We thank the reviewers for the constructive comments. Reviewers appreciate the effectiveness of the proposed ContraGAN (R1, R2, R3, R4), the novelty of the proposed 2C loss (R1, R2, R4), composability with modern regularization techniques (R2, R4), and usefulness of our software (R3). This rebuttal answers questions raised by reviewers. Every experiment and explanation in this rebuttal will be included in the paper.

(R1, R2, R4) Clarity. We will introduce the concept of data-to-data relations carefully. Symbols in Fig 1 will be polished, as suggested. cGAN with a projection discriminator [17] will be named as *ProjGAN* to avoid confusion. We will make smooth transitions among mutual information, metric learning, contrastive loss, and 2C loss.

(R1) Comparison with other metric learning losses shown in Fig. 1. P-NCA loss [24] does not explicitly look at data-to-data relations, and XT-Xent loss [25] (equivalent to Eq. 6) does not take account of data-to-label relations. Our 2C loss can take advantage of the strengths of both losses. Compared with Eq. 7 loss, 2C loss considers cosine similarities of negative samples. We conduct experiments to compare 2C loss with other losses. Every experiment is performed three times, and its mean$\pm$variance of FID [39] is reported below.

Dataset	Unconditional GAN [6]	with P-NCA loss [24]	with Eq.6 loss (XT-Xent) [25]	with Eq.7 loss	ContraGAN
CIFAR10 [21]	15.550$\pm$1.955	15.350$\pm$0.017	14.832$\pm$0.695	10.886$\pm$0.072	**10.597$\pm$0.273**
Tiny ImageNet [22]	56.297$\pm$1.499	47.867$\pm$1.813	54.394$\pm$9.982	33.488$\pm$1.006	**32.720$\pm$1.084**

(R3) Difference between XT-Xent loss and 2C loss. XT-Xent is intended for unsupervised learning, and XT-Xent only regards the augmented images as the positive samples. On the other hand, 2C loss utilizes weak supervision from label information. Therefore, compared with 2C loss, XT-Xent hardly gathers image embeddings of the same class. The table above shows the effectiveness of 2C loss. Besides, XT-Xent loss requires extra data augmentations and additional forward/backward propagations, which induce $15 \sim 20\%$ more training time than using 2C loss.

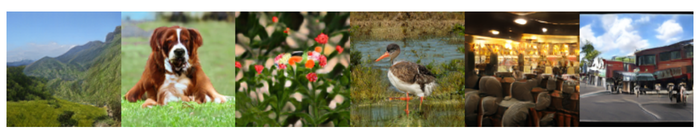
(R1, R2) Reliability of experiments. We provide updated Table 1 (left) and 2 (right) below after three times of experiments. To avoid the single-trial analysis, we will replace the original tables with these numbers. As pointed out by R2, we will mention that ProjGAN is on par with ContraGAN in CIFAR10 dataset.

Dataset/Batch Size/Res.	ACGAN [19]	ProjGAN [17]	ContraGAN
CIFAR10/64/32	10.697$\pm$0.129	10.739$\pm$0.016	**10.597$\pm$0.273**
Tiny ImageNet/256/64	88.628$\pm$5.523	37.563$\pm$4.390	**32.720$\pm$1.084**

Dataset/Batch Size/Res.	SNResGAN [4]	SAGAN [5]	BigGAN [6]	ContraGAN
CIFAR10/64/32	*17.5	17.127$\pm$0.220	10.739$\pm$0.016	**10.597$\pm$0.273**
Tiny ImageNet/1024/64	47.055$\pm$3.234	46.221$\pm$3.655	31.771$\pm$3.968	**29.492$\pm$1.296**

(R1, R2, R4) ImageNet. We perform ImageNet [18] experiments. It has not been completed within six days of the rebuttal period, and it reaches 160k iterations. We compare SAGAN and ProjGAN here, since we were able to get FID of those approaches for 160k iterations. Under the same iteration number, FID and synthesized images by ContraGAN is quite promising. Note that 160k iterations are pretty early stage, since 10M iterations are often applied.

Experiment with ImageNet with batch size 256. Image Res. 128 $\times$ 128, iteration 160k / 10M.		
SAGAN	ProjGAN	ContraGAN
40.96 [5]	32.456	**26.935**



(R2) Inconsistent SNResGAN results in Table 1 and 2. We found that SNResGAN [4] and SAGAN [5] can be improved by applying the moving average update (MAU) for the generator's weights (described in Sec. 3 in the supplement). We use our implementation for every result in Table 1, and we apply the MAU to report the best results. Table 2 takes the numbers from the original paper [4], so we did not use the MAU. This makes inconsistency.

(R3) Diversity of generated images. Since FID [39] can measure both fidelity and diversity of images, we claim that ContraGAN can generate more diverse images compared with the previous methods. For more analysis, Intra-FID [17] can be adopted to measure the degree of intra-class variation.

(R4) May generate images that are easily classifiable. ContraGAN can look at the condition of inside the batch and decide the authenticity of images using relations of examples. Thus, if the generator gives images from a restricted mode to the discriminator, the discriminator can recognize the generated samples as the fake using the relations. This procedure can lead the generator to create more diverse images to deceive the discriminator.

(R1, R2) Stability. The right figures show the singular values of convolutional layers. We observe mode-collapse of ProjGAN at 45k steps, whereas ContraGAN runs 72k steps without mode-collapse. We speculate that ContraGAN is harder to reach undesirable status, since ContraGAN jointly considers data-to-data and data-to-label relations. Tiny ImageNet dataset is used for this experiment.