Summary We thank the reviewers for providing important feedback to improve our work. We are delighted that the reviewers all recognized the relevance and novelty of the core contribution of our paper, the iREC compression algorithm (R1, R2, R3, R4), and noted its benefits in terms of compression performance both in the lossless (R1, R2, R3) and in the lossy domain (R3). We are also pleased that the reviewers found our paper to be clearly written (R1, R2, R4), our methodology correct (R2, R3, R4) and all judged our results to be reproducible. The reviewers had the following main concerns: **Performance:** (R1, R2) were concerned about the performance of our method against the state-of-the-art. **Datasets Used:** (R2, R4) voiced concerns with the datasets used in our experiments. **Motivation:** R3 asked us to expand on the motivation for our work and noted the lack of an **Ablation Study** to show the benefits of the beam search procedure discussed in Section 3.1.2. We address all four concerns below.

Performance We want to highlight that both Figure 3a and 3b show the performance on a logarithmic scale, which has the effect of over-emphasizing performance at high bitrates and under-emphasizing performance at low bitrates. Figure 1 shows the same data represented on a linear scale. We can see that iREC and Ballé (2018) outperform the competing baseline methods and that the difference between Ballé (2018) and iREC is imperceptible at all bitrates.

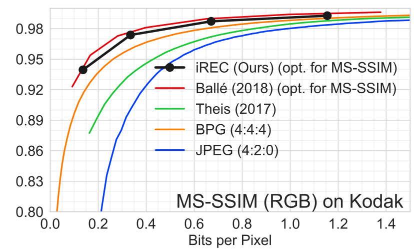

Figure 1: MS-SSIM rate-distortion curves

Datasets Used We agree with the reviewers that the CIFAR-10 and ImageNet32 datasets are inadequate to assess true performance well, hence the inclusion of the Kodak dataset in Table 1. We included the low-resolution comparisons to compare against competing bits-back methods that are yet to be scaled to high-resolution images. For the lossy compression comparisons, we opted to use the Kodak dataset due to its ubiquity as a benchmark test dataset for ML-based image compression methods. We will consider the Class B test sequences (suggested by R4) in our future works.

Motivation The motivation to our work is to bring the benefits of bits-back coding to the compression of single images and to the lossy compression domain. The first benefit is the ability to work with continuous latent distributions, which greatly expands the viable generative models for compression and simplifies training, and the second benefit is the theoretical guarantee on the codelength.

We agree with R3 that in the context of generative modelling, the use of continuous latent distributions has been studied extensively. In Section 2.1 we discuss the related approaches in the compression setting: we describe and contrast quantization approaches that use discrete latent distributions, and BB-ANS that uses continuous latent distributions. We are keen to expand on this discussion, please let us know if we missed any prominent works.

Ablation Study We thank R3 for pointing out the lack of mention of an ablation study in the main text. We performed several ablation studies, the results of which are reported in Figures 1 and 2 in the supplementary material, where we experimented with several settings for Ω, ϵ and B. In the special case of a single search beam ($B = 1$), the algorithm is equivalent to the importance sampling procedure described in Section 3.1.1. The most relevant details are in rows 1 and 3 of Figure 1, where both the coding overhead (row 1) and the residual overhead (row 3, a proxy for the bias of the sampling procedure) are shown to be much worse compared to using the beam search procedure, across a variety of different settings. In the final version of the paper we will include a clear discussion about these studies.

Further Questions and Issues

Provably close communication cost of iREC to the KL (R1): Since we draw $S \approx \exp(\mathrm{KL})$ samples, the index of a sample can be communicated in approx. $\log S = \mathrm{KL}$ nats (lines 170 - 175). We refer to Havasi et al. (2019) for the rigorous proof of correctness for the importance sampling procedure.

Gain from using continuous latent spaces (R1): We agree with the reviewer that we do not demonstrate significant performance benefits from this in the paper, and investigating how to best leverage this is left for future work.

Separating our method from bits-back methods in Tab. 1 (R2): We agree and we will separate them in the final version.

Typos and grammatical errors (R3): We thank the reviewer for bringing these to our attention, and we will correct these errors in the final version of the paper. Regarding line 160, we agree that it is confusing and potentially misleading and will clarify this point in the final version of the paper.

Including JPEG2000 as a benchmark in Figures 3a and 3b (R4): We opted against including more benchmarks in the figures to avoid clutter. We will make sure to include a more detailed figure in the supplementary material.

Thank you for reviewing our work. If this response adequately addressed your concerns, please consider adjusting your score.