An Empty Room is All We Want: Automatic Defurnishing of Indoor Panoramas

Supplementary Material

Mira Slavcheva[★] Dave Gausebeck[★] Kevin Chen[★] David Buchhofer Azwad Sabik Chen Ma
Sachal Dhillon Olaf Brandt Alan Dolhasz

Matterport

Visual results

In Figure 11, we provide larger versions of the defurnishing results from the main paper. In addition, we include results from the methods we compare to in Section 4.1, *i.e.* LaMa [43], LGPN-Net [13] and SD-inpainting [36], with a smaller mask dilation of 10 pixels as well as with no mask dilation.

LaMa and LGPN-Net demonstrate similar trends; a non-dilated mask is absolutely insufficient—large dark patches are inpainted and the outlines of the inpainting mask are recognizable. Mask dilation remedies this effect, but the inpainted region tends to look like a blurry blob with increasing mask size. Conversely, the textures are sharpest with the non-dilated masks.

In all examples, SD-inpainting hallucinates objects when the inpainting mask is not dilated. With a mask dilated by 10 pixels, objects are still hallucinated in the first two examples. With a mask dilated by 20 pixels, only the first example shows very noticeable hallucinations, *e.g.* the tables on the right, but larger shadows cast by the objects that are being removed, *e.g.* the couch in the second example, remain in the output image and may even be extended by the inpainting.

Note that in the main paper, we chose the best-looking result for each of these three methods and each example separately.

Ours-inpaint builds upon SD-inpainting but tackles the hallucinations—and indeed none of the examples have hallucinations. Ours-full makes sure that original textures are preserved wherever possible, which is valuable because of the original image has more detail, *e.g.* the kitchen island in the first example, and because our method may sometimes remove more details than necessary since it is specifically trained to remove shadows outside of the inpainting mask. Intricate textures may still be an issue for our method, like the floor in the last image, where one plank is inpainted in a notably darker color.

In addition, in Figure 12 we demonstrate that the inpainting component of our method is not influenced by mask dilation. There is no noticeable difference in the inpainted result in all examples but the first one, where the far away kitchen island is removed as the mask gets larger, while it should remain because it is built-in. This example demonstrates the usefulness of non-dilated masks for keeping far-away details intact.

Finally, in Figure 13 we show an example of an unfurnished space with synthetic furniture used for quantitative evaluation in Section 4.1.

Prompt set

The set of 32 prompts that we used for training is as follows:

$$\{X \;\; Y \;\; Z\}, \text{ where}$$
$$X = \{u \; V\}, \text{ where } u \in \{\emptyset, \text{“an”}\}, V \in \{\text{“empty”}, \text{“unfurnished”}\},$$
$$Y \in \{\text{“room”}, \text{“space”}, \text{“home”}, \text{“house”}\},$$
$$Z \in \{\emptyset, P \text{ if } u \neq \emptyset, PQ \text{ if } u \neq \emptyset\}, \text{ where } P = \text{“. uniformly blank”}, Q = \text{“, straight edges”}.$$

Note that during training, each image is assigned one prompt at random. Due to this randomness, the prompt in subsequent epochs for the same image might be different.

[★] denotes equal contribution.

✉ research@matterport.com

Figure 11. **Additional defurnishing comparisons.** The non-dilated mask is overlaid in blue. Image best viewed digitally.

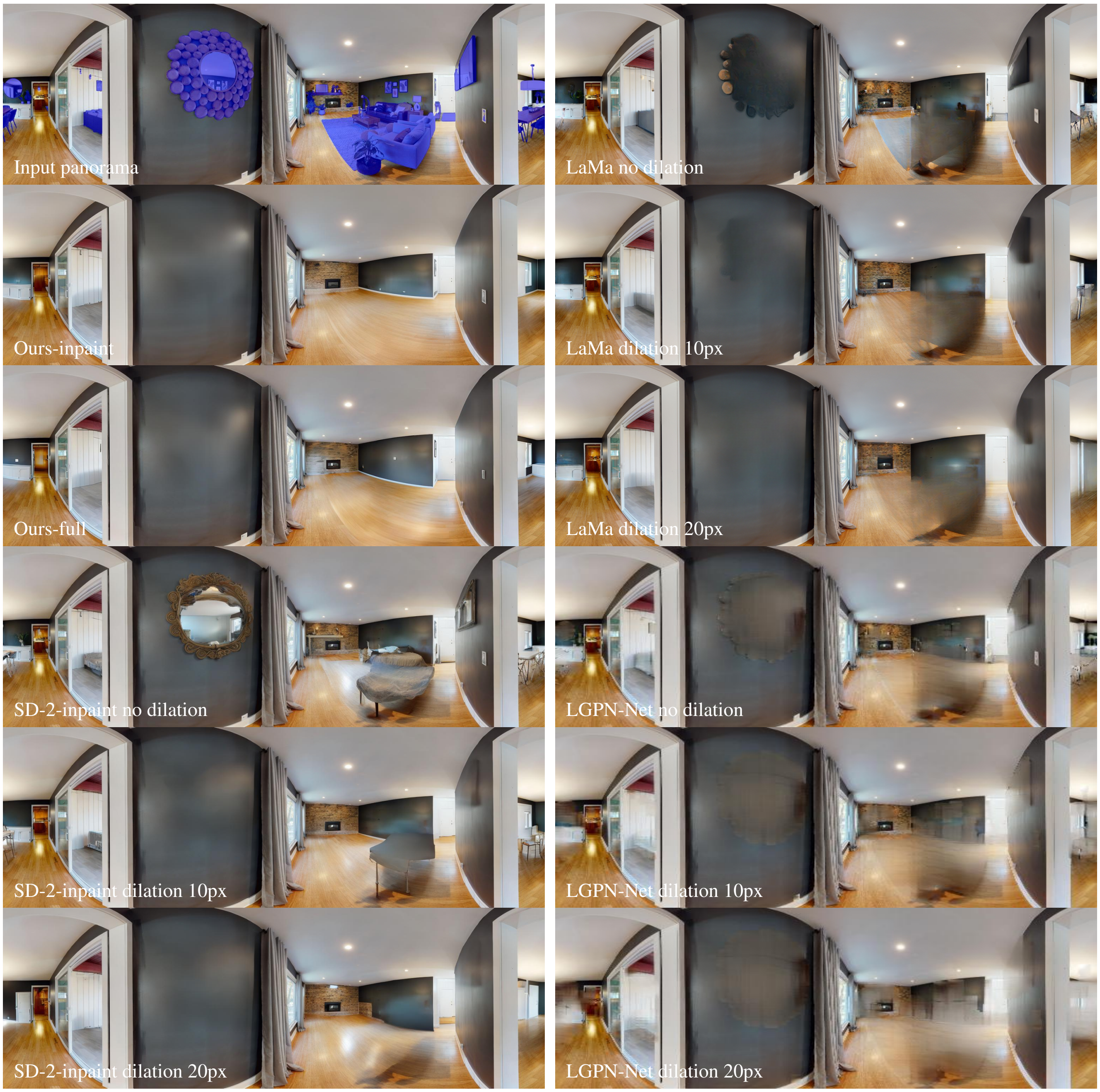

Figure 11. **Additional defurnishing comparisons.** The non-dilated mask is overlaid in blue. Image best viewed digitally.

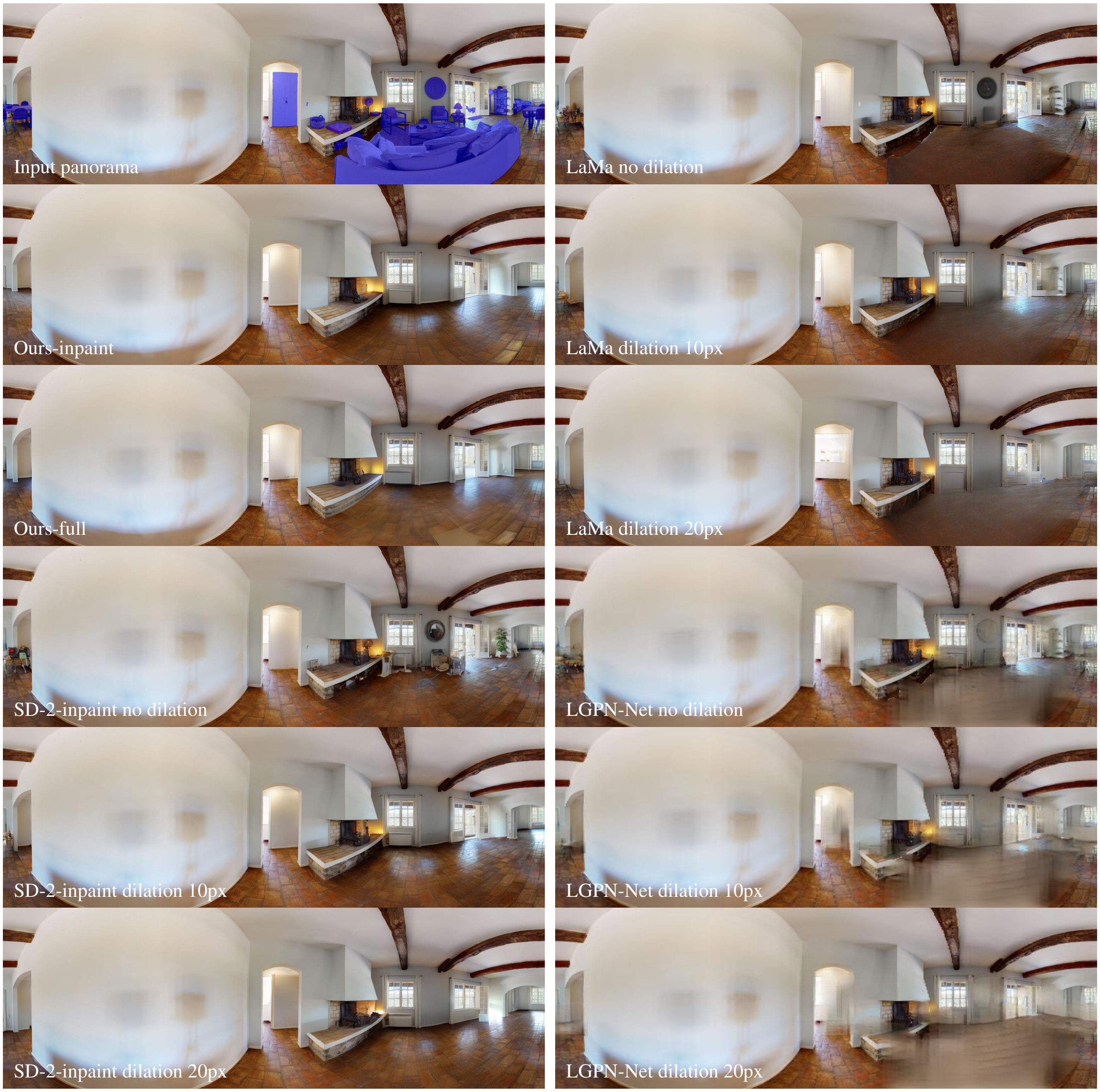

Figure 11. **Additional defurnishing comparisons.** The non-dilated mask is overlaid in blue. Image best viewed digitally.

Figure 11. **Additional defurnishing comparisons.** The non-dilated mask is overlaid in blue. Image best viewed digitally.

(a) No mask dilation (b) Dilation 10 pixels (c) Dilation 20 pixels

Figure 12. **Effect of mask dilation on our inpainting.** We show results from our inpainting component only. Our results are nearly identical regardless of the amount of mask dilation, apart from far away details like the kitchen island in the first example.

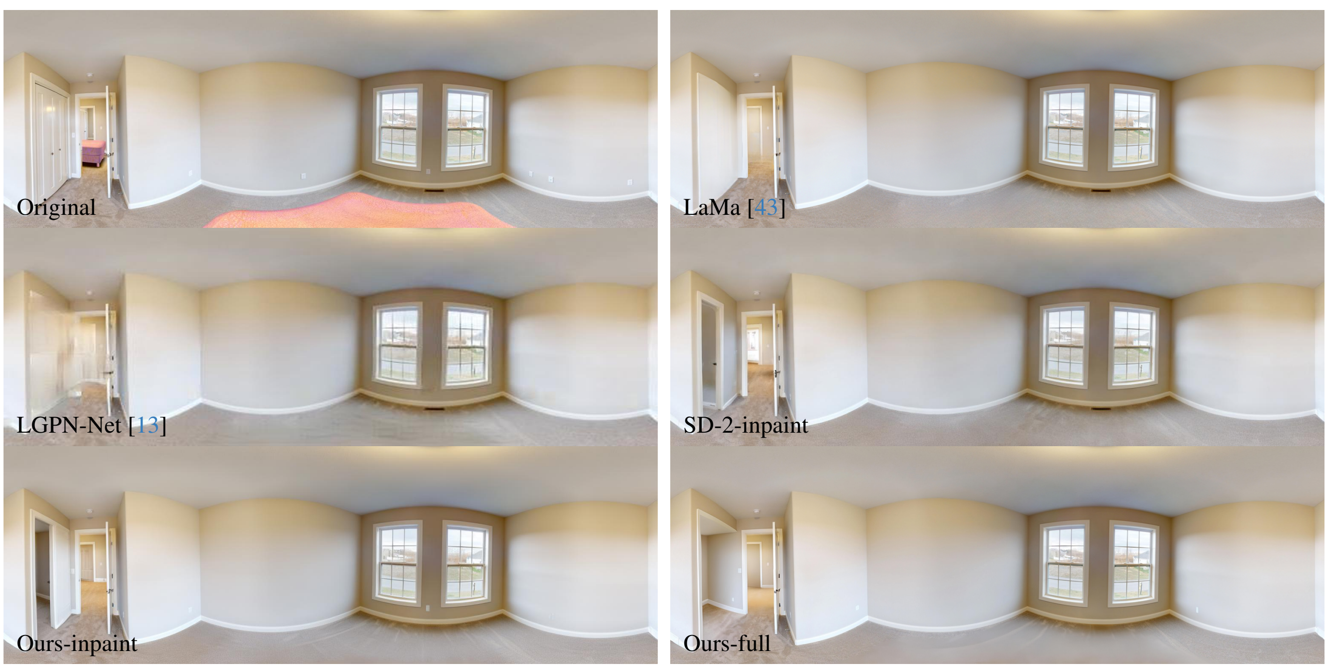

Figure 13. **Example of unfurnished space with synthetic furniture for quantitative evaluation.** The space is less complex than real furnished ones, letting LaMa and LGPN-Net produce higher frequency textures than SD and our inpainting, which is remedied by the blending in our full pipeline.