We wanted to thank the reviewers for their thoughtful, concrete suggestions, which we have worked to address.

Reviewer #1 presents two main suggestions. First, an improved discussion and comparison with related work, particularly comparing with inpainting, and studying how our method relates to those based on exploring latent dimensions to generate new samples. Secondly, Reviewer #1 requested a more in-depth exploration of SSNs, and concretely suggested adding qualitative and quantitative results of the model when varying the size of the blocks. We agree that we should have done a better job with the literature review, and based on your comments we expanded the related work section in these directions. Furthermore, we added two experiments (see below), including the one you suggested with a different block size, and one that elucidates how inpainting and LDBR differ. We hope that this addresses some of your concerns, and if you think so, it would be great if you could update your score accordingly :).

To Reviewer #2, we apologize for missing the mentioned references! We have added these along with a detailed paragraph discussion (see the second to last paragraph in the added related work). We hope that this, and the rest of the added literature review and experiments, addresses some of your concerns and shows how we build on and complement existing works. If you think so, we would really appreciate it if you could update your score accordingly!

To Reviewers #3 and #4, thank you so much for your comments! We hope that the added literature review and experiments improved or solidified your opinions on the paper. We fixed all the mentioned typos and inconsistencies (e.g. now use exclusively the StyleGAN-2 name, added the detail of blocks being 4x4x512, and so forth). We also added an experiment very similar to the one suggested by Reviewer #4 (see below), so thank you for that!

Added related work: “The relevant literature for manipulating the latent space of GANs can largely be partitioned into two major categories: methods that focus on manipulating global attributes such as age and gender, and methods that focus on making localized changes associated with segmentation maps and/or instance maps.

The first category involves approaches that aim to manipulate global attributes of an agnostic decoder’s latent space. For example, GANSpace (1) applies PCA to the latent space or feature space of a decoder to modify global attributes like the make of a car, background, or age. (2) similarly manipulates global attributes showing the latent can be disentangled after linear transformations, or (3) by performing optimization in the latent space of a decoder. However, these methods need many optimization steps for a single encoding or feature. Other methods train an encoder to factorize the latent in specific ways without modifying the generator, such as in (4), (5), or (6). Another class of approaches to manipulating global attributes trains an encoder jointly with the decoder, as done in ALAE (7), ALI (8), and BigBiGAN (9).

The second category involves modulating features via prior assumptions about which semantic features would be useful to modulate, such as faces in (10), (11), (12), or that there is only a single central object in each image as in (13), (14), (15). SSNs are in this category, aiming to modify local regions. If high resolution segmentation maps are available, the promising work of Bau et al (16), (17), and (18) (19), (20) have shown that it is possible to not only make geometric changes in the image by changing the segmentation maps, but also modify the textural properties within a given segmentation instance or class. Beyond being practical, the interpretable factorization in this line of work builds on similar approaches to understand individual units in classifiers in (21), (22), (23), (24), and provides insight into how these models are capturing or not capturing the distribution. The mode dropping phenomena highlighted by these segmentation-based approaches inspired our resampling work, particularly Bau et al.’s demonstration of how under-capacity GANs drop difficult classes such as humans in front of buildings from the support of the distribution (25).

We also believe that local resampling without explicit segmentations can be useful. For example, when segmentations are not available, or when the strong geometric prior of segmentation may be too restrictive. When coupled with segmentation, the latent representation tends to capture primarily textural details, whereas in our approach with SSNs the latent representation also captures geometric detail (it is more flexible at the cost of being less precisely controllable compared to segmentation approaches). In summary, our technique is useful when one does not have semantic segmentation available, or one wants to try out significant geometric changes not constrained by segmentation maps. ”

Added experiments: We include additional experiments to highlight two things: The distinction between inpainting and potential advantages in certain situations, and the workings of SSNs with new block resolutions.

In these experiments, we switch from 4x4 blocks to 8x8 blocks to showcase a more granular resampling. We resample the blocks constituting to the left eye in a picture three times (zoom to view well).

In A we obtain a local resampling: the left eye region is more lit and less shadowy. This is a typical desired case of LDBR. In B, we see change that spans outside the resampled region with glasses appearing across the face. This resampling adheres semantically since it would be out of distribution to have glasses appear only on one half of the face. In C we receive little to no change, which is also in distribution but arguably not the desired use case.


A B C

The distinction between LDBR and inpainting is very clear in case B. Inpainting by definition is not allowed to make changes to the area specified as conditioning, which includes the *right* eye. However, for SSNs, the other eye can be changed with added glasses. This example highlights the intrinsic trade-off between having a faithful (and diverse) resampling of the data distribution and low distortion.