1 We thank all reviewers for their constructive feedback. Below are our responses to the concerns raised by the reviewers.
2 Although we cannot respond to all the comments, we will do our best to reflect all of them in our final manuscript.

3 **[R1] Assumptions on CircleGAN.** We assume the real samples have maximum diversity and quality, and those
4 objectives are achieved at the largest isoline, which is the great circle. Thus, we place the real samples to the great
5 circle, which indicates the diversity of generated samples is lower than the real samples.

6 **[R1, R4] Comparison between CircleGAN (ours) and SphereGAN [22].** As R1 and R4 suggested, we created Fig. 1
7 below, where overall architecture and symbols are described and relevant concepts in training are compared between the
8 two models. The superscripts ('emb', 'cen', 'proj', 'rej', 'real', 'div') show the status of a variable, and the subscripts
9 ('i', 'r_i', 'f_i') denote the index or source of a variable. If clear from the context, we often omit the subscript i. For
10 notational simplicity, we omit tilde from $\tilde{p}$ to denote the unit pivotal vector.

(a) Overall Architecture

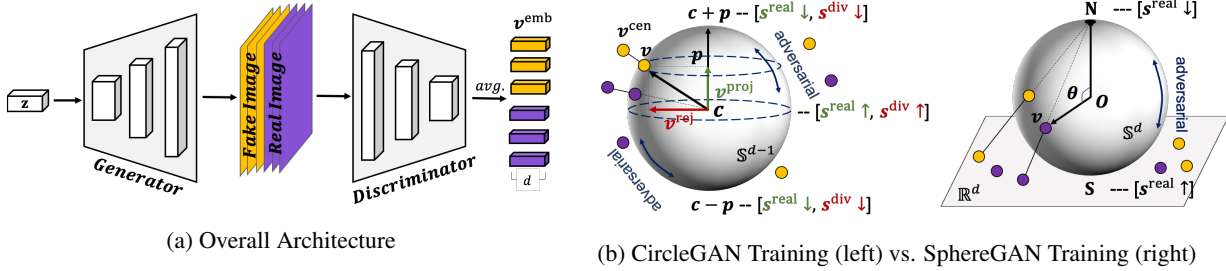(b) CircleGAN Training (left) vs. SphereGAN Training (right)

Figure 1: CircleGAN (ours) vs. SphereGAN [22]. (a) Given a randomly sampled noise vector z, the generator
synthesizes fake samples and the discriminator produces the embeddings v^{emb} from fake and real samples. (b)
CircleGAN projects these embeddings onto the hypersphere by centering (v^{cen}) and $\ell2$-normalization (v), whereas
SphereGAN performs inverse stereographic projection. Due to learnable pivotal p and center c vectors, CircleGAN is
more flexible and amenable to conditional settings. Note, however, that in SphereGAN they are fixed to the north pole
$\mathbf{N}$ and the origin $\mathbf{O}$, respectively. In training, CircleGAN performs adversarial learning based on the great circle using
the proposed score functions (s^{real} and s^{div}), and SphereGAN performs based on the point $\mathbf{N}$, which causes lack of
sample diversity. [R1: The isolines of spherical circles in CircleGAN are delineated with dashed lines.]

11 **[R2] Theoretical analysis and comparison of projection methods.** If we use the Wasserstein distance and replace
12 the cost function (Eq. 8 of [22]) with our proposed score functions (Eq. 3, 4 of ours), all the theoretical results of
13 SphereGAN should also hold for CircleGAN. In our experiments, however, we use the relativistic objective of Eq. 5
14 because the work of [11] has shown that the Wasserstein objective is a specific form of relativistic objectives and the
15 form of Eq.5 produces superior performance than the Wasserstein objective. We think similar effects of the theoretical
16 results in SphereGAN may also hold for our case because: (1) the projection function has nothing to do with the cost
17 function of the Wasserstein distance and thus does not change theoretical results of SphereGAN, (2) both CircleGAN
18 and SphereGAN use the same form of hypersphere embedding space, (3) the cost function is almost the same to the
19 angles (or $\ell2$-distances) from each particular reference, satisfying the conditions (non-negative & bounded) for the
20 Wasserstein space to be defined and allowing propositions 1 and 2 in SphereGAN to be applied, and (4) propositions 1
21 and 2 state that the convergence in the Wasserstein space is equivalent to that of the Wasserstein objective.

22 The superiority of CircleGAN mainly comes from learning based on the great circle. The inverse stereographic
23 projection (ISP) is problematic for this approach because ISP results in a great circle with a lower sample density
24 than the north pole and thus prevents the embedding space from representing the sample diversity. In this sense,
25 $\ell2$-normalization is adequate since it allows the great circle to have the highest sample density w.r.t. the pivotal point.

26 **[R3] Effectiveness of learning based on the great circle.** We conjecture the setting of [Table 1b, (4)] without other
27 techniques diminishes the efficacy of learning based on the great circle. To support this assumption, we directly ablate
28 the great circle based learning from the final models for both unconditional and conditional settings on CIFAR10. The
29 results worsen significantly on both unconditional settings (IS: 8.47 -> 8.57, FID: 12.9 -> 32.7) and conditional settings
30 (IS: 9.22 -> 8.41, FID: 5.83 -> 19.0). These results validate utilizing the great circle for training GANs.

31 **[R3] Additional computational cost.** For each iteration of the discriminator update, both center and pivotal vectors
32 are jointly updated using a control dependency. Thus, the increase in computational cost over the baseline is negligible.

33 **[R2, R4] Learning dynamics and Imagenet experiments.** In supplementary materials (L27-L49), we provide the
34 experiments on a large-scale Imagenet dataset with 128×128 resolution along with the learning curve of IS and FID.
35 CircleGAN outperforms recent methods [4, 18, 29] by a large margin using significantly fewer iterations. We expect the
36 stability and qualitative results to further improve by increasing the batch size as do other state-of-the-art methods.