

1 We thank all the reviewers for their valuable feedback and the positive evaluations of our paper. Below, we address the
 2 comments of each reviewer individually.

3 **[R1] Biases and Mode Drop:** The model of outliers we assume in our paper is that of large noise (Section 3). Hence,
 4 our model drops samples that are far from the true data distribution in the Wasserstein sense. As pointed by the reviewer,
 5 it is important not to drop rare modes, or have a mechanism to identify mode dropping when it happens. As shown in
 6 Table 1 of the main paper, there is no drop in FID scores on clean CIFAR-10 dataset, which suggests that no mode drop
 7 has occurred. To further understand the effect of biases, we train our robust Wasserstein GAN model on CelebA dataset
 8 with varying female:male ratio. We then measure the female:male ratio of the generated distribution obtained from the
 9 trained GAN (using an attribute classifier). Table 1 shows the results. We observe that even when fraction of males are
 10 as low as 2% in the input dataset, they are generated with a similar ratio in GANs. We will add these results to the paper.

Table 1: **Analyzing mode drop:** Training robust GAN on imbalanced CelebA. In each column, we report % of males generated by a GAN trained on the respective input dataset.

% Males - Input dataset	2 %	5 %	10 %	20 %
% Males - Vanilla GAN	5.23 %	7.42 %	12.16 %	21.29 %
% Males - Robust GAN	4.84 %	7.80 %	10.12 %	21.51 %

Table 2: **Partial Domain Adaptation Results:** Office-31 dataset. Accuracy in % is reported

Setting	W->A	D->W	D->A	A->W	W->D	A->D	Avg
Source only	71.7	94.5	73.1	54.5	94.2	65.6	75.6
Robust OT	92.8	95.7	93.4	87.9	97.2	85.5	92.1

11 **[R1] Identifying mode drop:** In some cases, as pointed by the reviewer, rare modes can potentially be dropped. In
 12 this case, we can use the weights estimated by our weight network $\mathbf{W}(x)$ to visualize which modes are dropped (Fig
 13 3 and 4 of the main paper). Samples with low weights are the ones that are dropped. We can use these weights in a
 14 boosting framework to train a mixture model to generate balanced datasets. We will add this discussion to the paper.

15 **[R2] Comparison with Sinkhorn Iterations:** We thank the reviewer for pointing us to the reference. The focus of
 16 our paper is to present relaxations to make unbalanced OT suited for deep learning applications. Hence, to compare our
 17 algorithm with [1], we trained a Resnet WGAN model on CIFAR-10 dataset as follows: We sample a batch of real and
 18 generated samples and consider these as empirical distribution of size n . For these distributions, we solve for the dual
 19 vectors u and v for K iterations using the Sinkhorn algorithm in [1]. Then, we use these solutions in the unbalanced OT
 20 dual objective and optimize for the generator. We then iterate between these two steps. We observed that the GAN
 21 model trained using this procedure produced poor generations (extremely blurred samples with mode collapse). Similar
 22 issues on using entropic GAN objectives for deep learning have been reported in [2].

23 [1] "On Unbalanced Optimal Transport: An Analysis of Sinkhorn Algorithm", ICML 2020.

24 [2] "Entropic GANs meet VAEs: A Statistical Approach to Compute Sample Likelihoods in GANs", ICML 2019.

25 **[R2] Optimization Objectives:** Objective (8) is easier to implement than (7) for neural networks because in (8),
 26 weights are obtained using a neural net. This makes the entire network end-to-end differentiable and the optimization
 27 can be approximately solved using SGD. In objective (7), however, we need to alternate between SGD steps for training
 28 GAN and a second-order cone optimization for estimating weights (which can be expensive for large datasets). Explicit
 29 algorithm for solving (7) and (8) (for GANs and DA) are provided in Alg. 1, 2 and 3 of Supplementary material.

30 **[R2] Asymptotic convergence:** Yes, the gap between empirical robust Wasserstein measure $\mathcal{W}^{rob}(\mathbb{P}_X^{(n)}, \mathbb{P}_Y^{(n)})$ and true
 31 robust Wasserstein $\mathcal{W}^{rob}(\mathbb{P}_X, \mathbb{P}_Y)$ goes to 0 as $n \rightarrow \infty$. The proof is very similar to the asymptotic convergence proof
 32 of unbalanced optimal transport as provided in reference [13]. We will explain it in the paper.

33 **[R3] Recent architectures and high resolution datasets:** In the main paper, we have provided results on Resnet-based
 34 GAN and Spectral Normalization GANs on various datasets including CIFAR (32×32 resolution) and CelebA (64×64
 35 resolution), which are benchmark datasets for GANs. We did not add perform experiments on large-scale higher
 36 resolution datasets like Imagenet due to the lack of computational resources.

37 **[R4] Special design choices:** Both GANs and domain adaptation experiments use distribution matching in their
 38 objective: GANs minimize distributional distance between real and generated samples, while in domain adaptation,
 39 we minimize distributional distance between source and target features. There are small changes in design choices,
 40 such as discriminators operating in feature vs image space, architectures, etc. A detailed description of algorithm and
 41 architectures are provided in supplementary material. Effectiveness on a wide range of problems shows the versatility
 42 of our proposed approach. We will further add a discussion in the revised draft.

43 **[R4] Office-31 Experiments:** Upon your suggestion, we performed experiments on partial domain adaptation problem
 44 on Office-31 dataset following the protocol provided in [3]. Results are obtained as shown in Table. 2. Our approach
 45 achieves good performance improvement compared to baseline.

46 [3] Cao et al., "Partial Adversarial Domain Adaptation", ECCV 2018.

47 **[R1-4] Typos.** Typos will be fixed and citations will be added.