

잠재 확산 모델의 잠재 공간에서 샘플링 과정에 대한 포괄적인 분석

A Comprehensive Analysis of the Sampling Process in the Latent Space of Latent Diffusion Models

Yu-Seung Roh, Chae-Hyun Kim, Yoon-Ryung Choi, Kwon-Soon Yong, Won-Yong Shin
Yonsei University

thftmd2@yonsei.ac.kr, kimchaehyun0315@yonsei.ac.kr, wendych@naver.com,
kwonsoon@yonsei.ac.kr, wy.shin@yonsei.ac.kr

요약

본 논문에서는 latent diffusion model (LDM)의 학습과정을 잠재 공간 (latent space) 관점에서 상세히 탐구한다. LDM은 기존 이미지를 autoencoder를 활용하여 저차원 공간인 잠재 공간으로 압축시키고 이를 denoising diffusion probabilistic model (DDPM)과 같이 학습시킴으로써 계산량을 줄이고 학습 시간, 샘플링 시간을 줄여 DDPM 대비 품질과 효율 면에서 압도적인 성능을 이끌어냈다. 하지만, 픽셀 단위로 학습과 샘플링을 진행하는 DDPM과 달리 LDM은 학습 혹은 샘플링 과정이 잠재 공간에서 진행되기 때문에 학습이나 샘플링이 어떻게 이루어지는지 제대로 확인할 수 없다. 따라서, 이를 확인하고자 잠재 공간에서 점들이 어떠한 분포를 가지는지 파악한다. 구체적으로, 주성분 분석을 이용하여 잠재 공간에서 각 벡터가 어떠한 분포를 가지는지 시각화하고, 샘플링 과정에서 현재 시간과 이전 시간 간 점들의 KL-divergence를 측정한다. 이를 통해 LDM의 학습과 샘플링이 효율적으로 진행되었는지 확인한다.

I. 서론

Latent diffusion model (LDM)[1]은 DDPM [2]의 순 과정과 역과정을 Autoencoder를 활용하여 저차원 공간인 잠재 공간에서 진행함으로써 계산량을 줄여 학습과 샘플링의 효율을 향상시켰다. 그러나 LDM의 잠재 공간에서는 학습이 효율적으로 일어나는지 확인할 수 없다. 우리는 이를 확인하기 위해 잠재 공간에서 각 시간 별 주성분 분석과 각 점의 분포를 비교하여 LDM의 샘플링 과정을 분석한다.

II. 제안 방법

본 논문에서는 특정 이미지가 잠재 공간에서 DDIM [3]을 통해 샘플링되는 과정을 시간 별로 분석한다. 구체적으로, 특정 시간의 latent vector를 주성분 분석을 통해 2차원으로 차원을 축소하고 시각화하여 어떠한 분포를 이루고 있는지 확인한다.

또한, 샘플링 과정에서 U-Net을 통과하기 전과 후의 latent vector에 대한 KL-divergence를 계산한다. 이 값들을 토대로 LDM의 샘플링이 효율적으로 진행되는지 분석한다.

III. 실험 결과 및 토의

제안 방법을 구현하기 위해 [1]의 사전 학습된 LDM을 이용하여 이미지를 샘플링하는 과정 중 몇가지 시간 단계에 대하여 주성분 분석을 하고, 이를 시각화한다.

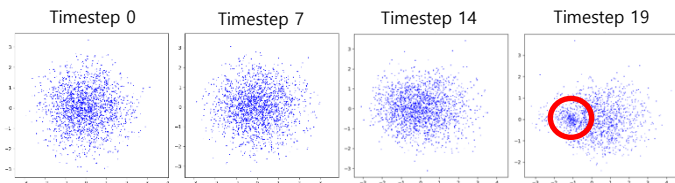


그림 1. 특정 시간 단계의 주성분 분석 시각화

그림 1에서 볼 수 있듯이, 초반 시간에서는 큰 변화가 없다가 후반 시간에 다다를수록 점들이 특정 구간에 집중되는 것을 확인할 수 있다.

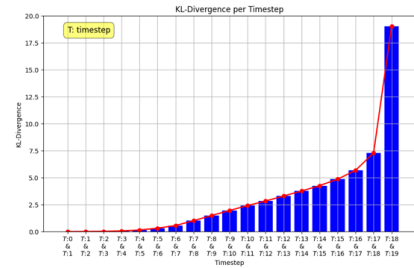


그림 2. 시간 단계 별 U-Net 통과 전후 KL-divergence

또한, 그림 2에서 볼 수 있듯이, 샘플링 과정에서 초반 시간 단계에서는 KL-divergence의 값이 0에 가깝고 단계가 지남에 따라 급격히 커지는 것을 알 수 있다. 이를 미루어 보아, LDM은 초반 시간보다 후반 시간에 집중적으로 학습이 이루어진 것을 확인할 수 있고, 이는 전반적인 학습이 불균형을 이루고 있음을 추론할 수 있다.

ACKNOWLEDGMENTS

This work was supported by the National Research Foundation of Korea (NRF), Republic of Korea through the Korean Government through MSIT under Grant 2021R1A2C3004345 and Grant RS-2023-00220762 by the Institute of Information and Communications Technology Planning and Evaluation (IITP), Republic of Korea through by the Korean Government through MSIT (6G Post-MAC—POsitioning and Spectrum—Aware intelligent MAC for Computing and Communication Convergence) under Grant 2021-0-00347.

참고 문헌

- [1] R. Rombach *et al.* High-resolution image synthesis with latent diffusion models. In CVPR 2022.
- [2] J. Ho *et al.* Denoising diffusion probabilistic models. In NeurIPS 2020.
- [3] J. Song *et al.* Denoising diffusion implicit models. In ICLR 2021.