We thank all the reviewers for their insightful and encouraging comments. We're encouraged by the reviewers' appreciation that 1) our method is well-motivated through the RL literature (R1, R2); 2) our empirical results on multiple tasks are comprehensive and promising (R1, R2, R4); and 3) the paper is well-written (R1, R2).

We emphasize that the **main technical novelty** of the paper lies in connecting GAN training with both TRPO/PPO and importance sampling through a new *principled variational GAN formulation* (Sec.3.1), which makes it possible to re-purpose probability ratio clipping and re-weighting for GAN training. We also devise an approximation technique to enable probability ratio estimation for implicit generative models. Our **empirical contributions** include the studies on a broad range of tasks (image generation, text generation, text style transfer) and our consistently improved performance.

For *theoretical analysis*, we show that the method is fully compatible with *Theorem 2* in [56] (ICML'19) which provides rigorous convergence analysis on GANs with Lipschitz discriminators and concludes 1) informative gradient pushes the model distribution to the real data distribution and 2) the only Nash-equilibrium is $p_{model} = p_{data}$.

Reviewer #1: Thanks for your positive comments on 1) our good motivation through RL and EBM, 2) improved performance on all 3 tasks, and 3) good paper writing. We'll add discussions and fix all issues in revision.

* *EMA:* EMA and our approach are *orthogonal*. EMA/MA averages generator parameters over time *outside* the training loop (Yazıcı et al., ICLR2019) to reduce the stochasticity of mini-batch training, and thus is independent of how GAN is trained. Moreover, EMA has to counter the generator's distributional shift issue by tuning hyper-parameters (window size and average ratio). Our work can come complementary to EMA by discouraging distribution shifts with the new surrogate loss, and can potentially make EMA easier to use. It's interesting to study the combination in the future.

* *Correctness:* In submission, we already compared with WGAN-GP under the same settings: image generation in Table1 and text generation in Table 3. So the contribution of gradient penalty is already ruled out for comparison. Since WGAN-GP alone on style-transfer has mode collapse issue, we did not discuss it.

* *Ratio-clipping only:* We emphasize that re-weighting and probability ratio clipping (KL regularization) are derived from the variational framework (Eq.2) in a *principled* way, from introduction of the variational distribution q. Discarding either of the two leads to improper handling of q and fails to conform to the framework (and the theoretical properties). We reported results of "reweighting-only for ablation study (despite its mathematical inappropriateness).

Reviewer #2: Thanks for appreciating that our method is well-motivated with good theoretical foundations, and shows promising results on all three tasks. We'll add details in appendix, discuss related work, and fix all other issues.

* *Human evaluation:* Thanks for the suggestion. Following the same setting in the RelGAN paper, we conducted human evaluation to compare RelGAN(1000) and our method. Ours obtained an average human score of 3.59, higher than 3.42 by RelGAN (Fleiss' Kappa score 0.61 showing *substantial* inter-rater agreement).

* *Hyperparameters:* Our method and WGAN-GP baseline use the same hyperparameter setting as RelGAN(1000)

Reviewer #3: We selected ϵ from $\{0.2, 0.4\}$, as they are typically used in PPO. We'll fix all other issues in the revision.

Reviewer #4: We first clarify for several concerns:

- In text generation, NLL_{gen} measures *diversity* (Line.235). Our model has better diversity than RelGAN (Table.3).
- In appendix 6.1, we meant it's *bounded* by a constant. The overall correctness is not affected. Also, please refer to the clarification of the theoretical analysis above. We will revise the statements for clarity.
- In Fig.3 (left), the update ratio of WGAN-GP is 5:1 (the best setup), the reweighting-only method used 5:1, and our full method used 5:5. We clarify that both WGAN-GP and ours used the *same amount of computations* (i.e., a 5:1 iteration is counted as 6 training batches, and a 5:5 iteration as 10 batches). We will make this clearer. The probability ratio clipping that discourages large generator updates allows us to update the generator more frequently.

* *PPO motivation and large-batch training:* Besides sample efficiency, PPO has a strong motivation/intuition to discourage excessively large model updates [45]. This suits well for stabilizing the generator in GANs, as acknowledged by R1 and R2. In practice, our surrogate loss achieves similar effect as the KL penalty in variational framework (Fig. 1 Left). The controlled update size also enables more frequent generator updates and better efficiency (Fig. 3 Left).

Large-batch training is effective for stabilization, but doesn't solve instability alone: Masson d'Autume et al. also used techniques including dense rewards and discriminator regularization; BigGAN used spectral normalization, truncation, and progressive scaling architecture. Our approach is orthogonal and can be combined with large-batch training.

* *Clarity:* As in Line.143 above Eq.(7), $\mathcal{L}_\phi$ is "the data log-likelihood of $q^{(t)}$ w.r.t ϕ", where $q^{(t)}$ is defined in Eq.(3). Z_ϕ in Eq.(8) is also estimated with importance sampling with $p_{\theta^{(t)}}$ as the proposal. We will make these clearer.

Please refer to our response to R1 for the clarification of "ratio clipping only".

In text generation we still used classifier C despite the explicit model (though it's not necessary).