Low-Rank Autoencoders: From Deterministic to Generative Shift

Alokendu Mazumder, Tirthajit Baruah, Bhartendu Kumar*, Rishab Sharma, Vishwajeet Pattanaik, Punit Rathore*

Indian Institute of Science — भारतीय विज्ञान संस्थान

RBCCPS Robert Bosch Centre for Cyber-Physical Systems

WACV 2024, JAN 4-8, WAIKOLOA HAWAII

Can a simple deterministic AE with low-rank latent space outperform a GAN in image generation?

Abstract & Motivation

- Autoencoders overlooks the fact that complex data (images) typically **resides in a lower-dimensional latent space**, which is essential for efficient representation.
- Our model uses a **low-rank regularizer** to learn an even **low-dimensional latent space**, while maintaining the autoencoder's fundamental goal.
- It is a simple autoencoder extension that learns low-rank latent space.

Architecture Overview

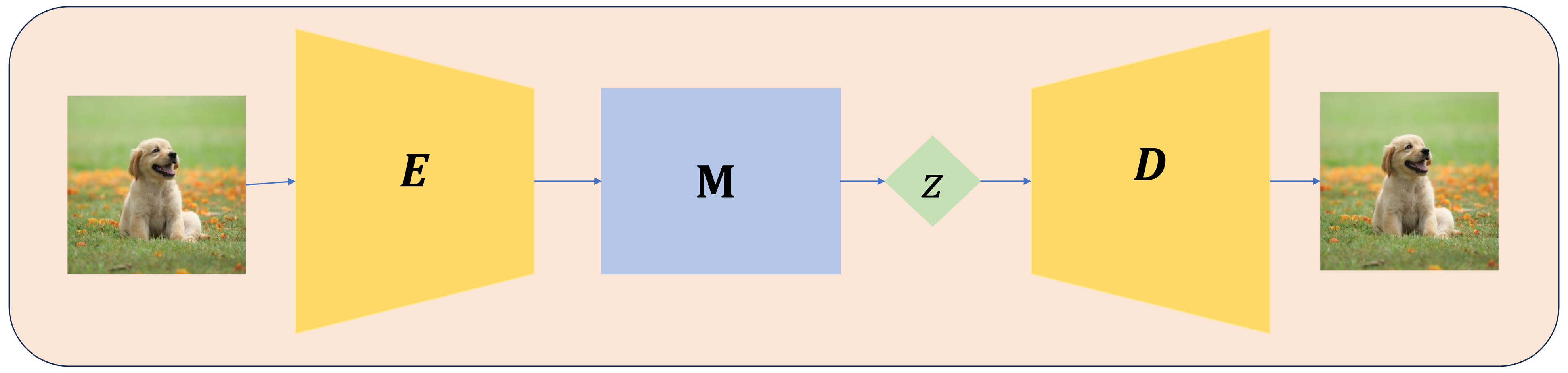

Fig 1: Low-Rank Autoencoder

- $\boldsymbol{E}: \mathbb{R}^{m \times n \times c} \to \mathbb{R}^{k}$ is encoder, $\boldsymbol{D}: \mathbb{R}^{k} \to \mathbb{R}^{m \times n \times c}$ is decoder and $\mathbf{M}$ is a linear layer sandwiched between encoder and decoder.
- The linear layer ($\mathbf{M}$) serves as a projection head into a **subspace with reduced rank.**
- **Loss Function:** $\boxed{\mathcal{L}(E, D, \mathbf{M}) = \| x - D\left(\mathbf{M}\left(E(x)\right)\right) \|_2^2 + \lambda . \| \mathbf{M} \|_*}$

Challenging Objectives Achieved

- Learns a better latent space with **generative capability** of simple AE.
- Generated images show **superior FID scores** compared to **VAE, GAN, and other SOTA AEs**.
- Outperforms baselines in downstream classification task.
- Proved the **convergence** of the learning algorithm under ADAM updates.
- Immune to **curse of dimensionality** effecting the latent space.

Experiments & Results

Interpolation & Dimensionality Reduction

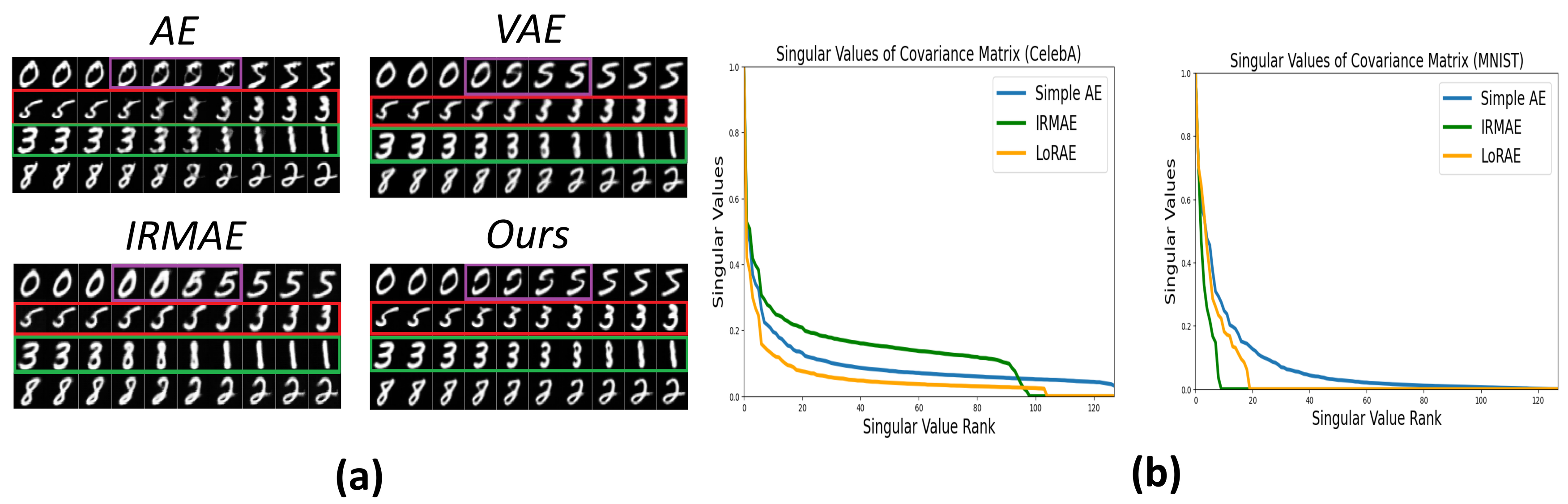

Fig 2: (a) Employing linear interpolation among data points on the MNIST dataset, (b) Singular value plot of the empirical covariance matrix of the latent space.

Image Generation

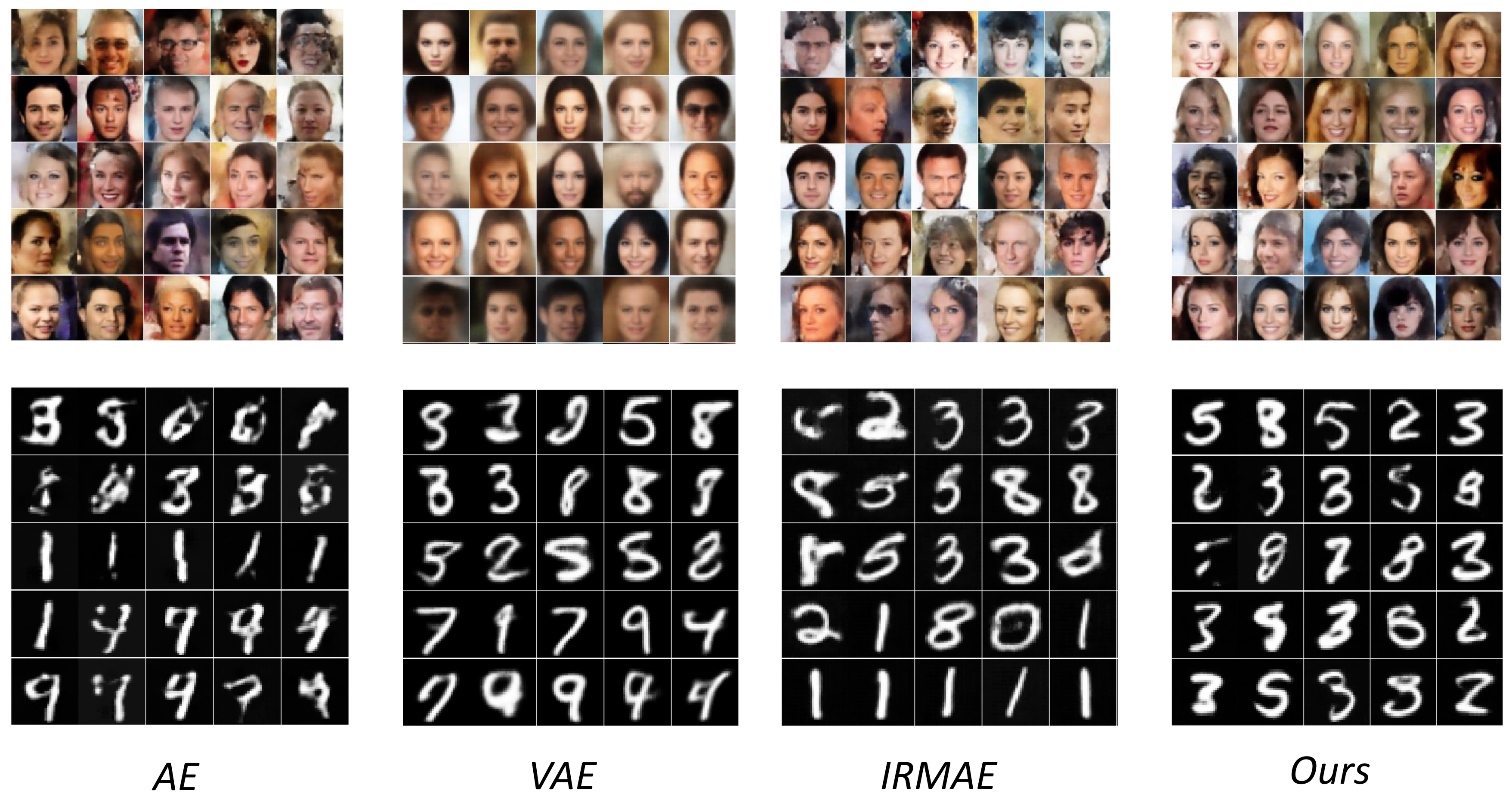

Fig 3: Images generated by AE, VAE, IRMAE and Our model on CelebA and MNIST

	MNIST		CelebA	
	$\mathcal{N}$	GMM	$\mathcal{N}$	GMM
AE	103.08	68.97	68.13	59.43
VAE	21.01	18.86	61.87	53.63
IRMAE	26.58	22.31	58.98	48.56
WAE	21.04	11.32	57.60	45.91
RAE	22.12	11.54	**50.31**	46.05
LoRAE	**19.50**	**11.09**	56.29	**45.43**

GAN	LS-GAN	NS-GAN	VAE + Flow	LoRAE
65.2	**54.1**	57.3	65.7	56.3

- Our model generates **high-quality images** as compared to other models.
- Our model achieves the **best FID score** among most of the **SOTA** models.
- Its success is due to **low-rank latent space** from **trace norm regularization**.

Downstream Classification

Size of training set	10	100	1000	10,000	60,000
AE	41.0	68.16	89.8	96.5	98.1
VAE	41.5	77.47	94.0	**98.5**	98.9
LoRAE	**46.6**	**89.02**	**95.4**	97.9	98.6
Supervised	37.8	73.59	94.2	98.3	**99.2**

- Our model beats AE and VAE in classification accuracy in low-data scenarios.

Table 1: Classification accuracy obtained from an AE, VAE and LoRAE on MNIST

Theoretical Analysis

Convergence of Learning Algorithm

Theorem 1: Let the loss function $\mathcal{L}(E, D, \mathbf{M})$ be K-Lipchitz and let $\| \mathcal{L} \|_2 \leq \gamma < \infty$. Then the following holds for deterministic ADAM iterates.

For any $\sigma > 0$, let $\alpha = \sqrt{2(\mathcal{L}(E_0, D_0, \mathbf{M}_0) - \mathcal{L}(E^*, D^*, \mathbf{M}^*))/K\delta^2 T}$, then there exists a natural number $T(\sigma, \delta)$ such that $\| \mathcal{L} \|_2 \leq \sigma$ for some $t \geq T(\sigma, \delta)$ where $\delta^2 = \frac{\gamma^2}{\epsilon^2}$

- When $T \geq \frac{2K\delta^2}{\sigma^4}\left(\mathcal{L}(E_0, D_0, \mathbf{M}_0) - \mathcal{L}(E^*, D^*, \mathbf{M}^*)\right) = T(\sigma, \delta)$, we have $\min_{0 \leq t \leq T-1} \| \mathcal{L}(E_t, D_t, \mathbf{M}_t) \|_2 \leq \sigma$.
- From above analysis, the rate of convergence is $\mathcal{O}\left(T^{-\frac{1}{4}}\right)$.

Lower Bound on Min-Max Distance Ratio

Theorem 2: Given any set of iid $x, x_1, x_2, \ldots, x_N \in \mathbb{R}^l$, we denote $d_{\max}^{E^*,\mathbf{M}^*} = \max_{0 \leq j \leq N} d^{E^*,\mathbf{M}^*}(x, x_j)$ and $d_{min}^{E^*,\mathbf{M}^*} = \min_{0 \leq j \leq N} d^{E^*,\mathbf{M}^*}(x, x_j)$, then we always have the conditional probability:

$$P\left(\frac{d_{\max}^{E^*,\mathbf{M}^*} - d_{min}^{E^*,\mathbf{M}^*}}{d_{min}^{E^*,\mathbf{M}^*}} \geq \Theta(\partial, \lambda) \mid \lambda > 0\right) = 1$$

where $d^{E^*,M^*}(x, x_j) = \frac{\mathbf{M}^*(E^*(x)) - \mathbf{M}^*\left(E^*(x_j)\right)\|_2}{rank(\mathbf{M}^*)}$, ∂ denotes the training dataset and $\Theta(\partial, \lambda)$ depends on training set and penalty λ.

- The nuclear norm regularization ensures a **minimum bound** for the **min-max distance ratio**.
- The lower bound on the min-max distance ratio enables **consistent data point discrimination**. Hence, making our model **immune to the curse of dimensionality**.

Rank of Latent Space

Proposition 1: The rank of latent space follows $\mathcal{O}(\lambda^{-1})$.

- Stronger regularization implies smaller rank of latent space.

PMRF Prime Minister's Research Fellowship

SERB INDIA

* denotes equal contribution