Received: 7 September 2023 | Revised: 14 December 2023 | Accepted: 23 January 2024

DOI: 10.4218/etrij.2023-0375

ORIGINAL ARTICLE

ETRI Journal WILEY

Generative autoencoder to prevent overregularization of variational autoencoder

YoungMin Ko[1,2] | SunWoo Ko[1,2] | YoungSoo Kim[1,2]

[1]Department of Artificial Intelligence, Jeonju University, Jeonju, Republic of Korea

[2]Artificial Intelligence Research Laboratory, Jeonju University, Jeonju, Republic of Korea

Correspondence
YoungSoo Kim, Department of Artificial Intelligence, Jeonju University, Republic of Korea.
Email: pineland@jj.ac.kr

Funding information
This study was supported by a research grant of Jeonju University in 2022.

Abstract

In machine learning, data scarcity is a common problem, and generative models have the potential to solve it. The variational autoencoder is a generative model that performs variational inference to estimate a low-dimensional posterior distribution given high-dimensional data. Specifically, it optimizes the evidence lower bound from regularization and reconstruction terms, but the two terms are imbalanced in general. If the reconstruction error is not sufficiently small to belong to the population, the generative model performance cannot be guaranteed. We propose a generative autoencoder (GAE) that uses an autoencoder to first minimize the reconstruction error and then estimate the distribution using latent vectors mapped onto a lower dimension through the encoder. We compare the Fréchet inception distances scores of the proposed GAE and nine other variational autoencoders on the MNIST, Fashion MNIST, CIFAR10, and SVHN datasets. The proposed GAE consistently outperforms the other methods on the MNIST (44.30), Fashion MNIST (196.34), and SVHN (77.53) datasets.

KEYWORDS

autoencoder, data augmentation, dimensionality reduction, generative model, variational autoencoder

1 | INTRODUCTION

Machine learning is intended to accurately predict a population using observations from that population for training [1–3]. However, data scarcity, which occurs in situations such as limited access to data, difficulty in data collection, and class imbalance, is common in the real world and imposes challenges to machine learning [4–7]. In particular, if data are difficult to observe and few samples are obtained, the performance of machine learning cannot be guaranteed. Generative models can handle data scarcity by synthesizing data belonging to a population [8, 9]. Specifically, generative models can improve the prediction performance of a machine learning model on a population by generating data from a source dataset, as illustrated in Figure 1.

Data can be generated by estimating the underlying distribution and then synthesizing data that follows the estimated distribution. However, when observed data are represented in a high-dimensional input space, such as video, image, sound, or natural language, a meaningful distribution is difficult to estimate from the observations. This is because as the data dimensionality increases, the number of samples required to estimate the distribution increases exponentially [10–12].

A variational autoencoder (VAE) [13] performs variational inference and uses an autoencoder to estimate a low-dimensional posterior distribution given

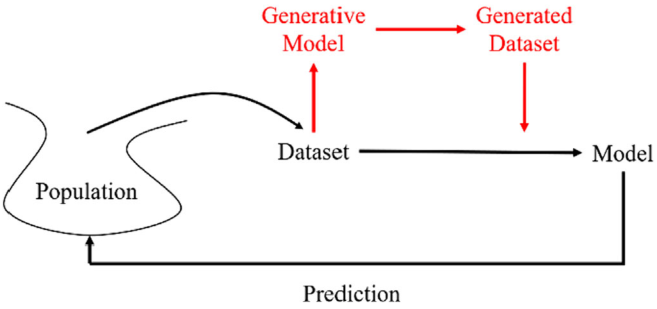


FIGURE 1 The generative model used to train the machine learning model.

high-dimensional data based on the manifold hypothesis [14–16]. The VAE is a generative model that estimates a distribution in a relatively dense low dimension instead of the original high dimension. In practice, a VAE implements the evidence lower bound (ELBO), which is the lower bound for variational inference, as the objective function [13, 17, 18]. The ELBO consists of a reconstruction term for recovering high-dimensional input data and a regularization term for regularizing the distribution of the latent space. However, during optimization of these two terms, imbalance problems occur [19–22]. In particular, overregularization can occur when the reconstruction term is not suitably minimized [23]. Other VAE-based models also have objective functions with reconstruction and regularization terms and are susceptible to imbalance. For instance, if the reconstruction term is not sufficiently small, the generation performance of the VAE cannot be guaranteed.

We propose a generative autoencoder (GAE) that first minimizes the reconstruction term using an autoencoder to ensure that this term suitably represents the population. It then estimates the distribution using latent vectors mapped onto a relatively dense low-dimensional space.

The main contributions of this study are summarized as follows:

- We ensure that the reconstruction term is sufficiently small to represent the population for the generative model to estimate low-dimensional latent distributions.
- We propose a GAE to prevent imbalance between the reconstruction and regularization terms. Experimental results confirm that the proposed GAE provides a better Fréchet inception distance (FID) than other VAE models.

The remainder of this paper is organized as follows. In Section 2, we describe various studies on VAE emphasizing reconstruction and regularization. Section 3 describes the situation in which the reconstruction term is not sufficiently minimized and the use of the proposed GAE to address this problem. In Section 4, VAE experiments are presented to illustrate overregularization, and the performance of the proposed GAE is compared with that of nine VAE models. Finally, the conclusions are presented in Section 5.

2 | RELATED WORK

2.1 | ELBO objective function for VAEs

The VAE proposed in Kingma and Welling [13] is intended for variational inference to estimate low-dimensional posterior distributions and uses the ELBO in (1) as the objective function because the posterior distribution is unknown and the latent vector of the posterior distribution is not directly observed [13, 17, 18]. The ELBO consists of regularization (left) and reconstruction (right) terms:

$$\min D_{KL}\left[q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}_i)\middle\|p(\boldsymbol{z})\right] - E_{\sim q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}_i)}[\log p_{\boldsymbol{\theta}}(\boldsymbol{x}_i|\boldsymbol{z})], \quad (1)$$

where $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}_i)$ denotes an encoder that represents the distribution of low-dimensional latent vector $\boldsymbol{z}$ given sample $\boldsymbol{x}_i$ in higher dimensions and $\boldsymbol{\phi}$ denotes the encoder parameters. The regularization term regulates the space of latent dimensions using the Kullback–Leibler divergence, which measures the difference between the distributions of $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}_i)$ and prior distribution $p(\boldsymbol{z})$. In the reconstruction term, $p_{\boldsymbol{\theta}}(\boldsymbol{x}_i|\boldsymbol{z})$ is the decoder that maps $\boldsymbol{z}$ onto $\boldsymbol{x}_i$ because of sampling from $q_{\boldsymbol{\phi}}(\boldsymbol{z}|\boldsymbol{x}_i)$, where $\boldsymbol{\theta}$ denotes the decoder parameters.

The VAE aims to generate high-quality data by estimating low-dimensional latent distributions. Ideally, both goals of the ELBO can be achieved with a sufficiently flexible neural network. However, this is difficult in practice owing to the finite capacity of such networks [24]. Numerous studies have been conducted to overcome various limitations, which can be described considering the objective function with reconstruction and regularization terms, as listed in Table 1.

2.2 | Other objective functions for VAEs

The conventional ELBO assumes that the prior distribution obeys a normal distribution [13]. This approach maps a wide variety of input data onto a single distribution, hindering the training of a neural network. VaDE [25], GMVAE [26], VampVAE [27], and other methods [28–30] mitigate the problems caused by simple prior distributions by using the more flexible Gaussian

TABLE 1 Comparison of methods using VAE-based objective functions.

Method	Objective function		Method to mitigate imbalance between reconstruction and regularization terms
	Reconstruction term	Regularization term	
VaDE [25]	Same	Flexible prior distribution (GMM)	Pretrained model with adjusted hyperparameters as weights for both terms
GMVAE [26]			Limit effect of regularization term below a threshold and adjust hyperparameters as weights for both terms
VampVAE [27]			Adjusted hyperparameters as weights for both terms
Other models [28–30]			
Beta-VAE [31]		Weights	
Info-VAE [24]		Maximum mean discrepancy	
WAE [32]		Weights	
Vector-quantized VAE [33]		Regularization term for discretized latent vector	
Hyperspherical VAE [34]		Von Mises–Fisher function	
Autoencoder + regularization [35]		Constraints such as L2-norm and gradient penalty	
VAE-GAN [36–38]	GAN objective function		

Abbreviations: GAN, generative adversarial network; VAE, variational autoencoder.

mixture model (GMM). Clusters from GMMs have been demonstrated to group similar data in an unsupervised manner.

Other studies have weighted the regularization term in different ways or introduced new objective functions. BetaVAE [31] induces weight β in the regularization term by adding a constraint such that the regularization term of the ELBO is less than a threshold. By adjusting β, the representation of the latent vector can be adjusted to better capture the data characteristics. InfoVAE [24] uses an efficient data reconstruction method by introducing a maximum mean discrepancy as the regularization term into the existing ELBO, thereby increasing the mutual information between the latent vector and input data. WAE [32] introduces a Wasserstein distance as the objective function for the reconstruction term, and the objective function has a lambda weight hyperparameter for regularization. VAE-generative adversarial network (GAN) [36–38] combines the objective functions of a GAN, vector-quantized VAE [33], and hyperspherical VAE [34], and it can regularize the shape of the latent space. On the other hand, different regularization terms were explored in Ghosh et al. [35], such as the L2-norm and gradient penalty, to regularize the latent vector. Under common distribution assumptions, learning a probabilistic encoder–decoder pair of VAEs is equivalent to learning a deterministic structure that adds noise to the decoder input.

The abovementioned studies consider regularization and reconstruction terms. To guarantee the performance of the generative model, the reconstruction term should be sufficiently small to describe the target population. However, an imbalance between the two terms occurs in practice during optimization, leading to an important problem for VAEs [19–23]. Remarkably, overregularization [23] prevents the reconstruction term to be sufficiently reduced owing to the large influence of the regularization term. This problem appears not only in VAEs but in many other methods [25, 26]. To mitigate overregularization, VaDE [25] uses pretrained stacked autoencoders as weights, whereas GMVAE [26] uses the method in Kingma et al. [22] to modify the objective function and thus limit the effect of the regularization term below a threshold. However, these approaches do not completely solve overregularization. Most studies on VAEs consider weights as hyperparameters that control the regularization-to-reconstruction term ratio [24–27, 31, 32]. However, if the model structure should be changed to fit a specific application or dataset, no appropriate practical guidelines are available. For a generative model to work, the reconstruction term must be sufficiently small to suitably describe a population.

3 | PROPOSED METHOD

Consider dataset $X=\{\boldsymbol{x}_i\}_{i=1}^{m}$ of m observed samples (independent and identically distributed random variables) in high-dimensional input space $\mathcal{X}$. Our objective is to estimate the distribution in a relatively dense low-dimensional latent space $\mathcal{Z}$ for X in high-dimensional space $\mathcal{X}$:

$$\begin{gathered}\min L\left(X,\tilde{X}\right)\\ \text{subject to}\, Z=W_e^{(L)}\cdots\left(\sigma_e^{(1)}\left(W_e^{(1)}X\right)\cdots\right)\equiv f_e(X),\\ \tilde{X}=W_d^{(L)}\cdots\left(\sigma_d^{(1)}\left(W_d^{(1)}Z\right)\cdots\right)\equiv f_d(Z).\end{gathered} \tag{2}$$

For estimation, we first use an autoencoder to minimize (2) such that reconstruction error $L\left(X,\tilde{X}\right)$ is less than some boundary value δ belonging to the population. In (2), $\tilde{X}$ is the data reconstructed by the autoencoder for input X. Specifically, encoder f_{e} and decoder f_{d} iteratively perform linear transformations $W_{\mathrm{e}}^{(l)}, W_{\mathrm{d}}^{(l)}, l=1,\ldots,L$ and apply activation functions $\sigma_{\mathrm{e}}^{(l)}, \sigma_{\mathrm{d}}^{(l)}, l=1, \ldots, L-1$. In addition, f_{e} maps input X onto Z in low-dimensional space $\mathcal{Z}$, and f_{d} maps Z onto $\tilde{X}$ in high-dimensional space $\mathcal{X}$.

Given parameters $W^* \in W_{\mathrm{e}}^{(l)*}, W_{\mathrm{d}}^{(l)*}$ for f_{e}^* and f_{d}^* satisfying $L\left(X, \tilde{X}\right)<\delta$, we assume that the autoencoder satisfies the following properties:

- Spaces $\mathcal{X}$ and $\mathcal{Z}$ are continuous if the activation function is continuous because every linear transformation is continuous.
- Given metric spaces $(\mathcal{X}, d_{\mathcal{X}})$ and $(\mathcal{Z}, d_{\mathcal{Z}})$, f_{d}^* satisfies a K-Lipschitz continuous function on arbitrary vectors $\boldsymbol{z}_i, \boldsymbol{z}_j \in \mathcal{Z}$ for a reasonable positive real number $K \geq 0$:

$$d_X\left(f_{\mathrm{d}}^*(\boldsymbol{z}_i), f_{\mathrm{d}}^*(\boldsymbol{z}_j)\right) \leq K d_Z\left(\boldsymbol{z}_i, \boldsymbol{z}_j\right). \tag{3}$$

This ensures that the values do not change considerably when the data sampled from the distribution estimated from $\mathcal{Z}$ are mapped onto $\mathcal{X}$.

Figure 2 illustrates an autoencoder that satisfies the abovementioned assumptions. When the image of a digit from a population is input into space $\mathcal{X}$, it is mapped back onto an image belonging to the population in space $\mathcal{X}$ via f_{e}^* and f_{d}^*.

We use Z in dense continuous low-dimensional space $\mathcal{Z}$ to estimate the distribution. Then, we apply maximum likelihood estimation to obtain a parameter by assuming a distribution. Given a certain probability distribution D (for example, normal distribution), the likelihood function of parameter $\boldsymbol{\vartheta}$ for $Z=\{\boldsymbol{z}_i\}_{i=1}^{m}$ is given by the following joint probability distribution:

$$\mathcal{L}(\boldsymbol{\vartheta})=D_{\boldsymbol{\vartheta}}(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_m). \tag{4}$$

When parameter $\widehat{\boldsymbol{\vartheta}}$ that maximizes the likelihood function in (4) is found, a new sample $\boldsymbol{z}_{\text{new}}$ can be obtained from estimated distribution $D_{\widehat{\boldsymbol{\vartheta}}}$ in (5). Then, using f_{d}^*, $\boldsymbol{x}_{\text{new}}$ is mapped from $\mathcal{X}$ onto the data belonging to the population described by (5), as illustrated in Figure 3. Sample $\boldsymbol{z}_{\text{new}}$ from distribution $D_{\widehat{\boldsymbol{\vartheta}}}$ estimated in continuous dense space $\mathcal{Z}$ is mapped onto $\boldsymbol{x}_{\text{new}}$ in high-dimensional space $\mathcal{X}$ as follows:

$$\boldsymbol{x}_{\text{new}}=f_{\mathrm{d}}^*(\boldsymbol{z}_{\text{new}}), \boldsymbol{z}_{\text{new}} \sim D_{\widehat{\boldsymbol{\vartheta}}}. \tag{5}$$

Hence, we can create N new samples using distribution $D_{\widehat{\boldsymbol{\vartheta}}}$ estimated in the lower-dimensional space and map them onto space $\mathcal{X}$ via f_{d}^*.

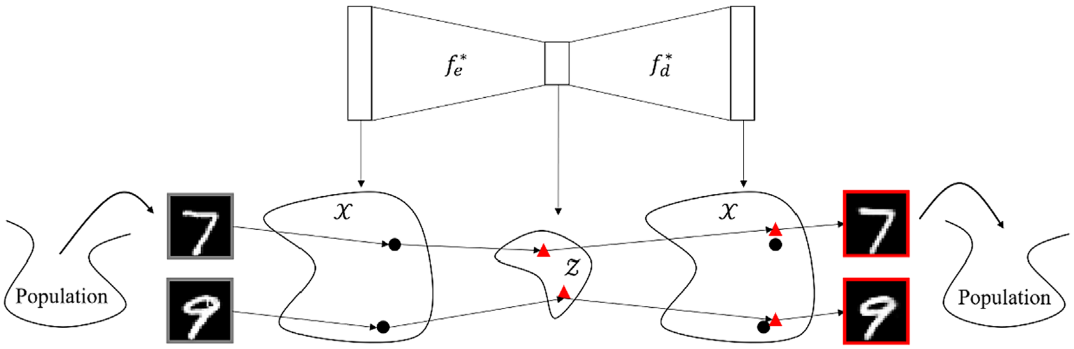


FIGURE 2 Autoencoder with suitably minimized reconstruction error to describe a population. Images of digits are given in high-dimensional space $\mathcal{X}$ and mapped onto lower-dimensional space $\mathcal{Z}$ by encoder f_{e}^*. The images reconstructed to high-dimensional space $\mathcal{X}$ by decoder f_{d}^* represent the population.

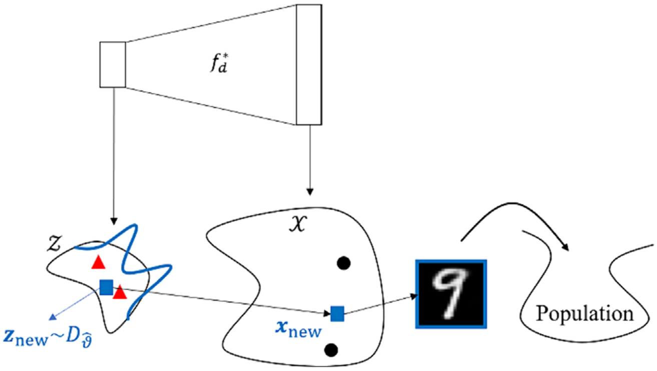


FIGURE 3 Data generation from estimated distribution $D_{\hat{\vartheta}}$. Sample z_{new} from $D_{\hat{\vartheta}}$ is mapped onto x_{new} belonging to the target population through decoder f_{d}^*.

The proposed GAE is described in Algorithm 1. In practice, boundary value δ should be determined by an expert with knowledge of the relevant data or based on experimental experience. The algorithm is used as a hyperparameter with a reasonably small value.

4 | EXPERIMENTS

We conducted experiments to evaluate the importance of the reconstruction error and generation performance of the proposed GAE.

ALGORITHM 1 GAE.

```
Set $X = \{x_i\}_{i=1}^m$ (observed data)
Set encoder $f_e$ and decoder $f_d$ for autoencoder
Set $W$ (parameters of autoencoder)
Set $L$ (reconstruction error)
Set boundary value $\delta$: 0.1 (hyperparameter)
While $L(X, \tilde{X}) \geq \delta$:
    Compute $Z = f_e(X)$
    Compute $\tilde{X} = f_d(Z)$
    Compute $L(X, \tilde{X})$
    Compute gradient $\frac{\partial L}{\partial W}$
    Update $W$ using learning rate $\rho$:
    $W = W - \rho \frac{\partial L}{\partial W}$
Obtain $f_e^*$ and $f_d^*$ satisfying $L(X, \tilde{X})$
Compute $Z = f_e^*(X)$
Set $D_\vartheta$ (specific distribution with parameter $\vartheta$)
Set $\mathcal{L}(\vartheta)$ (likelihood function for $Z$ given $D_\vartheta$)
Compute $\hat{\vartheta}$ that maximizes $\mathcal{L}(\vartheta)$
Sample $z_{new} \sim D_{\hat{\vartheta}}$
Compute $x_{new} = f_d^*(z_{new})$
```

4.1 | Evaluation of reconstruction error

We performed a simple visualization experiment to evaluate the VAE when the distribution was estimated without properly minimizing the reconstruction error. The dataset used in the experiment contained 50 equally spaced datapoints from a truncated cosine function (first column of Figure 4). The cosine function expressed in two dimensions was represented as a manifold, which is a one-dimensional curve, and a one-dimensional latent space was trained using a VAE and the proposed GAE with one hidden layer for the encoder and decoder, 256 nodes, and rectified linear unit (ReLU) activation. Both models considered the mean squared error for reconstruction. The prior distribution of the VAE was a normal distribution. The proposed GAE minimized the reconstruction error and estimated the parameters of a normal distribution in a one-dimensional latent space. Both models were tested in five trials over 2000 epochs per trial.

Table 2 lists the loss function values for the GAE and VAE as the mean and standard deviation across the five trials. The reconstruction errors show that the GAE outperforms the VAE. The reconstruction error is high for the VAE owing to overregularization. Figure 4 shows the reconstruction results of the 50 datapoints for the VAE and GAE after training as well as the data generated from the estimated distribution. The reconstructed data in Figure 4 show that the VAE suffers from overregularization because the regularization error converges to zero in the five trials, but the reconstruction error remains high. A VAE using the ELBO objective function cannot guarantee a sufficiently small reconstruction error because an adequate balance between the reconstruction and regularization terms cannot be achieved. Consequently, the VAE reconstructs data unrelated to the cosine function. On the other hand, the proposed GAE has a smaller reconstruction error than the VAE in the five trials. In trials 1, 3, and 4, the GAE reconstruction is representative of the population.

We then performed decoder reconstruction of 10 000 data samples from the estimated distribution in Figure 4. The VAE provides samples from a normal distribution and produces data that are completely unrelated to the cosine function in the five trials. This is because the VAE encoder suitably represents the normal distribution, but the decoder does not minimize the reconstruction error. In contrast, the proposed GAE perfectly generates the cosine function for trials 3 and 4 owing to the good reconstruction of the training data. Thus, the GAE reflects the properties of the population, and the corresponding low-dimensional latent space is correctly established. Nevertheless, in trials 1, 2, and 5, the data belonging to the population are not suitably generated

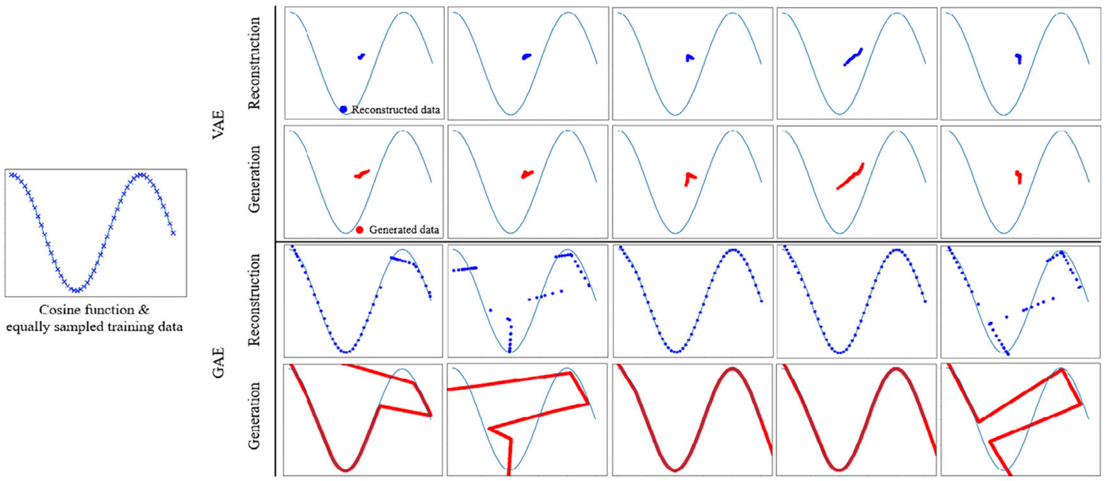


FIGURE 4 Experimental results of reconstruction and generation using variational autoencoder (VAE) and generative autoencoder (GAE). The first column shows the training data obtained from a cosine function for evaluation. Data were reconstructed using the VAE and GAE and generated from the estimated distribution, with columns 2–6 showing the results for the five trials.

TABLE 2 Loss function values for GAE and VAE expressed as mean ± standard deviation across five trials.

Method	Reconstruction error	Regularization error
GAE	0.0101 ± 0.0106	–
VAE	0.5131 ± 0.0041	0.0006 ± 0.0004

Abbreviations: GAE, generative autoencoder; VAE, variational autoencoder.

because the reconstruction error is large compared with the training data. This appears to be a common optimization problem for deep neural networks such as the VAE and GAE.

From the experiment, we observed that if the reconstruction error is not sufficiently small to describe the population, meaningless data are reconstructed from the estimated distribution.

4.2 | Evaluation of GAE data generation

To verify the data-generation capability of the proposed GAE, we compared it with nine different methods, including the VAEs mentioned in Section 2. Two main types of methods were evaluated, namely, (1) those that use the GMM as a prior distribution and (2) those that use modified regularization terms. From the first category, we evaluated VaDE [25], GMVAE [26], and VampVAE [27]. From the second category, we evaluated BetaVAE [31], InfoVAE [24], WAE [32], VAE-GAN [36], and RAE-L2 (regularized autoencoder using the L2-norm) [35]. The VAE in Kingma and Welling [13] was also evaluated.

The MNIST, Fashion MNIST, CIFAR10, and SVHN datasets were used for the evaluation experiments. The FID was used as the evaluation metric for the generative model. It was calculated from 10 000 generated datapoints per test dataset. The same model structure was used for all the datasets. Specifically, we adopted a fully connected autoencoder, and the encoder and decoder had three hidden layers, with 500, 500, and 2000 nodes in the encoder and 2000, 500, and 500 nodes in the decoder. The activation function was ReLU, and the latent dimension was 10 for the MNIST and Fashion MNIST datasets and 32 for the CIFAR10 and SVHN datasets. In all cases, 30 epochs were considered, each with minibatch size of 100. In addition, Adam optimization was applied, and the binary cross-entropy was used as the reconstruction error. A GMM was used to estimate the GAE distribution, and 10 clusters were used for the GMM of GMVAE, VaDE, VampVAE, and RAE-L2. To prevent overregularization, we used the hyperparameter settings suggested in the original studies to weigh the regularization terms. Nevertheless, overregularization occurred for InfoVAE and GMVAE, and thus, we experimentally determined the appropriate value for balancing the two terms.

Table 3 lists the obtained FIDs. The proposed GAE has the best FID score (a lower FID is better) on all three datasets, except for the CIFAR10 dataset. In particular, for the SVHN dataset, the difference between the FIDs of the VampVAE and GAE was considerable.

Figure 5 shows the reconstruction results of the models after randomly extracting 10 test images from the MNIST dataset to visualize the reconstruction performance. Compared with the original test images, the 10 evaluated models show reconstructions that are

TABLE 3 FIDs of evaluated models on target datasets.

Dataset	VAE	BetaVAE	InfoVAE	WAE	VAE-GAN	VaDE	GMVAE	VampVAE	RAE-L2	GAE (ours)
MNIST	123.44	51.53	116.67	89.79	366.39	47.58	106.92	82.59	55.81	**44.30**
Fashion MNIST	490.57	283.41	448.18	435.81	723.49	200.95	449.20	325.93	231.76	**196.34**
SVHN	123.32	106.33	176.99	117.09	268.01	100.89	474.25	1603.11	100.8	**77.53**
CIFAR10	493.47	**441.65**	601.74	443.27	1003.24	470.33	891.39	655.38	516.38	478.08

Note: The bolded values highlight the best results among the 10 models compared for the four datasets. Lower is better.
Abbreviations: FIDs, Fréchet inception distance; GAE, generative autoencoder; GAN, generative adversarial network; VAE, variational autoencoder.

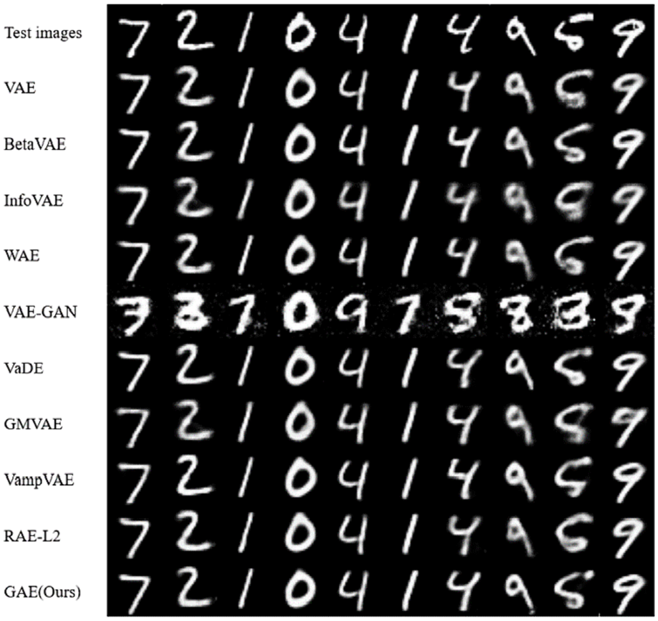


FIGURE 5 Reconstructed test images.

sufficiently good to belong to the population. However, for the reconstructed images of digit 5, the models provide relatively poor reconstructions compared with the other digit images. Notably, VAE-GAN reconstructs thick and noisy digits.

Figures 6 and 7 show 100 samples generated from the MNIST and Fashion MNIST datasets. VAE, BetaVAE, InfoVAE, WAE, and VAE-GAN generated data from a normal distribution, whereas VaDE, GMVAE, VampVAE, RAE-L2, and GAE were trained using GMM, and the estimated GMM was used for generation. In Figures 6 and 7, each row of images generated using VaDE, GMVAE, VampVAE, RAE-L2, and GAE represents 10 samples from each GMM cluster. In Figure 6, VAE and WAE show generated images largely outside of the population compared with reconstruction (Figure 5). The models that used the GMM to estimate the distribution generally produce similar images in each cluster, except for GMVAE (Figures 6 and 7). Overall, the

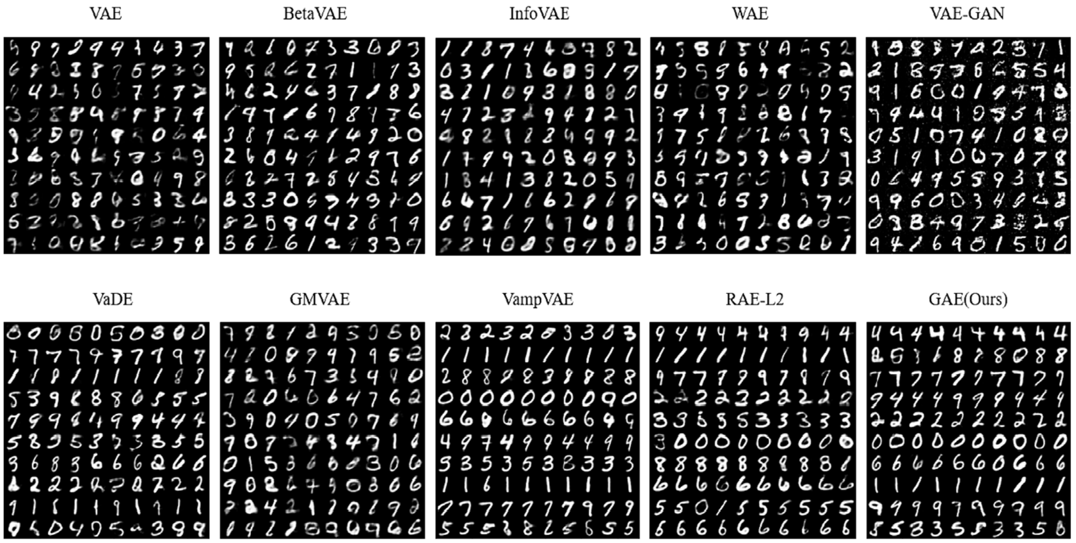


FIGURE 6 Data generated from evaluated models on MNIST dataset. The first row was generated from a normal distribution, and the second row was generated from a Gaussian mixture model (GMM). In the generated images of the second row, each line contains 10 images generated from the different GMM clusters.

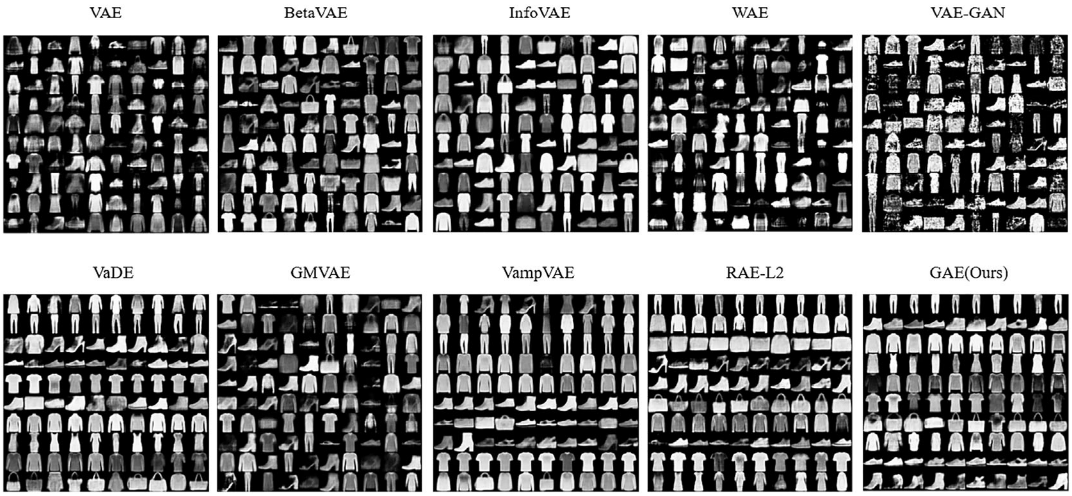


FIGURE 7 Data generated from evaluated models on Fashion MNIST dataset. The first row was generated from a normal distribution, and the second row was generated from a Gaussian mixture model (GMM). In the generated images of the second row, each line contains 10 images generated from the different GMM clusters.

TABLE 4 Total training time (in seconds) of evaluated models on target datasets.

Dataset	VAE	BetaVAE	InfoVAE	WAE	VAE-GAN	VaDE	GMVAE	VampVAE	RAE-L2	GAE (ours)
MNIST	135.04	132.99	147.85	138.65	1122.81	270.07	661.64	1154.87	133.22	**125.22**
Fashion MNIST	142.29	130.33	156.04	129.15	1343.71	284.89	690.54	1443.72	142.12	**125.75**
SVHN	224.71	**205.87**	476.00	411.16	1294.30	544.17	1649.37	1655.77	326.02	291.79
CIFAR10	143.58	118.20	281.78	204.76	1073.37	247.43	640.93	1025.30	163.43	**116.38**

Note: The bolded values highlight the best results among the 10 models compared for the four datasets. Lower is better.
Abbreviations: GAE, generative autoencoder; GAN, generative adversarial network; VAE, variational autoencoder.

generative power of GAEs is high owing to the minimization of the reconstruction error by setting an appropriately small dimensionality.

The total training time of the models in a computer equipped with an NVIDIA A100 graphics processor is listed in Table 4. The GAE required less training time than the other models. Hence, our method of simply minimizing the reconstruction error and then estimating the distribution requires less training time than similar methods.

5 | CONCLUSION

For a generative model, such as the VAE, the reconstruction error must be sufficiently small to suitably describe the population and guarantee a high generation performance. However, many studies on VAEs have shown reconstruction errors that are not sufficiently small owing to an imbalance in the objective function. We propose a GAE model that first sufficiently minimizes the reconstruction error and then maps it onto a low-dimensional latent vector via an encoder to estimate the data distribution. Nevertheless, various limitations require the rigorous verification of the theoretical assumptions. Specifically, we need a specific theory of reconstruction errors that are sufficiently small to represent a population. In addition, the properties that satisfy the K-Lipschitz continuous function should be determined. In future work, we will consider generative models that adhere to these two properties more rigorously.

ACKNOWLEDGMENTS

This study was supported by a research grant of Jeonju University in 2022.

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest.

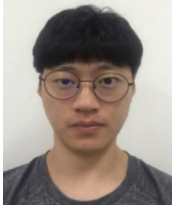
ORCID

YoungMin Ko https://orcid.org/0000-0003-2779-3170

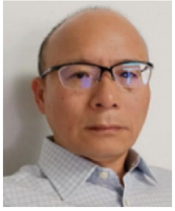
REFERENCES

1. C. M. Bishop, *Pattern recognition and machine learning*, 1st ed., Springer, New York, NY, USA, 2006.
2. I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*, MIT Press, Cambridge, MA, USA, 2016.
3. M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of machine learning*, MIT Press, Cambridge, MA, USA, 2018.
4. R. Babbar and B. Schölkopf, *Data scarcity, robustness and extreme multi-label classification*, Mach. Learn. **108** (2019), 1329–1351.
5. Y. T. Dile and R. Srinivasan, *Evaluation of CFSR climate data for hydrologic prediction in data-scarce watersheds: an application in the Blue Nile River Basin*, J. Am. Water Resour. Assoc. **50** (2014), 1226–1241.
6. R. Longadge and S. Dongre, *Class imbalance problem in data mining review*, arXiv Preprint, 2013, DOI 10.48550/arXiv.1305.1707
7. Z.-H. Zhou and X.-Y. Liu, *Training cost-sensitive neural networks with methods addressing the class imbalance problem*, IEEE Trans. Knowl. Data Eng. **18** (2005), 63–77.
8. I. Goodfellow, *NIPS 2016 tutorial: generative adversarial networks*, arXiv preprint, 2016, DOI 10.48550/arXiv.1701.00160
9. L. Ruthotto and E. Haber, *An introduction to deep generative modeling*, GAMM-Mitteilungen **44** (2021), e202100008.
10. G. Hughes, *On the mean accuracy of statistical pattern recognizers*, IEEE Trans. Inf. Theory **14** (1968), 55–63.
11. G. V. Trunk, *A problem of dimensionality: a simple example*, IEEE Trans. Pattern Anal. Mach. Intell. **3** (1979), 306–307.
12. G. J. McLachlan, *Discriminant analysis and statistical pattern recognition*, Wiley-Interscience, Hoboken, NJ, USA, 2004.
13. D. P. Kingma and M. Welling, *Auto-encoding variational Bayes*, arXiv Preprint, 2013, DOI 10.48550/arXiv.1312.6114
14. L. Cayton, *Algorithms for manifold learning, eScholarship*, University of California, 2008.
15. S. T. Roweis and K. S. Lawrence, *Nonlinear dimensionality reduction by locally linear embedding*, Science **290** (2000), 2323–2326.
16. E. Elhamifar and R. Vidal, *Sparse subspace clustering: algorithm, theory, and applications*, IEEE Trans. Pattern Anal. Mach. Intell. **35** (2013), 2765–2781.
17. C. Doersch, *Tutorial on variational autoencoders*, arXiv preprint, 2016, DOI 10.48550/arXiv.1606.05908
18. D. P. Kingma and M. Welling, *An introduction to variational autoencoders*, Vol. **12**, Found. Trends Mach. Learn, 2019, 307–392.
19. C. K. Sønderby, T. Raiko, L. Maaløe, S. K. Sønderby, and O. Winther, *Ladder variational autoencoders*, Adv. Neural Inf. Process Syst. **29** (2016).
20. X. Chen, D. P. Kingma, T. Salimans, Y. Duan, P. Dhariwal, J. Schulman, I. Sutskever, and P. Abbeel, *Variational lossy autoencoder*, arXiv Preprint, 2016, DOI 10.48550/arXiv.1611.02731
21. B. Dai and D. Wipf, *Diagnosing and enhancing VAE models*, arXiv Preprint, 2019, DOI 10.48550/arXiv.1903.05789
22. D. P. Kingma, T. Salimans, R. Jozefowicz, X. Chen, I. Sutskever, and M. Welling, *Improved variational inference with inverse autoregressive flow*, Adv. Neural Inf. Process Syst. **29** (2016).
23. B. Dai, L. Wenliang, and D. Wipf, *On the value of infinite gradients in variational autoencoder models*, Adv. Neural Inf. Process Syst. **34** (2021), 7180–7192.
24. S. Zhao, J. Song, and S. Ermon, *InfoVAE: balancing learning and inference in variational autoencoders*, (Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA), 2019, pp. 5885–5892.
25. Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, *Variational deep embedding: An unsupervised and generative approach to clustering*, (Proceedings of the 26th International Joint Conference on Artificial Intelligence, Melbourne, Australia), 2017, pp. 1965–1972.
26. N. Dilokthanakul, P. A. Mediano, M. Garnelo, M. C. Lee, H. Salimbeni, K. Arulkumaran, and M. Shanahan, *Deep unsupervised clustering with Gaussian mixture variational autoencoders*, ArXiv Preprint, 2016, DOI 10.48550/arXiv.1611.02648
27. J. Tomczak and M. Welling, *VAE with a VampPrior*, (Proceedings of the 21st International Conference on Artificial Intelligence and Statistics, Lanzarote, Spain), 2018, pp. 1214–1223.
28. D. B. Lee, D. Min, S. Lee, and S. J. Hwang, Meta-GMVAE: mixture of Gaussian VAE for unsupervised meta-learning, (International Conference on Learning Representations), 2021.
29. L. Tran, M. Pantic, and M. P. Deisenroth, *Cauchy-Schwarz regularized autoencoder*, J. Mach. Learn. Res. **23** (2022), 5010–5046.
30. C. Guo, J. Zhou, H. Chen, N. Ying, J. Zhang, and D. Zhou, *Variational autoencoder with optimizing Gaussian mixture model priors*, IEEE Access **8** (2020), 43992–44005.
31. I. Higgins, L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed, and A. Lerchner, *beta-VAE: learning basic visual concepts with a constrained variational framework*, (International Conference on Learning Representations), 2016.
32. I. Tolstikhin, O. Bousquet, S. Gelly, and B. Schoelkopf, *Wasserstein auto-encoders*, arXiv preprint, 2017, DOI 10.48550/arXiv.1711.01558
33. A. van Den Oord and O. Vinyals, *Neural discrete representation learning*, Adv. Neural Inf. Process Syst. **30** (2017).
34. T. R. Davidson, L. Falorsi, N. De Cao, T. Kipf, and J. M. Tomczak, *Hyperspherical variational auto-encoders*, arXiv preprint, 2018, DOI 10.48550/arXiv.1804.00891
35. P. Ghosh, M. S. Sajjadi, A. Vergari, M. Black, and B. Schölkopf, *From variational to deterministic autoencoders*, arXiv preprint, 2019, DOI 10.48550/arXiv.1903.12436
36. A. B. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, *Autoencoding beyond pixels using a learned similarity metric*, (International Conference on Machine Learning, New York, NY, USA), 2016, pp. 1558–1566.
37. R. Gao, X. Hou, J. Qin, J. Chen, L. Liu, F. Zhu, Z. Zhang, and L. Shao, *Zero-VAE-GAN: generating unseen features for generalized and transductive zero-shot learning*, IEEE Trans. Image Process. **29** (2020), 3665–3680.
38. Y. Xian, S. Sharma, B. Schiele, and Z. Akata, *F-VAEGAN-D2: A feature generating framework for any-shot learning*, (Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA), 2019, pp. 10275–10284.

AUTHOR BIOGRAPHIES



YoungMin Ko received his MS degree in engineering from the School of Artificial Intelligence, Jeonju University, Jeonju, Republic of Korea, in 2022. Currently, he is pursuing his PhD degree in artificial intelligence at Jeonju University, Jeonju, Republic of Korea. His main research interests are unsupervised learning and generative models in artificial intelligence.



SunWoo Ko received his MS degree in industrial engineering from the Korea Advanced Institute of Science and Technology, Republic of Korea, in 1988. He received his PhD degree in industrial engineering from the Korea Advanced Institute of Science and Technology, Republic of Korea, in 1992. Currently, he is a professor with the Department of Artificial Intelligence, Jeonju University, Republic of Korea. His main research interests are unsupervised learning and generative models in artificial intelligence.

YoungSoo Kim received his BS degree in computer science from the Republic of Korea Air Force Academy, Republic of Korea, in 1994; MS degree in computer science engineering from Sogang University, Republic of Korea, in 2001; and PhD degree in computer science engineering from the Korea Advanced Institute of Science and Technology, Republic of Korea, in 2009. He worked as a researcher in the Battlefield Informatization Lab of the Military Development Research Center, Korea National Defense Research Institute, Republic of Korea, until 2021. Since then, he has been an assistant professor with the Department of Artificial Intelligence, Jeonju University, Republic of Korea. His main research interests are artificial intelligence, Internet of Things, embedded systems, and edge/fog computing.

How to cite this article: Y. Ko, S. Ko, and Y. Kim, *Generative autoencoder to prevent overregularization of variational autoencoder*, ETRI Journal (2024), 1–10, DOI 10.4218/etrij.2023-0375