We thank the reviewers for their useful feedback. Overall, we see a positive reception of our work. Reviewer 1 points out that the paper is "well-written", reviewer 4 says that "the paper has an originality" and notices the "very nice results on the MRI reconstruction task", and reviewer 5 finds that "this paper would be very useful for Computer Vision community". We start by emphasizing our main contributions to help clarify some of the issues raised by reviewers 4 & 5 about structure and focus of the paper. We have added the following statement of contributions to the Introduction:

1. **The first invertible "learning to infer" model.** We draw inspiration from generative models to modify an existing "learning to infer" approach called "Recurrent Inference Machines" to be fully invertible. This allows us to train more expressive (deeper) models than before because we can overcome memory constraints during training using invertible learning [16]. Because the presented model is invertible it will allow for semi- and un-supervised training in the future [18]. This is especially relevant for domains where the signal of interest $\mathbf{p}$ is always unknown, such as in synthesis imaging in radio astronomy.
2. **Stable invertible learning of very deep models.** To the best of our knowledge, we present the deepest network that has been successfully trained with invertible learning[1]. In practice, invertible learning can be unstable due to numerical errors that can accumulate in very deep networks [16]. In our experience, common invertible layers [18-20] suffered from this problem. We give intuitions why these layers might introduce training instabilities, and present a new layer that addresses these issues and enables stable invertible learning of very deep networks (400 layers).
3. **Scale to large observations.** We demonstrate in experiments that our model can be trained on large volumes in MRI (3d). Previous "learning to infer" models were only able to perform reconstruction on 2d slices. For data that has been acquired with 3d sequences, however, these approaches are not feasible anymore. Our approach overcomes this issue. This result is relevant for other domains with large observations such as synthesis imaging in radio astronomy.

**Reviewer 1: Generality of our approach and the question how specialized MRI really is.** Undersampled MRI image reconstruction is a difficult deconvolution problem and, as such, it is a prime example of an inverse problem. Our approach is an extension of the RIM, as recognized by reviewer 5, which has proven successful in a number of imaging tasks (l. 196-198). Here, we wanted to focus on the most challenging and practical problems where observation size is an issue. We plan to apply our approach to synthesis imaging in radio astronomy in the future, however, this application would have been beyond the scope of the presented paper. We modified the paper to emphasize the significance of accelerated MRI as an inverse problem.

**Reviewer 1: Paging memory between GPU and CPU** That is indeed infeasible. The Nvidia V100 used for training has a memory bandwidth of 900 GB/s, while the PCIe 3.0 x16 bus that connects the card to main memory has transfer rates < 16 GB/s. Data transfer between main memory and GPU memory is generally considered a bottleneck.

**Reviewer 1: Inversion of Eq. 6** For improved clarity we have changed eq. 6 to show the forward computation on the left, and the reverse computation on the right:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{x}_1 & \qquad \mathbf{x}_2 &= \mathbf{y}_2 - \mathcal{G}(\mathbf{y}_1) \\ \mathbf{y}_2 &= \mathbf{x}_2 + \mathcal{G}(\mathbf{y}_1) & \qquad \mathbf{x}_1 &= \mathbf{y}_1 \end{aligned} \tag{1}$$

We did the same for eq. 14. Please note, that the inversion holds independent on the parametrisation of $\mathcal{G}(\cdot)$, i.e. $\mathcal{G}(\cdot)$ does not have to be invertible.

**Reviewer 4: Increase font size in figure 2** Fixed.

**Reviewer 4 & 5: Reviewer 5 emphasized that the invertible layer appeared to be "somewhat apart from the proposed method", and concluding that "the coherence of the paper was blurred from adding this"** We see why the reviewer finds this unclear and we hope that the above contribution section clarifies our reasons for including the invertible layer in the paper. To elaborate more, we originally had contributions 1 and 3 as our goals. But since invertible learning for "learning to infer" models is new territory, it was not immediately clear whether established layers could guarantee stable invertible learning in this scenario. After all, it is well recognized that great care has to be taken in initialization and parametrization of Glow and real-NVP layers for stable training. In experiments, we verified that these layers indeed made training of the i-RIM unstable. Hence, we introduced a new layer that addresses the issues we identified with other layers (l.136-147), and that allowed us to train models in a stable fashion. We therefore understand our invertible layer as a crucial component to guarantee stable training of the i-RIM. To better justify the invertible layer as a component for training the i-RIM we have added more context in the introduction of section 2.2.

[1] `http://bayesiandeeplearning.org/2018/papers/37.pdf` presents a 300 layer deep network trained on CIFAR10