Journal of KIIT. Vol. 24, No. 8, pp. 1-10, Aug. 31, 2026. pISSN 1598-8619, eISSN 2093-7571
http://dx.doi.org/10.14801/jkiit.2026.24.8.1

저차원 VAE 잠재공간의 생성 성능 향상을 위한 FID 기반 VAE-GMM 모델 선택

고영민*[1], 이연승*[2], 고선우*[3]

Improving Generative Quality in Low-Dimensional VAE Latent Spaces via FID-Guided VAE-GMM Model Selection

Yeongmin Ko*[1], Yeonseung Lee*[2], and Sunwoo Ko*[3]

본 과제(결과물)은 2026년도 교육부 및 전북특별자치도의 재원으로 전북RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE) 의 결과임.(2026-RISE-13-JJU)
This research was supported by the Regional Innovation System & Education(RISE) program through the Jeonbuk RISE Center, funded by the Ministry of Education(MOE) and the Jeonbuk State, Republic of Korea.(2026-RISE-13-JJU)

요 약

변분 오토인코더(VAE)는 고차원 입력 데이터를 연속적인 저차원 잠재공간으로 표현할 수 있어 잠재공간 기반 생성에 널리 사용된다. 그러나 복원 오차나 ELBO(Evidence Lower Bound)의 개선이 곧 생성 품질 향상을 의미하지는 않는다. 본 논문은 각 epoch의 VAE 잠재평균 집합에 가우시안 혼합 모델(GMM)을 적합하여 VAE-GMM 생성기를 구성하고, 생성 샘플과 검증 데이터 집합 사이의 FID(Fréchet Inception Distance)를 이용해 검증 집합 FID가 가장 낮은 epoch를 선택하는 방법을 제안한다. MNIST, Fashion-MNIST, SVHN, CIFAR10 실험에서 제안 방법은 마지막 epoch까지 학습된 VAE-GMM 기준선보다 일관되게 낮은 시험 FID를 보였고, Fashion-MNIST와 SVHN에서는 비교한 VAE 계열 모델 중 가장 낮은 FID를 달성하였다.

Abstract

Variational Autoencoders (VAEs) are widely used for latent-space-based generation because they can represent high-dimensional input data in a continuous low-dimensional latent space. However, improvements in reconstruction error or the Evidence Lower Bound (ELBO) do not necessarily indicate improved generation quality. This paper proposes a method that constructs a VAE-GMM generator by fitting a Gaussian Mixture Model (GMM) to the set of VAE latent means at each epoch and selects the epoch with the lowest validation-set Fréchet Inception Distance (FID), measured between generated samples and the validation dataset. Experiments on MNIST, Fashion-MNIST, SVHN, and CIFAR-10 show that the proposed method consistently achieves lower test FID than the final-epoch VAE-GMM baseline. In particular, on Fashion-MNIST and SVHN, it achieves the lowest FID among the compared VAE-family models.

Keywords
variational autoencoder, generation quality, latent space validation, gaussian mixture model, fréchet inception distance

* 전주대학교 인공지능학과(*[3] 교신저자)
- ORCID[1]: https://orcid.org/0000-0003-2779-3170
- ORCID[2]: https://orcid.org/0009-0003-4507-7255
- ORCID[3]: https://orcid.org/0000-0002-6328-5440

· Received: May 14, 2026, Revised: Jun. 11, 2026, Accepted: Jun. 14, 2026
· Corresponding Author: Sunwoo Ko
Dept. of Artificial Intelligence, Jeonju University, Jeonju, Republic of Korea
303 Cheonjam-ro, Wansan-gu, Jeonju-si, Jeonbuk State 55069, South Korea
Tel.: +82-63-220-2303, Email: godfriend@jj.ac.kr

Ⅰ. 서 론

VAE(Variational Autoencoder)는 입력 데이터를 연속적인 잠재변수로 압축하고 다시 복원함으로써 데이터 분포를 학습하는 대표적인 비지도 생성모델이다[1][2]. VAE가 갖는 가장 큰 의미는 고차원 관측 데이터를 바로 다루기 어려운 경우, 그 구조를 보다 다루기 쉬운 저차원 연속 잠재공간으로 옮겨 샘플링, 보간, 군집화, 생성 등의 후속 작업을 가능하게 한다는 점이다[3]. 따라서 VAE는 단순한 차원축소 기법이 아니라 확률적 표현 학습기이자 잠재공간 기반 생성의 출발점으로 볼 수 있다.

그러나 생성 관점에서 적절하게 학습되었는지를 무엇으로 판단할 것인지는 여전히 쉽지 않다. 지도학습에서는 정확도나 검증 손실을 통해 일반화 성능을 직접 확인할 수 있지만, 비지도 생성모델에서는 복원 오차나 ELBO(Evidence Lower Bound)가 낮다고 해서 생성 샘플의 품질까지 좋아졌다고 단정할 수 없다[4]. 특히 학습 후반부에 모델이 학습 데이터의 복원 구조에 더 민감하게 적합되면, 잠재공간에서 무작위 샘플링하여 복원한 결과는 오히려 참인 분포와 멀어질 수 있다. 본 논문은 이러한 현상을 비지도 학습 관점의 잠재공간 과적합으로 해석한다. 즉, 모델이 학습 데이터에 대한 복원 오차는 최소화되지만, 생성 관점에서의 일반화 능력은 약화될 수 있다는 것이다.

최근 latent diffusion이나 flow matching과 같이 잠재표현을 기반으로 하는 생성 프레임워크가 주목받고 있다[5][6]. 또한 확산모델과 점수기반 생성모델 역시 생성모델 연구의 중요한 흐름을 형성하고 있다[7][8]. 이처럼 잠재표현을 전제로 하는 생성 프레임워크가 확산될수록, 잠재공간이 실제로 샘플링에 적합한지 검증하는 문제는 더욱 중요해지고 있다. 그러나 일반적인 VAE 학습 파이프라인에는 이를 직접 측정하는 절차가 없다.

이 문제를 다루기 위해 본 연구는 각 epoch에서 형성된 VAE 잠재공간을 별도로 평가하는 간단한 검증 절차를 제안한다. 구체적으로 encoder가 출력한 잠재평균 집합에 GMM(Gaussian Mixture Model)을 적합하여 VAE-GMM 생성기를 구성하고, 여기서 생성한 샘플과 검증 데이터 집합 사이의 FID (Fréchet Inception Distance)를 계산한다. 가장 낮은 검증 집합 FID를 기록한 epoch를 선택함으로써, 최종 epoch가 아니라 검증 FID가 가장 낮은 시점의 잠재공간을 사용할 수 있다.

본 논문의 기여는 다음과 같다. 첫째, 생성모델에서 복원 기준과 생성 일반화 기준이 어긋날 수 있음을 잠재공간 과적합 관점에서 정리한다. 둘째, 각 epoch의 VAE를 VAE-GMM 생성기로 변환하고 검증 집합 FID로 평가하는 단순한 모델 선택 절차인 VAE-GMM-FID를 제안한다. 셋째, MNIST, Fashion-MNIST, SVHN, CIFAR10 실험에서 제안 방법이 최종 epoch VAE-GMM 기준선보다 일관되게 낮은 FID를 달성함을 보인다.

Ⅱ. 관련 연구

2.1 VAE의 의미와 잠재공간 표현

VAE의 목적함수는 식 (1)에 나타나는 음의 ELBO를 최소화한다. encoder $q_\phi(\boldsymbol{z}|\boldsymbol{x})$와 decoder $p_\theta(\boldsymbol{x}|\boldsymbol{z})$로 구성되는 잠재변수 생성모델이다. 식 (1)의 왼쪽 항을 복원 오차, 오른쪽 항을 KL 발산항이라고 한다. KL 발산항에서 사전분포 $p(\boldsymbol{z})$는 일반적으로 표준다변량정규분포를 가정한다. 식 (2)와 같이 encoder는 입력 $\boldsymbol{x}$가 주어지면 잠재분포의 평균 벡터 $\mu_\phi(\boldsymbol{x})$와 분산 벡터 $\sigma_\phi(\boldsymbol{x})^2$를 출력하고 decoder는 식 (3)과 같이 재매개변수화로 샘플링된 잠재벡터 $\boldsymbol{z}$로부터 원본 데이터를 복원하는 데에 사용된다. VAE의 핵심은 단순한 autoencoder와 달리 잠재공간에 확률구조를 부여한다는 점이며, 이는 잠재공간에서의 샘플링과 생성 가능성을 확보한다.

$$E_{q_\phi(\boldsymbol{z}|\boldsymbol{x})}[-\log p_\theta(\boldsymbol{x}|\boldsymbol{z})] + D_{KL}(q_\phi(\boldsymbol{z}|\boldsymbol{x}) \parallel p(\boldsymbol{z})) \quad (1)$$

$$q_\phi(\boldsymbol{z}|\boldsymbol{x}) = N(\boldsymbol{z}; \mu_\phi(\boldsymbol{x}), diag(\sigma_\phi(\boldsymbol{x})^2)) \quad (2)$$

$$\boldsymbol{z} = \mu_\phi(\boldsymbol{x}) + \sigma_\phi(\boldsymbol{x}) \odot \boldsymbol{\epsilon},\ \boldsymbol{\epsilon} \sim N(\boldsymbol{0}, I) \quad (3)$$

식 (3)의 $\sigma_\phi(\boldsymbol{x})$는 잠재분포의 표준편차 벡터를 나타내고, $\boldsymbol{\epsilon}$은 각 차원이 독립인 표준정규분포를

따르는 잡음 벡터이며, ⊙는 원소별 곱을 의미한다.

VAE의 목적함수는 복원 오차와 KL 발산 정규화 항의 균형을 학습하지만, 실제 생성에 사용할 잠재공간의 품질을 직접적으로 측정하지 않는다. 따라서 낮은 복원 오차 또는 낮은 음의 ELBO는 복원 관점에서 성능이 우수한 VAE를 의미할 수는 있으나, 생성 샘플의 분포가 실제 데이터 분포와 유사하다는 것을 보장하지는 않는다.

2.2 비지도 학습에서의 과적합과 모델 선택

지도학습의 과적합은 대개 학습 데이터에 대한 정확도는 상승하지만 모집단을 대표한다고 가정한 시험 데이터에 대한 정확도는 하락하는 현상으로 설명된다. 비지도 생성모델에서도 유사한 현상을 생각할 수 있지만, 직접적인 클래스 정확도가 없기 때문에 이를 포착할 검증 기준이 필요하다. VAE의 경우 학습 후반부에 ELBO 또는 복원 기준은 계속 개선되더라도 잠재공간에서 생성한 샘플의 분포가 시험 데이터와 멀어질 수 있다.

따라서 비지도 생성모델의 모델 선택은 단순히 마지막 epoch를 사용하는 것이 아니라, 검증 데이터를 사용하여 잠재공간의 생성 성능에 대한 과적합을 측정하여 개선할 수 있다. 본 논문은 이 일반화 성능을 생성 샘플과 검증 데이터의 분포 유사도로 평가하며 이를 위해 검증 집합 FID를 도입한다. 이 접근은 복원 중심 학습과 생성 중심 선택을 분리한다는 점에서 실용적이다.

본 연구에서 말하는 잠재공간 과적합은 학습 손실 감소 자체를 의미하지 않는다. 이는 학습 데이터 복원에는 더 잘 맞아가지만, 해당 epoch의 잠재분포에서 샘플링한 생성 결과가 검증 데이터 분포와 멀어질 수 있는 현상을 의미한다. 따라서 검증 집합 FID는 잠재공간 자체의 수학적 최적성을 판정하는 기준이라기보다, decoder를 거친 최종 생성 샘플이 검증 데이터와 얼마나 유사한지를 관찰하는 경험적 생성 일반화 기준으로 해석할 수 있다.

2.3 VAE-GMM의 의미 해석

학습된 VAE로부터 실제 샘플을 생성하는 한 방법은 잠재벡터의 경험적 분포를 명시적 밀도모델로 근사하는 것이다[9]. 또한 잠재분포를 혼합 구조로 모델링하는 대표적 접근으로 VaDE와 GMVAE가 있다[10][11]. 본 논문에서의 VAE-GMM은 학습된 VAE의 encoder가 만든 잠재평균 벡터 집합을 GMM으로 근사한 뒤, 그 분포로부터 샘플링하여 decoder에 통과시키는 생성기를 의미한다. 즉, VAE가 잠재공간의 좌표계를 학습하는 역할을 한다면 VAE-GMM은 그 공간에서 실제 데이터가 밀집한 영역을 보다 직접적으로 반영하는 샘플러라고 해석할 수 있다. 이러한 해석에서 GMM은 새로운 학습 목표가 아니라 잠재공간을 점검하는 역할을 한다. 잠재공간이 생성에 적합한 구조로 형성되었다면 그 잠재평균 벡터 집합에 적합된 GMM으로부터 샘플링한 벡터를 decoder에 통과시켰을 때 실제 데이터 분포와 유사한 샘플이 생성되어야 한다. 반대로 그렇지 않은 상황이라면 잠재공간은 생성용으로 충분히 일반화되지 못한 것으로 해석할 수 있다.

기존의 VaDE와 GMVAE는 혼합 prior 또는 군집 구조를 모델 내부에 포함시켜 잠재분포와 군집을 함께 학습하는 접근이다. 반면 본 연구의 VAE-GMM-FID는 학습 목적함수나 prior 구조를 변경하지 않고, 각 epoch에서 이미 학습된 VAE의 잠재평균 집합을 GMM으로 근사한 뒤 검증 집합 FID를 이용해 생성 관점에서 가장 적합한 checkpoint를 선택하는 절차이다. 또한 BIC, AIC, log-likelihood와 같은 기준은 잠재공간에서의 밀도 적합도를 평가하지만, 본 연구의 기준은 GMM에서 샘플링한 잠재벡터를 decoder에 통과시켜 얻은 최종 생성 이미지가 검증 데이터 분포와 얼마나 가까운지를 평가한다. 따라서 본 연구의 차별성은 새로운 혼합 잠재모형의 제안이 아니라, 동일한 VAE 기반 생성 파이프라인에서 생성 일반화 기준에 따라 checkpoint를 선택하는 실용적 모델 선택 규칙을 제안한다는 점에 있다.

2.4 Fréchet Inception Distance

FID는 ImageNet으로 사전학습된 InceptionV3 네트워크의 풀링 레이어 출력을 특징 벡터로 사용한다. 대규모 데이터로 학습된 이 네트워크는 이미지의 의

미적 구조를 효과적으로 포착하며, 고차원 픽셀 공간에서 직접 분포 차이를 측정하는 것보다 안정적이고 의미 있는 평가를 제공한다. 실제 이미지와 생성 이미지의 특징 통계를 이용하면 FID는 다음과 같이 정의된다[12].

$$FID = \| \mu_r - \mu_g \|_2^2 + Tr(\Sigma_r + \Sigma_g - 2(\Sigma_r \Sigma_g)^{1/2}) \quad (4)$$

식 (4)는 실제 이미지 특징 분포와 생성 이미지 특징 분포 사이의 Frechet 거리를 계산하는 FID 정의식이다. 여기서 μ_r, Σ_r, μ_g, Σ_g는 각각 실제 이미지와 생성 이미지 특징의 평균 및 공분산을 의미한다. $Tr(\cdot)$는 trace 연산자이며, 마지막 항은 두 공분산 행렬 곱의 제곱근을 나타낸다. 따라서 FID 값이 낮을수록 생성 샘플의 특징 분포가 실제 데이터 분포에 더 가깝다는 의미이다. FID는 resize 방식이나 feature extraction 파이프라인에 민감할 수 있으므로 [13], 본 연구에서는 검증 집합과 시험 집합 평가에 동일한 전처리 및 계산 절차를 사용하였다. 또한 FID를 최종 정량지표이면서 동시에 모델 선택을 위한 검증 기준으로 사용하였다.

III. 제안 방법

제안 방법의 목적은 저차원 VAE 잠재공간의 생성 성능을 개선하는 것이다. 이를 위해 마지막 epoch를 고정적으로 사용하는 대신, 각 epoch의 잠재분포를 GMM으로 근사하여 생성 샘플을 만들고 검증 집합 FID가 가장 낮은 epoch의 모델을 선택한다(그림 1).

3.1 데이터 분할과 잠재평균 벡터 추출

전체 데이터셋은 학습, 검증, 시험 데이터 집합으로 분할된다. 검증 데이터 집합은 checkpoint 선택에만 사용하며 모델 파라미터 업데이트에는 사용하지 않는다. 각 t epoch에서 i 번째 학습 샘플에 대한 encoder가 출력하는 잠재평균 벡터를 식 (5)와 같이 표현할 수 있다.

$$z_i^{(t)} = \mu_\phi^{(t)}(x_i) \quad (5)$$

이후에 t epoch에서 GMM으로 분포를 추정할 때 잠재평균 벡터 $z_i^{(t)}$을 사용한다.

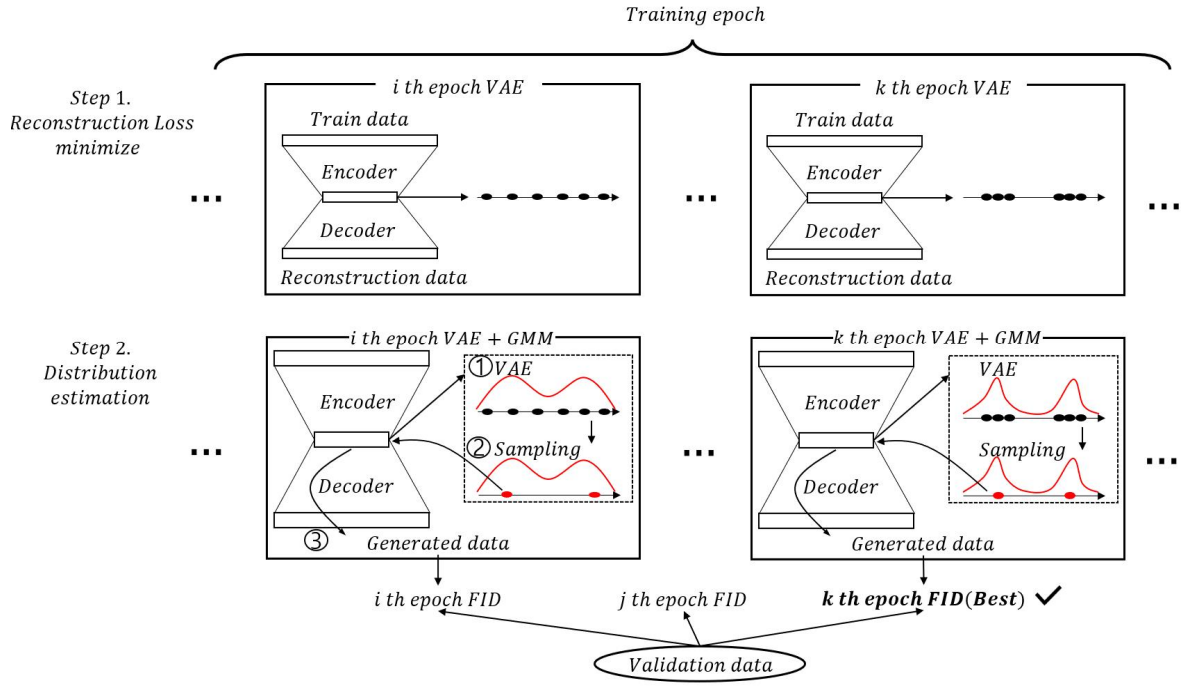


그림 1. FID 기반 VAE 검증 및 VAE-GMM 모델 선택 절차
Fig. 1. FID-based VAE validation and VAE-GMM model selection procedure

3.2 epoch 별 잠재분포의 GMM 근사

epoch t에서 학습 데이터로부터 얻은 전체 잠재 평균 벡터들을 다음과 같이 정의한다.

$$Z^{(t)} = \left\{ \boldsymbol{z}_i^{(t)} \right\}_{i=1}^{N_{train}} \tag{6}$$

여기서 N_{train}은 학습 데이터의 개수이며, $Z^{(t)}$는 epoch t 에서 형성된 잠재공간에 대한 경험적 표본 집합을 의미한다. 본 연구에서는 이 집합의 경험적 분포를 해당 epoch의 잠재분포에 대한 근사로 보고 GMM을 적합한다. epoch t에서 잠재분포 근사 모형 $q^{(t)}(\boldsymbol{z})$는 식 (7)과 같이 정의된다.

$$q^{(t)}(\boldsymbol{z}) = \sum_{k=1}^{K} \pi_k^{(t)} \mathcal{N}(\boldsymbol{z}; \boldsymbol{m}_k^{(t)}, \Sigma_k^{(t)}) \tag{7}$$

여기서 K는 혼합 분포 성분 수, $\pi_k^{(t)}$는 k번째 가우시안 성분의 혼합 가중치, $\boldsymbol{m}_k^{(t)}$는 평균 벡터, $\Sigma_k^{(t)}$는 공분산 행렬을 의미한다. 즉, epoch t에서 얻어진 잠재평균 벡터 집합 $Z^{(t)}$를 K개의 가우시안 성분으로 이루어진 혼합분포로 근사하는 것이다.

이와 같이 얻어진 $q^{(t)}(\boldsymbol{z})$는 epoch t에서 형성된 잠재공간의 밀도 구조를 반영하는 잠재 생성 분포로 해석할 수 있다. 각 Gaussian 성분은 잠재공간 내 데이터 군집에 대응하므로 성분별로 명시적 샘플링이 가능하여 데이터가 실제로 밀집한 영역을 직접 활용할 수 있다. 또한 성분별 평균과 분산을 통해 해당 epoch의 잠재공간에서 데이터가 어떤 구조로 분포하는지를 해석 가능한 형태로 확인할 수 있다.

3.3 FID 기반 샘플 생성과 모델 선택

GMM을 적합한 뒤 epoch별 혼합분포에서 잠재벡터를 샘플링하고, 이를 decoder에 통과시켜 생성 샘플을 얻는다. 그 다음 생성 샘플과 검증 집합 사이의 FID를 계산한다. 검증 집합 FID가 가장 낮은 epoch를 선택하고 해당 encoder, decoder, 적합된 GMM을 최종 모델로 사용한다. 즉, 마지막 epoch를 그대로 사용하는 대신 검증 집합 FID가 가장 낮은 checkpoint를 선택한다. GMM을 적합한 뒤 epoch별 혼합분포에서 잠재벡터를 샘플링하고, 이를 decoder에 통과시켜 생성 샘플을 얻는다. 그 다음 생성 샘플과 검증 집합 사이의 FID를 계산한다. 검증 집합 FID가 가장 낮은 epoch를 선택하고 해당 encoder, decoder, 적합된 GMM을 최종 모델로 사용한다. 즉, 마지막 epoch를 그대로 사용하는 대신 검증 집합 FID가 가장 낮은 checkpoint를 선택한다. 알고리즘 1은 제안한 FID 기반 VAE-GMM 모델 선택 절차를 단계별로 정리한 것이다. 먼저 전체 데이터셋을 학습 집합과 검증 집합으로 분할한 뒤, 각 epoch마다 VAE를 학습하고 학습 데이터의 encoder 평균벡터를 추출한다. 이후 epoch별 잠재벡터 집합에 K-성분 GMM을 적합하고, 적합된 GMM에서 샘플링한 잠재벡터를 decoder에 통과시켜 생성 샘플을 얻는다. 마지막으로 생성 샘플과 검증 집합 사이의 FID를 계산하여 현재 epoch의 검증 FID가 이전까지의 최솟값보다 낮으면 해당 VAE checkpoint와 GMM을 저장한다. 최종적으로 알고리즘 1은 검증 FID가 가장 낮은 checkpoint와 이에 대응하는 GMM을 최종 모델로 반환한다.

알고리즘 1. FID 기반 VAE-GMM 모델 선택 절차
Algorithm 1. FID-based VAE-GMM model selection procedure

Input: full dataset, maximum epoch, number of GMM components, number of generated samples
Output: selected VAE checkpoint and fitted GMM

1. Split the full dataset into training and validation sets.
2. Train the VAE for each epoch.
3. Extract encoder mean vectors from the training data at each epoch.
4. Fit a K-component GMM to the epoch-wise latent vectors.
5. Sample latent vectors from the fitted GMM and generate samples through the decoder.
6. Compute the FID between generated samples and the validation set.
7. If the current validation FID is the lowest so far, store the VAE checkpoint and the fitted GMM.
8. Return the selected checkpoint and fitted GMM as the final model.

Ⅳ. 실험 설정

4.1 데이터셋

실험은 MNIST, Fashion-MNIST, SVHN, CIFAR10의 네 가지 이미지 데이터셋에서 수행하였다. MNIST와 Fashion-MNIST는 28×28 회색조 이미지이며, SVHN과 CIFAR10은 32×32 색상 이미지이다. 각 데이터셋의 학습 세트 일부는 epoch별 모델 선택을 위한 검증 집합으로 분리하였다.

4.2 네트워크 구조 및 학습 설정

비교의 공정성을 위해 비교 모델들은 가능한 한 유사한 환경에서 학습하였다. encoder와 decoder는 각각 3개의 은닉층을 갖는 fully connected network로 구성하였고, encoder의 은닉층 노드 수는 500, 500, 2000개, decoder의 은닉층 노드 수는 2000, 500, 500개로 설정하였다. 활성화 함수는 ReLU를 사용하였다. 잠재차원은 데이터 복잡도를 고려하여 MNIST와 Fashion-MNIST에서는 10, SVHN과 CIFAR10에서는 32로 설정하였다. VAE 학습은 Adam(learning rate 1e-4), batch size 100, 100 epoch로 수행하였다. 잠재벡터의 경험적 분포를 근사하기 위해 모든 데이터셋에서 혼합 성분 수 K=10인 GMM을 사용하였으며, 이는 사용한 네 데이터셋이 모두 10개의 의미적 범주를 갖는다는 점을 반영한 실용적 설정이다. 다만 각 GMM 성분이 데이터 클래스와 일대일로 대응한다고 가정하지는 않는다. FID 계산을 위해 각 모델에서 10,000개의 생성 샘플을 사용하였다. 기준선 VAE-GMM은 학습 마지막 epoch의 잠재공간에 GMM을 적합한 뒤 생성 샘플을 얻는 방식이며, 제안한 VAE-GMM-FID는 동일한 구조를 사용하되 마지막 epoch 대신 검증 집합 FID가 가장 낮은 epoch를 선택한다. 추가적인 비교 모델로는 Beta-VAE, InfoVAE, WAE, VAE-GAN, VaDE, GMVAE, RAE-L2, VampVAE를 사용하였고, 가능한 한 동일한 encoder/decoder 구조, optimizer, batch size, epoch 수 아래에서 학습하였다. 따라서 본 비교는 각 모델의 최적 튜닝 성능 자체보다는 동일 실험 조건에서의 상대적인 생성 성능 비교로 해석하는 것이 적절하다.

4.3 평가 절차

FID는 검증 기준과 시험 평가지표로 모두 사용하되, 검증 집합은 epoch 선택에만 사용하고 최종 보고는 시험 집합 기준으로 수행하였다. 회색조 이미지는 InceptionV3 입력 요구사항에 맞추기 위해 단일 채널을 3회 복제하여 3채널 이미지로 변환하고, 모든 이미지는 75×75로 크기 조정한 뒤 특징을 추출하였다. InceptionV3는 include_top=False, pooling='avg' 설정을 사용하였다. 보고한 결과는 모두 5회 반복 실험의 평균과 표준편차로 제시하였다.

Ⅴ. 실험 결과 및 논의

5.1 정량적 비교

표 1은 비교 모델들의 시험 집합 FID 결과를 정리한 것이다. 제안한 VAE-GMM-FID는 네 개 데이터셋 모두에서 최종 epoch VAE-GMM 기준선보다 더 낮은 평균 FID를 기록하였다. 구체적으로 MNIST에서는 약 5.0%, Fashion MNIST에서는 약 12.6%, SVHN에서는 약 4.3%, CIFAR10에서는 약 8.0%의 FID 감소를 보였다. 이 결과는 VAE 기반 구조를 바꾸지 않고도 검증 기반 모델 선택만으로 생성 성능을 개선할 수 있음을 보여준다. Fashion-MNIST와 SVHN에서는 제안 방법이 비교한 VAE 계열 모델 중 가장 낮은 FID를 달성하였다. MNIST와 CIFAR10에서도 전체 비교 모델 중 최저는 아니었지만, 동일한 VAE 기반 생성기인 VAE-GMM 기준선과 비교하면 분명한 개선이 있었다. 이 결과는 제안 방법의 핵심 가치가 새로운 기반 구조를 설계하는 데 있지 않고, 이미 학습된 VAE를 어떻게 검증하고 선택할 것인가에 있음을 시사한다. 특히 본 방법은 encoder, decoder, prior 학습 목적함수를 새로 설계하지 않아도 된다. 즉, 기존 VAE 학습 파이프라인 뒤에 GMM 적합과 검증 FID 계산만 추가하면 되므로 구현 비용이 낮고, 현재 사용 중인 VAE 기반 생성 실험에 바로 부착할 수 있는 실용적 장점이 있다.

표 1. 비교 모델들의 시험 FID 결과 (5회 반복 실험의 평균 ± 표준편차)
Table 1. Test FID results of compared models (mean ± standard deviation over 5 runs)

Model	MNIST	FASHION MNIST	SVHN	CIFAR10
VAE	90.96 ± 3.07	369.95 ± 9.18	98.37 ± 1.72	450.13 ± 8.28
Beta-VAE	59.96 ± 2.82	277.93 ± 5.43	89.30 ± 1.12	445.62 ± 3.92
InfoVAE	90.52 ± 2.35	371.23 ± 8.98	100.77 ± 0.91	495.48 ± 2.99
WAE	73.87 ± 3.65	344.51 ± 11.40	93.58 ± 1.62	**402.12 ± 5.08**
VAE-GAN	70.18 ± 9.21	512.38 ± 11.37	273.68 ± 14.23	820.23 ± 3.28
VaDE	46.35 ± 6.89	206.48 ± 6.54	92.25 ± 1.28	482.34 ± 8.50
GMVAE	93.95 ± 6.41	433.40 ± 16.15	388.07 ± 12.81	733.24 ± 6.76
RAE-L2	**40.83 ± 4.89**	197.32 ± 6.28	116.30 ± 29.07	470.64 ± 29.32
VampVAE	85.65 ± 4.75	331.93 ± 13.11	1550.61 ± 27.34	660.12 ± 7.24
VAE-GMM	45.88 ± 6.03	209.52 ± 10.45	84.21 ± 1.01	504.20 ± 10.26
VAE-GMM-FID	43.57 ± 7.54	**183.16 ± 2.39**	**80.55 ± 2.05**	463.94 ± 7.68

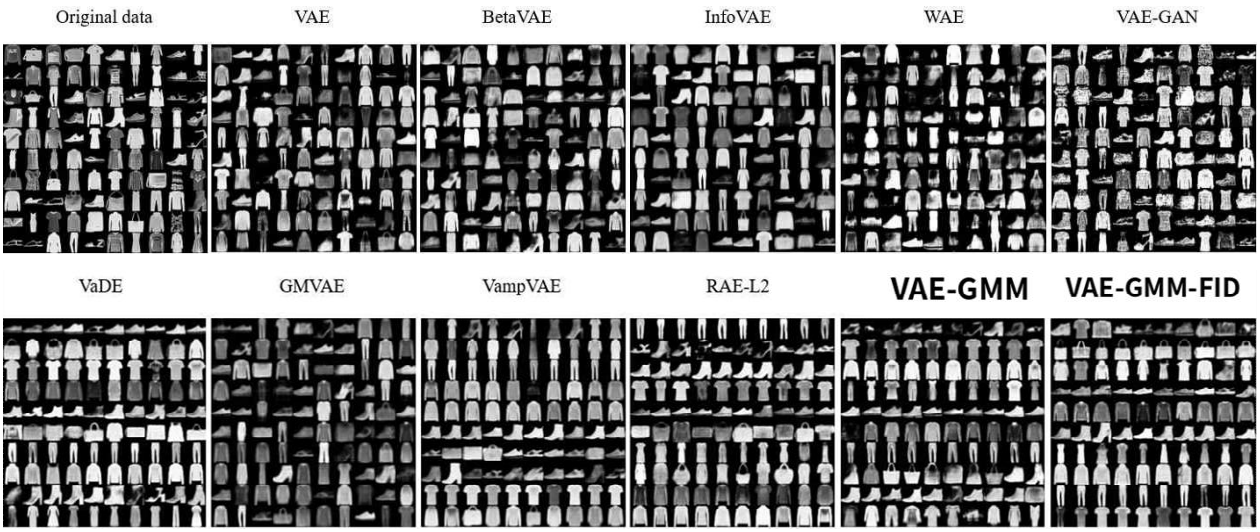


그림 2. Fashion-MNIST 샘플 비교
Fig. 2. Comparison of Fashion-MNIST samples

일부 비교 모델, 특히 VAE-GAN과 GMVAE가 높은 FID를 보인 것은 adversarial loss의 안정성 또는 혼합 잠재구조 학습이 공통 fully connected 설정에서 충분히 발현되지 않았기 때문일 수 있다. 따라서 본 결과는 동일 실험 조건에서의 상대 비교로 이해하는 것이 적절하다. 또한 본 논문은 현재 완료된 실험 범위까지만 결과를 정리하고, 보다 큰 규모의 확장 실험은 향후 과제로 분리하였다.

제안 방법의 성능 향상은 새로운 encoder, decoder, prior 또는 학습 목적함수에서 비롯된 것이 아니라, 동일한 학습 파이프라인에서 검증 집합 FID가 가장 낮은 epoch를 선택한 결과로 해석할 수 있다. 학습 후반부에는 복원 기준이 계속 개선되더라도 잠재분포에서 샘플링한 생성 결과는 검증 데이터와 멀어질 수 있으며, 검증 집합 FID 기반 선택은 이러한 불일치를 탐지한다. 따라서 동일한 VAE 학습 결과에서도 최종 epoch가 아니라 검증 집합 FID가 가장 낮은 checkpoint를 사용함으로써 더 낮은 시험 집합 FID를 달성할 수 있다.

본 연구의 목적은 최신 대규모 생성모델과의 직접적인 성능 경쟁보다, 동일한 VAE 기반 생성 파이프라인 내에서 검증 집합 FID가 가장 낮은 checkpoint를 선택하는 모델 선택 기준을 제안하는 데 있다.

5.2 정성적 결과

그림 2는 Fashion-MNIST에서의 정성적 생성 결

과를 보여준다. 제안한 VAE-GMM-FID는 그림 2에서 확인할 수 있듯이 최종 epoch VAE-GMM 기준선보다 더 일관된 형상과 범주적 구조를 보이는 샘플을 생성하는 경향을 보인다.

이는 검증 FID 기반 모델 선택이 수치상의 개선뿐 아니라 생성 샘플 분포가 검증 데이터 분포에 더 가까운 checkpoint를 선택할 수 있음을 시사한다.

그림 3은 서로 다른 epoch에서 생성된 샘플 예시를 보여준다. 복원 품질이 개선되더라도 생성 샘플의 시각적 품질이 단조롭게 향상되지 않음을 확인할 수 있다. 이 현상은 잠재공간 과적합의 직관적 예로 해석할 수 있다. 더 늦은 epoch은 학습 데이터 구조를 더 충실히 반영할 수 있지만, 샘플링된 잠재벡터의 다양성과 시각적 일관성은 저하될 수 있다. 검증 FID 기반 선택은 이러한 불일치를 검증 데이터 분포를 통해 탐지한다. 정성적으로도 제안 방법은 윤곽 붕괴, 이질 범주의 특징 혼합, 흐릿한 형상과 같은 현상을 줄이는 경향을 보였으며, 이는 선택된잠재공간이 최종 epoch보다 검증 집합FID 관점에서 생성에 더 적합함을 뒷받침한다.

그림 3. VAE 학습 과정의 서로 다른 epoch에서 생성된 샘플

Fig. 3. Examples of generated samples at different epochs during the VAE training process

5.3 비지도 학습 관점의 해석

VAE의 의미는 입력을 낮은 차원으로 압축하는 데만 있지 않다. VAE는 데이터 분포를 연속 잠재변수로 설명하고, 그 공간에서의 샘플링 가능성을 열어 주는 확률적 표현학습기이다. 따라서 생성 관점에서 적절한 VAE란 단지 복원 오차가 낮은 모델이 아니라, 잠재공간에서 샘플링한 결과가 검증 데이터 분포와 유사한 샘플을 생성하는 모델이어야 한다. VAE-GMM은 이러한 VAE 잠재공간을 실제 생성기로 해석하기 위한 연결고리이다. VAE가 잠재표현을 학습하고, GMM이 그 표현 위의 경험적 밀도를 근사함으로써, VAE-GMM은 학습된 잠재공간에서 어떤 분포를 기반으로 샘플링할 것인가에 대한 실용적인 샘플링 기준을 제공한다. 본 연구의 VAE-GMM-FID는 여러 epoch의 VAE-GMM 중에서 검증 FID가 가장 낮은 모델을 선택하는 절차로 해석될 수 있으며, 이는 비지도 학습에서의 과적합 완화 전략으로 이해할 수 있다.

Ⅵ. 결 론

본 논문에서는 저차원 VAE 잠재공간의 생성 성능 향상을 위해, epoch별 잠재평균 벡터 집합에 GMM을 적합하고 검증 FID가 가장 낮은 checkpoint를 선택하는 VAE-GMM-FID 방법을 제안하였다. 제안 방법은 기반 구조나 학습 목적함수를 바꾸지 않고도, 검증 FID 기준으로 더 낮은 값을 보인 잠재공간을 선택할 수 있다는 점에 의의가 있다.

MNIST, Fashion-MNIST, SVHN, CIFAR10 실험에서 제안한 VAE-GMM-FID는 최종 epoch VAE-GMM 기준선보다 일관되게 낮은 시험 집합 FID를 달성하였다. 특히 Fashion-MNIST와 SVHN에서는 비교한 VAE 계열 모델 중 가장 낮은 FID를 보였다. 이러한 결과는 검증 FID가 VAE 잠재공간의 생성 적합성을 평가하고, 검증 집합에 대해 더 낮은 FID를 달성하는 VAE-GMM을 선택하는 실용적 기준이 될 수 있음을 시사한다.

본 연구는 VAE를 확률적 잠재표현 학습기로, VAE-GMM을 그 표현 위의 경험적 밀도 생성기로, VAE-GMM-FID를 검증 기반 모델 선택 규칙으로 해석하였다. 현재 단계에서는 완료된 FID 중심 정량·정성 실험을 우선 보고하였으며, 향후에는 혼합성분 수 K, 잠재차원, 추가 생성지표(IS, Precision/Recall) 및 후속 잠재 생성기와의 결합을 포함한 확장 분석을 검토할 예정이다.

References

[1] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes", Proc. Int. Conf. Learn. Represent. (ICLR), Banff, Canada, Apr. 2014. https://doi.org/10.48550/arXiv.1312.6114.

[2] D. J. Rezende, S. Mohamed, and D. Wierstra, "Stochastic Backpropagation and Approximate Inference in Deep Generative Models", Proc. 31st Int. Conf. Mach. Learn. (ICML), Beijing, China, pp. 1278-1286, Jun. 2014. https://doi.org/10.5555/3044805.3045035.

[3] C. Doersch, "Tutorial on Variational Autoencoders", arXiv preprint, arXiv:1606.05908, Jun. 2016. https://doi.org/10.48550/arXiv.1606.05908.

[4] R. Wei, C. Garcia, A. El-Sayed, V. Peterson, and A. Mahmood, "Variations in Variational Autoencoders - A Comparative Evaluation", IEEE Access, Vol. 8, pp. 153651-153670, Aug. 2020. https://doi.org/10.1109/ACCESS.2020.3018151

[5] P. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-Resolution Image Synthesis With Latent Diffusion Models", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 10684-10695, Jun. 2022. https://doi.org/10.1109/CVPR52688.2022.01042.

[6] Y. Lipman, R. T. Q. Chen, H. Ben-Hamu, M. Nickel, and M. Le, "Flow Matching for Generative Modeling", Proc. Int. Conf. Learn. Represent. (ICLR), Kigali, Rwanda, May 2023. https://doi.org/10.48550/arXiv.2210.02747.

[7] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models", Adv. Neural Inf. Process. Syst., Vol. 33, pp. 6840-6851, Dec. 2020. https://doi.org/10.5555/3495724.3496298.

[8] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole, "Score-Based Generative Modeling through Stochastic Differential Equations", Proc. Int. Conf. Learn. Represent. (ICLR), Online, May 2021. https://doi.org/10.48550/arXiv.2011.13456.

[9] Y. Ko, S. Ko, and Y. Kim, "Generative autoencoder to prevent overregularization of variational autoencoder", ETRI Journal, Vol. 47, No. 1, pp. 80-89, Feb. 2025. https://doi.org/10.4218/etrij.2023-0375.

[10] Z. Jiang, Y. Zheng, H. Tan, B. Tang, and H. Zhou, "Variational Deep Embedding: An Unsupervised and Generative Approach to Clustering", Proc. 26th Int. Joint Conf. Artif. Intell. (IJCAI), Melbourne, Australia, pp. 1965-1972, Aug. 2017. https://doi.org/10.24963/ijcai.2017/273.

[11] N. Dilokthanakul, P. A. M. Mediano, M. Garnelo, M. C. H. Lee, H. Salimbeni, K. Arulkumaran, and M. Shanahan, "Deep Unsupervised Clustering with Gaussian Mixture Variational Autoencoders", arXiv preprint, arXiv:1611.02648, Nov. 2016. https://doi.org/10.48550/arXiv.1611.02648.

[12] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium", Adv. Neural Inf. Process. Syst., Vol. 30, pp. 6627-6638, Dec. 2017. https://doi.org/10.5555/3295222.3295408.

[13] G. Parmar, R. Zhang, and J.-Y. Zhu, "On Aliased Resizing and Surprising Subtleties in GAN Evaluation", Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, pp. 11410-11420, Jun. 2022. https://doi.org/10.1109/CVPR52688.2022.01112.

[14] I. Higgins, et al., "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework", Proc. Int. Conf. Learn. Represent. (ICLR), Toulon, France, Apr. 2017.

[15] S. Zhao, J. Song, and S. Ermon, "InfoVAE: Balancing Learning and Inference in Variational Autoencoders", Proc. AAAI Conf. Artif. Intell. (AAAI-19), Honolulu, HI, USA, pp. 5885-5892, Jan.-Feb. 2019. https://doi.org/10.1609/aaai.v33i01.33015885.

[16] I. Tolstikhin, et al., "Wasserstein Auto-Encoders", Proc. Int. Conf. Learn. Represent. (ICLR), Vancouver, BC, Canada, Apr.-May 2018. https://doi.org/10.48550/arXiv.1711.01558.

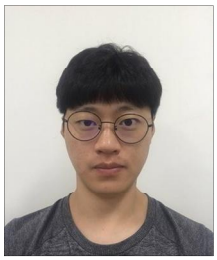
[17] A. B. L. Larsen, S. K. Sønderby, H. Larochelle, and O. Winther, "Autoencoding beyond pixels using a learned similarity metric", Proc. 33rd Int. Conf. Mach. Learn. (ICML), New York, NY, USA, pp. 1558-1566, Jun. 2016. https://doi.org/10.5555/3045390.3045555.

[18] J. M. Tomczak and M. Welling, "VAE with a VampPrior", Proc. 21st Int. Conf. Artif. Intell. Stat. (AISTATS), Playa Blanca, Lanzarote, Canary Islands, Spain, pp. 1214-1223, Apr. 2018. https://doi.org/10.48550/arXiv.1705.07120.

[19] P. Ghosh, M. S. M. Sajjadi, A. Vergari, M. Black, and B. Schölkopf, "From Variational to Deterministic Autoencoders", Proc. Int. Conf. Learn. Represent. (ICLR), Addis Ababa, Ethiopia, Apr. 2020. https://doi.org/10.48550/arXiv.1903.12436.

저자소개

고 영 민 (Yeongmin Ko)



2020년 8월 : 전주대학교 경영학과(학사)
2022년 8월 : 전주대학교 인공지능학과(석사)
2026년 2월 : 전주대학교 인공지능학과(박사)
2026년 2월 ~ 현재 : 전주대학교 RISE 산학협력단 연구원
관심분야 : Data Science & Artificial Intelligence

이 연 승 (Yeonseung Lee)

2021년 3월 ~ 현재 : 전주대학교 인공지능학과 학부과정, 전주대학교 RISE 산학협력단 학생연구원
관심분야 : Artificial Intelligence, Deep Neural Network model, Autoencoder

고 선 우 (Sunwoo Ko)

1985년 8월 : 고려대학교 산업공학과(학사)
1988년 2월 : KAIST 산업공학과(석사)
1992년 8월 : KAIST 산업공학과(박사)
2005년 3월 ~ 현재 : 전주대학교 인공지능학과 교수
관심분야 : Data Science & Artificial Intelligence