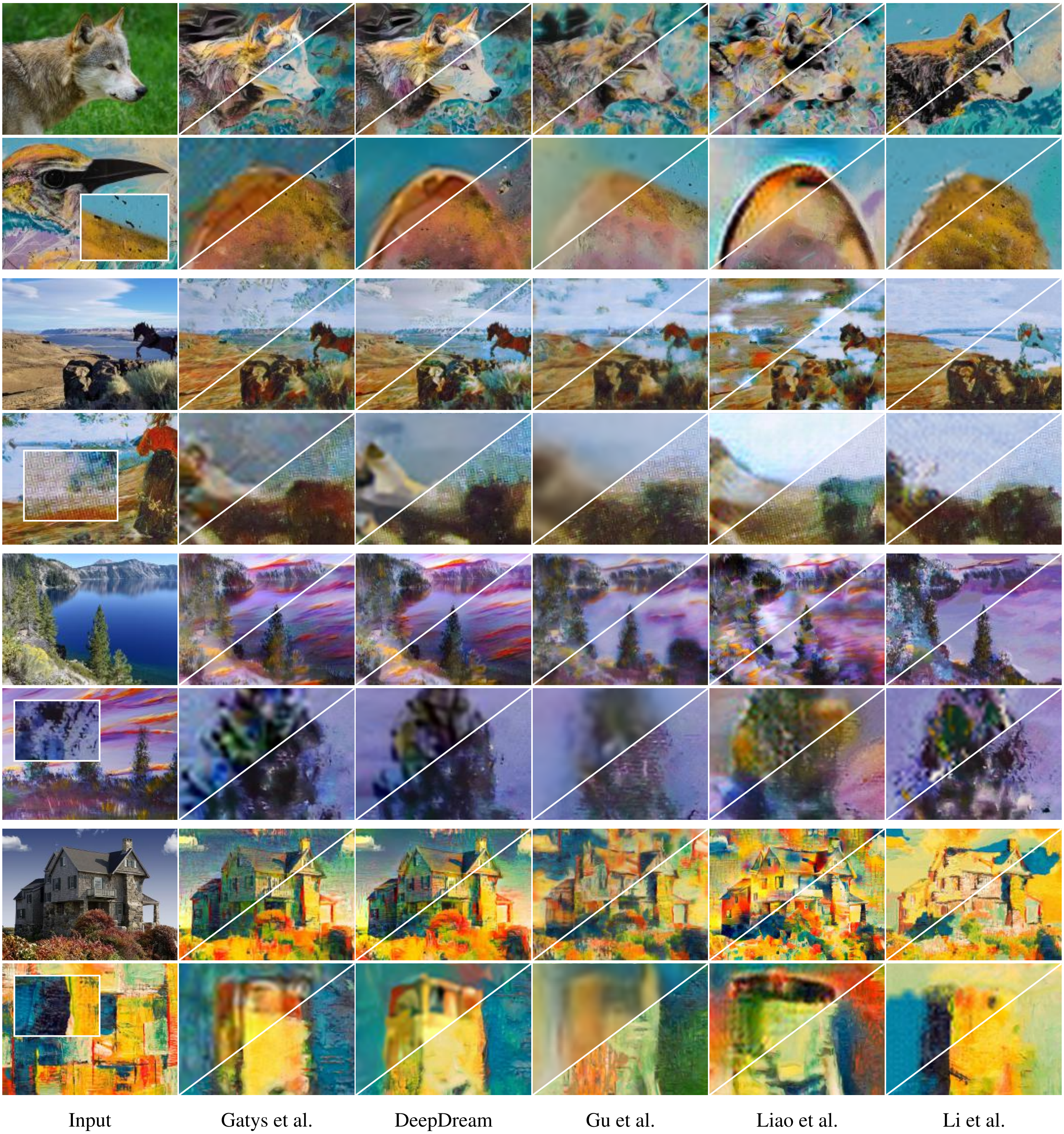

Fig. 9: Our method enhancing the results of five different state-of-the-art neural-based approaches: The leftmost column shows content images and style exemplars (with zoomed patches). Next, left-to-right, are the result of DeepArt based on [1], DeepDream, Gu et al. [6], Liao et al. [26], and Li et al. [22]. The top-left triangle shows the result of the underlying neural-based approach (bicubically up-sampled from a typical size of 600 × 400 px to the target resolution), while the bottom-right shows result enhanced by our method (top row–entire stylized images, bottom row–zoom-in). We encourage the reader to zoom-in into the figure extensively or look into our supplementary material to better appreciate the difference. Our results not only have significantly higher resolution but also better preserve the original colors and canvas structure as well as brush strokes visible in the exemplar painting. Various artifacts caused by the neural approach are significantly suppressed, and contrast is representative of the original artwork. The results thus appear distinctly more faithful. All of the content and target images shown in this figure are of resolution ranging from 4000 × 2200 to 6000 × 4000 px.

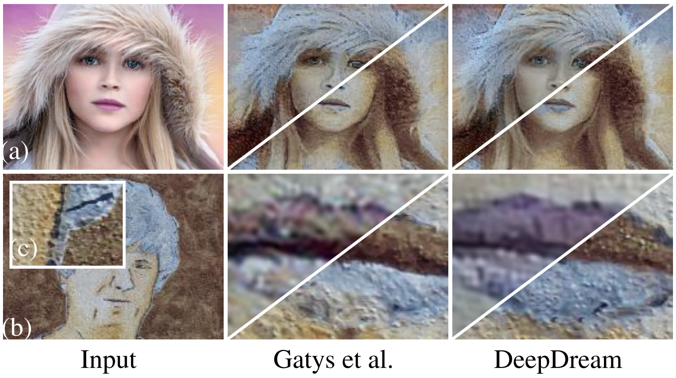

Fig. 10: Portrait on a wall: (a) target content of resolution 4000 × 3000 px, (b) style exemplar of a painting on a wall having the same resolution, (c) 10x zoom-in to the (b) to show fine artistic attributes and structure of the canvas–wall/plaster. Our method is entirely independent of the used artistic medium as well of a canvas the style exemplar is presented on. The results are presented in the same fashion as in Fig. 9.

two promising solutions. First, extending our pipeline in a way that patch-based synthesis is guided by a neural network trained for segmentation on both natural and artistic images to encourage more semantically correct matching of patches. Second, incorporate mask-based loss function as described in [31]. Although, this might not be feasible for all neural-network approaches we use or in a case when it is desired to treat an underlying neural-network as a black box.

Our technique helps to restore high-frequency details and essential attributes of used artistic media; however, in some cases, this process might destroy some of the important content details. We see a promising solution in the work of Calvo [32], where they introduce a technique to intensify or reduce the stylization strength locally.

Another interesting follow-up of our work could be an extension to videos. This might seems straightforward, but even if the video delivered by the underlying neural-based style transfer method is stable in time, randomness in the patch-based step of our pipeline will most likely introduce disturbing temporal inconsistency. To solve this, one could use techniques described in [16] or [5].

Another area for future work worth exploring would be adding interactions to control the result. Also, some of the neural-based approaches support multiple style exemplars; we suggest to explore possibilities of using multiple styles in our enhancing scenario.

7. Conclusion

We have presented a new approach that combines neural and patch-based style transfer techniques, and proposed a way to utilize the generality of the former, while achieving the texture quality of the latter. We introduced a computationally inexpensive algorithm for upscaling the synthesis output to obtain its high-resolution version and a new approach to neural-based style transfer that can use responses of the neural network directly as a guide for patch-based synthesis. Thanks to those advances, we can produce style transfer results with notably larger resolutions than previous neural-based techniques and

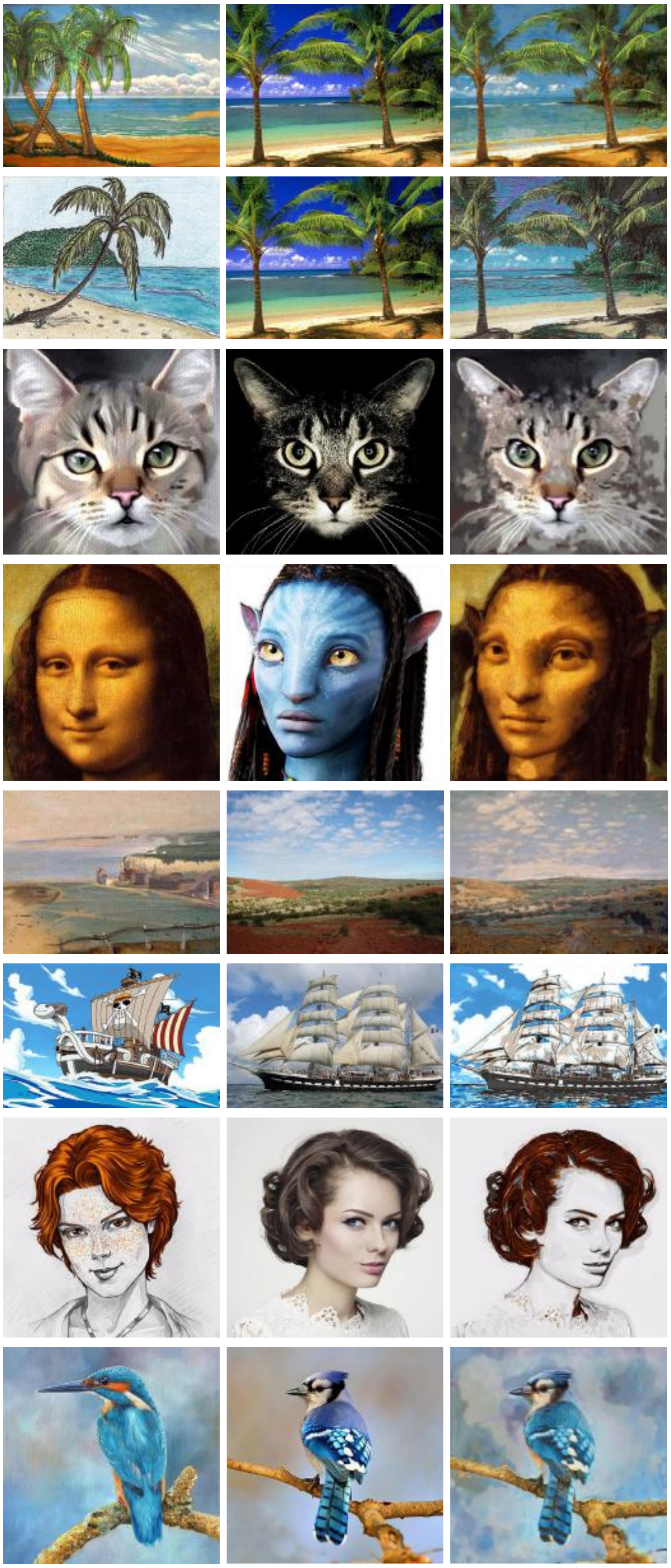

Fig. 11: Results produced by our VGG-guided style transfer algorithm (from left to right): style exemplar, target image, and our result. Our method works well namely in cases when style and target images depict similar content, i.e., when they have compatible VGG activations.

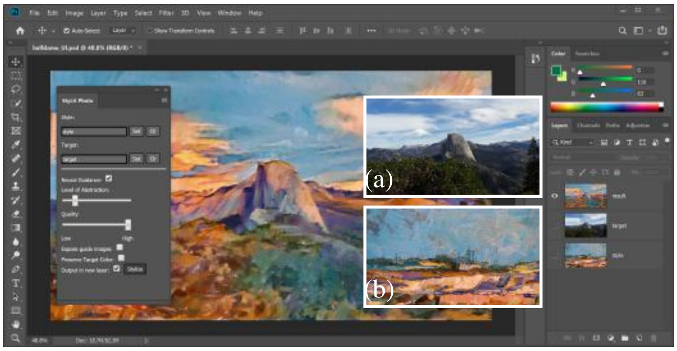

Fig. 12: A screenshot of our method running in Adobe Photoshop: (a) zoom of a target layer, (b) zoom of a style layer; the visible layer is the result of DeepDream enhanced by our method.

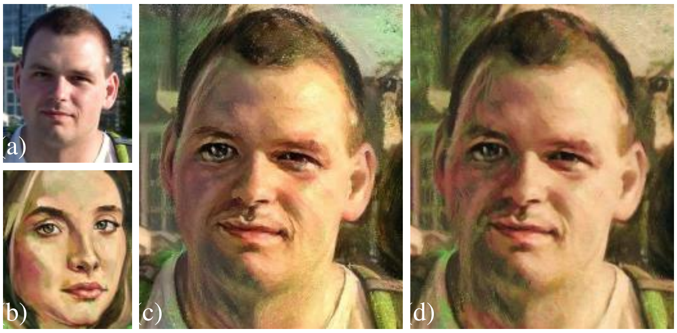

Fig. 13: Large-scale artifact limitation: (a) content image, (b) style exemplar, (c) result of Gatys et al., distortions in eye region are visible, (d) ours, colors and high-frequency details are reproduced well; however, in our current pipeline, large-scale artifacts produced by the underlying neural approach are not fixed. Thus distortion in the eye region is still apparent.

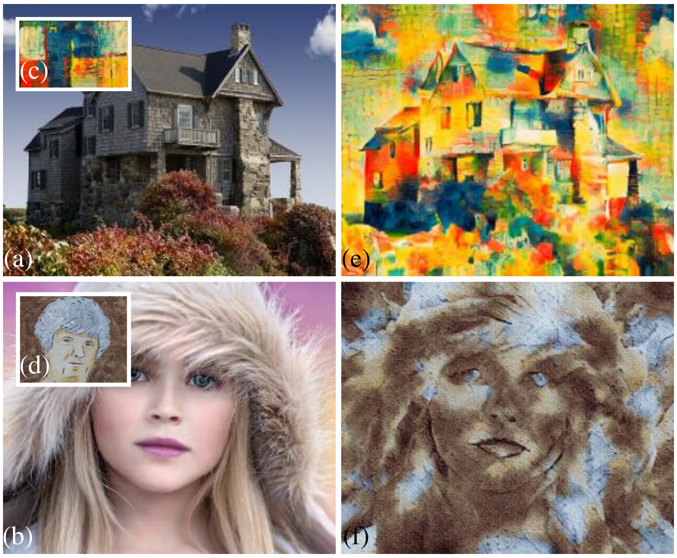

Fig. 14: A limitation common to neural-based approaches: (a-b) content image, (c-d) style exemplar, (e-f) result of [22] enhanced by our method. The content of the original image is not preserved well. In the first case, the similar mixture of colors is used to paint bushes, house, and also the sky. In the second case, all colors appearing in the style exemplar are used to stylize the target regardless of its content. However, high-frequency content is reproduced well. To address this limitation, we propose to incorporate a neural network trained for image segmentation into our pipeline.

significantly reduce the computational overhead while retaining comparable visual quality. We believe our method could enable broader applicability of style transfer methods in commercial practice. To that end, we integrated our approach into Adobe Photoshop in the form of a plug-in.

Acknowledgements

We would like to thank all anonymous reviewers for insightful comments and suggestions. This research began as an internship by Ondřej Texler at Adobe. It was funded by Adobe and has been supported by the Technology Agency of the Czech Republic under research program TE01020415 (V3C – Visual Computing Competence Center), by the Grant Agency of the Czech Technical University in Prague, grant No. SGS19/179/OHK3/3T/13 (Research of Modern Computer Graphics Methods), and by the Research Center for Informatics, grant No. CZ.02.1.01/0.0/0.0/16_019/0000765.

References

[1] Gatys, LA, Ecker, AS, Bethge, M. Image style transfer using convolutional neural networks. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2016, p. 2414–2423.

[2] Fišer, J, Jamriška, O, Lukáč, M, Shechtman, E, Asente, P, Lu, J, et al. StyLit: Illumination-guided example-based stylization of 3D renderings. ACM Transactions on Graphics 2016;35(4):92.

[3] Hertzmann, A, Jacobs, CE, Oliver, N, Curless, B, Salesin, DH. Image analogies. In: SIGGRAPH Conference Proceedings. 2001, p. 327–340.

[4] Johnson, J, Alahi, A, Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of European Conference on Computer Vision. 2016, p. 694–711.

[5] Fišer, J, Jamriška, O, Simons, D, Shechtman, E, Lu, J, Asente, P, et al. Example-based synthesis of stylized facial animations. ACM Transactions on Graphics 2017;36(4):155.

[6] Gu, S, Chen, C, Liao, J, Yuan, L. Arbitrary style transfer with deep feature reshuffle. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2018, p. 8222–8231.

[7] Wexler, Y, Shechtman, E, Irani, M. Space-time completion of video. IEEE Transactions on Pattern Analysis and Machine Intelligence 2007;29(3):463–476.

[8] Kyprianidis, JE, Collomosse, J, Wang, T, Isenberg, T. State of the "art": A taxonomy of artistic stylization techniques for images and video. IEEE Transactions on Visualization and Computer Graphics 2013;19(5):866–885.

[9] Hertzmann, A. Painterly rendering with curved brush strokes of multiple sizes. In: SIGGRAPH Conference Proceedings. 1998, p. 453–460.

[10] Schmid, J, Senn, MS, Gross, M, Sumner, RW. Overcoat: an implicit canvas for 3D painting. ACM Transactions on Graphics 2011;30(4):28.

[11] Lu, J, Barnes, C, DiVerdi, S, Finkelstein, A. RealBrush: painting with examples of physical media. ACM Transactions on Graphics 2013;32(4):117.

[12] Zheng, M, Milliez, A, Gross, MH, Sumner, RW. Example-based brushes for coherent stylized renderings. In: Proceedings of International Symposium on Non-Photorealistic Animation and Rendering. 2017, p. 3.

[13] Kaspar, A, Neubert, B, Lischinski, D, Pauly, M, Kopf, J. Self tuning texture optimization. Computer Graphics Forum 2015;34(2):349–360.

[14] Jamriška, O, Fišer, J, Asente, P, Lu, J, Shechtman, E, Sýkora, D. LazyFluids: Appearance transfer for fluid animations. ACM Transactions on Graphics 2015;34(4):92.

[15] Sýkora, D, Jamriška, O, Texler, O, Fišer, J, Lukáč, M, Lu, J, et al. StyleBlit: Fast example-based stylization with local guidance. Computer Graphics Forum 2019;38(2):83–91.

[16] Jamriška, O, Šárka Sochorová, , Texler, O, Lukáč, M, Fišer, J, Lu, J, et al. Stylizing video by example. ACM Transactions on Graphics 2019;38(4).

[17] Simonyan, K, Zisserman, A. Very deep convolutional networks for large-scale image recognition. CoRR 2014;abs/1409.1556.

[18] Gatys, LA, Ecker, AS, Bethge, M, Hertzmann, A, Shechtman, E. Controlling perceptual factors in neural style transfer. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2017, p. 3985–3993.

[19] Ulyanov, D, Lebedev, V, Vedaldi, A, Lempitsky, V. Texture networks: Feed-forward synthesis of textures and stylized images. In: Proceedings of International Conference on Machine Learning. 2016, p. 1349–1357.

[20] Dumoulin, V, Shlens, J, Kudlur, M. A learned representation for artistic style. CoRR 2016;abs/1610.07629.

[21] Chen, D, Yuan, L, Liao, J, Yu, N, Hua, G. Stylebank: An explicit representation for neural image style transfer. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition 2017;:2770–2779.

[22] Li, Y, Fang, C, Yang, J, Wang, Z, Lu, X, Yang, MH. Universal style transfer via feature transforms. In: Advances in Neural Information Processing Systems. 2017, p. 385–395.

[23] Huang, X, Belongie, SJ. Arbitrary style transfer in real-time with adaptive instance normalization. Proceedings of IEEE International Conference on Computer Vision 2017;:1510–1519.

[24] Lu, M, Zhao, H, Yao, A, Xu, F, Chen, Y, Lin, X. Decoder network over lightweight reconstructed feature for fast semantic style transfer. Proceedings of IEEE International Conference on Computer Vision 2017;:2488–2496.

[25] Li, C, Wand, M. Combining markov random fields and convolutional neural networks for image synthesis. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2016, p. 2479–2486.

[26] Liao, J, Yao, Y, Yuan, L, Hua, G, Kang, SB. Visual attribute transfer through deep image analogy. ACM Transactions on Graphics 2017;36(4):120.

[27] Futschik, D, Chai, M, Cao, C, Ma, C, Stoliar, A, Korolev, S, et al. Real-time patch-based stylization of portraits using generative adversarial network. In: Proceedings of the 8th ACM/EG Expressive Symposium. 2019, p. 33–42.

[28] Fišer, J, Lukáč, M, Jamriška, O, Čadík, M, Gingold, Y, Asente, P, et al. Color Me Noisy: Example-based rendering of hand-colored animations with temporal noise control. Computer Graphics Forum 2014;33(4):1–10.

[29] Turk, M, Pentland, A. Eigenfaces for recognition. Journal of Cognitive Neuroscience 1991;3(1):71–86.

[30] Odena, A, Dumoulin, V, Olah, C. Deconvolution and checkerboard artifacts. Distill 2016;.

[31] Reimann, M, Klingbeil, M, Pasewaldt, S, Semmo, A, Trapp, M, Döllner, J. Locally controllable neural style transfer on mobile devices. The Visual Computer 2019;:1–17.

[32] Calvo, S, Serrano, A, Gutierrez, D, Masia, B. Structure-preserving style transfer. In: Proceedings of Spanish Computer Graphics Conference. 2019, p. 25–30.