We are glad that the reviewers found **<u>MetaSDF</u>** to be a “novel approach [with] many potential benefits" **(R1)**, the "motivation [...] very convincing and perfectly pitched to the reader" **(R3)**, and "providing interesting insights about the learning of neural implicit shape representations" **(R2)**. Neural implicit representations are an emerging field with applications across vision, graphics, and other branches of ML. We propose a novel approach to generalization across these emerging representations, with clear benefits on inference time and out-of-distribution generalization (Sec. 4). We stress that we *do not* claim a new state-of-the-art approach to shape reconstruction **(R2)**, merely that the proposed approach performs *on par* with state-of-the-art approaches (see ln. 47, ln. 246). Instead, our key contribution is establishing the connection between the emerging fields of neural implicit representations and gradient-based meta-learning. This is the very first step, and a wealth of existing work on meta-learning may provide significant further improvements. We believe that this will spur follow-up work benefitting both of these promising research directions.

Further ShapeNet classes **(R1, 3)** We have trained models on the ShapeNet "benches" class—please see qualitative test-time reconstructions for dense SDF samples (a), and surface samples only (b). Quantitatively, we achieve a Chamfer Distance (CD) of mean $3.01\mathrm{e}{-4}$, median $8.54\mathrm{e}{-5}$, stddev $1.37\mathrm{e}{-3}$ (ours) vs. mean $3.11\mathrm{e}{-4}$, median $7.36\mathrm{e}{-5}$, stddev $1.7\mathrm{e}{-3}$ (DeepSDF) while outperforming the PointNet-encoder model. We will benchmark another class with fine features for the final manuscript. We have done our best to benchmark the proposed approach extensively, but note that each class takes several days to train on two of the largest available NVIDIA GPUs, the 48 GB RTX8000, which we share with other researchers. We note that 2D results (Sec. 4) are entirely consistent with our 3D results. Together, this conclusively demonstrates the potential of MetaSDF. Research on gradient-based meta-learning is progressing quickly, and it is poised to become more computationally affordable in the near future (see ln. 283).

Reconstruction from partial observations **(R1)** We ran an experiment to reconstruct SDFs from depth maps as in Fig. 8 of DeepSDF—see qualitative result in (c)—with no further fine-tuning or heuristics. Quantitatively, we achieve a Chamfer Distance of $6.55\mathrm{e}{-4}$ for planes, where DeepSDF reports $1.16\mathrm{e}{-3}$. We use the results supplied by the authors of DeepSDF, as the authors do not provide code or data for this experiment and we did not succeed in reproducing their results. We will add experiments and comparisons with further classes to the final manuscript.

Related work & IM-NET **(R2)** We will discuss DISN in-depth. IM-NET reconstructs shapes using either an image encoder, which is of different modality than the 3D samples MetaSDF is focused on, or a PointNet encoder with conditioning via concatenation. We benchmark against this architecture (see submission Table 3, Fig. 8).

MIOU **(R1)** We have computed MIOU for planes, tables: 0.87, stddev $9.26\mathrm{e}{-2}$ and 0.85, stddev $1.43\mathrm{e}{-1}$ (MetaSDF) vs. 0.85, stddev $8.37\mathrm{e}{-2}$ and 0.85, stddev $1.30\mathrm{e}{-1}$ (DeepSDF). We will add MIOU for all classes to the paper.

Out-of-Distribution generalization **(R2)** We have reconstructed benches with the model trained on tables. Please find qualitative results in (d). We will add a quantitative benchmark to the final manuscript.

Standard deviations **(R3)** We have computed the CD std. dev. on planes and tables: $1.29\mathrm{e}{-4}$ and $2.39\mathrm{e}{-4}$ (MetaSDF) vs. $2.07\mathrm{e}{-4}$ and $1.48\mathrm{e}{-3}$ (DeepSDF). We will include further classes and MIOU std. dev. in the final paper.

Convergence comparison **(R3)** Please see (e) for a plot of test-time iterations (log scale) vs. test loss. We define convergence as the number of optimization steps until the test-time loss does not decrease further.

Memory Complexity **(R3)** MAML requires the computation of second-order gradients, whose memory complexity depend on the auto-differentiation algorithm, which are evolving rapidly (see Bettencourt et al., NeurIPS 2019 workshops). We can state that MAML's memory complexity scales linearly with inner-loop gradient-descent iterations. With the architectures described in the paper, MAML requires about twice the memory of the auto-decoder.

Novelty in light of MAML **(R3)** We do not claim any fundamental contributions to gradient-based meta-learning itself, and will highlight this in the paper. However, we found several technical challenges specific to applying gradient-based meta-learning to generalization across implicit neural representations. First, vanilla MAML underfits the data, and it is necessary to leverage per-parameter, per-step tunable learning rates as proposed by Li et al. (2019), Antoniou et al. (2018). Gradient warping Flennerhag et al. (2019) and related methods are likely to improve performance further. Next, performing gradient descent with truncated SDF values as proposed in DeepSDF leads to unstable training & catastrophic failure that the model does not recover from. This was addressed by the proposed composite loss function.

Applications to GANs **(R3)** Applications to GANs are an exciting avenue of future work which we will discuss.

Improve exposition **(R4)** The motivation for our work is indeed to find a good initialization that allows fitting an implicit neural representation in few gradient descent steps. We will include an overview figure to better motivate our proposed meta-learning approach. ℓ_1 in ln. 133 refers to the L1 loss. A discussion of hypernetworks can be found in “Scene Representation Networks”, Sitzmann et al. (2019). Ln. 174 indeed describes the proposed architecture.