# Abstract

Keywords: Deep Latent Variable Generative Models; Regularized Autoencoders; Latent Space; Few-shot Generation.

The rise of deep neural networks has significantly advanced unsupervised generative modeling. Numerous Deep Generative Models (DGMs) have been proposed, including Variational Autoencoders (VAE) [KW14], Generative Adversarial Networks (GAN) [GPAM+14], Wasserstein Autoencoders (WAE) [TBGS18], Adversarial Autoencoders (AAE) [MSJG16], Autoregressive Models [vdOKE+16, OKK16], Normalizing Flow-based Models [KD18], Energy-based Models [LCH+06, DM19], and Diffusion Models [HJA20]. These models may be categorized along various dimensions, such as their architectural variations, presence or absence of explicit latent representation, training methodology and stability, density estimation capability, and time taken for sampling. In this thesis, our focus is on *Deep Latent Variable Generative Models* (DLVGMs), which refer to generative autoencoders (VAE, WAE, AAE) and GANs. This class of models uses low-dimensional latent representations of high-dimensional data, allowing learning informative low-dimensional representations for various downstream tasks (clustering, classification, and disentangling generative factors) while facilitating novel data generation. Interestingly, GANs are particularly notable for their high-quality generation but suffer from unstable training, mode collapse, and hyperparameter sensitivity. In contrast, Autoencoders with regularized latent spaces (VAE, WAE, AAE) offer stable training, interpretable inference, and efficient sampling, though their generated images are visually less impressive than those from GANs. In this thesis, we intend to address some of the complementary strengths and limitations of these DLVGMs. The thesis is organized in several parts, as outlined below.

In Part I (Prologue) of this thesis, we introduce various deep generative frameworks, define DLVGMs, highlight challenges (such as poor generation quality of RAEs, representation learning issues, and the need for large-scale data) associated with DLVGMs, and review existing solutions.

In Part II, 'Optimizing the Latent Space of RAEs for Improved Generation,' we diagnose the reasons behind the poor generation quality of generative AE frameworks by exploring two questions: 1. What is the 'optimal' latent dimension [MCJ+20], and 2. What is the 'optimal' latent prior [MASP21a] for a good generation? We hypothesize natural data generation as a two-step process involving a true low-dimensional latent space and a non-linear mapping to a high-dimensional data space. We show that under the assumption

of a Gaussian prior, the best generation quality is achieved when the dimensionality of the generative AE's bottleneck layer matches the true latent dimensionality [MCJ+20]. In [MASP21a], we relax the Gaussian prior assumption to learn the prior flexibly, considering the true latent dimensionality.

In Part III, 'Optimizing the Latent Space of RAEs for Task-Specific Representation Learning,' we focus on representation learning in an RAE framework. First, we study the impact of bias-variance trade-off due to fixed prior distribution versus learnable priors on representation quality and demonstrate that learning the prior flexibly helps the model discover the actual data structure, improving clustering performance [MASP21b]. While [MASP21b] uses uni-modal data, we address disentangled representation learning in a multi-modal setting in [MSSA23]. We decompose the joint latent space into continuous and discrete components, each with domain-specific and domain-invariant representations. We demonstrate the effectiveness of these disentangled joint representations in downstream tasks like classification and generation. In our subsequent work [MST+23], we combine the representation learning aspect with the generation ability of RAEs to develop a framework for class-imbalance mitigation to enhance discriminating performance. Precisely, we propose a minority oversampling method that is distance-metric-free and class-preserving by design.

In Part IV, 'Few-shot Generation Using DLVGMs,' we address the issue that while DLVGMs offer a plethora of applications, they are data-hungry, limiting their applicability in real-world scenarios with data scarcity. Specifically, we develop techniques to transfer a source DLVGM built with large-scale data to a 'close' target domain with limited data. In [MTSA23], we perform few-shot 'generative domain adaptation' via inference-time latent-code learning by prepending a latent adapter network. While [MTSA23] achieves high-quality generation, it requires considerable time to generate due to inference-time optimization. In [MTSA24], we address this by learning to sample the parameters of the latent adapter network using a hypernetwork.

Finally, in Part V, we summarize our contributions and propose directions for future work based on the techniques and methods introduced in this thesis.