# SUPPLEMENTARY: DIFFUSION-BASED COMPRESSION QUALITY TRADEOFFS WITHOUT RETRAINING

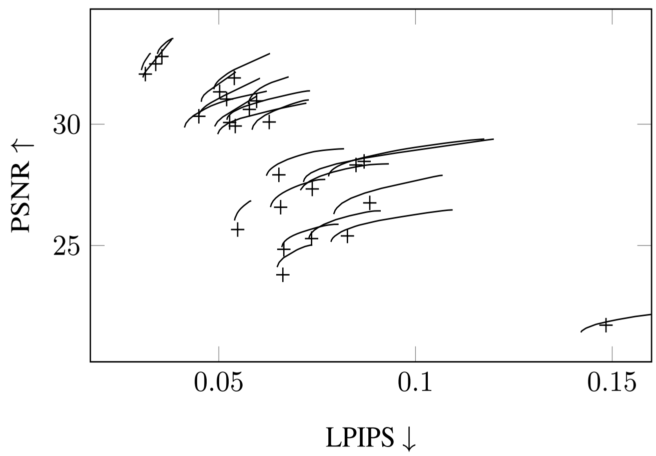



**Fig. A.1**: **Optimal configurations for CDC $x_0$ [1]** according to $\mathcal{T}$ with $\alpha \in [0, 1]$ and the corresponding default configuration for the 24 Kodak images. With the default configuration marked, and the best possible results shown as a black line.

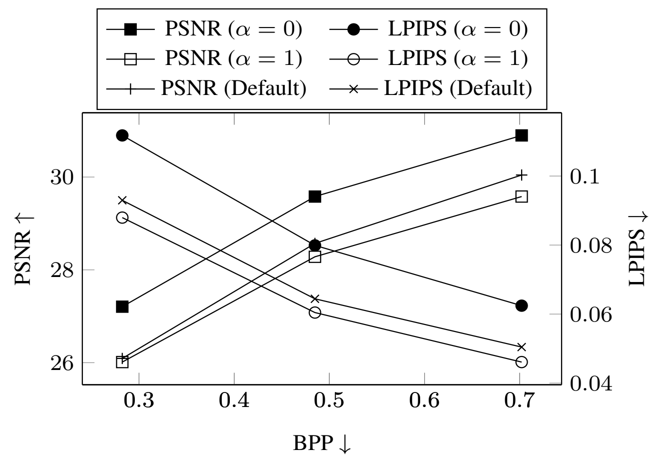



**Fig. A.2**: **Extreme tradeoffs for CDC $x_0$ [1]** with $\alpha = 0$ and $\alpha = 1$ and the default configuration. For the optimized configurations we employed Bayesian Optimization using Gaussian Processes and 30 iterations. The proposed method allows to select a tradeoff between PSNR and LPIPS performance without any retraining. However, when optimizing only for a single metric (PSNR or LPIPS), performance in the other metric drops.

## A. POTENTIAL QUALITY IMPROVEMENTS ON KODAK

In addition to fig. 3 of the main paper, which only shows the first Kodak image, we provide results for all of the 24 Kodak images in fig. A.1, which shows the best configurations according to $\mathcal{T}$ for any $\alpha$, as determined by the grid search, and the corresponding default configuration of the CDC $x_0$ model [1].

As discussed in section 4.2 of the main paper, we find that it is possible to achieve performance improvements for any of the tested images. The achievable improvement strongly depends on the specific image and targeted metric. In many cases, the default configuration of CDC $x_0$ [1] already scores quite well on LPIPS, and only improvements in PSNR are possible.

In fig. A.2 we compare the extreme tradeoffs with $\alpha = 0$ and $\alpha = 1$ to the default configuration of CDC $x_0$ [1] for the Kodak dataset for various bitrates. As expected [2], optimizing for either PSNR ($\alpha = 0$) or LPIPS ($\alpha = 1$) results in a drop in the other metric compared to the default configuration.

### A.1. Comparison regarding the optimization criterion $\mathcal{T}$

In fig. A.3 we show the performance of different blackbox optimization techniques in regards to $\mathcal{T}$ compared against the number of evaluations. As mentioned in the main paper, Bayesian Optimization using Gaussian Processes performs slightly better than the other tested approaches.

## B. RESULTS ON THE CLIC2022 DATASET

In this section we provide additional results for the **CLIC2022** [4] dataset, which contains 30 images with the larger side being 2048 pixels. As shown in fig. B.4 we observe similar performance improvements for this dataset as compared to the other datasets. With 50 steps of Bayesian Optimization using Gaussian Processes we achieve an increase in PSNR of 0.138dB and reduction in LPIPS of $-0.003$.

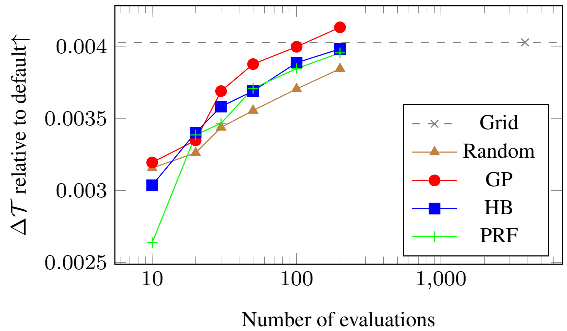



**Fig. A.3**: **Differences for various optimization methods to the CDC $x_0$ default [1]** configuration in regards to the optimization criterion $\mathcal{T}$ vs. number of iterations on Kodak [3].

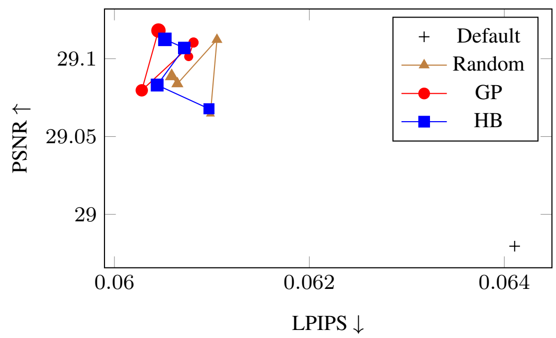



**Fig. B.4**: **CDC $x_0$ default [1] vs. CDC optimized** with different methods with 10, 20, 30, or 50 iterations in terms of average PSNR and average LPIPS on CLIC2022 [4].

## C. LIMITING THE NUMBER OF DDIM STEPS

The number of DDIM steps used during decompression is a parameter that will influence the decoding speed of the learned image compression method. In many cases, it might be beneficial to perceptual quality to increase the number of DDIM steps. As a consequence, optimizing the parameter configuration for LPIPS will often result in a higher number of DDIM steps than the default (given the parameter bounds selected in table 1). For example, as shown for the first Kodak image in fig. 3, using significantly fewer DDIM steps than the default (17) harms performance in terms of LPIPS. Furthermore, most of the best configurations use more than the default 17 steps. However, increasing the number of steps also slows down the decoding process, which can be undesirable when decoding speed is a critical factor.

Therefore, in fig. C.5 we compare our results with an optimization run, which was limited to a maximum of 17 DDIM steps (using 10/20/30 optimization iterations). Limiting the maximum number of decoding steps leads to worse quality in terms of PSNR and LPIPS for the tested CDC model, compared to the results achieved using the parameter space detailed in table 1. However, even when constraining the number of DDIM steps, the optimization process still leads to an increase in performance compared to the default configuration. Specifically, the choice of our optimization criterion $\mathcal{T}$ leads to slightly worse LPIPS, but an increase in PSNR.

For future work, it could also be worthwhile to include the number of DDIM steps in the optimization criterion, balancing possible performance improvements against the increase in decoding time.

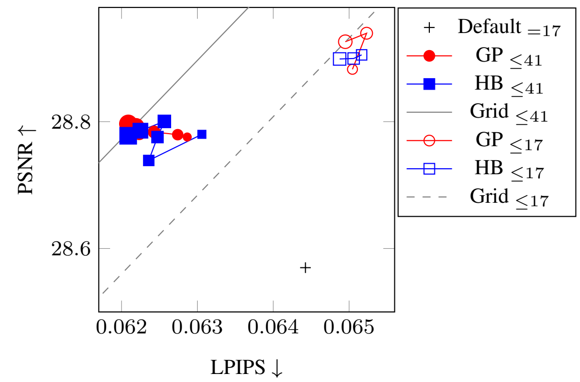



**Fig. C.5**: **CDC $x_0$ default [1] vs. our CDC configurations** optimized using Gaussian Processes or Hyperbands in terms of average PSNR and average LPIPS on Kodak. Comparing the normal optimization settings to limiting the number of DDIM sampling steps to 17.

## D. DATASET OPTIMIZED PARAMETERS

It is possible to outperform the default configuration of CDC by optimizing the sampling parameters for the entire test dataset. This configuration achieves better LPIPS compared to the default configuration ($-0.02$), with a similar PSNR ($+0.03$dB). However, it still falls short of the performance achieved by our proposed per-image optimization approach.

## E. ADDITIONAL VISUAL COMPARISONS

To expand on the visual comparisons of the main paper, we provide a larger and more detailed version of the figure in fig. E.6.

In fig. E.7 we provide a visual overview of how the tradeoff parameter $\alpha$ affects the generated image for $0.9 \leq \alpha \leq 1.0$. Values $\alpha < 0.9$ lead to the same configuration during the optimization for the selected Kodak image and CDC checkpoint and will therefore generate the same image.

We provide additional visual comparisons for images generated with PerCo [5] (fig. E.8), as well as the lowest bitrate checkpoint of CDC $x_0$ [1] (figs. E.9 to E.14). We show a selection of cropped reconstructions from the Kodak [3] dataset optimized for different tradeoffs ($\alpha$).

As mentioned in the original PerCo paper, PSNR is not a suitable metric for very low bitrate scenarios, where realism is most important. Therefore, as expected, optimizing for PSNR does not yield visually pleasing results [5]. The images with $\alpha = 0$ are usually very washed out and lack contrast. Optimizing for LPIPS is much more reasonable in this case. However, other optimization targets might prove to be better suitable for this model.

For CDC, optimizing for PSNR with $\alpha = 0$ generally results in very smooth images, often lacking some details. When increasing $\alpha \to 1$ the model will generate sharper and more detailed images. However, this can sometimes lead to slightly noisy images.

## F. REFERENCES


[1] R. Yang and S. Mandt, "Lossy image compression with conditional diffusion models," *NeurIPS*, vol. 36, 2024.

[2] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *Proceedings of the IEEE CVPR*, 2018, pp. 6228–6237.

[3] E. K. Company, "Kodak lossless true color image suite," https://r0k.us/graphics/kodak/.

[4] CLIC, "Workshop and challenge on learned image compression (clic)," http://clic.compression.cc/2022/, 2022.

[5] M. Careil, M. J. Muckley, J. Verbeek, and S. Lathuilière, "Towards image compression with perfect realism at ultra-low bitrates," in *ICLR*, 2023.

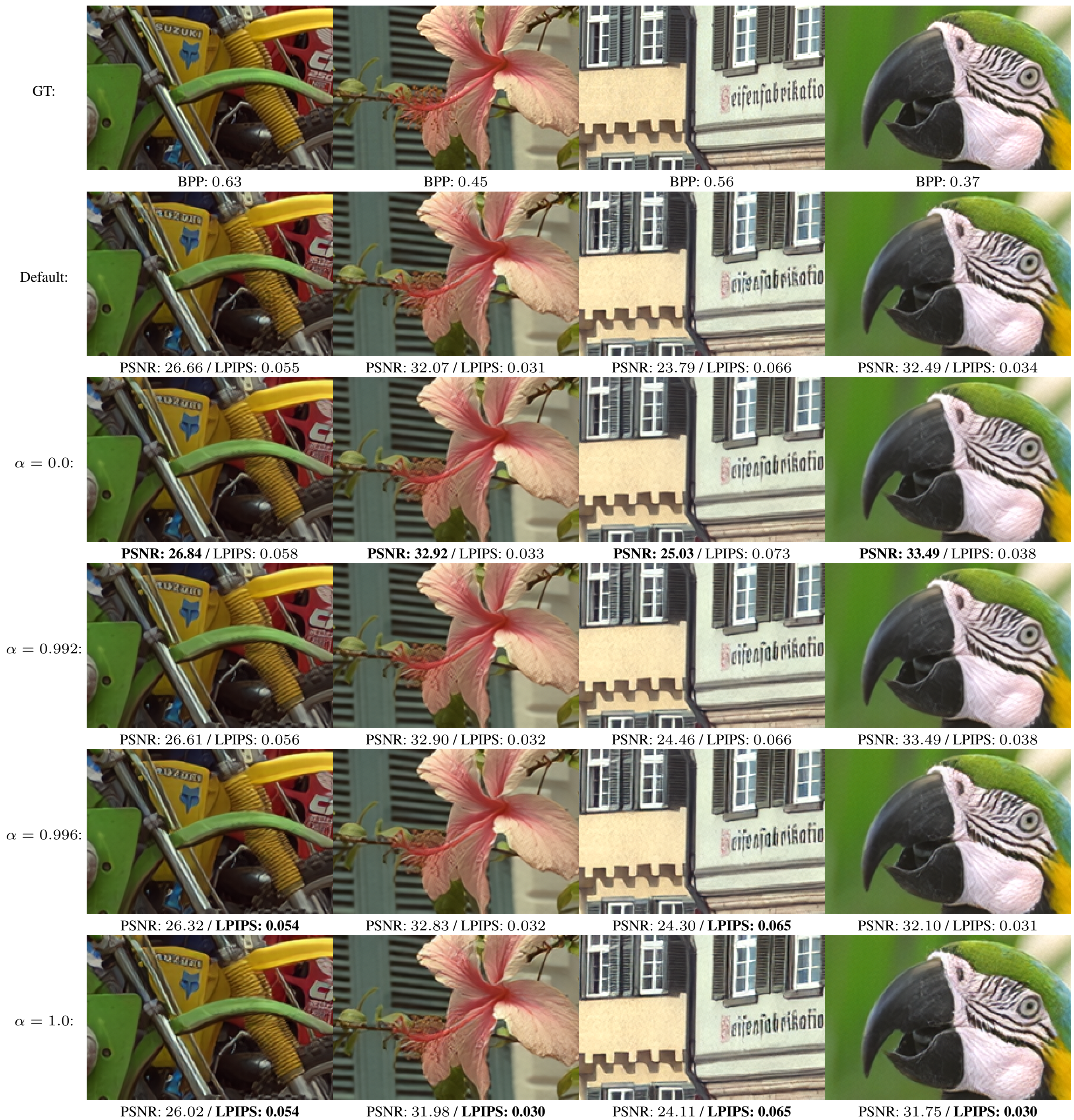



**Fig. E.6**: **Qualitative comparison of different tradeoffs between LPIPS and PSNR as generated by CDC $x_0$ [1]** for different crops of Kodak images. Images optimized for PSNR tend to be more blurry while optimizing for LPIPS can result in more noisy images.

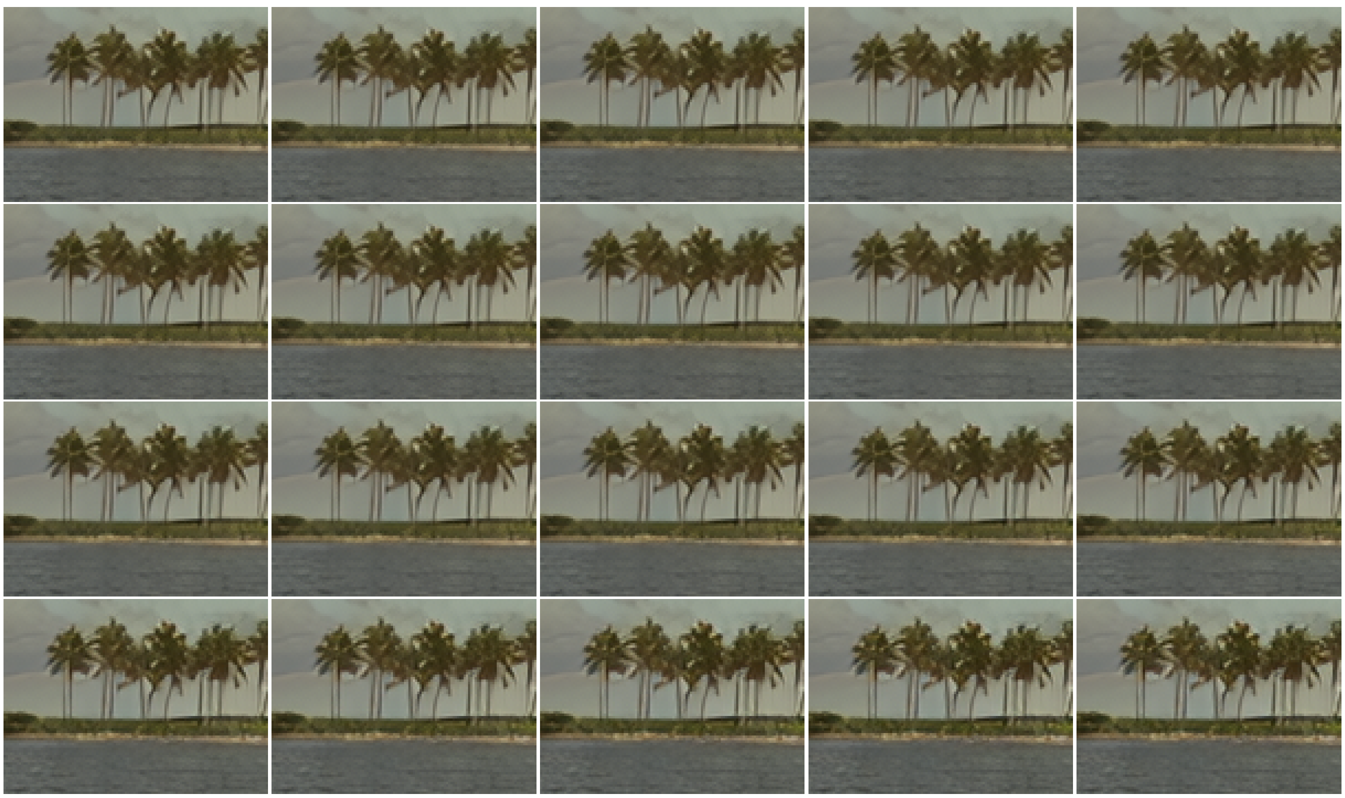
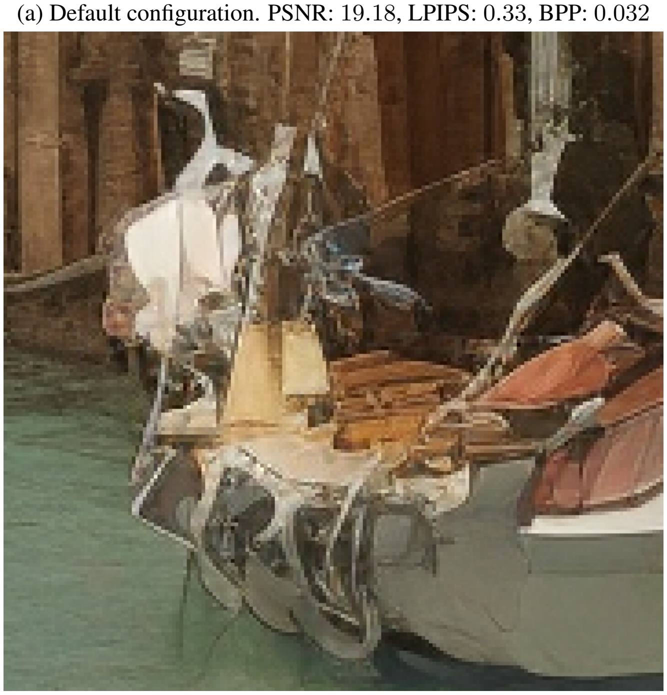


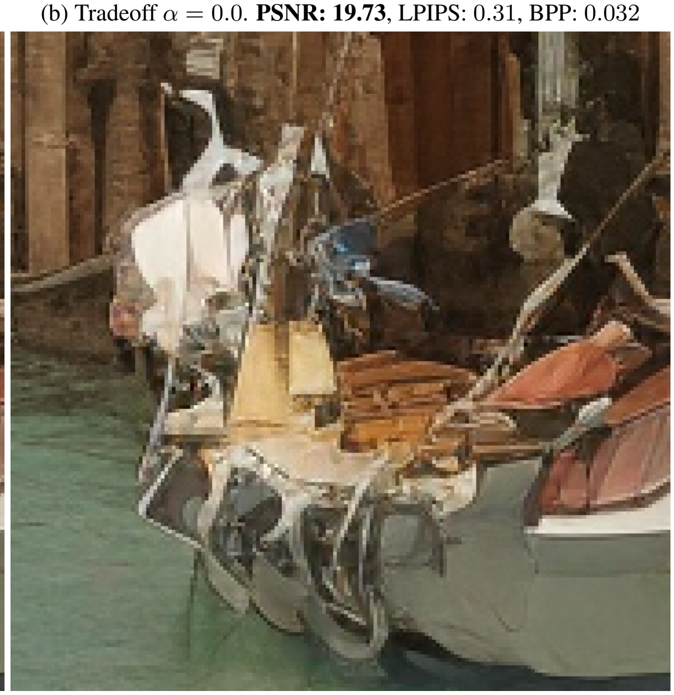
**Fig. E.7**: **Qualitative comparison of tradeoffs uniformly distributed from $\alpha = 0.9$ (top-left) to $\alpha = 1.0$ (bottom-right) as generated by CDC $x_0$ [1]**. As is the case for most Kodak images for CDC $x_0$ [1], optimizing with $\alpha < 0.9$ results in the same configuration as when optimizing for $\alpha = 1.0$. The tradeoff parameter $\alpha$ is most sensitive near $1.0$.

(a) Default configuration. PSNR: $19.18$, LPIPS: $0.33$, BPP: $0.032$

(b) Tradeoff $\alpha = 0.0$. **PSNR: 19.73**, LPIPS: $0.31$, BPP: $0.032$



(c) Tradeoff $\alpha = 0.98$. PSNR: $19.46$, **LPIPS: 0.30**, BPP: $0.032$



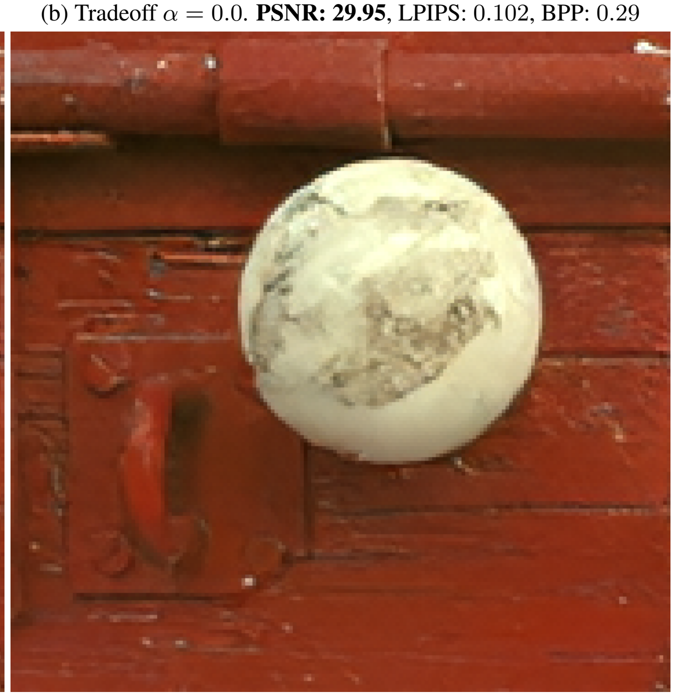
(d) Tradeoff $\alpha = 1.0$. PSNR: $19.19$, **LPIPS: 0.30**, BPP: $0.032$

**Fig. E.8**: **Qualitative comparison of different tradeoffs for PerCo [5].** Optimizing this method for PSNR (b) leads to images with very low contrast. When optimizing for LPIPS (c, d) the generated images increase in sharpness and contrast.

(a) Default configuration. PSNR: 28.91, LPIPS: 0.094, BPP: 0.29

(b) Tradeoff $\alpha = 0.0$. **PSNR: 29.95**, LPIPS: 0.102, BPP: 0.29

(c) Tradeoff $\alpha = 0.985$. PSNR: 29.59, LPIPS: 0.094, BPP: 0.29



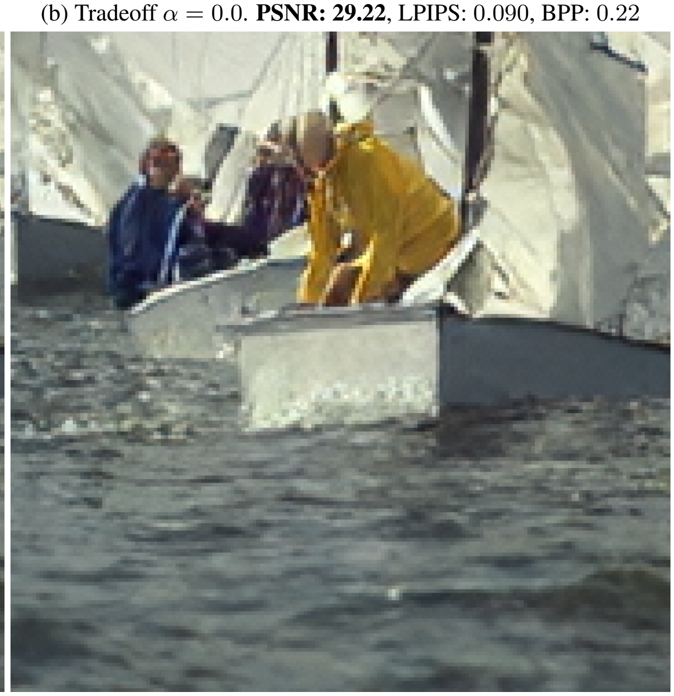
(d) Tradeoff $\alpha = 1$. PSNR: 28.91, **LPIPS: 0.089**, BPP: 0.29

**Fig. E.9**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated images tend to be more blurry. However, the generated image improves over the default configuration (a) by 1dB in PSNR. With increasing $\alpha$ (c, d), the sharpness of the image increases and texture details, such as the structure in the wood, become clearer.

(a) Default configuration. PSNR: 28.47, LPIPS: 0.080, BPP: 0.22

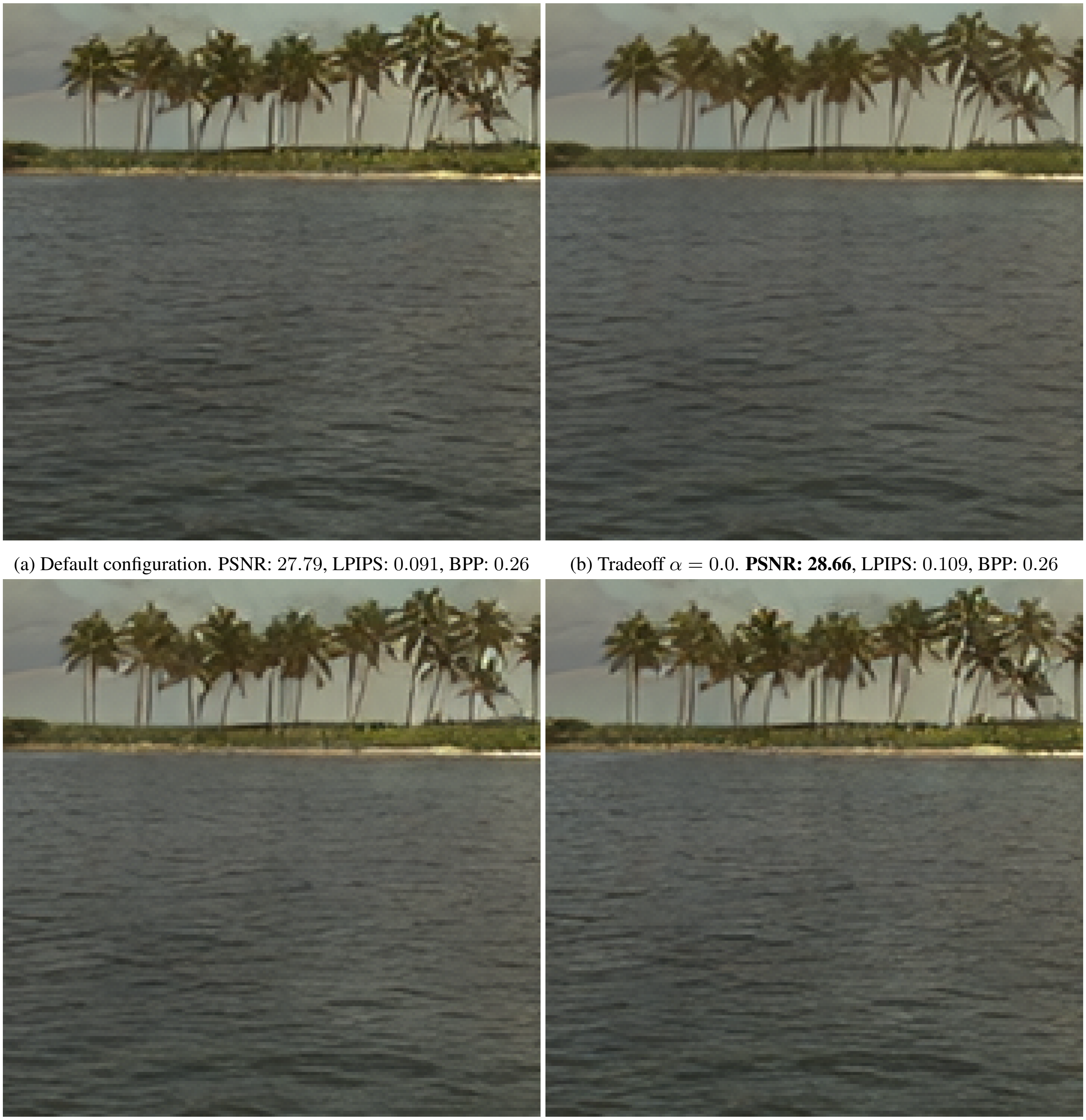
(b) Tradeoff $\alpha = 0.0$. **PSNR: 29.22**, LPIPS: 0.090, BPP: 0.22

(c) Tradeoff $\alpha = 0.985$. PSNR: 28.14, LPIPS: 0.070, BPP: 0.22



(d) Tradeoff $\alpha = 1$. PSNR: 27.92, **LPIPS: 0.068**, BPP: 0.22

**Fig. E.10**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated image is lacking detail in areas such as the water. However, the reconstruction improves over the default configuration (a) by almost 0.8 dB in PSNR. With increasing $\alpha$ (c, d) sharpness increases and the water has a much more realistic look. However, especially (d) introduces some noise artifacts, for example in the faces.



(a) Default configuration. PSNR: 27.79, LPIPS: 0.091, BPP: 0.26

(b) Tradeoff $\alpha = 0.0$. **PSNR: 28.66**, LPIPS: 0.109, BPP: 0.26

(c) Tradeoff $\alpha = 0.98$. PSNR: 28.08, LPIPS: 0.087, BPP: 0.26

(d) Tradeoff $\alpha = 1$. PSNR: 27.69, **LPIPS: 0.083**, BPP: 0.26

**Fig. E.11**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated image is lacking detail in areas such as the water and vegetation. However, the reconstruction improves over the default configuration (a) by about 0.8 dB in PSNR. With increasing $\alpha$ (c, d), the sharpness of the image increases and the water and vegetation have a much more realistic look.

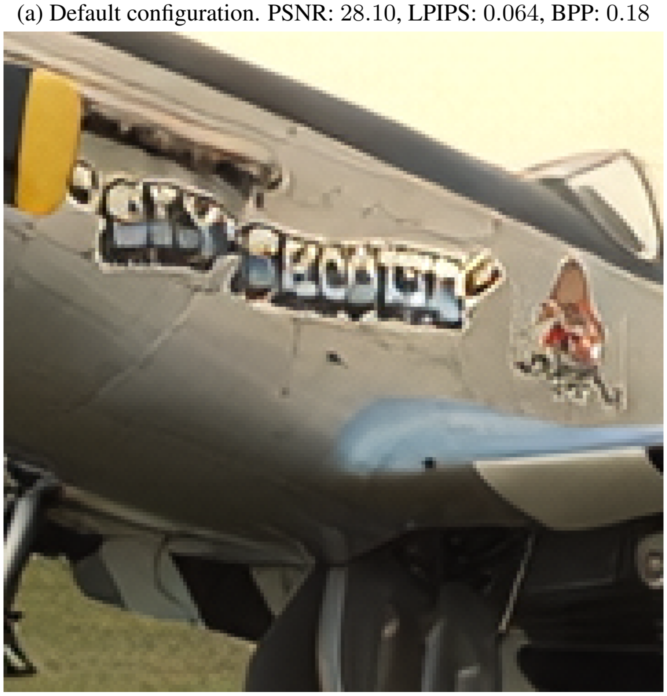
(a) Default configuration. PSNR: 25.50, LPIPS: 0.101, BPP: 0.27 (b) Tradeoff $\alpha = 0.0$. **PSNR: 26.64**, LPIPS: 0.139, BPP: 0.27

(c) Tradeoff $\alpha = 0.98$. PSNR: 25.63, LPIPS: 0.095, BPP: 0.27 (d) Tradeoff $\alpha = 1$. PSNR: 25.43, **LPIPS: 0.093**, BPP: 0.27

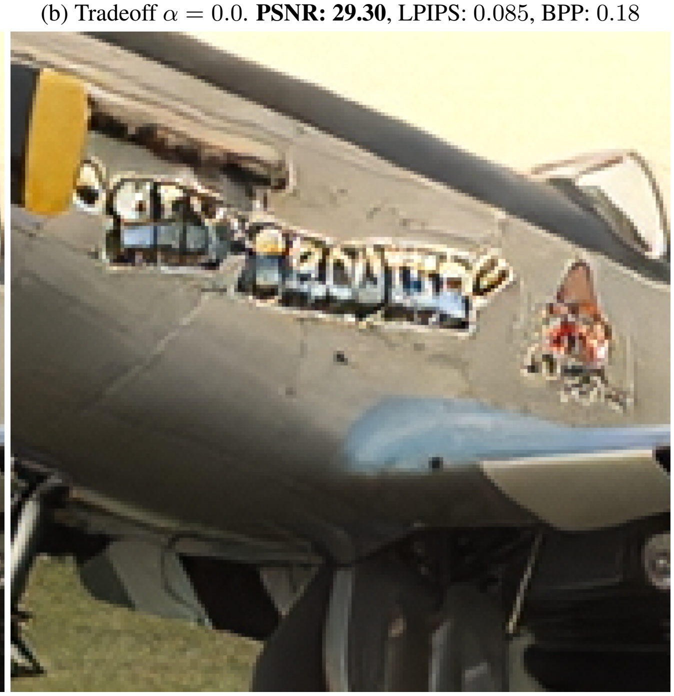
**Fig. E.12**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated image is very smooth, and especially the vegetation is missing detail. However, the reconstruction improves over the default configuration (a) by about 1.1 dB in PSNR. With increasing $\alpha$ (c, d), the sharpness of the image increases and the vegetation has a much more realistic look. Additionally, some smaller details, such as the shadows on the ring, are more visible.

(a) Default configuration. PSNR: 28.10, LPIPS: 0.064, BPP: 0.18

(b) Tradeoff $\alpha = 0.0$. **PSNR: 29.30**, LPIPS: 0.085, BPP: 0.18



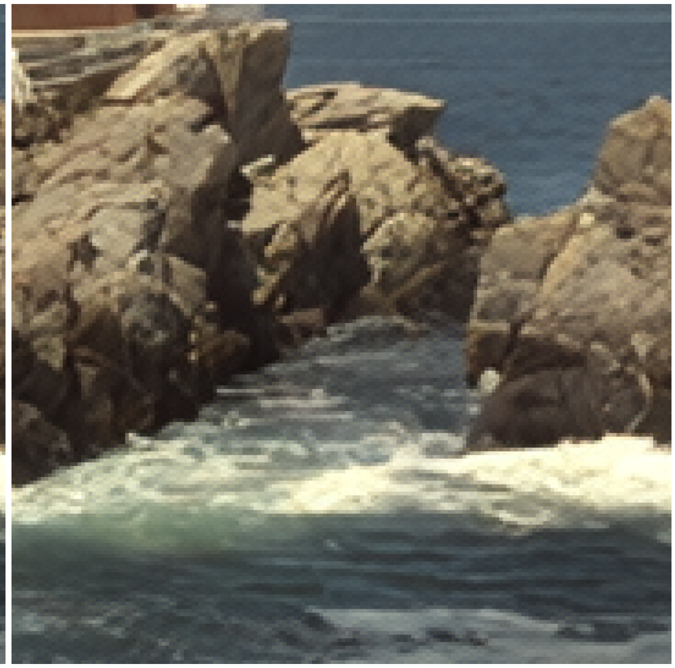
(c) Tradeoff $\alpha = 0.98$. PSNR: 28.67, LPIPS: 0.070, BPP: 0.18



(d) Tradeoff $\alpha = 1$. PSNR: 27.73, **LPIPS: 0.061**, BPP: 0.18

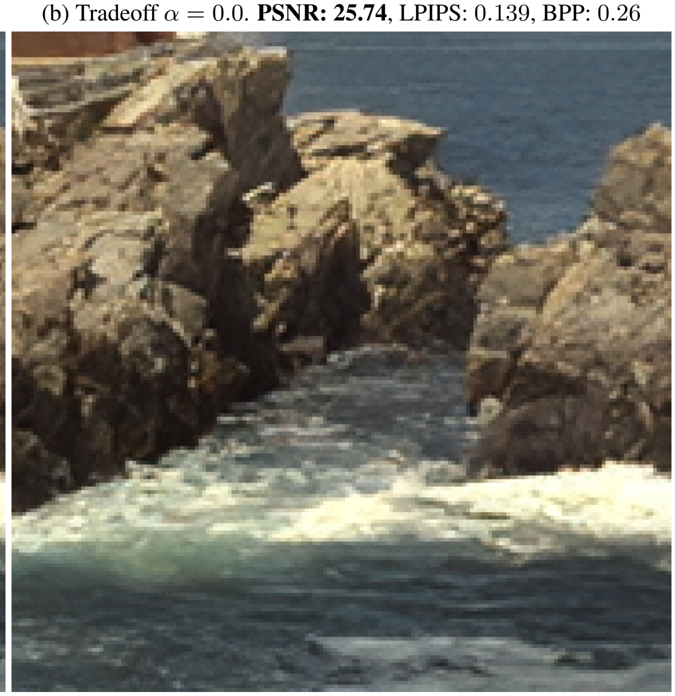
**Fig. E.13**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated image is very smooth, and missing some details in areas. However, the reconstruction improves over the default configuration (a) by about 1.2 dB in PSNR. With increasing $\alpha$ (c, d), the sharpness of the image increases. However, for high values of $\alpha$ (d), this comes at the cost of some noisy artifacts, especially noticeable in the lettering on the side of the plane.

(a) Default configuration. PSNR: 24.52, LPIPS: 0.120, BPP: 0.26



(b) Tradeoff $\alpha = 0.0$. **PSNR: 25.74**, LPIPS: 0.139, BPP: 0.26

(c) Tradeoff $\alpha = 0.98$. PSNR: 24.88, LPIPS: 0.110, BPP: 0.26



(d) Tradeoff $\alpha = 1$. PSNR: 24.48, **LPIPS: 0.106**, BPP: 0.26

**Fig. E.14**: **Qualitative comparison of different tradeoffs for CDC $x_0$ [1].** When optimizing this method for PSNR (b) the generated image is very smooth, and missing texture details for the rocks and water. However, the reconstruction improves over the default configuration (a) by about 1.2 dB in PSNR. With increasing $\alpha$ (c, d), the sharpness of the image increases, which is especially noticeable on the rock face.