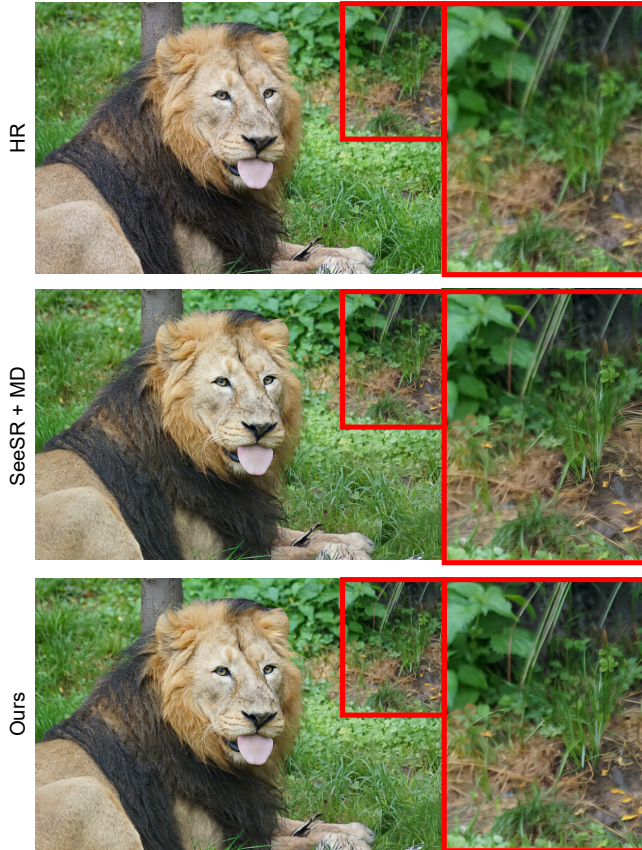


## Appendix

Here we report a selection of materials that provides further insight into our experiments, such as classical comparisons (see Fig. 7 and Fig. 9) and those used in our user study (see Fig. 8, Fig. 10, and Fig. 11).



**Fig. 7.** Qualitative comparison (809; DIV2K Val) between SeeSR+MD (PSNR $\uparrow$ :25.107, SSIM $\uparrow$ :0.598, LPIPS $\downarrow$ :0.076) and our approach (PSNR $\uparrow$ :**25.627**, SSIM $\uparrow$ :**0.621**, LPIPS $\downarrow$ :**0.069**). The global tags were “animal, break, floor, grass, green, lay, lion, lush, man, mane, mouth, relax, tree”. Similarly, SeeSR+MD hallucinates fur-like patterns in the brown dirt, leading to artifacts that degrade the visual quality and contribute to its inferior performance compared to our approach, which preserves the natural texture of the dirt more effectively. Once again, our method generates finer, sharper details that surpass the level found in the HR image.



**Fig. 8.** More qualitative comparison on 512 $\times$ 512 center-cropped and zoomed-in images that were used in our user study.



**Fig. 9.** Qualitative comparison of a 2K image (899; DIV2K Val) between LR, SeeSR+MD (PSNR $\uparrow$ :25.273, SSIM $\uparrow$ :0.769, LPIPS $\downarrow$ :0.130), our method (PSNR $\uparrow$ :**26.217**, SSIM $\uparrow$ :**0.794**, LPIPS $\downarrow$ :**0.103**), and HR. In general, we observe that our method reconstructs details in background objects better than SeeSR+MD (see light patterns in the lower left corner).



**Fig. 10.** More qualitative comparison on  $512 \times 512$  center-cropped and zoomed-in images that were used in our user study.



**Fig. 11.** More qualitative comparison on  $512 \times 512$  center-cropped and zoomed-in images that were used in our user study.