Relevance   I would first like to start by commenting on the relevance of this work to the NeurIPS community, as this was a concern raised by two reviewers. NeurIPS has a long history of accepting 'application' papers, and I believe that our paper falls well within the category 'Applications -> Health' as listed in the call-for-papers. The workshop 'MEDICAL IMAGING MEETS NEURIPS' was held in 2017,2018 and 2019; and was well attended. We decided to submit this work to NeurIPS because it directly addresses the banding issue that was first identified in the 2019 workshop, and we believe that many researchers working on the fastMRI dataset that we use will be attending the conference (virtually) this year as well, as a second fastMRI competition has been announced as part of the 2020 workshop.

I would like to urge the reviewers to discuss this relevance issue with each other, and contact the area chair for guidance if necessary. This is a difficult issue to decide, and I hope you can take into account that the NeurIPS community is large and diverse.

Response to reviewer 1   Thank you for the detailed feedback, it is much appreciated. You're absolute right that it is very difficult to access the quality of the images generated by a reconstruction method when there are conflicting concerns. A method that produces visually pleasing results but introduces additional artifacts is of little use in practice. The focus of our approach from the beginning was to avoid introducing artifacts, which is why we took the approach of an orientation adversary, rather a more traditional GAN style adversary. Our early experiments with a GAN approach showed that it introduced false detail into the reconstruction.

We believe our approach is actually motivated by the process producing the artifacts. As we conjecture, the artifacts are due to the orientation of the sampling process. This is why a loss that penalizes orientation-aligned artifacts is sensible.

We did not ask the panel of radiologists to directly assess if new artifacts had been introduced, and in retrospect I believe you are right, it would strengthen our work if we had. I can say however that we have not seen any evidence that our method introduces new artifacts, and we have looked at a large number of reconstructions from the model. We have also regularly consulted with a radiologist (separate from the panel) who did not identify any issues. We would be more than happy to add a discussion of this to the paper to address your concerns.

Also, we would like to note that Table 1 doesn't shown that our method retains less detail. The difference in detail scores is actually within the statistic margin of error, as shown by the p-value listed in the table. Note the p-value is above 1 due to the Bonferroni correction, but it would still be much larger than 0.05 even without the correction.

Response to reviewer 2   We have not been able to answer all your questions due to lack of space. In regards to the use of 1.5T scans for the experiments, we do believe it is the signal-to-noise ratio that primarily contributes to the greater degree of banding compared to 3T scans. Our focus on 1.5T scans is motivated by clinical practice and the direction the field is taking as a whole. They are still very common due to the lower cost, and they must be used in cases where the patient has implants that have not been certified as safe for 3T scanning. Using machine-learning to enable even lower field scanners is currently an active research area, including portable scanners which are at the prototype stage.

You are absolutely correct that a larger scale study of the ability of radiologists to still detect pathologies in the reconstructed images should be considered the gold-standard. However, this sort of comparison is extremely expensive as it requires a much larger amount of radiologist time to achieve statistical significance, it's just not possible for us to do such a comparison.

Unfortunately, our method can not be applied in cases where the data is undersampled. We have spent a lot of time trying to find a solution and we have not been able to come up with any method that works without full data for training.

Response to reviewer 3   We hope we have addressed your concerns with relevance above, and we hope you can reconsider your rating in light of our comments. We agree that we need to provide a more detailed introduction to the terms and concepts used so that our paper is more approachable. We can fully commit to improving the readability of our work for the camera ready.

We will also add a discussion of how our approach builds upon classical compressed sensing. As you note, there is some surprising differences between deep-learning approaches and classical approaches with regards to the masks. Equispaced masks out perform random masks for deep-learning reconstructions (SSIM 0.926 vs 0.915 for our model). The recently released brain portion of the fastMRI dataset moved to equispaced masks from the random masks used in the initial knee release for this reason.

Response to reviewer 4   Please see our response to reviewer 2. Our focus on reducing this particular artifact was initiated by our regular consultation with practicing radiologists, they have told us that it is a very major concern they have with the reconstructions, second only to the retention of anatomical detail and pathologies.