Supplementary Material for "Scaling Painting Style Transfer"

Bruno Galerne[1,2], Lara Raad[3], José Lezama[3†], and Jean-Michel Morel[4]

[1]Institut Denis Poisson, Université d'Orléans, Université de Tours, CNRS [2]Institut Universitaire de France (IUF)
[3]Instituto de Ingeniería Eléctrica, Facultad de Ingeniería, Universidad de la República [4]City University of Hong Kong

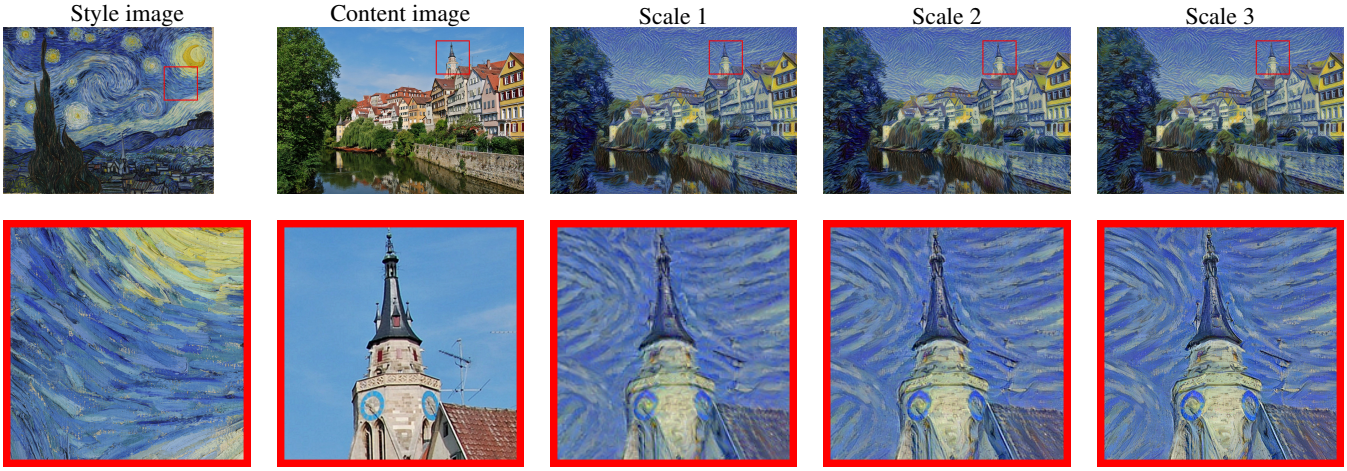

Figure 1: *UHR multiscale style transfer. Top row from left to right: style (3906×4933), content (3906×5800), transfer at scale 1 (976×1450), 2 (1953×2900), 3 (3906×5800). Bottom row: zoomed in detail for each image (from* 200^2 *to* 800^2*). While the transfer is globally stable from one scale to the other, each upscaling allows to add fine pictural details which give an authentic painting aspect to the final output image.*

1. Implementation details and reproducibility

Our implementation is based on PyTorch and our source code is available here: https://github.com/bgalerne/scaling_painting_style_transfer

It mainly uses a class dealing with tensor spatial partitioning to compute the localized gradient block by block.

In most figures the UHR images have been downscaled and saved in jpeg format. UHR images can be downloaded from the dedicated project website https://www.idpoisson.fr/galerne/spst/index.html.

As mentioned in the paper, the values of all the weights λ_c, w_L, w'_L, w''_L, $L \in \mathcal{L}_s$, have been fixed for all images, hence all figures are reproducible.

[†]José Lezama is now at Google Research. Contributed to this work while at Universidad de la República.

2. Visualization of Gram loss correction

A "mean plus std" corrective term is added to the Gram loss to avoid loss of contrast artifacts and grayish color alteration that may occur when minimizing this loss, as previously documented [SC17; RWB17; HVCB21]. Figure 2 shows comparative experiment when using the original Gram loss $E^L_{\text{style}}(x;v)$ instead of our proposed augmented style loss $\tilde{E}^L_{\text{style}}(x;v)$ (see the main paper for equations).

3. Additional style transfer results

3.1. Multiscale example

Figure 1 presents a second example of ultra-high resolution (UHR) style tranfer with each intermediate results of the multiscale algorithm. One can see how the texture of the painting and the stroke details are gradually refined.

3.2. Additional comparison between SPST and SPST-fast

We introduce a faster version of SPST, called SPST-fast that uses less and less iterations as the scale grows. SPST-fast is five times

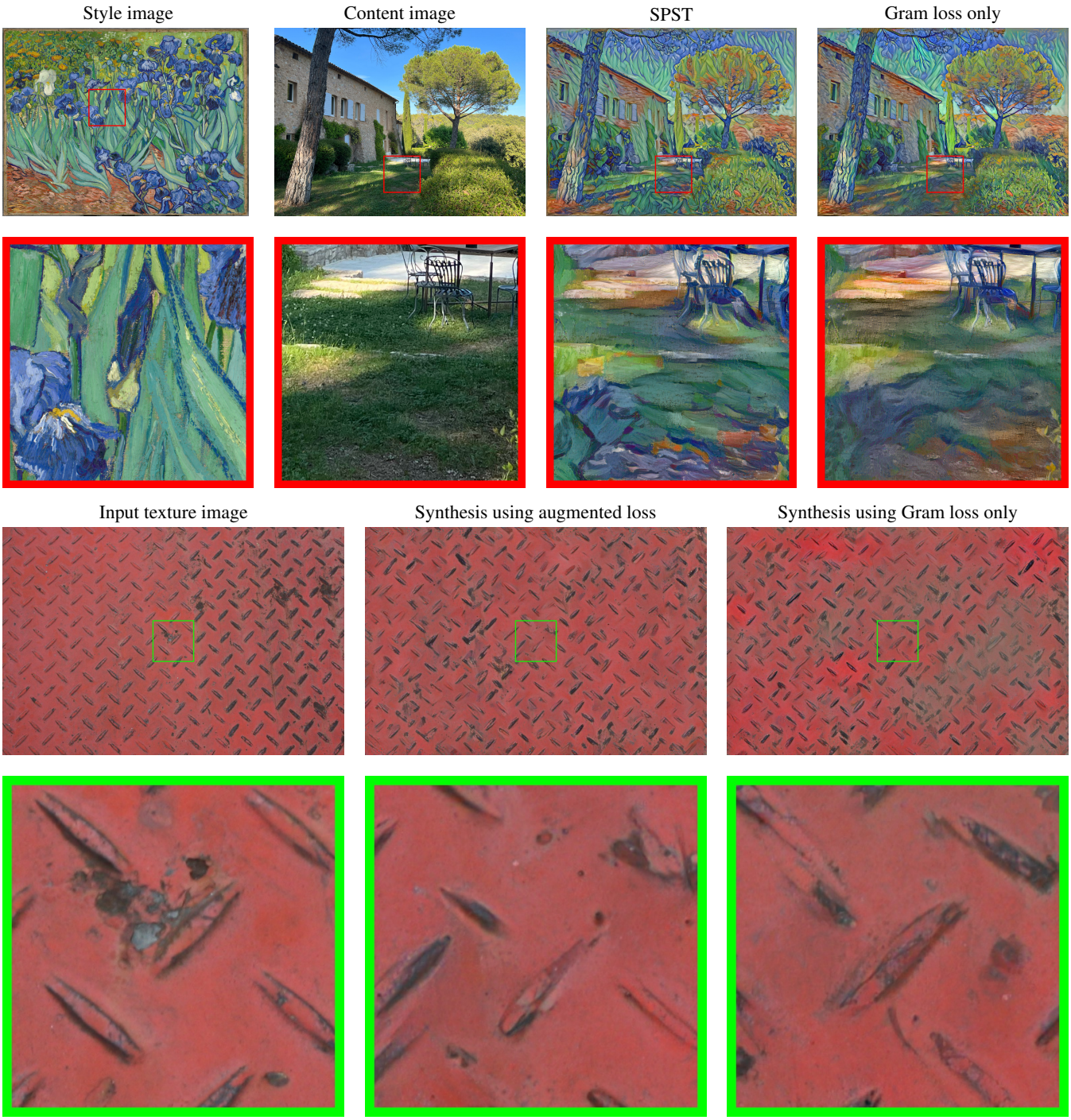

Figure 2: *Interest of the augmented style loss for style transfer and texture synthesis: Top part: Style transfer with style of size 6048×7914, content and transfer results of size 6048×8064. Bottom part: Texture synthesis results for an image of size 2848×4272. Using only the original Gram loss produces gray color areas. Note that for the style transfer example this results in an image with darker colors and less brushstrokes.*

faster than SPST. It produces results that are really close to our SPST baseline. The main difference is that some textural details at the highest scale are less aligned with the UHR content since we only use only a few iterations for the last scale (eg 66 or 30 iterations instead of 300). Figure 3 shows the three examples of style transfer used as comparison in the main paper. One can see that both results are visually very close. A close inspection shows some slight variations such as for the eye detail in the last example.

For the sake of completeness, the results of SPST-fast for the two examples of identity tests are shown in Figure 4. One can observe on the top part of the first close-up that some brush strokes texture are better reproduced with SPST.

3.3. Additional comparison experiments

Figure 5 presents some additional comparison examples. These results show that the competing methods suffer from color imbalance and do not match the fine texture and strokes of the style painting. This is due to the fact that neither URST [CWX*22] nor CD [WLW*20] take into account details at different scales. On the other hand, we are able to convey the appearance of fine details and painting strokes to the content image. In the first example, neither CD nor URST is consistent with the size of the brushstroke visible in the style image. In the second example, CD doubles the frequency of details (e.g., branches), resulting in structural inconsistency, while URST loses the branches completely. In the third example, CD has a halo effect around the bell tower and URST has inconsistent color compared to the input style and diffuminated edges.

Figure 6 shows an additional examples of style transfer on a portrait. Once again, we can observe that our method is the only one able to transfer the texture and strokes of the painting to the portrait content. In this example, URST has inconsistent colors compared to the style image and neither URST nor CD is consistent with the fine scale details of the style brush strokes.

3.4. Failure cases due to content-style mismatch

Figure 7 shows some examples of results in case of content-style mismatches that underline the importance of correctly choosing the couple content/style for a correct style transfer. The first row example uses the same style as in Figure 6 but with a different content image where the proportion of the face space is much larger than in the style image. As result, some red dots from the background are spread over the face. Note also that the lack of a beard in the content produces an undesirable effect on the face. The content image of the second row example lack of details and the style transfer creates undesirable noisy patterns in the sky and grass area. The third example is more extreme and results in loss of details from the content image and synthesis of phantom portrait silhouettes in the flat areas of the image.

As said in the conclusion of the main paper, allowing local control [GEB*17] for UHR style transfer is an interesting direction for future developments.

3.5. Comparison between SPST and AST Using Neurally-Guided Patch-Based Synthesis

Our paper focuses on neural style transfer methods. As said in the main paper, [TFF*20] propose a hybrid method combining neural style transfer and patch-based transfer. Figure 8 shows a comparison between our algorithm and the one described in [TFF*20]. It can be observed that compared to the UHR fast UST methods, the style details coherence at different scales is by far of better quality. Despite that, the content image is only used at the lowest scale for neural style transfer. This yields a final stylized image with high-resolution details unaligned with the original content image. For instance, in the example of figure 8, the chairs present in the content image are no longer present in the stylized result of [TFF*20] (third column) contrary to our result (second column).

3.6. Full resolution images

To limit the main paper file size full resolution UHR images were not included. All UHR images have been downscaled by a factor $\times 4$, with highlighted details included with the true resolution.

In the following figures Figure 9, 10, 11, and 12, we include four experiments with the style and content images with a better resolution (only downgraded by a factor $\times 2$) and our style transfer result in full resolution. Note that images have been compressed using jpeg quality 85 to limit the size of this supplementary material.

4. Painting images dataset for the *identity test* experiment

Figure 13 shows the 79 UHR painting images used for the *identity test* where we evaluate whether a method was able to reproduce a painting when the content and style image were identical.

5. Additional UHR texture synthesis results

Figures 14, 15 and 16 present additional UHR texture synthesis results.

References

[CWX*22] CHEN, ZHE, WANG, WENHAI, XIE, ENZE, et al. "Towards Ultra-Resolution Neural Style Transfer via Thumbnail Instance Normalization". *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 1. June 2022, 393–400. DOI: 10.1609/aaai.v36i1.19916 3, 6, 7.

[GEB*17] GATYS, LEON A., ECKER, ALEXANDER S., BETHGE, MATTHIAS, et al. "Controlling Perceptual Factors in Neural Style Transfer". *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 2017 3.

[HVCB21] HEITZ, ERIC, VANHOEY, KENNETH, CHAMBON, THOMAS, and BELCOUR, LAURENT. "A Sliced Wasserstein Loss for Neural Texture Synthesis". *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2021, 9412–9420 1.

[RWB17] RISSER, ERIC, WILMOT, PIERRE, and BARNES, CONNELLY. *Stable and Controllable Neural Texture Synthesis and Style Transfer Using Histogram Losses*. 2017. DOI: 10.48550/ARXIV.1701.08893 1.

[SC17] SENDIK, OMRY and COHEN-OR, DANIEL. "Deep Correlations for Texture Synthesis". *ACM Trans. Graph.* 36.5 (July 2017). ISSN: 0730-0301. DOI: 10.1145/3015461 1.

Figure 3: *Comparison of SPST and SPST-fast style transfers (same images as the comparison figure of the main paper). First row: content (3168×4752), style (2606×3176), SPST and SPST-fast use three scales; third row: content (3024×4032), style (3024×3787), SPST and SPST-fast use three scales; fifth row: content (4480×5973), style (6000×4747), SPST and SPST-fast use four scales. Both algorithm produce visually similar style transfer. Some details of SPST are slightly better such as the eye contours in the last example.*

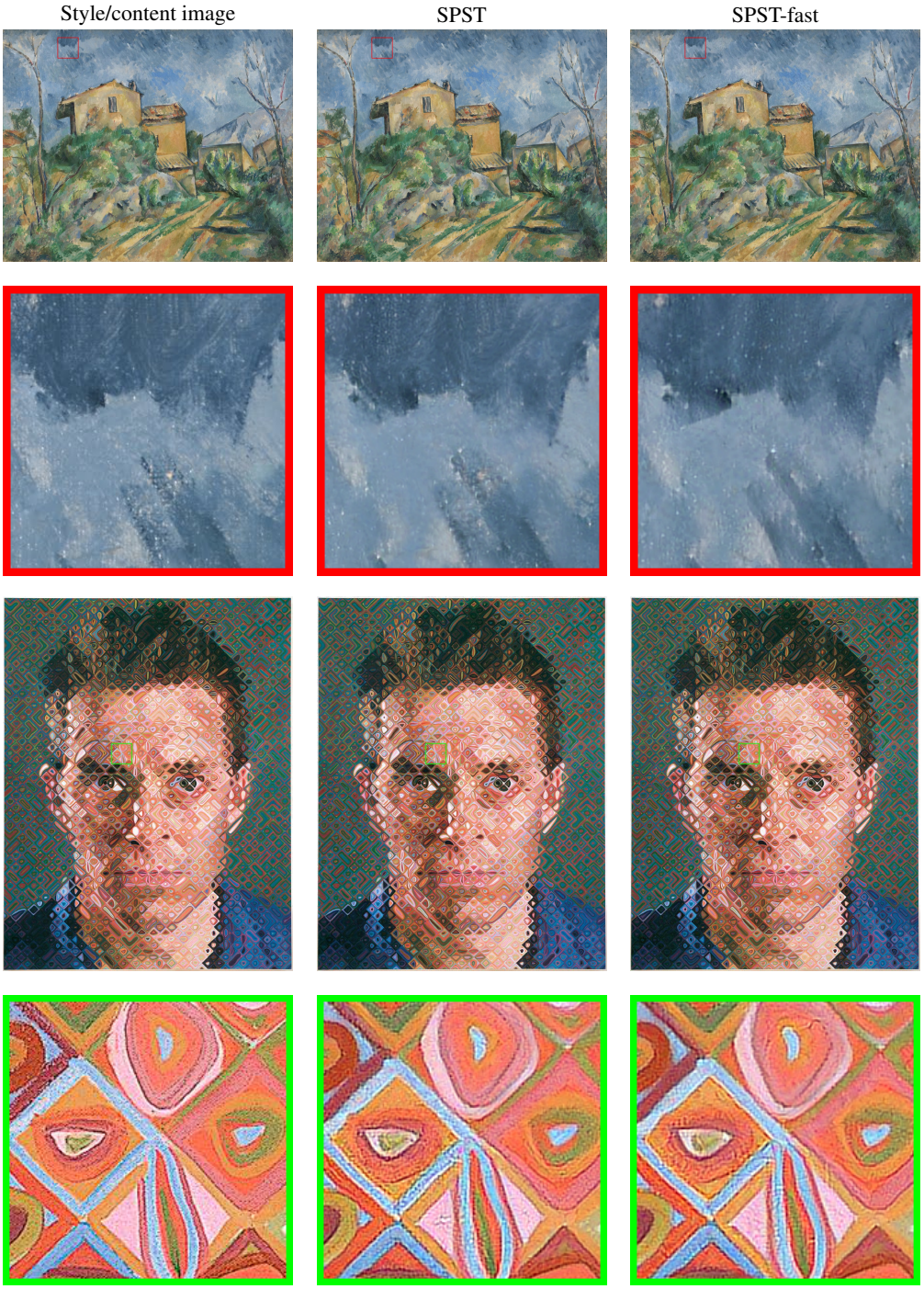

Figure 4: *Identity test: Comparison between SPST and SPST-fast results. First row: The style image has resolution 3375×4201; Third row: The style image has resolution 3095×4000 (UHR images have been downscaled by ×4 factor to save memory). Second and fourth row: Close-up view with true resolution. SPST-fast outputs are visually close to SPST results but have slightly less details.*

Figure 5: *Comparison of UHR style transfers. For each example, top row, left to right: style, content, our result (SPST), CD [WLW*20], URST [CWX*22]. Bottom row: zoom in of the corresponding top row. First row: content (3024×4032), style (3024×4477). Third row: content (3024×4032), style (3024×4738). Fifth row: content (3024×4032), style (3655×2836). We used three scales for all our results. Observe the loss of details and the unrealistic looks of the outputs produced by both fast methods.*

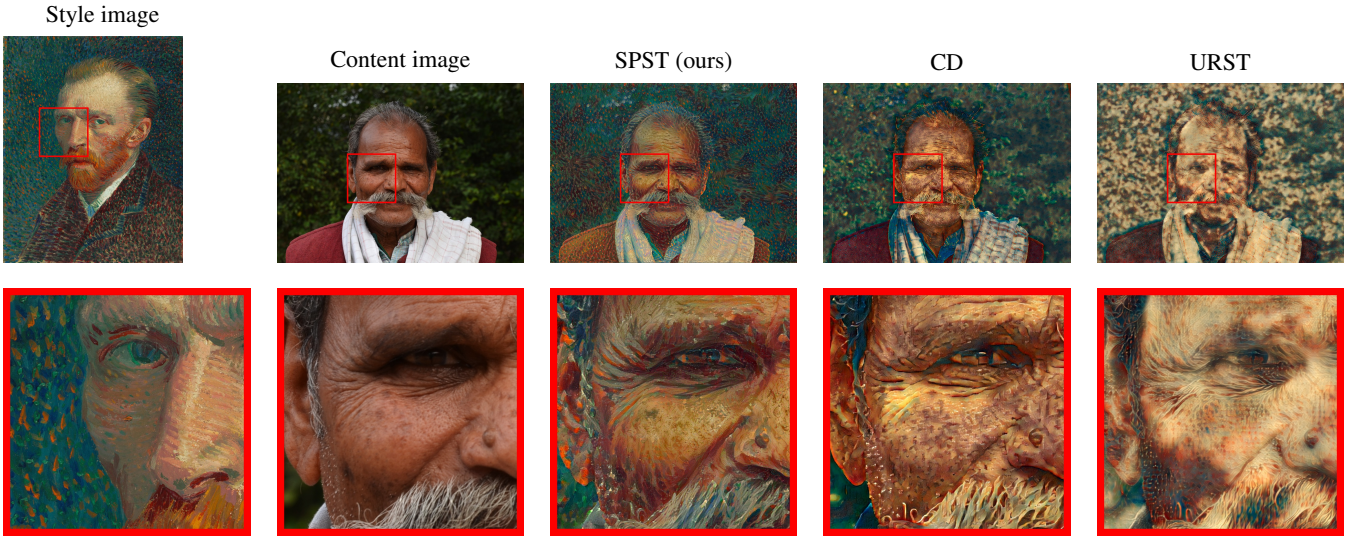

Figure 6: *Comparison of UHR style transfer on portraits. For each example, top row, left to right: style, content, our result (SPST), CD [WLW*20], URST [CWX*22]. Bottom row: zoom in of the corresponding top row. Content (3264×4512), style (4126×3264), SPST uses three scales. Observe the loss of details and the unrealistic looks of the outputs produced by both fast methods, notably in the background and skin texture.*

[TFF*20] TEXLER, ONDŘEJ, FUTSCHIK, DAVID, FIŠER, JAKUB, et al. "Arbitrary style transfer using neurally-guided patch-based synthesis". *Computers & Graphics* 87 (2020), 62–71. ISSN: 0097-8493. DOI: https://doi.org/10.1016/j.cag.2020.01.002 3, 9.

[WLW*20] WANG, HUAN, LI, YIJUN, WANG, YUEHAI, et al. "Collaborative Distillation for Ultra-Resolution Universal Style Transfer". *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2020 3, 6, 7.

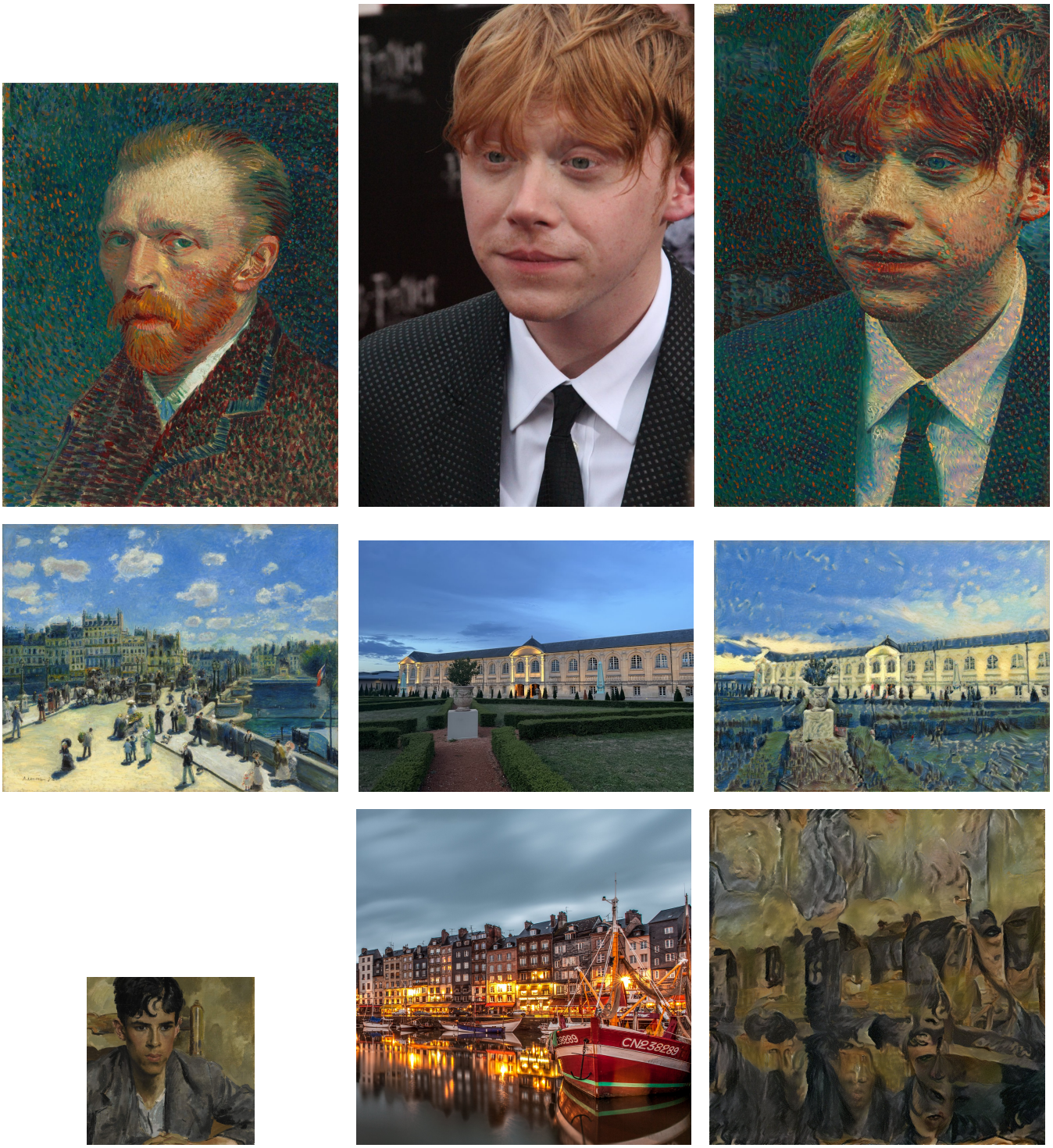

Figure 7: *Examples of UHR style transfer with content-style mismatch. For each example, from left to right: style, content and our result. First row: content (5184×3456), style (4368×3456). Second row: content (3024×4032), style (3024×3787). Third row: content (4096×4096), style (2048×2048). We used three scales for the first two results and four scale for the last example.*

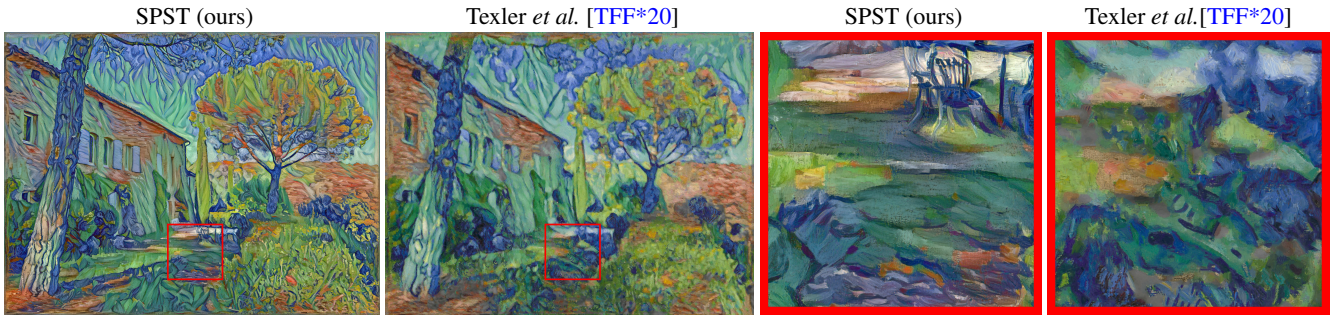

Figure 8: *Comparison to AST Using Neurally-Guided Patch-Based Synthesis: From left to right: Our result, result of [TFF*20] (after a Gatys style transfer of size 384×512), and close up detail. By design the result of [TFF*20] is not faithful to the HR details of the content image. Here the chair simply disappear in the output result while it is reproduced using fine brushstrokes in the SPST output*

Figure 9: *UHR style transfer in full resolution (from Figure 1 of the main paper). Top row: style image (4226×5319), content image (6048×8064). Bottom: result (6048×8064). Observe that fine details such as the canvas texture in the sky is well transferred.*

Figure 10: *UHR style transfer in full resolution (from Figure 2 of the main paper). Top row: style image (6048×7914), content image (6048×8064). Bottom: result (6048×8064). Observe how very fine details such as the chairs look as if painted.*

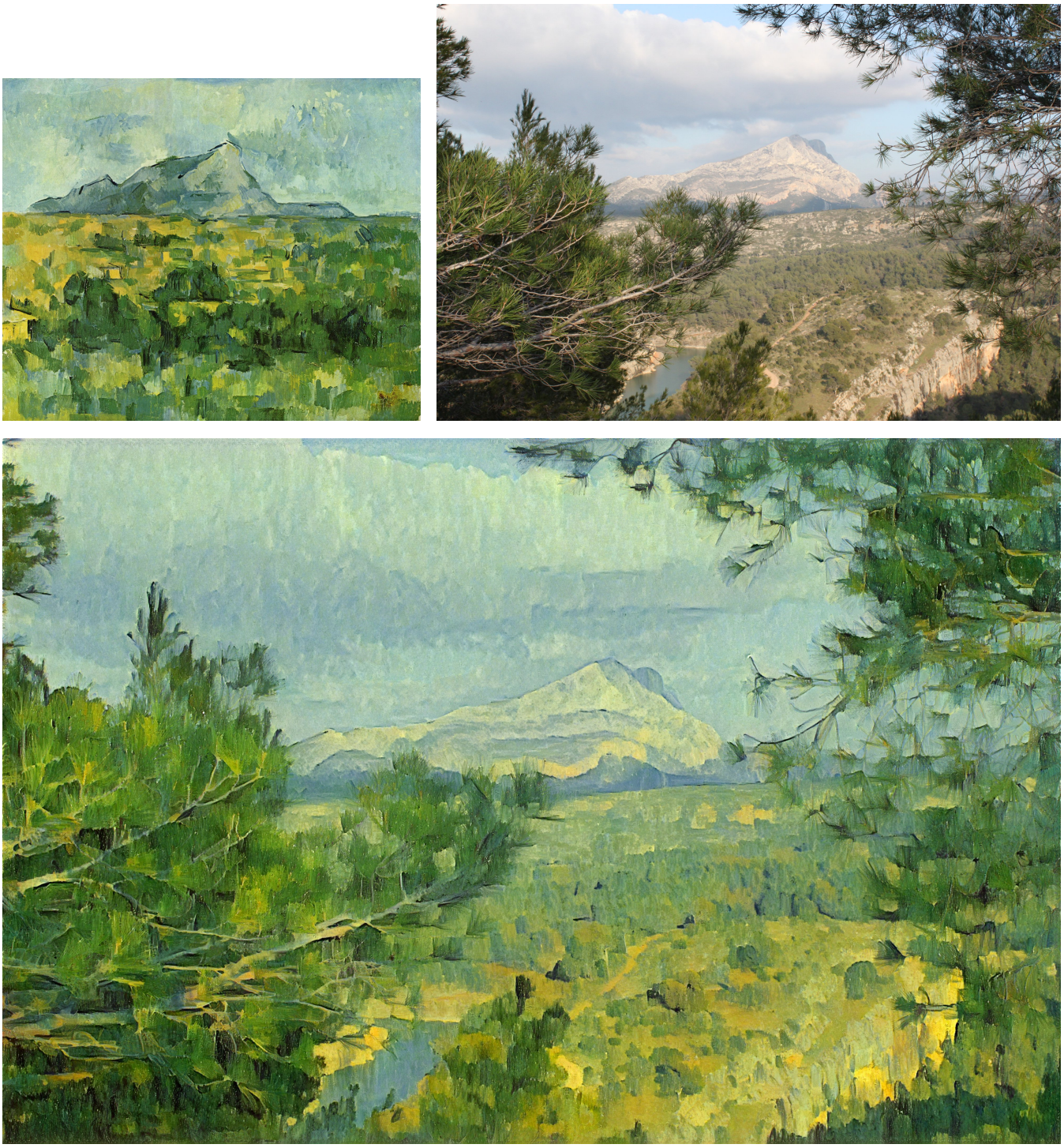

Figure 11: *UHR style transfer in full resolution (from Figure 4 of the main paper). Top row: style image (2606×3176), content image (3168×4752). Bottom: result (3168×4752).*

Figure 12: *UHR style transfer in full resolution (from Figure 4 of the main paper). Top row: style image (3024×3787), content image (3024×4032). Bottom: result (3024×4032).*

Figure 13: *Overview of the* 79 *UHR painting images used for the* identity test.

Figure 14: *UHR texture synthesis: From left to right: Input texture image, synthesis using 5 scales, synthesis using 3 scales. Image have size (3024×4032) (downscaled by a factor 4 for inclusion in the .pdf) and true resolution details have size (512×512).*

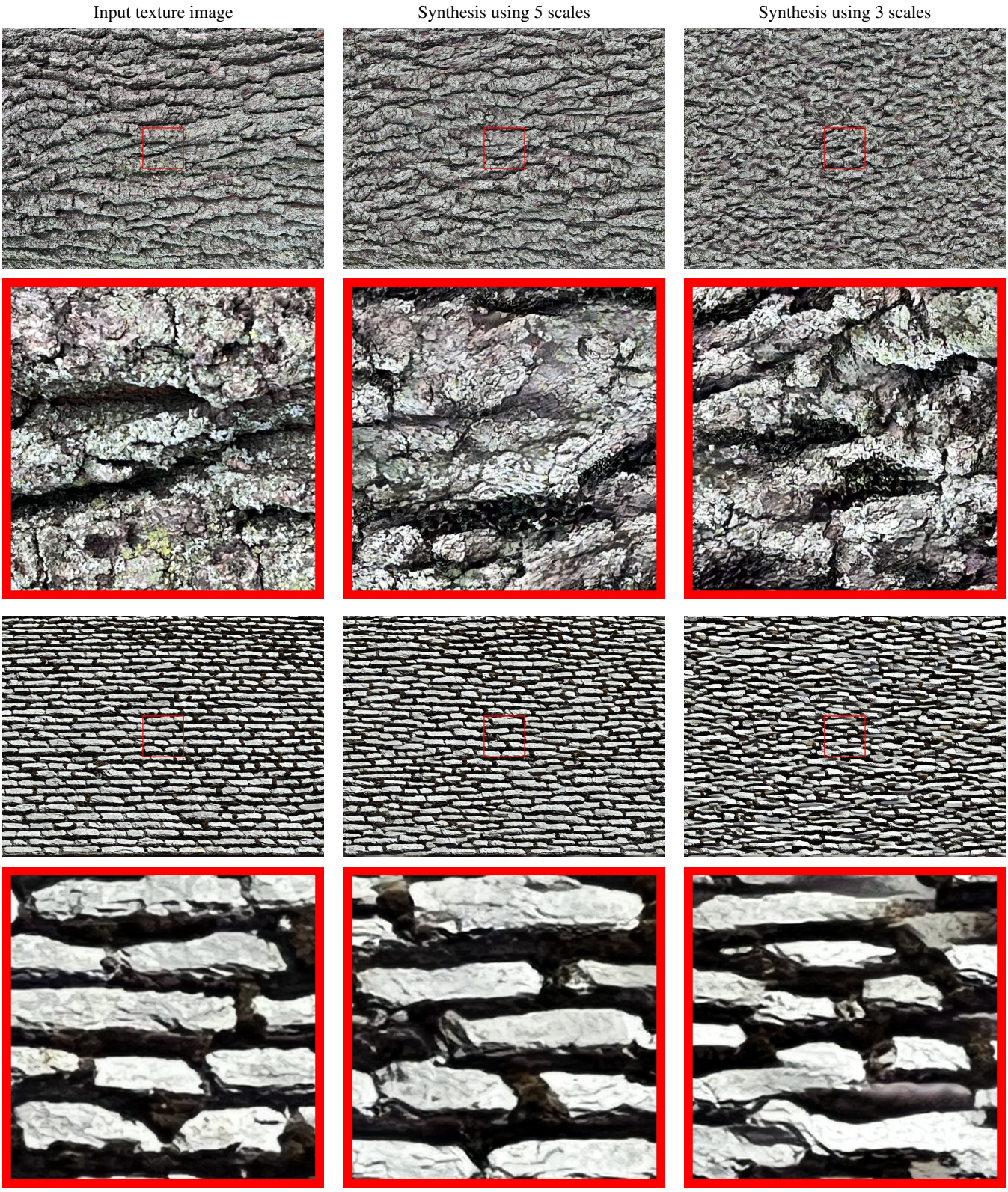

Figure 15: *UHR texture synthesis (same as Figure 14): From left to right: Input texture image, synthesis using 5 scales, synthesis using 3 scales. Image have size (3024×4032) (downscaled by a factor 4 for inclusion in the .pdf) and true resolution details have size (512×512).*

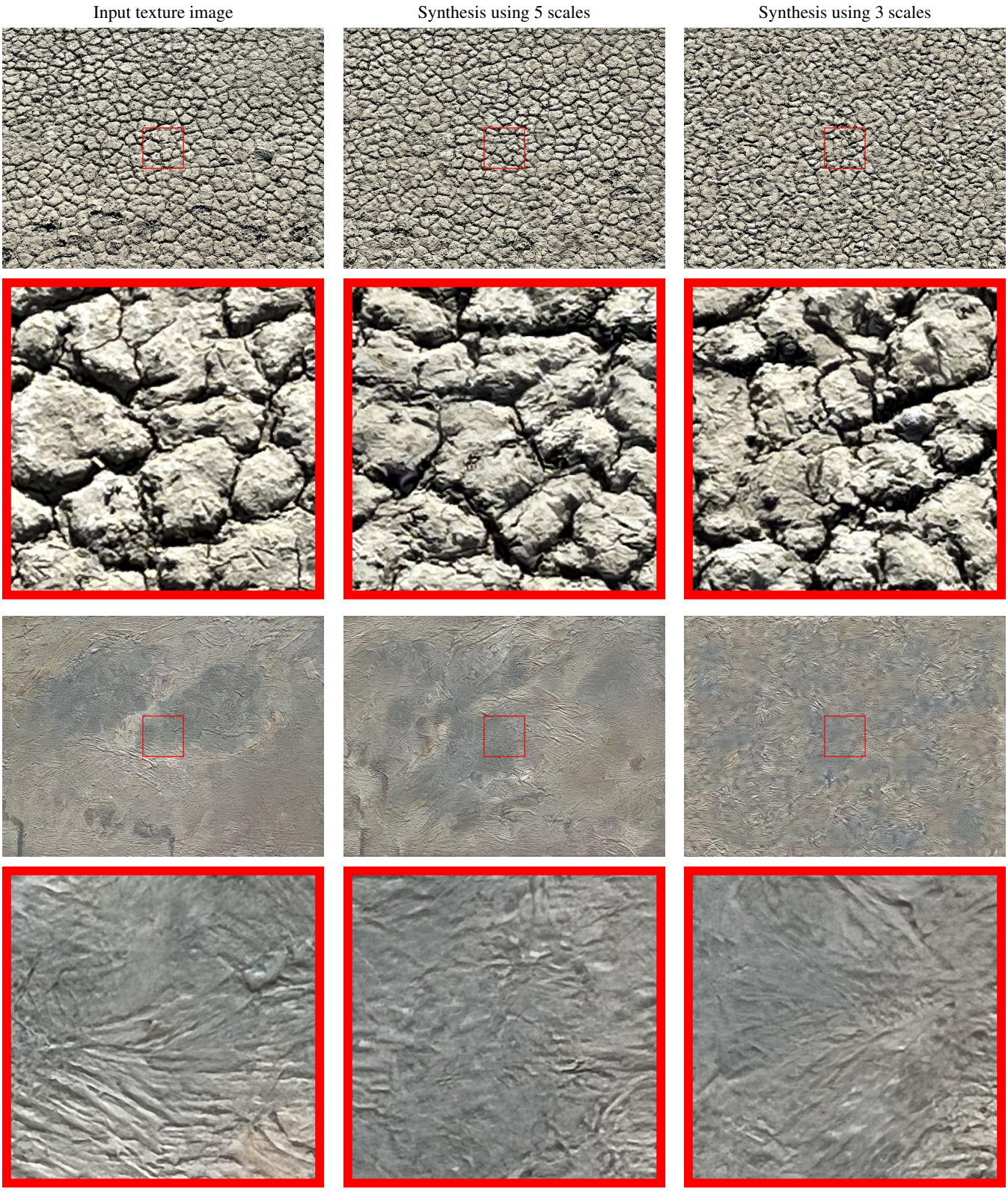

Figure 16: *UHR texture synthesis (same as Figure 14): From left to right: Input texture image, synthesis using 5 scales, synthesis using 3 scales. Image have size (3024×4032) (downscaled by a factor 4 for inclusion in the .pdf) and true resolution details have size (512×512).*