

## A. Implementation Details

**Training** We set the global batch size to 32. If the loss is calculated on data  $x$  directly, we run the training for 0.1M iterations. Otherwise, 1M iterations will be used following Eq. 6. We use the Adam optimizer and apply a constant learning rate  $1e^{-4}$ . Most existing works train DDM to predict the noise in the input. In contrast, our model predict the output, *i.e.*  $x_0$ , directly in the first intrinsic iteration. Empirical results show that this leads to a much faster convergence ( $\approx 90\%$  faster) without affecting the final output quality compared with predicting noise.

**Model** The model architecture follows DDPM<sup>5</sup>. Conditional signals are concatenated to the U-Net and attentions are added to 8 - 32 resolutions.

**Loss** In Eq. 6, by choosing  $p = 1, 2$ , we have some interesting observations.  $p = 1$  demonstrates much better stability in terms of color faithfulness. In contrast,  $p = 2$  yields obviously bad results, as shown in Figure 10. For each input, we randomly sample four outputs. It could be due to optimizer settings and data fitting issues. Currently, we don't have a convincing explanation and we leave it for the future study.

## B. More examples

We provide more restored examples on both original CelebA-HQ (Figure 11) and synthetically degraded CelebA-HQ test datasets (Figure 12), and also two real-world datasets WebPhoto (407 internet images) and LFW datasets (1711 images) [43], as shown in Figure 14 and Figure 13. Consistently, the proposed method can generate much more realistic, natural and faithful faces.

---

<sup>5</sup><https://github.com/hojonathanho/diffusion>

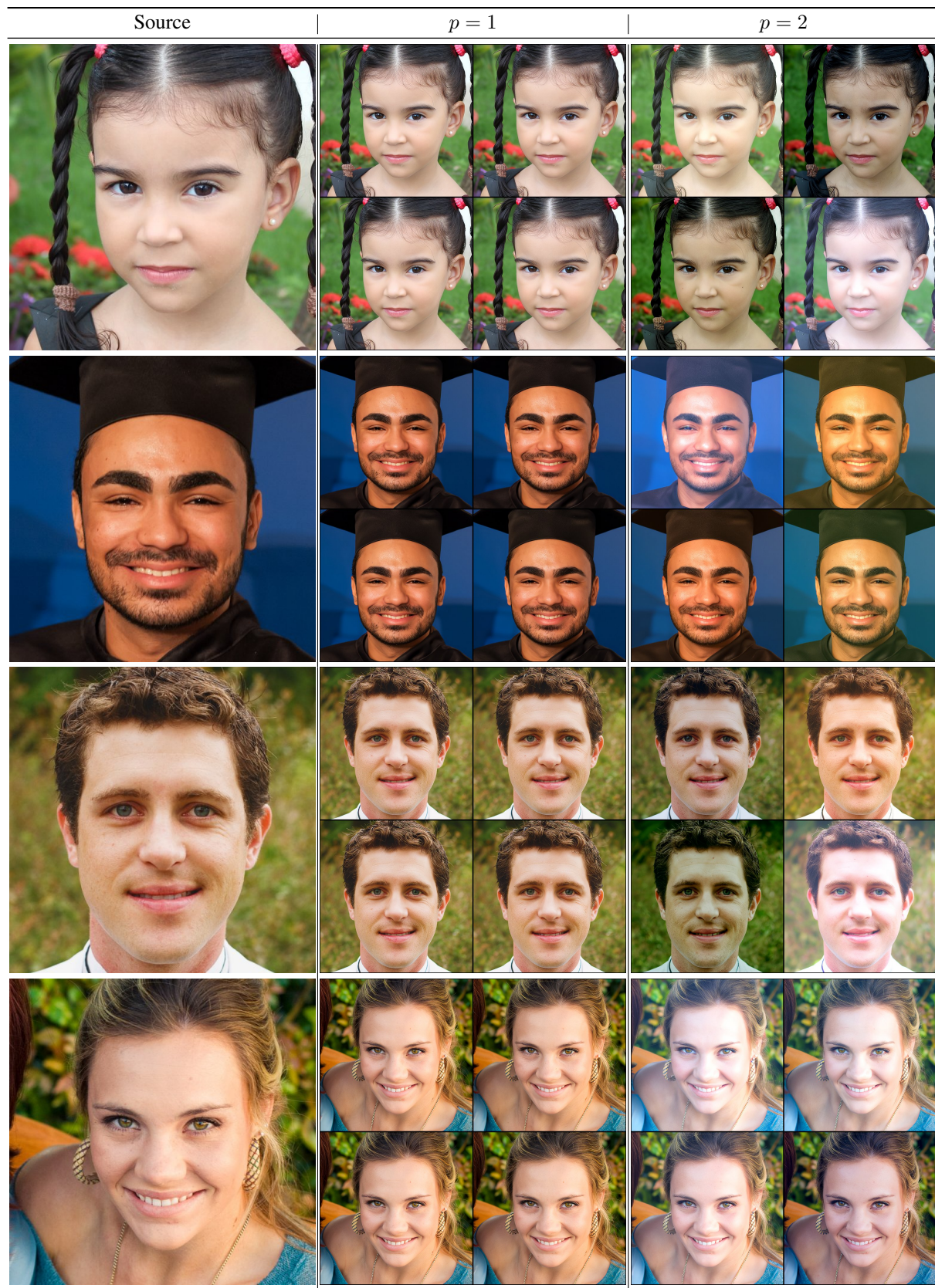


Figure 10: On the impact of  $L_1$  and  $L_2$ . Examples come from the FFHQ dataset.



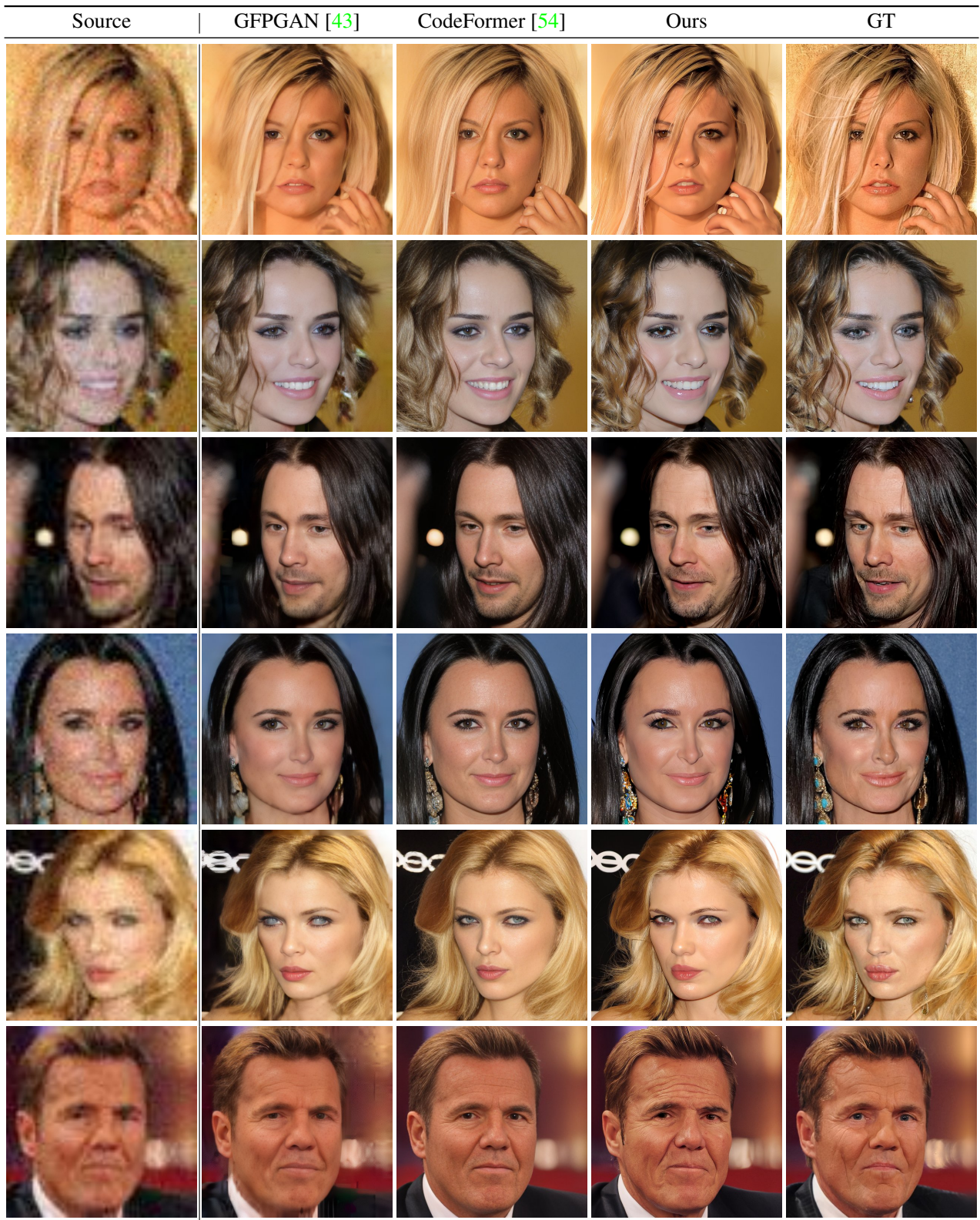


Figure 11: Qualitative comparison with state-of-the-art restoration models on synthetically degraded CelebA-HQ Test.



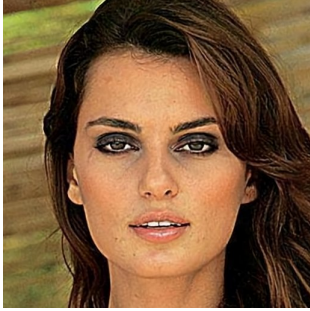

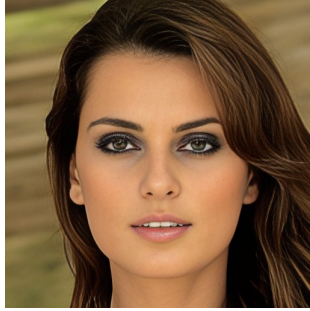
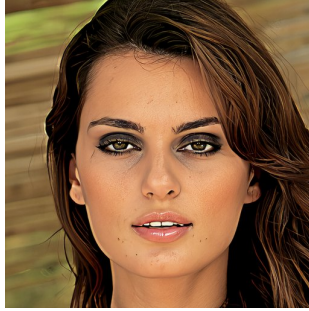

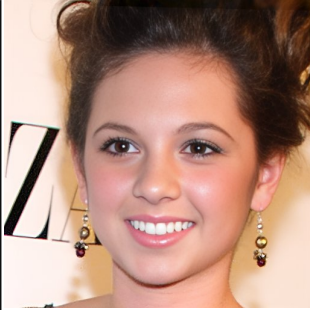


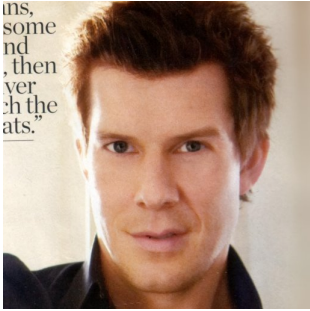
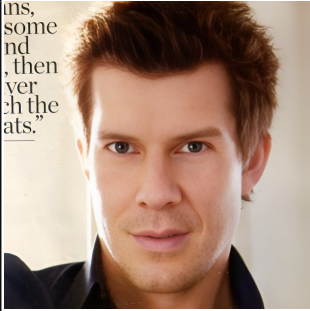
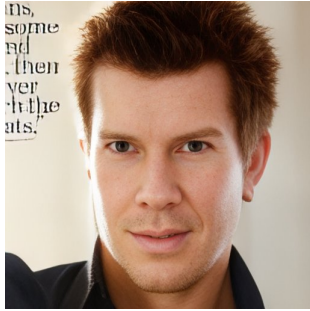
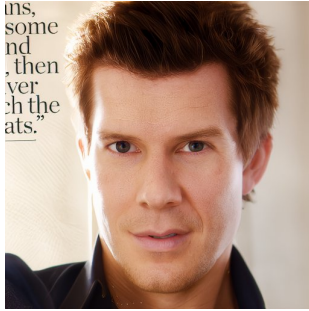




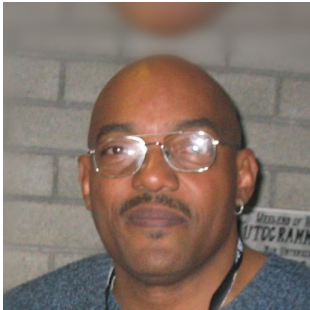
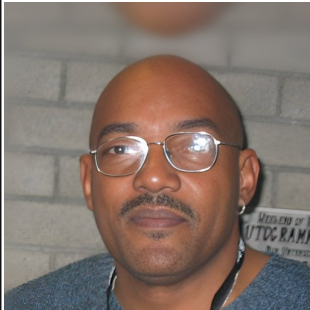

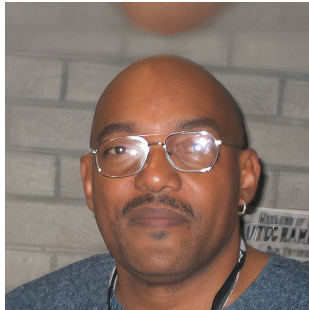
Source	GFPGAN [43]	CodeFormer [54]	Ours
			
			
 <p>ins, some nd , then ver ch the ats."</p>	 <p>ins, some nd , then ver ch the ats."</p>	 <p>ins, some nd , then ver ch the ats."</p>	 <p>ins, some nd , then ver ch the ats."</p>
			
			

Figure 12: Qualitative comparison with state-of-the-art restoration models on original CelebA-HQ Test.





Figure 13: Qualitative comparison with state-of-the-art restoration models on LFW test set.









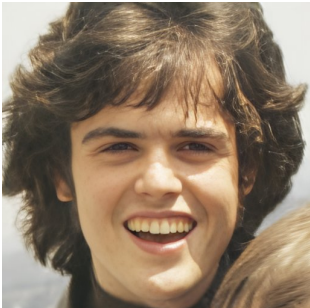

Source	GFPGAN [43]	CodeFormer [54]	Ours
			
			
			
			
			

Figure 14: Qualitative comparison with state-of-the-art restoration models on WebPhoto test set.