Supplementary Material for
Preserving Photographic Defocus in Stylised Image Synthesis

Hong-Yi Wang [1] and Yu-Ting Wu [†2]

[1] s711283120@gm.ntpu.edu.tw, National Taipei University, Taiwan
[2] yutingwu@mail.ntpu.edu.tw, National Taipei University, Taiwan

This supplementary document outlines the implementation details of the layer-based defocus rendering algorithm and presents additional results of our method.

1. Implementation details of layer-based defocus rendering algorithm

---
Algorithm 1 Layer-based defocus rendering algorithm.

Input: Stylized image I with estimated blur map B, optimal maximum filter size K, and the range of blur magnitude $[b_{\min}, b_{\max}]$ in the blur map.

Output: Stylized image I_b with defocus blur

1: $t = (b_{\max} - b_{\min})/(K+1)$
2: **for** $l \leftarrow 0$ to K **do**
3:     $M^l = (B \geq (b_{\min} + l \cdot t)) \wedge (B \leq (b_{\min} + (l+1) \cdot t))$
4:     $I^l = M^l \odot I$
5:     $B^l = M^l \odot B$
6:     $b^l = \text{Mean}(B^l)$
7: **end for**
8: $b^{\min} = b^0$
9: $b^{\max} = b^K$
10: $I_s = [0]$
11: $M_s = [0]$
12: **for** $l \leftarrow 0$ to K **do**
13:     $r^l = K \cdot (b^l - b^{\min})/(b^{\max} - b^{\min})$
14:     $M_b^l = M^l * G(r^l)$
15:     $I_b^l = I^l * G(r^l)$
16:     $M_s = M_s \odot (1 - M_b^l) + M_b^l$
17:     $I_s = I_s \odot (1 - M_b^l) + I_b^l$
18: **end for**
19: $I_b = I_s \oslash M_s$

The step-by-step procedure of our layer-based defocus rendering is outlined in Algorithm 1. The algorithm takes as input a stylized image I generated by a style transfer method, the estimated blur map B, and the optimal maximum filter size K predicted from the kernel mapping network. In addition, the minimum and maximum blur magnitudes, $b_{\min}$ and $b_{\max}$, are derived from the blur map and provided as inputs to the algorithm.

In our formulation, given a maximum filter size K, the blur map is quantized into $K+1$ discrete blur levels, where each level corresponds to a blur magnitude interval of $t = (b_{\max} - b_{\min})/(K+1)$. Each pixel in the content image is then assigned to one of these blur levels based on its blur magnitude in the blur map. Every blur level is associated with a Gaussian filter of size ranging from 0 to K, reflecting the corresponding degree of blur. Collectively, these $K+1$ filters constitute the Gaussian filter bank.

Next, in lines 2-7, we construct per-level data for each blur level. For a given level l, a level mask M^l is generated by identifying all pixels associated with that specific blur level, that is, whose blur magnitudes lie within the interval $[b_{min} + l \cdot t, b_{min} + (l+1) \cdot t]$. Using this mask, we obtain the partial stylized image I^l, which contains only the pixels of that level, through Hadamard multiplication: $I^l = M^l \odot I$. We also compute the average blur magnitude b^l for each level. After all levels are processed, the average blur magnitudes at the minimum (level 0) and maximum (level K) levels are denoted as $b^{\min}$ and $b^{\max}$, respectively.

Lines 12-18 implement the layer-based defocus rendering, progressing from the lowest to the highest blur level. Two accumulation buffers, M_s and I_s, are maintained to store the progressively blurred mask and image, respectively. For each level l, the Gaussian filter size r^l is computed as $r^l = K \cdot (b^l - b^{\min})/(b^{\max} - b^{\min})$. This guarantees that the minimum blur level is assigned a filter radius of zero, while the maximum blur level is assigned the maximum filter radius K. A Gaussian filter with this size is then applied to both the level mask M^l and the partial corresponding stylized image I^l, producing blurred outputs M_b^l and I_b^l. These are then accumulated into M_s and I_s, restricted to pixels where the mask is nonzero. This strategy prevents out-of-focus regions from bleeding into in-focus areas, thereby mitigating blur clipping. After processing all levels, the final stylized image with defocus blur, I_b, is obtained by normalizing I_s with M_s.

2. Additional results

Figure 1-6 present additional results of applying our method to various existing style transfer techniques, including aesthetic-aware [HJL*20] (Figure 1), depth-aware [IM22] (Figure 2 and 3), attention-aware [ZXW*24] (Figure 4 and 5), and brushstroke-

based [KWHO21] (Figure 6) approaches. Across all cases, our method effectively preserves the defocus effects from the original content images in the stylized outputs and consistently achieves lower perceptual error (LPIPS).

References

[HJL*20]   HU Z., JIA J., LIU B., BU Y., FU J.: Aesthetic-aware image style transfer. In *Proc. ACM International Conference on Multimedia* (2020), p. 3320–3329. `doi:10.1145/3394171.3413853`. 1, 3

[IM22]   IOANNOU E., MADDOCK S.: Depth-aware neural style transfer using instance normalization. In *Proc. Computer Graphics and Visual Computing* (2022). `doi:10.2312/cgvc.20221165`. 1, 4, 5

[KWHO21]   KOTOVENKO D., WRIGHT M., HEIMBRECHT A., OMMER B.:   Rethinking style transfer: From pixels to parameterized brush-strokes. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (June 2021), pp. 12191–12200. `doi:10.1109/CVPR46437.2021.01202`. 2, 8

[ZXW*24]   ZHANG C., XU X., WANG L., DAI Z., YANG J.: S2WAT: image style transfer via hierarchical vision transformer using strips window attention. In *Proc. International Conference on Neural Information Processing Systems (AAAI)* (2024). `doi:10.1609/aaai.v38i7.28529`. 1, 6, 7

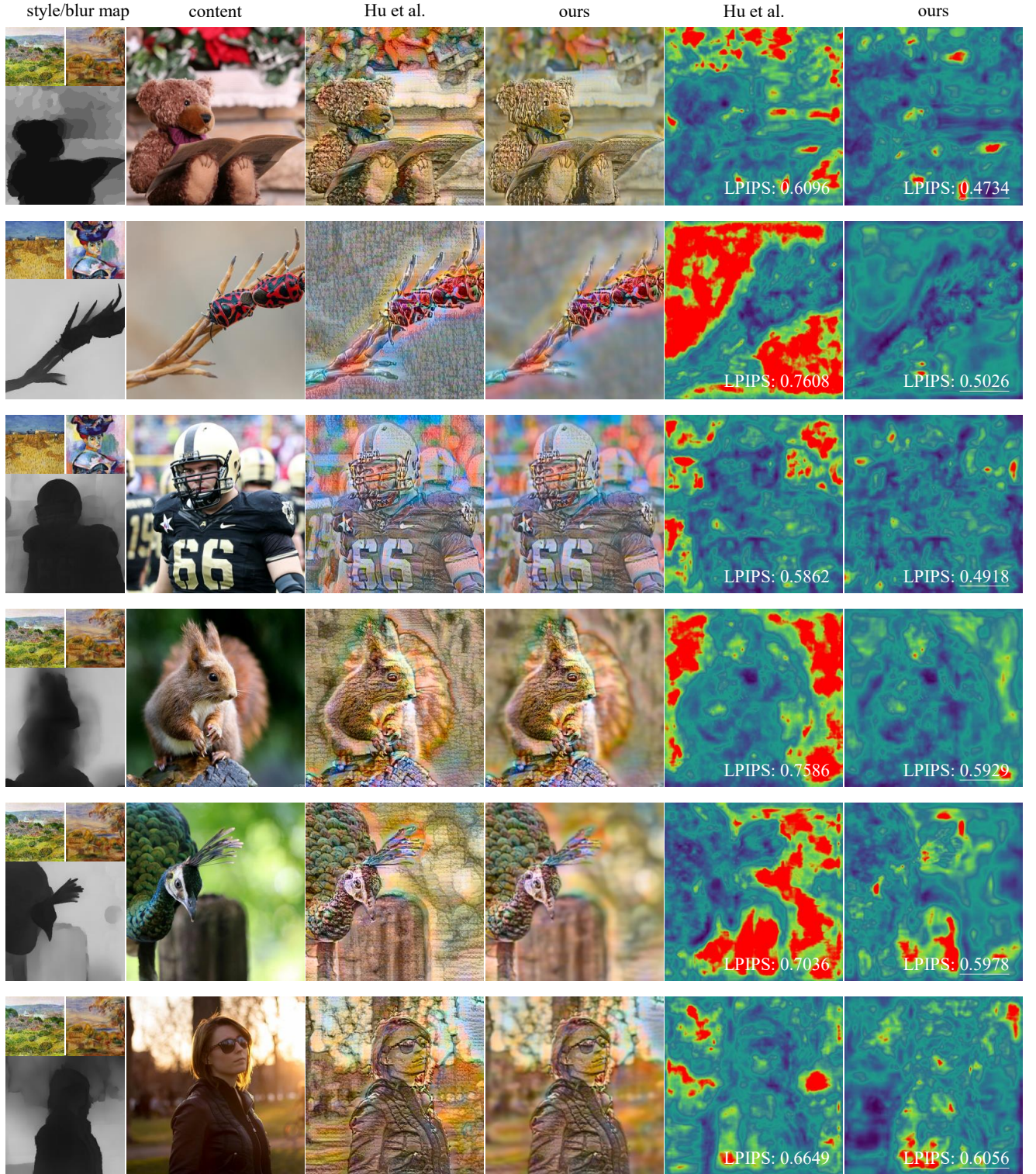| style/blur map	content	Hu et al.	ours	Hu et al.	ours

Figure 1: Additional results using our method with the aesthetic-aware style transfer approach [HJL*20]. The first column shows the texture (top left) and color (top right) style reference images, together with the blur map (bottom) estimated from the content image in the second column. Columns three and four present the stylized output of Hu *et al.* [HJL*20] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.
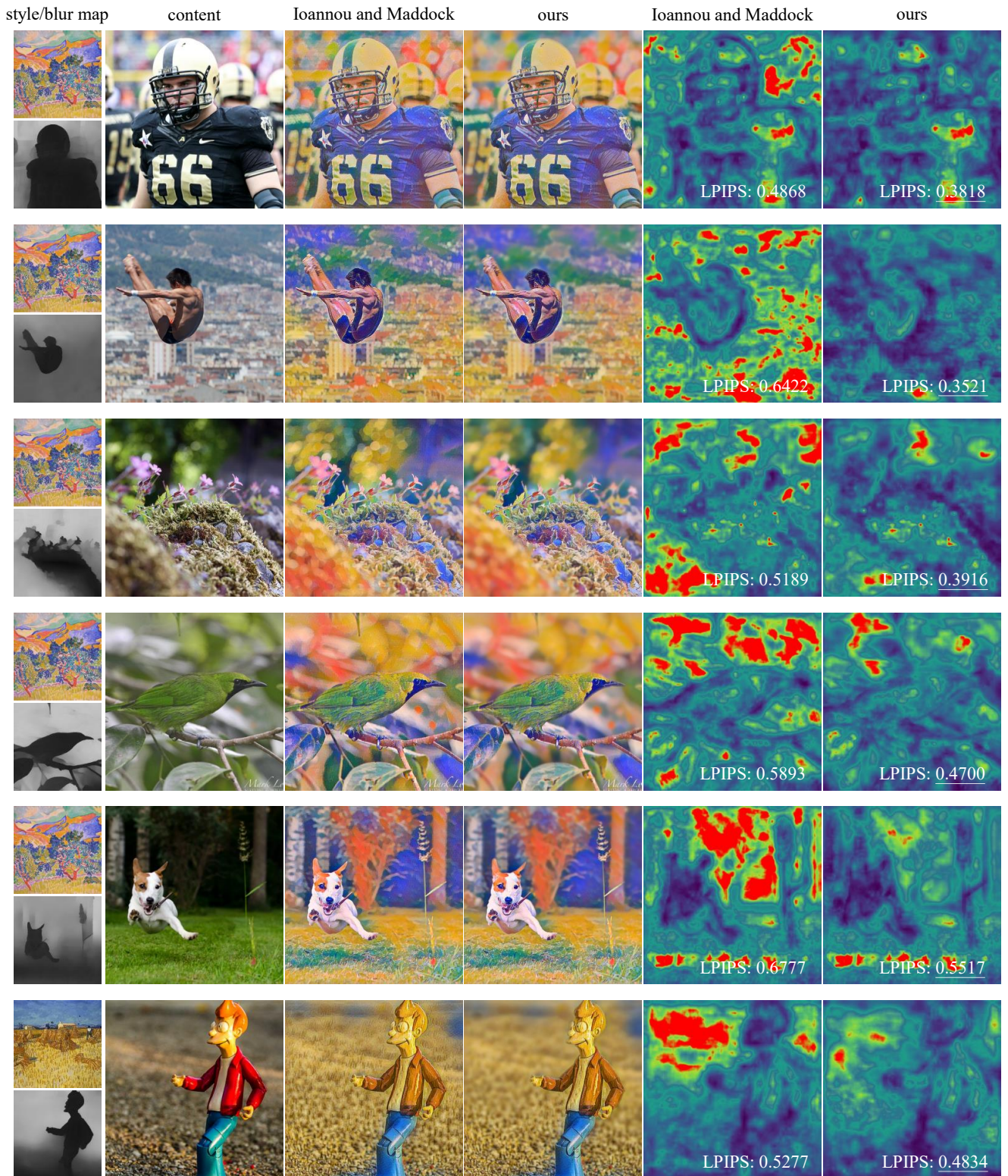
Figure 2: *Additional results using our method with the depth-aware style transfer approach [IM22]. The first column shows the style image (top) alongside the blur map (bottom), estimated from the content image in the second column. Columns three and four present the stylized output of Ioannou and Maddock [IM22] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.*
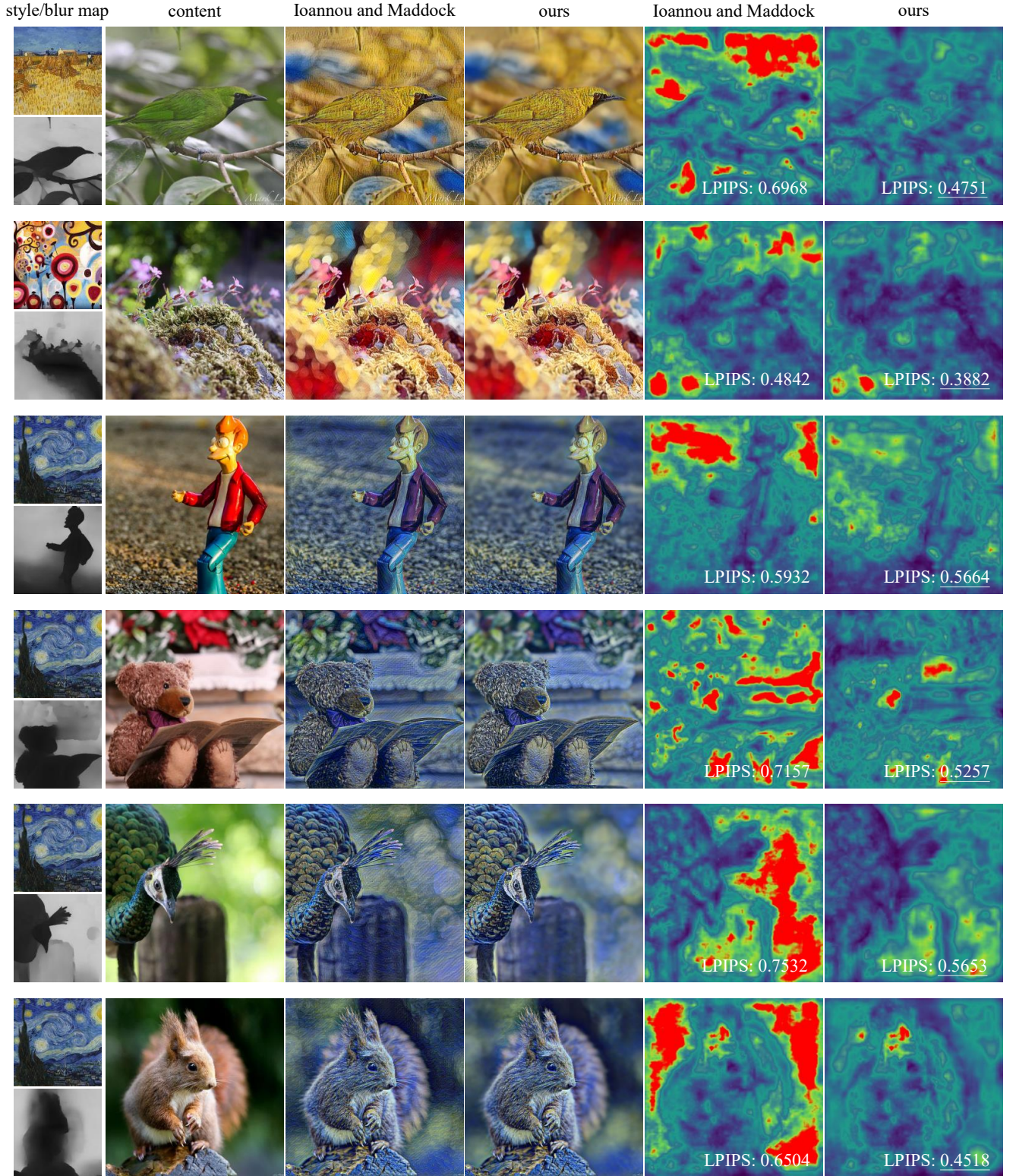
Figure 3: Additional results using our method with the depth-aware style transfer approach [IM22]. The first column shows the style image (top) alongside the blur map (bottom), estimated from the content image in the second column. Columns three and four present the stylized output of Ioannou and Maddock [IM22] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.
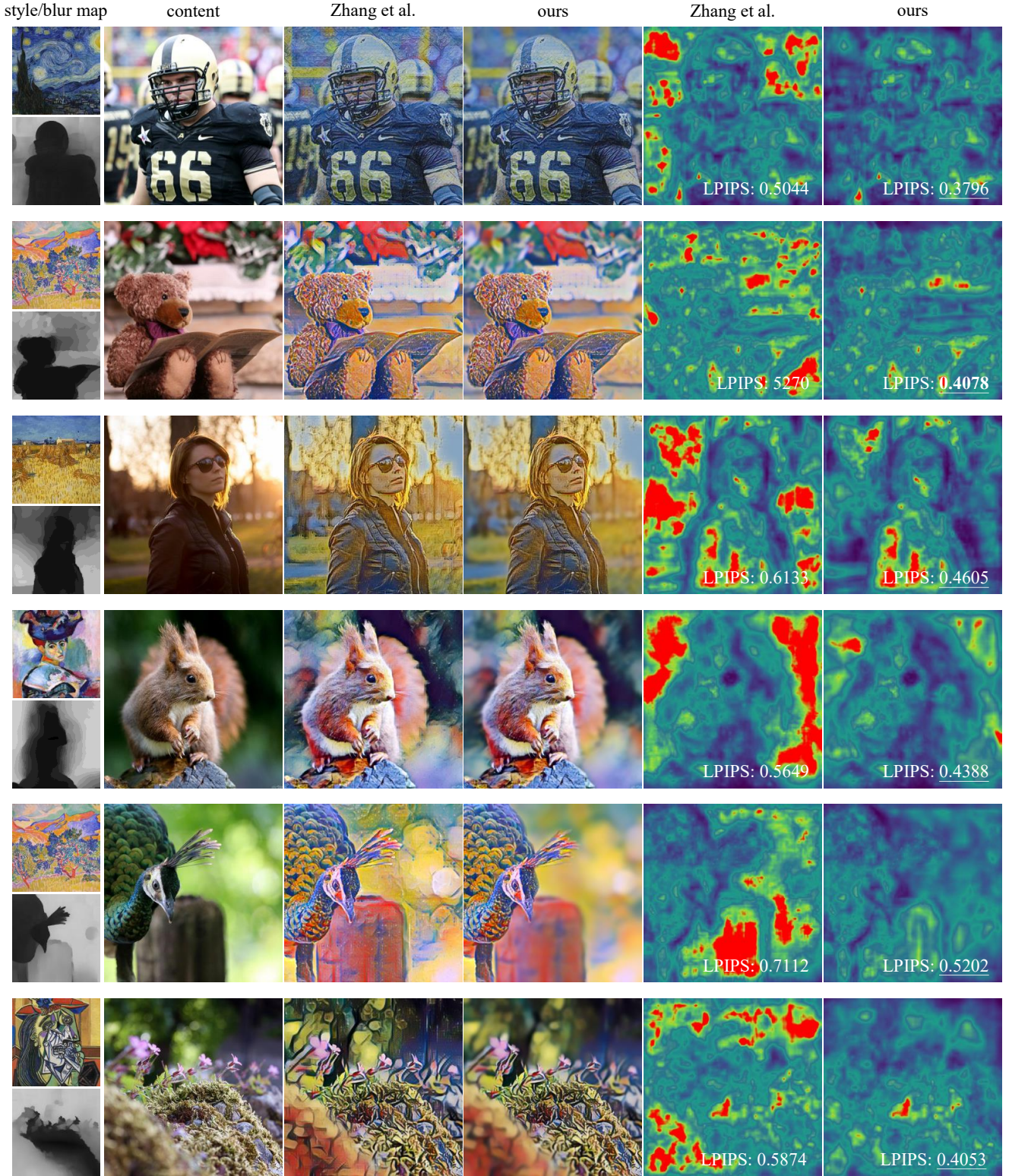
Figure 4: Additional results using our method with the attention-aware style transfer approach [ZXW*24]. The first column shows the style image (top) alongside the blur map (bottom), estimated from the content image in the second column. Columns three and four present the stylized output of Zhang *et al*. [ZXW*24] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.

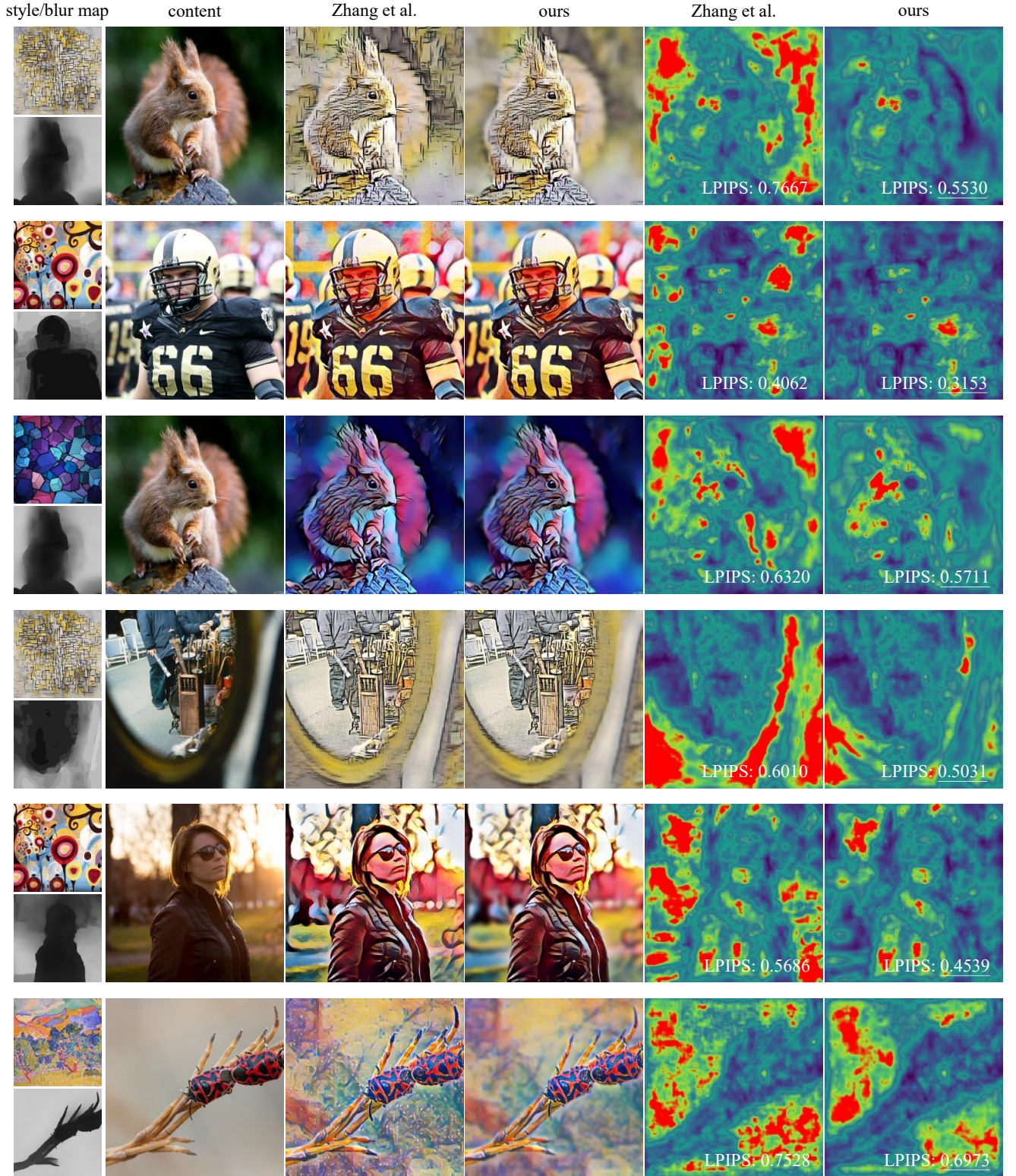| style/blur map	content	Zhang et al.	ours	Zhang et al.	ours

Figure 5: Additional results using our method with the attention-aware style transfer approach [ZXW*24]. The first column shows the style image (top) alongside the blur map (bottom), estimated from the content image in the second column. Columns three and four present the stylized output of Zhang *et al*. [ZXW*24] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.

H.-Y. Wang & Y.-T. Wu / Supplementary Material forPreserving Photographic Defocus in Stylised Image Synthesis
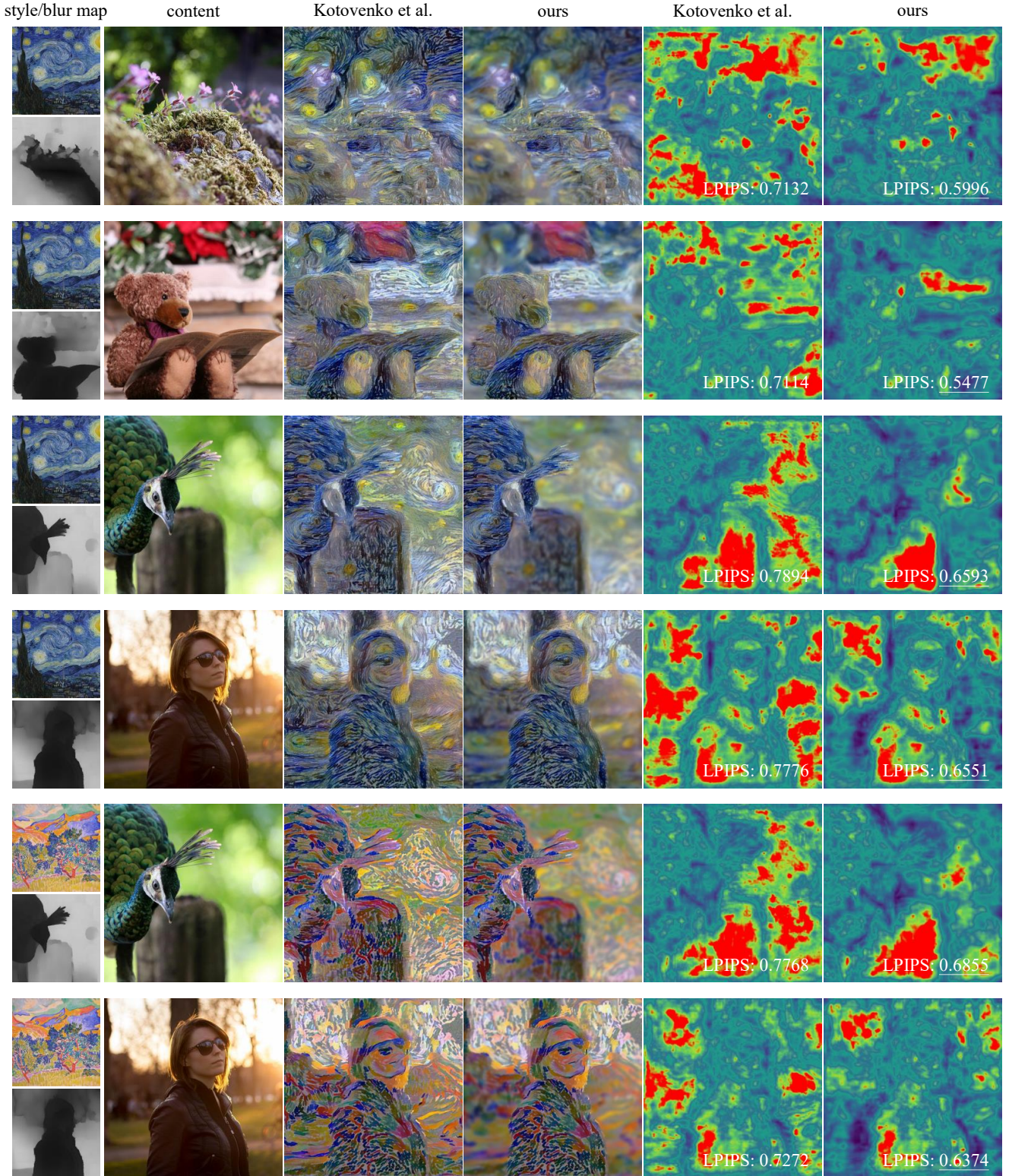

Figure 6: Additional results using our method with the brushstroke-based style transfer approach [KWHO21]. The first column shows the style image (top) alongside the blur map (bottom), estimated from the content image in the second column. Columns three and four present the stylized output of Kotovenko *et al*. [KWHO21] and our enhanced result, respectively. The final two columns display the corresponding LPIPS scores and their visualizations.