The 8th ACM/EG Expressive Symposium
C. Kaplan, A. Forbes, and S. DiVerdi (Editors)

Enhancing Neural Style Transfer using Patch-Based Synthesis

O. Texler[1], J. Fišer[2], M. Lukáč[2], J. Lu[2], E. Shechtman[2], D. Sýkora[1]

[1]Czech Technical University in Prague, Faculty of Electrical Engineering, Czech Republic
[2]Adobe Research, USA

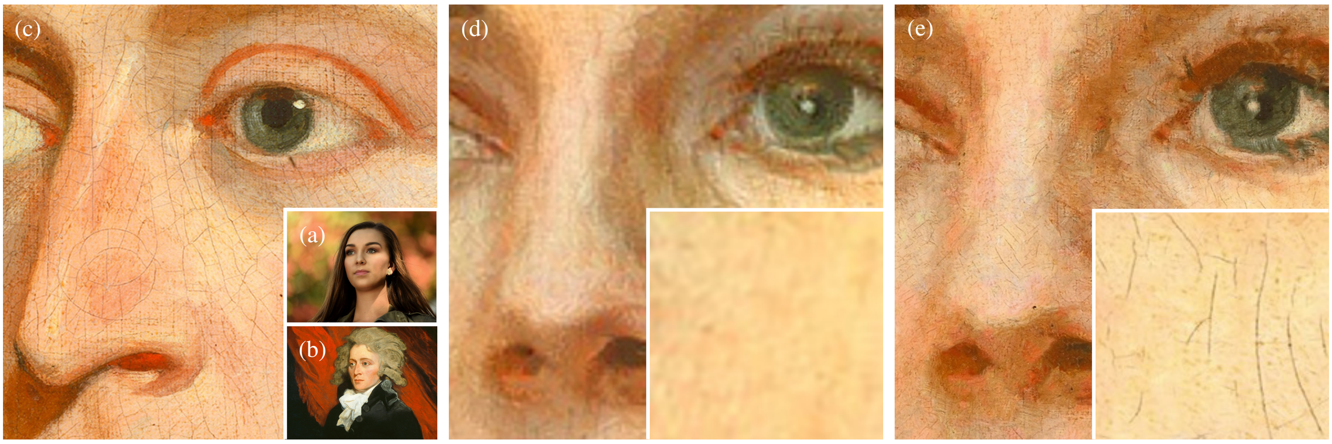

Figure 1: *An example of enhancing the result of neural-based approach using our method: (a) target photograph, (b) style exemplar of the same size, (c) 6× zoom in to the style exemplar, (d) the output of neural-based method DeepArt [GEB16] is capable to perform convincing stylization; nevertheless, the image contains artifacts caused by parametric nature of the used neural network. High-frequency details like a 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 faithfuly, 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 even further.*

Abstract

We present a new approach to example-based style transfer which combines 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 also 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.

CCS Concepts
• ***Computing methodologies*** → ***Non-photorealistic rendering;*** *Image processing;*

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 [JAFF16], DeepArt [GEB16], StyLit [FJL*16], FaceStyle [FJS*17]) 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 which accurately preserves the textural details of the original artistic exemplar.

DOI: 10.2312/exp.20191075

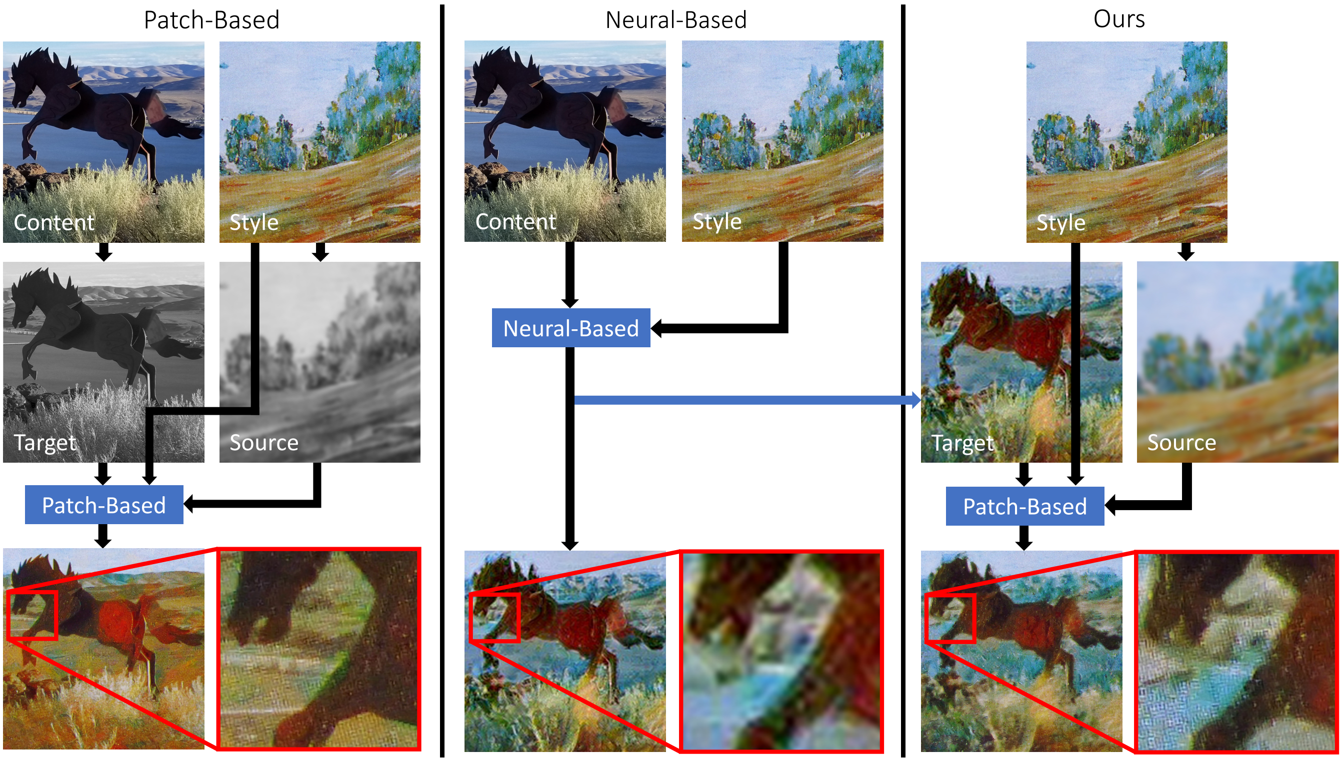

Figure 2: *Simplified scheme of a patch-based, neural-based, and our hybrid style transfer method: The left column shows a patch-based approach [FJL*16] with guidance based on blurred grayscale images as proposed in the original Image Analogies method [HJO*01]. The resulting image has high texture quality 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 [GEB16], 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 guidance channel for patch-based style transfer. Our result attenuates the neural artifacts and restores the original color and texture of the style exemplar.*

Though guided patch-based synthesis approaches [FJL*16, FJS*17] 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 [FJS*17], but designing meaningful guidance automatically in general case remain a difficult problem. On the other hand, neural-based style transfer [GEB16, GCLY18] do 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. Similar limiation holds also 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 aforementioned limitations by combining neural style transfer, patch-based synthesis and dense correspondence field upscape. 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. 1. 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. For 346Mpix (26400×13100 px) example, see Fig. 7. Our approach is general and can utilize any existing neural stylization method. 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

Non-photorealistic rendering [KCWI13] is a well-established field of computer science. Stroke-based approaches are commonly used where a set of predefined strokes are rotated and translated according to guiding information, such as image gradients. This line of approach can be applied in 2D [Her98] or 3D [SSGS11] stylization. The main drawback is the restriction of using a pre-defined set of strokes, textures or patterns which limit the variety and fidelity of the stylization output. This limitation is partly overcome by introducing example-based brushes by [LBDF13], and later extended by [ZMGS17]; however, it is still limited only to a specific domain–brush strokes.

A more robust and general example-based approach called *Image Analogies* was pioneered by Hertzmann et al. [HJO*01]. Given an arbitrary style exemplar and a set of guidance channels, the stylized image can be produced using guided patch-based synthesis [WSI07, KNL*15, FJL*16]. This approach has been applied to various stylization scenarios including fluid animations [JFA*15], 3D renders [FJL*16], or facial animations [FJS*17]. Nevertheless, a common drawback of this method is that it requires the preparation of custom-tailored guidance to deliver compelling stylization quality. Also when applying patch-based synthesis at a higher resolution, huge computational power is required which makes its use in a real production environment hardly accessible.

Neural-based style transfer approaches recently became popular due to advances made by Gatys et al. [GEB16], they successfully applied pre-trained convolutional neural network VGG [SZ14] to the problem of style transfer. The core idea of their method is to match statistics in the domain of VGG [SZ14] features of both the content and style images. While the technique produces impressive stylization results for some particular style exemplars, it usually suffers from loss of high-frequency details of the style exemplar which is caused by the convolutional nature of the underlayin neural network. Moreover, neural techniques usually require non-negligible computational and memory overhead. Although a feed-forward network can be pre-trained to speed up the stylization [JAFF16, ULVL16, DSK16, CYL*17], every new style requires additional costly training. Recently, adoption of encoder–decoder scheme was proposed [LFY*17, HB17, LZY*17] to enable arbitrary style transfer in a feed-forward fashion. 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 which leads to a lower quality of the synthesized imagery at a pixel level.

Recently, attempts to combine patch-based and neural-based techniques were proposed. Li et al. [LW16] search local neural patches from the style image concerning the structure of a content image, which leads to a better reproduction of local textures. Liao et al. [LYY*17] later extended this idea in their *Deep Image Analogy* framework which adapts the concept of *Image Analogies* [HJO*01] in the domain of VGG features. Gu et al. [GCLY18] recently proposed to perform reshuffle in spirit of [KNL*15] to reduce overuse of particular features. 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 [WSI07] 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 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 a initial result (see Fig. 2 middle column). The resulting image on its own lacks high-frequency details of the style exemplar, contains artifacts such as geometric distortions, colors that are not present in the original style, 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 transfer result as guiding channel for patch-based synthesis. This enables us to combine advantages of both techniques and addresses the aforementioned limitations (see Fig. 2 right column). In particular, a pair of guidance channels *Source* and *Target* are needed for guided patch-based synthesis. We use blurred style exemplar as the *Source* guide and the low-resolution neural transfer result as the *Target* guide. After running the guided patch-based synthesis our result (Fig. 2 right column, bottom) effectively attenuates the neural artifacts and restores the original color and texture of the original style exemplar.

Fig. 3 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 paragraphs:

(1) Neural-based style transfer. Both $Style$ (Fig. 3a) and $Content$ (Fig. 3b) images are sub-sampled by a coefficient α. This subsampling step is necessary not only to overcome the resolution restrictions but more importantly it allows us 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. 3c) is then used as a guide for patch-based synthesis method. Its resolution will be improved later in our pipeline.

(2) Guided patch-based synthesis. The output from the neural method (Fig. 3c) 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 [FJL*16] algorithm for synthesis, however, we adpot its original error metric for measuring patch similarity to our needs. Let $\mathcal{S}$ be a style exemplar, $\mathcal{O}$ an output image, and $G^{\mathcal{S}}$

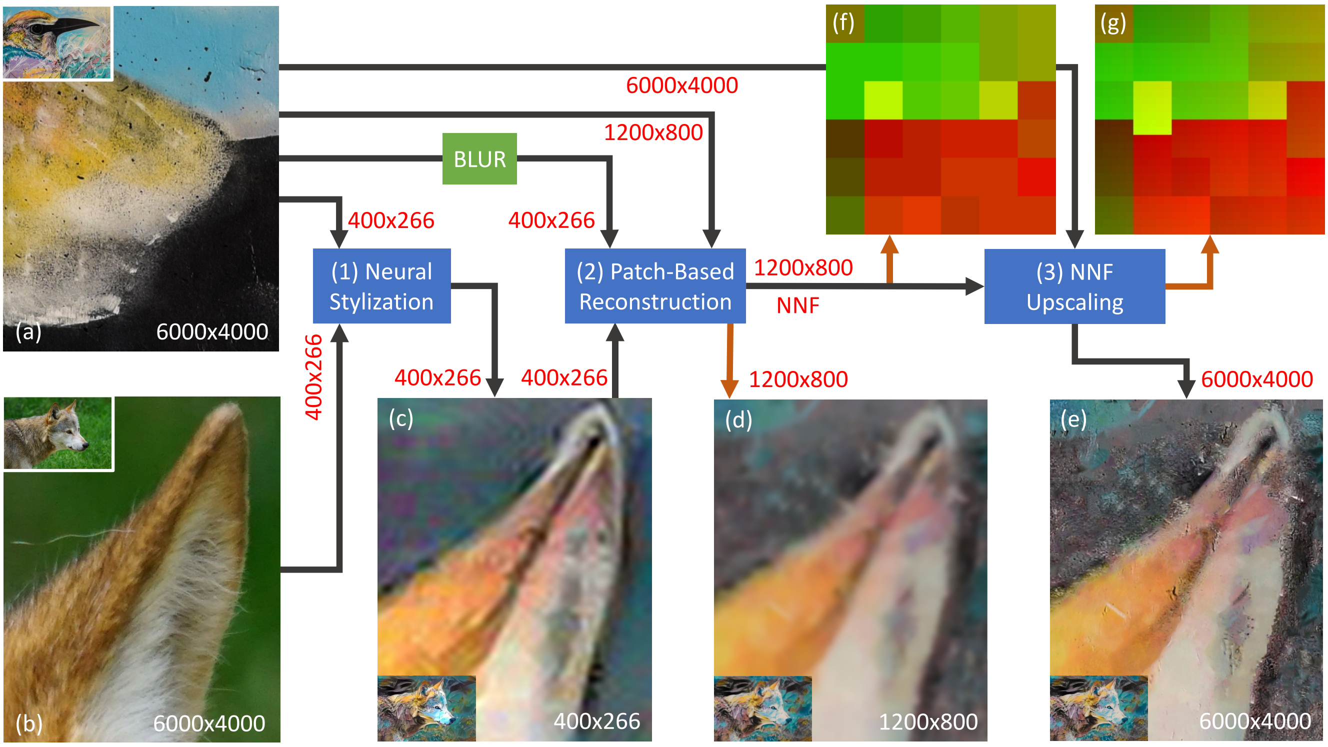

Figure 3: *Proposed pipeline: (a) style exemplar and (b) content image are both subsampled α–times and processed by a neural-based style transfer method (1) which results in low resolution image (c) where fine details are missing and artifacts are apparent (see green and purple checkerboard artifacts). Next, α–times subsampled (a) with β–times subsampled (a) and (c) are used as an input to a patch-based synthesis algorithm (2) which outputs dense nearest neighbour field (NNF) (f) from which the corresponding image (d) can be produced using voting step [WSI07]. Finally, in NNF upscaling (3) is performed, where 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 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 (f) and the style exemplar in its original resolution (a), high-resolution, perfectly sharp final result is created using voting step (e).*

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:

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

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 $\mathcal{O}$. Only $\mathcal{O}$ is iteratively updated during the optimization process described in StyLit [FJL*16].

To obtain *Source* guide image, we use the already sub-sampled style image which was used in the step (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* [FLJ*14], 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. The result of a patch-based synthesis is shown in Fig. 3d; however, in the next step of our pipeline, the output of a patch-based synthesis is used in its dense nearest neighbour field representation (Fig. 3f) and will be turned into an image later.

(3) NNF upscaling. Given the computed NNF–nearest neighbour field (Fig. 3f) and the style exemplar in its original resolution (Fig. 3a), a *voting step* (c.f. [WSI07]) needs to be performed in order to reconstruct the final image. To reduce the computational overhead, we perform the patch-based synthesis (2) at β–times lower resolution than is the original target resolution (β essentially defines the *working resolution* of a patch-based method). Next, the resulting **nnf** (Fig. 3f) is upscaled by a factor of β to obtain the **NNF** (Fig. 3g) of the same resolution as the target image as follows:

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

Finally, we perform a voting step using **NNF** to produce the final high-resolution and sharp image representing the original artistic media and canvas precisely (Fig. 3e).

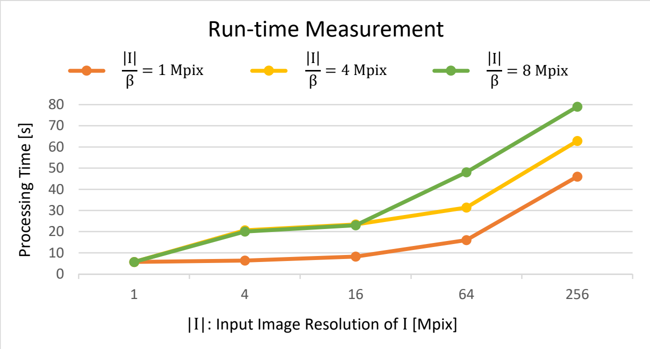

Figure 4: *Performance of our method (full pipeline–Fig. 3, excluding the neural part) on images ranging from resolution of 1Mpx, (i.e.* 1000×1000 *px) to extremely large resolution of 256Mpix (i.e.,* 16000×16000 *px). Orange, yellow, and green lines show a case where parameter* β *was set that the patch-based 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.*

4. 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 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 (26400×13100 px) or even larger, and to get the final result much faster (see an extreme-resolution result in Fig. 7 and our supplementary material). We observed that if the β parameter is in range 1–4, the perceived loss in the quality is almost negligible. If the β parameter 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 4. On a desktop PC computational overhead is even lower, e.g., on NVIDIA Quadro M2000, stylizing the image of size 160 Mpix takes between 3–30 seconds depending on the selection of parameter β. Increasing the parameter β causes that the computational time increases linearly while the number of pixels is growing exponentially. Our method requires a few hundred of MBs of RAM/GPU memory. The exact amount depends on the resolution of the input images and the value of 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 from hundreds of milliseconds to 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 which are then processed separately and stitched together, the resulting image would still suffer from the convolutional nature of used neural network which introduces disturbing high-frequency artifacts and colors that are 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. 5 and 6). In all cases, applying patch-based synthesis with neural transfer result 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 [ODO16], and (3) results are sharper containing important high-frequency details of the original brush strokes and underlying canvas structure. Fig. 7 demonstrates stylization of a 346Mpix (26400×13100 px) image. Despite 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 from neural approach for guiding of the patch-based synthesis we compared our approach to the guidance based only on blurred gray scale images (Fig. 2 left column) as proposed in the original Image Analogies method [HJO*01], the result does not properly respect the content semantics, causing trees to become pink.

Fig. 8 shows UI prototype of our method running in Photoshop.

5. 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. 9). 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 a 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. 10). To address this issue, we suggest extending our pipeline in a way that patch-based synthesis can be guided by a neural network trained for segmentation on both natural and artistic images to encourage more semantically correct matching of patches.

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.

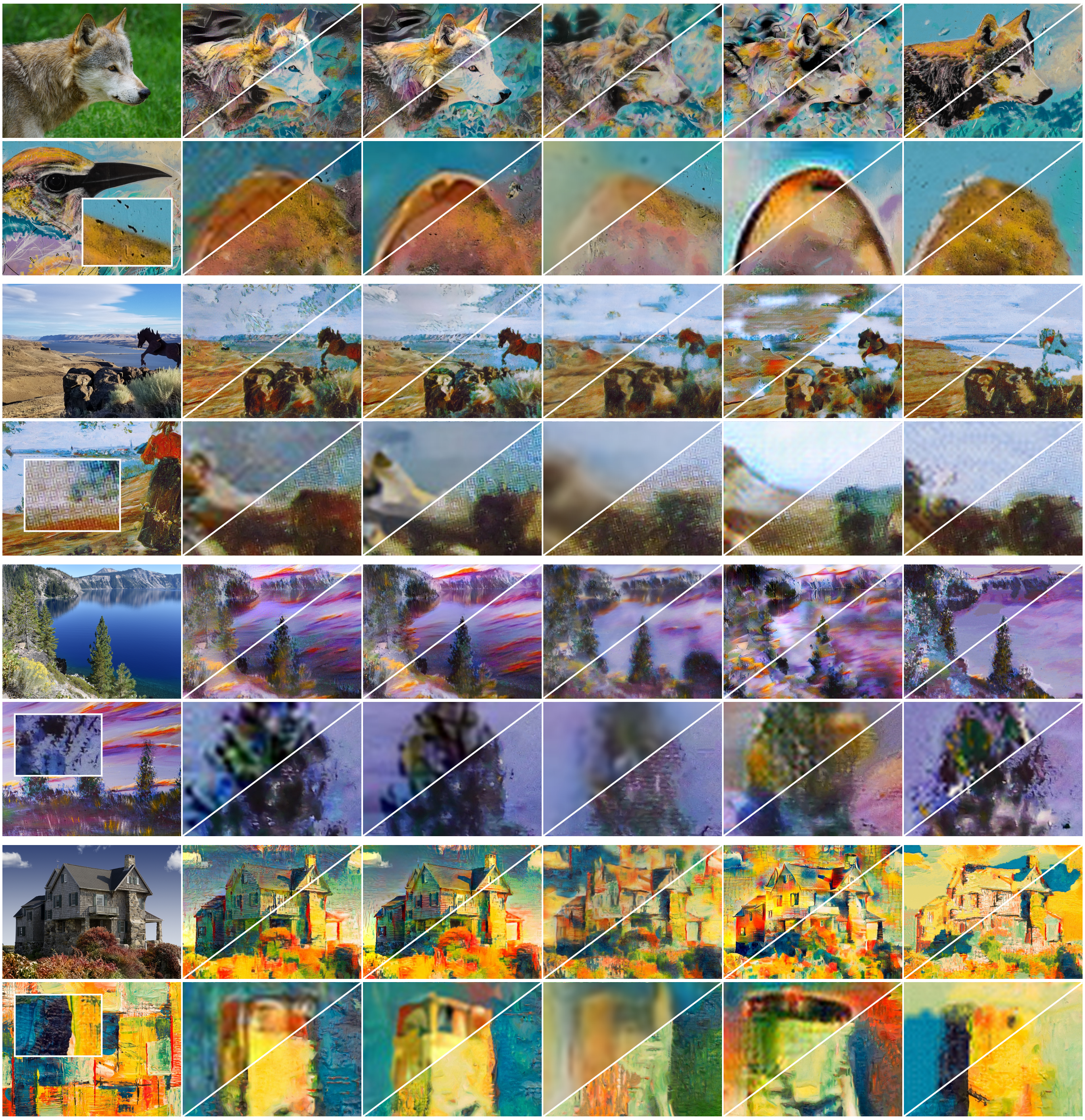

Figure 5: *Our method enhancing the results of five different state-of-the-art neural-based approaches: The leftmost column shows content images and style exemplars (with zoomed patches). Next, left-to-right, are the result of DeepArt based on [GEB16], DeepDream, Gu et al. [GCLY18], Liao et al. [LYY* 17], and Li et al. [LFY* 17]. Top-left triangle shows result of the underlying neural-based approach (bi-cubically 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 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.*

Figure 6: *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 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 5.*

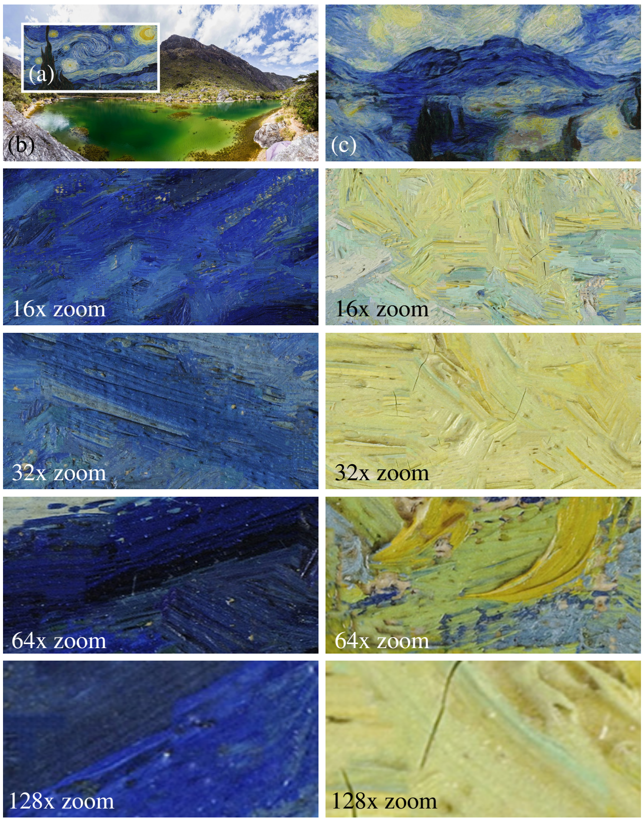

Figure 7: *Extremely high-resolution example: (a) style exemplar of* 26400×13100 *px, (b) content image of the same resolution, (c) result of [GEB16] enhanced by our method. Below, zoom-in patches of different parts of (c) up to zoom of* $128\times$ *are shown; see all the individual brush strokes and its sharp boundaries. Also, notice how well the structure of the original canvas and little cracks of the painting are preserved.*

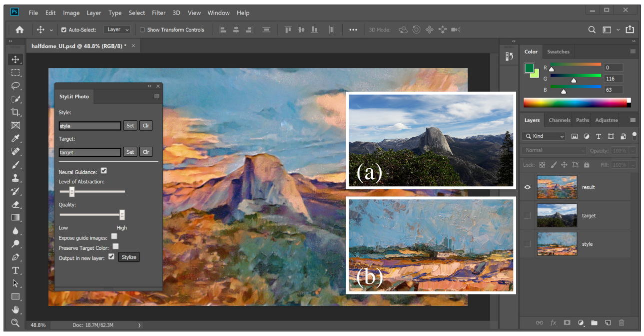

Figure 8: *A screenshot of our method running in Adobe Photoshop: (a) zoom of a target layer, (b) zoom of a style layer; visible layer is result of DeepDream enhanced by our method.*

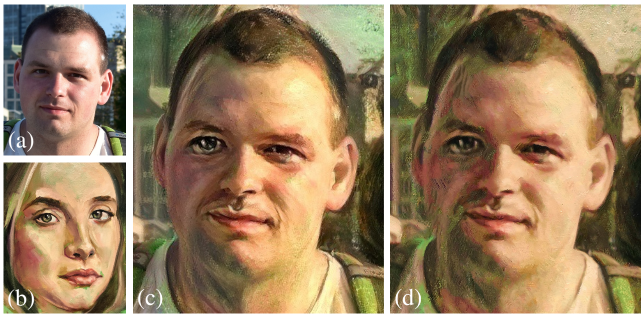

Figure 9: *Large-scale artifact limitation: (a) content image, (b) style exemplar, (c) result of Gatys et al., distortions in eye region are visible, (d) ours, colors and high-frequency details are reproduced well; however, in our current pipeline large-scale artifacts produced by the underlying neural approach are not fixed. Thus distortion in the eye region is still apparent.*

6. Conclusion

We have presented a new approach that combines a 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 also introduced a computationally inexpensive algorithm for upscaling the synthesis output to obtain its high-resolution version. Because of that, we are able to produce results one or two orders of magnitude larger than previous approaches with comparable computational overhead, and thus 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 Uni-

Figure 10: *Limitation common to neural-based approaches: (a-b) content image, (c-d) style exemplar, (e-f) result of [LFY*17] enhanced by our method. 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 its content. However, high-frequency content is reproduced well. To address this limitation, we propose to incorporate neural network train for image segmentation into our pipeline.*

versity 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

[CYL*17] 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. 3

[DSK16] Dumoulin V., Shlens J., Kudlur M.: A learned representation for artistic style. *CoRR abs/1610.07629* (2016). 3

[FJL*16] Fišer J., Jamriška O., Lukáč M., Shechtman E., Asente P., Lu J., Sýkora D.: StyLit: Illumination-guided example-based stylization of 3D renderings. *ACM Transactions on Graphics 35*, 4 (2016), 92. 1, 2, 3, 4

[FJS*17] Fišer J., Jamriška O., Simons D., Shechtman E., Lu J., Asente P., Lukáč M., Sýkora D.: Example-based synthesis of stylized facial animations. *ACM Transactions on Graphics 36*, 4 (2017), 155. 1, 2, 3

[FLJ*14] Fišer J., Lukáč M., Jamriška O., Čadík M., Gingold Y., Asente P., Sýkora D.: Color Me Noisy: Example-based rendering of hand-colored animations with temporal noise control. *Computer Graphics Forum 33*, 4 (2014), 1–10. 4

[GCLY18] 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), pp. 8222–8231. 2, 3, 6

[GEB16] Gatys L. A., Ecker A. S., Bethge M.: Image style transfer using convolutional neural networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2414–2423. 1, 2, 3, 6, 7

[HB17] Huang X., Belongie S. J.: Arbitrary style transfer in real-time with adaptive instance normalization. *Proceedings of IEEE International Conference on Computer Vision* (2017), 1510–1519. 3

[Her98] Hertzmann A.: Painterly rendering with curved brush strokes of multiple sizes. In *SIGGRAPH Conference Proceedings* (1998), pp. 453–460. 3

[HJO*01] Hertzmann A., Jacobs C. E., Oliver N., Curless B., Salesin D. H.: Image analogies. In *SIGGRAPH Conference Proceedings* (2001), pp. 327–340. 2, 3, 5

[JAFF16] 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), pp. 694–711. 1, 3

[JFA*15] Jamriška O., Fišer J., Asente P., Lu J., Shechtman E., Sýkora D.: LazyFluids: Appearance transfer for fluid animations. *ACM Transactions on Graphics 34*, 4 (2015), 92. 3

[KCWI13] Kyprianidis J. E., 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 19*, 5 (2013), 866–885. 3

[KNL*15] Kaspar A., Neubert B., Lischinski D., Pauly M., Kopf J.: Self tuning texture optimization. *Computer Graphics Forum 34*, 2 (2015), 349–360. 3

[LBDF13] Lu J., Barnes C., DiVerdi S., Finkelstein A.: RealBrush: painting with examples of physical media. *ACM Transactions on Graphics 32*, 4 (2013), 117. 3

[LFY*17] Li Y., Fang C., Yang J., Wang Z., Lu X., Yang M.-H.: Universal style transfer via feature transforms. In *Advances in Neural Information Processing Systems* (2017), pp. 385–395. 3, 6, 8

[LW16] 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), pp. 2479–2486. 3

[LYY*17] Liao J., Yao Y., Yuan L., Hua G., Kang S. B.: Visual attribute transfer through deep image analogy. *ACM Transactions on Graphics 36*, 4 (2017), 120. 3, 6

[LZY*17] 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. 3

[ODO16] Odena A., Dumoulin V., Olah C.: Deconvolution and checkerboard artifacts. *Distill* (2016). doi:10.23915/distill.00003. 5

[SSGS11] Schmid J., Senn M. S., Gross M., Sumner R. W.: Overcoat: an implicit canvas for 3D painting. *ACM Transactions on Graphics 30*, 4 (2011), 28. 3

[SZ14] Simonyan K., Zisserman A.: Very deep convolutional networks for large-scale image recognition. *CoRR abs/1409.1556* (2014). 3

[ULVL16] 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), pp. 1349–1357. 3

[WSI07] Wexler Y., Shechtman E., Irani M.: Space-time completion of video. *IEEE Transactions on Pattern Analysis and Machine Intelligence 29*, 3 (2007), 463–476. 3, 4

[ZMGS17] Zheng M., Milliez A., Gross M. H., Sumner R. W.: Example-based brushes for coherent stylized renderings. In *Proceedings of International Symposium on Non-Photorealistic Animation and Rendering* (2017), p. 3. 3