We would like to thank reviewers for their time and the effort they put into reviewing our submission. We welcome their thoughtful suggestions, which we will incorporate into the final improved version of this submission.

Q1: (R1,3,4) Real datasets are too simple. A1: We argue that small and controlled experiments provide more insight into the inner workings and failure modes of novel objectives. We are the first to introduce a purely likelihood-based alignment objective, and we empirically verify that resulting models preserve local structure of aligned domains better than non-parametric models, are free from mode collapse and training instability of GANs, and theoretically justify why they suffer from vanishing generator gradients in higher dimensions as much as their adversarial counterparts. We argue that this empirically verified result adds value to the scientific community, freeing future researchers from having to re-discover this confusing failure mode (vanishing gradients in higher dimensions) of a seemingly proper likelihood-ratio based alignment objective. Such findings are better explained with simpler, controlled experiments. We will add results on higher resolution images in other domains.

Q2: (R4) No comparison to / improvement upon existing GAN-based domain adaptation methods in terms of classification accuracy. A2: Recent improvements in the performance of classifiers adapted using adversarial alignment largely rely on additional losses utilizing source labels, such as semantic consistency in CyCADA [3] or classifier discrepancy in [4], added on top of the unsupervised adversarial alignment. In this work, we are primarily interested in improving the robustness of the alignment itself without making any assumptions about the downstream task. We will extend the related work section with recent advances in adaptation of classification models and explicitly acknowledge that the comparison to methods that make active use of source labels is beyond the scope of this work.

Q3: (R1) Local optima of LRMF wrt Th 2.3 (3), when does the equality hold? A3: The inequality (3) holds for any θ (shared model parameters) and ϕ (transformation parameters), but is tight (equality holds) only when the bias term is zero, and the shared model is optimal. However, the existence of local minima for theta might indeed prevent LRMF from converging to the actual value of LR-distance, *i.e.* the statement of the theorem always holds, but the final model produces a useful alignment only when a sufficiently "deep" minimum is found, otherwise the method fails, and the loss value is indicative of this failure. We will explicitly acknowledge this in the paper.

Q4. (R4) Claims without proper citations. A4: "difficult to quantitatively reason about the performance of [GANs]" - we will add [1; 2; 5]; "but rarely on density alignment" - we will cite AlignFlow and PointFlow here, as the only published attempts at doing that, to our knowledge; "In general, this would require solving an adversarial optimization problem" - we will append with "We show in Section 2 (Eq 1) that, in general, this would . . . ".

Q6: (R3) Extended description of results and setup on real data. How were transformation hyperparameters chosen? We choose M that gives high validation likelihood on A and B separately, since optimal θ_S has to approximate A well. The choice of Real NVP vs GLOW is dictated by the same principle. We also show that in lower dimensions LMRF works even in the over-parameterized regime. We will extend the description of the setup and the discussion of results obtained on real data with an illustrated version of the following observation: in higher dimensions, more complex shared models quickly "envelop" two datasets before they become aligned, forming two disjoint "bubbles" of density that fail to merge (objective landscape is flat wrt ϕ), whereas simpler shared models result in the alignment of means and variances only.

Q7: (R4) Effect on the number of parameters vs AlignFlow. A8: Indeed, compared to the AlignFlow-inspired baseline, our approach requires 2-3 times more parameters, but preserves the local structure of aligned domains better, as shown in the paper. Higher number of trainable parameters does not cause overfitting, as shown in Setup 1. However, GPU memory becomes an issue when training complex LRMF flows. We were able to fit a GLOW-LRMF with three scales and eight affine coupling layers on each scale with batch size = 4. This should be enough for medium-sized images from natural domains. We will explicitly acknowledge this in the related work and limitation sections.

Q8: (R4) Inductive bias from two datasets in the shared model. A9: Upon convergence the shared model has to fit the target dataset. It will be biased towards same kinds of patterns as the underlying flow model on target data.

Q9. (R1,3,4) Clarifications and corrections: *1) negative cross-entropy in line 80* - correct, we apologize for this typo, we double-checked, all other uses of this term in the paper are consistent with the literature; *2)* $F \circ G^{-1}(A)$ *composition refers to an AlignFlow-inspired baseline?* - correct; *3) should this be likelihood ratio than log-likelihood ratio?* - technically, you are right, but the term "log-likelihood ratio test" seems to be well adopted in the literature, though, technically, incorrect, since this is a "log of the likelihood ratio".

References

[1] Shane Barratt and Rishi Sharma. A note on the inception score. *arXiv preprint arXiv:1801.01973*, 2018.

[2] Ali Borji. Pros and cons of gan evaluation measures. *Computer Vision and Image Understanding*, 2019.

[3] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A. Efros, and Trevor Darrell. CyCADA: Cycle consistent adversarial domain adaptation. In *ICML*, 2018.

[4] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018.

[5] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2015.