AS-IntroVAE: Adversarial Similarity Distance Makes Robust IntroVAE Supplementary Material

Changjie Lu LUCHA@KEAN.EDU
Wenzhou-Kean University, Wenzhou, China

Shen Zheng SHENZHEN@ANDREW.CMU.EDU
Carnegie Mellon University, Pittsburgh, USA
Wenzhou-Kean University, Wenzhou, China

Zirui Wang LX1936@126.COM
Zhejiang University, Hangzhou, China

Omar Dib ODIB@KEAN.EDU
Wenzhou-Kean University, Wenzhou, China

Gaurav Gupta GGUPTA@KEAN.EDU
Wenzhou-Kean University, Wenzhou, China

0.1. Introduction

In this supplementary material, we first show experiment results with different weights for Adversarial Similarity Distance (AS-Distance) and KL Divergence, and then proceeds to more visual comparisons on image generation tasks on various benchmark dataset.

0.2. AS-Distance and KL Divergence

In this section, we use a visual comparison of image generation and image reconstruction tasks to show that the following hyperparameter combinations are worse than the weight annealing method introduced in the paper. The hyperparameter combinations are (1) AS-IntroVAE with a weight of 1.0 for AS-Distance and 0 for KL Divergence and (2) AS-IntroVAE with a weight of 0.5 for both AS-Distance and KL Divergence.

Figure 1: AS-IntroVAE performance at CelebA-128, using only AS-Distance and no KL divergence. The upper/middle/bottom two row refer to real/reconstructed/generated images. We can see that the images are over-smoothed and looks blurry without the help of KL divergence.

Figure 2: S-IntroVAE performance at CelebA-128, when the weight for KL divergence and AS-Distance are both 0.5. The upper/middle/bottom two rows refer to real/reconstructed/generated images. We can see that the images are with significant blur.

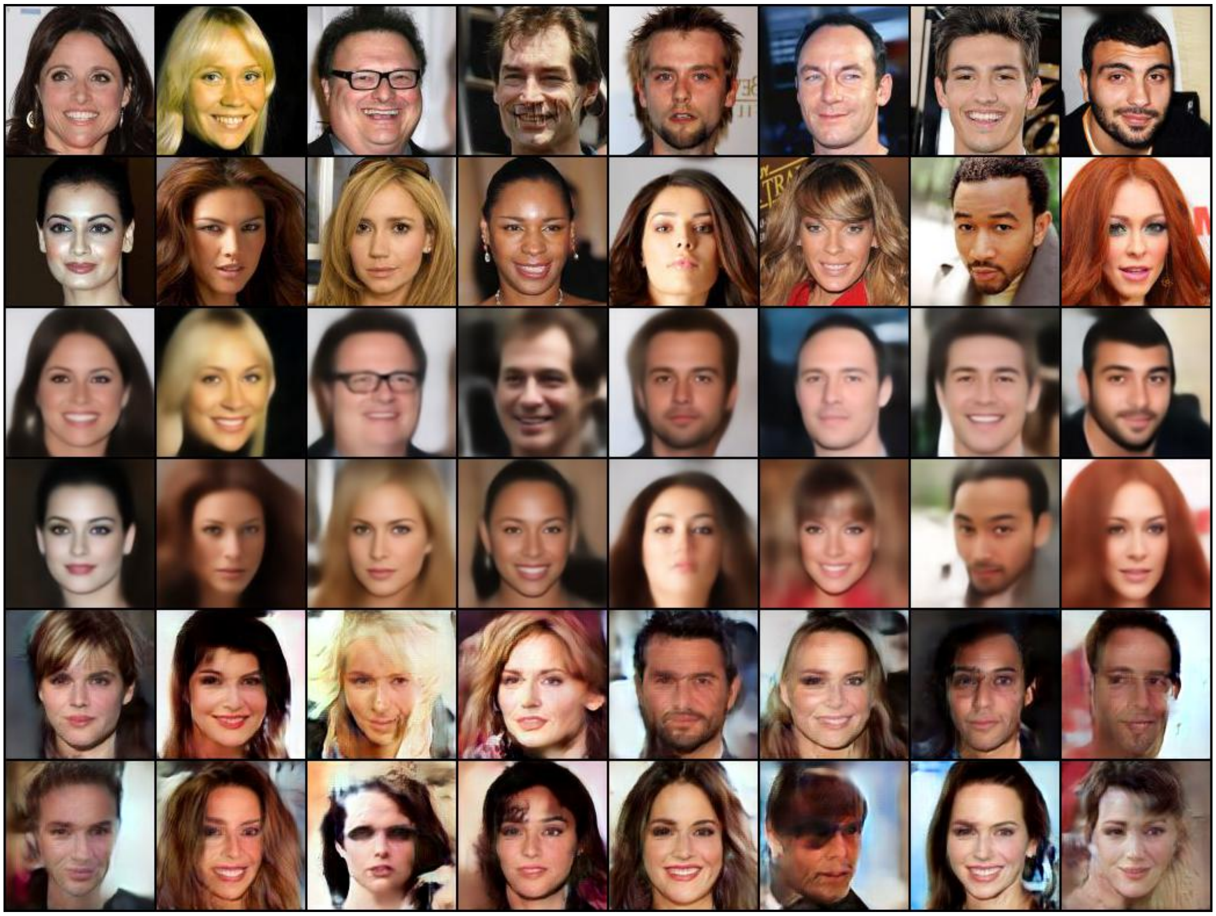

Figure 3: AS-IntroVAE performance at CelebA-128, when the weight for KL divergence and AS-Distance are both 0.5. The upper/middle/bottom two rows refer to real/reconstructed/generated images. From this figure and the figure above, we note that different images display different levels of sharpness and blur. Therefore, we conclude that this hyperparameter combination causes the model to have unstable training and fluctuating performances.

1. Visual Comparison for Image Generation

This section shows the additional visual comparison for image generation tasks. Specifically, we display the results on four datasets, including CelebA-128, CelebA-256, MNIST, and CIFAR10. For each dataset, we randomly select 16 images from each model's output dataset. In each figure, the upper left images are from AS-IntroVAE, the upper right images are from S-IntroVAE, and the bottom images are from WGAN-GP. Note that

Figure 4: Image generation visual comparisons at CelebA-128 dataset (resolution: 128×128).

Figure 5: Image generation visual comparisons at CelebA-256 dataset (resolution: 256×256).

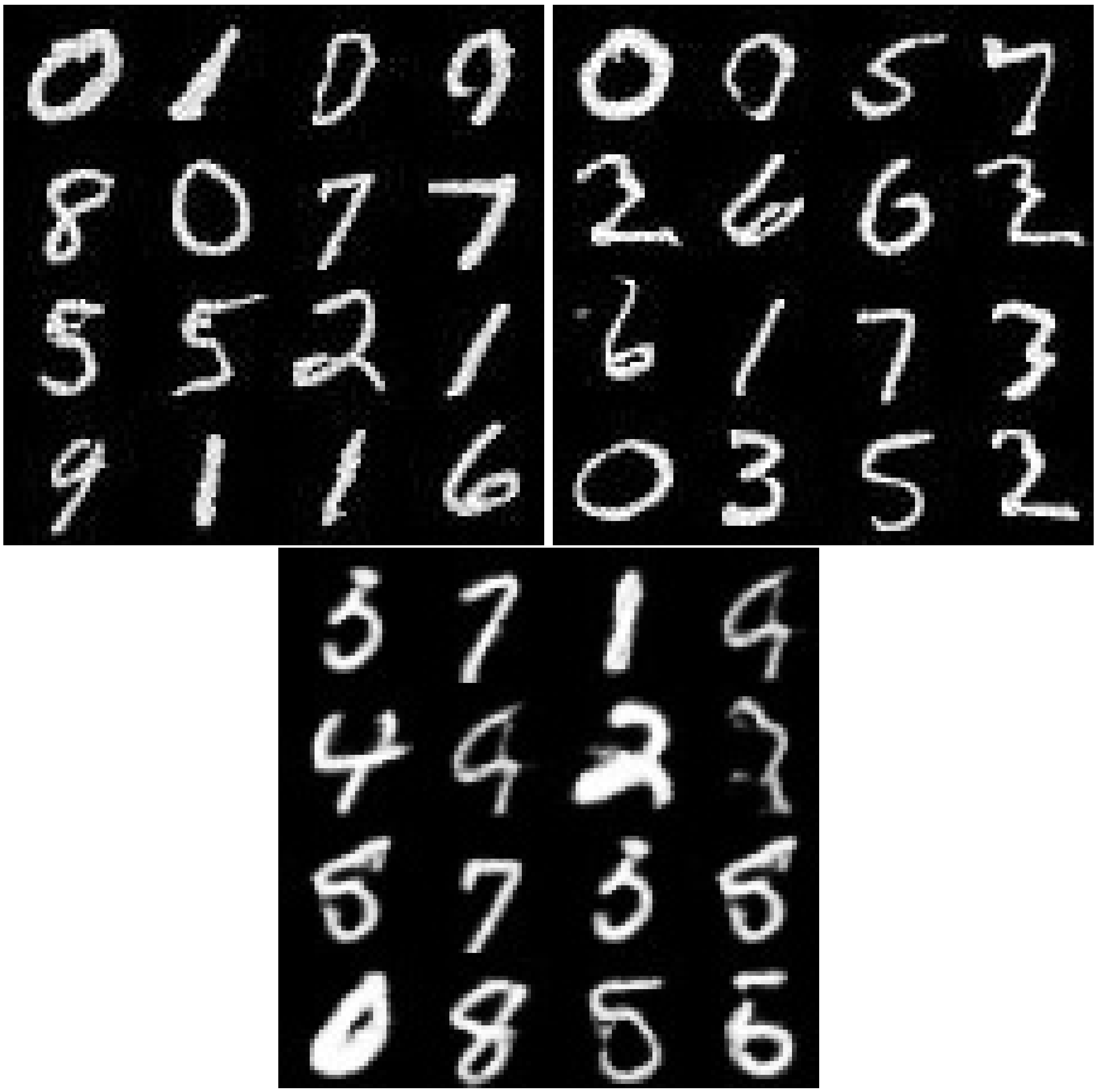

Figure 6: Image generation visual comparisons at MNIST dataset (resolution: 28×28).

Figure 7: Image generation visual comparisons at CIFAR10 dataset (resolution: 32×32)

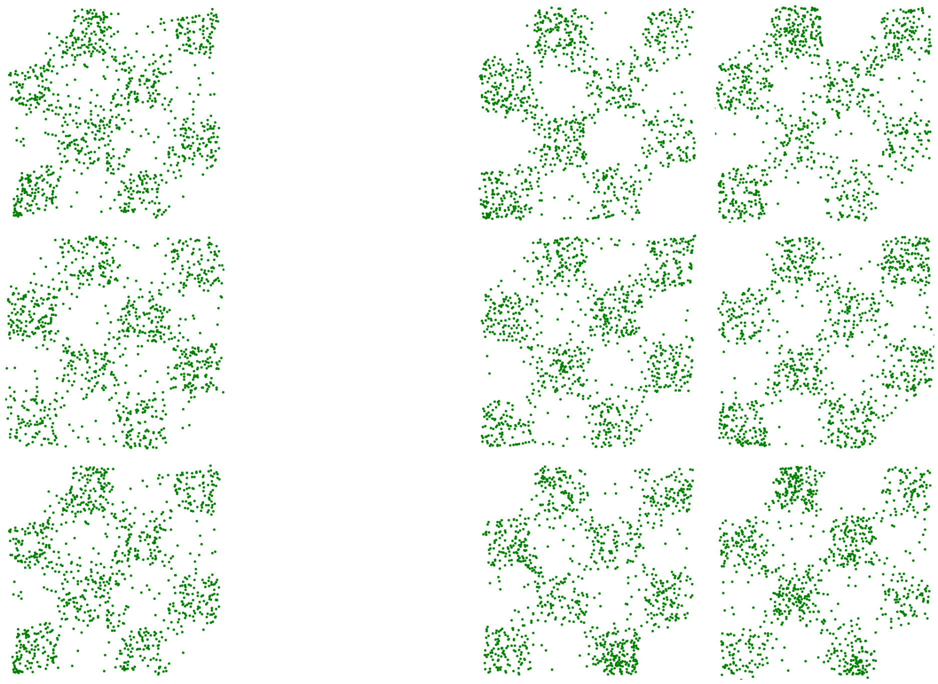

Figure 8: Visual Comparison on 2D Toy Dataset Checkerboard. From top to bottom row: results with different hyperparameters. From left to right column: VAE, IntroVAE, S-IntroVAE, Ours. The results show that AS-IntroVAE has a slight advantage over S-IntroVAE in terms of point clustering and centroid convergence.

		VAE	IntroVAE	S-IntroVAE	Ours
C1	KL	22.1	NaN	20.7	**20.4**
	JSD	10.8	–	**9.6**	**9.6**
C2	KL	21.2	NaN	21.0	**20.6**
	JSD	9.9	–	10.0	**9.6**
C3	KL	21.7	NaN	21.2	**20.9**
	JSD	10.7	–	10.3	**9.9**

Table 1: 2D Toy Dataset Checkerboard KL↓/JSD↓ Score Table. The Table shows that the proposed AS-IntroVAE has the best score for KL and JSD under all hyperparameter combinations.