VARIATIONAL DIFFUSION AUTO-ENCODER: LATENT SPACE EXTRACTION FROM PRE-TRAINED DIFFUSION MODELS

GEORGIOS BATZOLIS

As a widely recognized approach to deep generative modeling, Variational Auto-Encoders (VAEs) still face challenges with the quality of generated images, often presenting noticeable blurriness. This issue stems from the unrealistic assumption that approximates the conditional data distribution, p(x|z), as an isotropic Gaussian. In this paper, we propose a novel solution to address these issues. We illustrate how one can extract a latent space from a pre-existing diffusion model by optimizing an encoder to maximize the marginal data log-likelihood. Furthermore, we demonstrate that a decoder can be analytically derived post encoder-training, employing the Bayes rule for scores. This leads to a VAE-esque deep latent variable model, which discards the need for Gaussian assumptions on p(x|z) or the training of a separate decoder network. Our method, which capitalizes on the strengths of pre-trained diffusion models and equips them with latent spaces, results in a significant enhancement to the performance of VAEs.