We thank all reviewers for acknowledging the novelty and contributions of our work. We thank all reviewers for their constructive comments for improving our paper. In this rebuttal, a) We improved our generative tasks experiments including comparing IRMAE to modern AEs and comparing them with varying latent dimensions. b) We want to emphasize the importance of the superior performance of our model on semi-supervised classification tasks. This shows an advantage of our approach on representation learning for downstream tasks which was considered difficult for AEs. As this is the first work of applying implicit regularization method, there could be many follow-up questions to explore.

R1: We added an experiment of our model using different initial variance settings. See Table 1 below. It's interesting that the regularization effect varies corresponding to the initial condition. We will study this effect in our future work.

The ablation study in the appendix is to test whether tying linear matrices can help reduce the number of parameters, which however results in worse performance. This shows the importance of having redundant degree of freedom for the implicit regularization dynamics to function.

R2: Thanks for the experiment suggestions. We first added a comparison of our model to several modern AEs on CelebA. See Table 2 below. Our model outperforms strong baselines such as WAE [1] and RAE [2]. We agree that AEs perform differently with varying latent dimension. We compare IRMAE with AE with different latent dimension settings in Table 4 and Figure 1 below. IRMAE outperforms AEs with optimal dimensionality.

We will reorganize the content as suggested. We will discuss deep linear generators papers in the related works.

R3: We added a new experiment of comparing our method against bottleneck AE in Table 4 and Figure 1 below. This justifies our method over explicit low dimensional setting. We want to emphasize that using an explicitly selected latent dimension requires prior knowledge. Our method, like many other regularization methods, does not guarantee finding optimal latent dimension but reduces the effort of manually searching or requirement of prior knowledge.

The purpose of this work is to propose a genetic representation learning method instead of specific state-of-the-art feature e.g. disentanglement by beta-VAE. Applying our method over these models will remain our future work.

Regarding L.143, ablation study: We fix the weight during training. This proves that the regularization effect comes from the gradient descent dynamics instead of just the architecture.

We claim our method can have a stronger regularization effect by adding more linear layers. It does not guarantee theoretical minimum rank. The number of linear layers is a hyperparameter that needs to be optimized. We admit we lack enough experiments comparing the effect of different depths. Therefore, we added the experiment in Table 3 below.

The PCA experiment proves that IRMAE learns a dense latent space and solves the problem that naive deterministic AEs have holes in their latent space.

R4: Regarding L.76-77, L.107-109, we agree that it's inappropriate to claim a superior performance related to smaller intrinsic latent dimensions. VAE tends to use the entire prior latent space, while IRMAE, on the other hand, tends to use smaller latent dimensions due to the regularization effect. It is possible that VAE with a proper selected latent dimension can achieve better results. IRMAE and VAE have quite different mechanisms. And this is an quite interesting phenomena of our approach compared to existing literature. Nonetheless, we believe a simple idea of inserting new layers to achieve comparable results as widely-used VAE is a sufficient contribution.

Regarding L.131-132, IRMAE significantly outperforms VAE on low-data semi-supervised settings. These types of tasks are important as AEs are usually considered less competitive in representation learning for downstream tasks [3].

We admit that we lack the comparison of different number of linear layers. Hence, we added a experiment in Table 3.

Table 1: Effect of different initial variance of linear matrices. MNIST.

Variance	1x	2x	4x
Latent Rank	8	43	66
FID	37.4	33.8	49.0

Table 2: Effect of different number of linear layers. MNIST.

Depth (l)	2	4	8	12
Latent Rank	70	39	8	4
FID	44.0	30.1	37.4	62.6

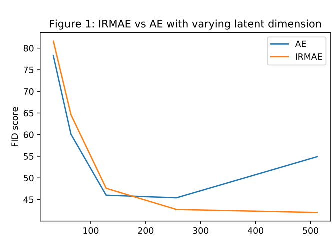

Table 3: IRMAE vs modern AEs. FID on CelebA.

WAE [1]	53.7
RAE [2]	44.7
IRMAE	**42.0**

Table 4: IRMAE vs AE with different latent dimension. FID on CelebA.

Latent dimension	32	64	128	256	512
IRMAE (l=4)	81.6	64.6	47.6	42.7	42.0
AE	78.2	60.1	46.0	45.4	53.9

[1] "Wasserstein Auto-Encoders", I. Tolstikhin et al. ICLR 2018
[2] "From Variational To Deterministic Autoencoders" P. Ghosh et al. ICLR 2020
[3] "Large Scale Adversarial Representation Learning" J. Donahue et al. NeurIPS 2019