VQ-Flows: Vector Quantized Local Normalizing Flows

Sahil Sidheekh*[1], Chris B. Dock*[2], Tushar Jain[1], Radu Balan[2], Maneesh K. Singh[1]

*−Equal Contribution,[1]−Verisk Analytics, [2]−University of Maryland, College Park, Partial support by NSF-DMS 2108900 and Simons Foundation 818333

Objectives

- Improve the expressively of normalizing flow to model data distributions on (finite sets of) differentiable manifolds. Handle arbitrary topology of the data space, and different dimensionality than the latent space.
- Provide a statistical framework for normalizing flows as manifold chart maps and an trainable DNN model for the same.
- Empirical validation of improved density estimation and sample generation, scaling to high dimensional image data (MNIST).

Introduction

Unlike other generative models, normalizing flows (NFs) have the advantage of allowing for exact density estimation [1]. Unfortunately, this benefit comes at the cost of requiring the flow to be a diffeomorphism (a.e.), restricting the applicability of NFs to data manifolds that are diffeomorphic to the latent space $\mathcal{Z}$. In particular, the data must have the same dimension as $\mathcal{Z}$ if the NF is to perform well. We overcome these limitations by viewing NFs as chart maps of the data manifold, thus allowing for data manifolds with more complex topology. Our main contributions are:

- A scalable method of learning good charts using a vector quantized auto encoder.
- A statistical and scalable (to high dimensions) framework for combining normalizing flows from different charts.

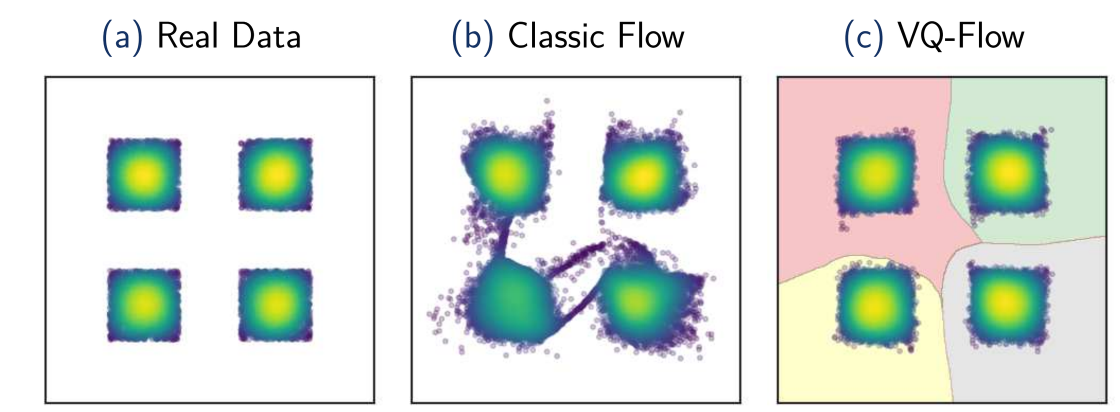

Figure: Augmentation of our framework (c) enables a classic flow (b) to better model the discontinuities in the data manifold through a learned atlas of charts(shaded region).

The chart regions $U_1 \dots, U_K$

A VQ-AE learns an encoder map $E : \mathcal{X} \to \mathcal{V}$, a decoder map $D : \mathcal{V} \to \mathcal{X}$, and a collection of "chart centers" $Q = \{v_k\}_{k=1}^K \subset \mathcal{V}$ that minimize the error $\mathcal{L}(D(\text{argmin}_{v \in Q}||v - E(x)||_2), x)$. Once D, E, and Q are learned we compute $d_k(x) = ||E(x) - v_k||_2$ for $k = 1, \dots K$. We would like charts to overlap, but also to be sparse in the sense that no x has too many charts. Fix $\epsilon > 0$, let $\tilde{d}_1 \leq \dots \leq \tilde{d}_K$ be the sorted permutation of $d_1, \dots, d_K$ then define $U_k = \{x : ||E(X) - v_k||_2 < (1+\epsilon)\tilde{d}_m(x)\}$.

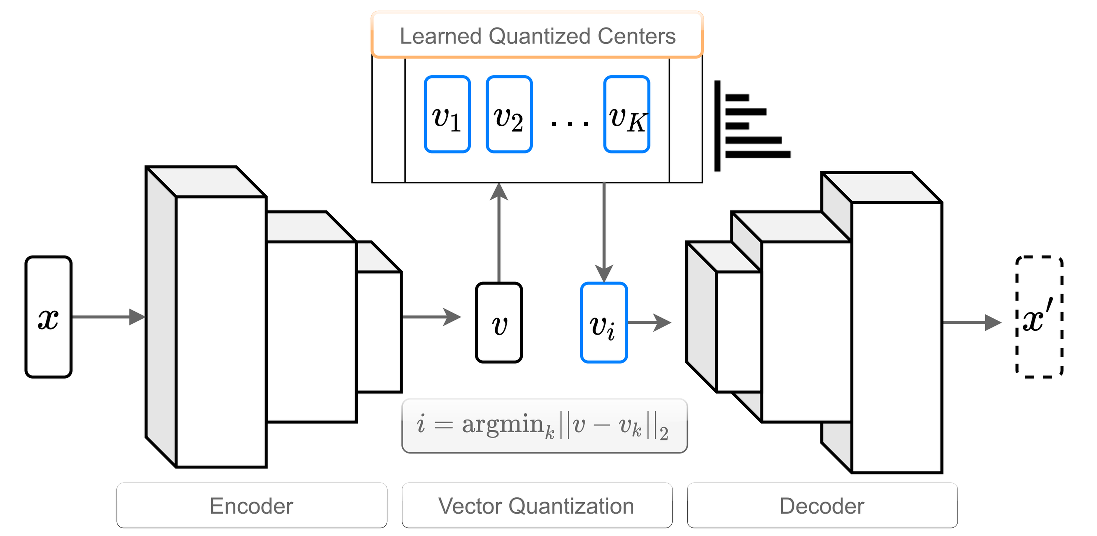

Figure: Learning quantized centers on the low dimensional data manifold using a vector quantized auto-encoder.

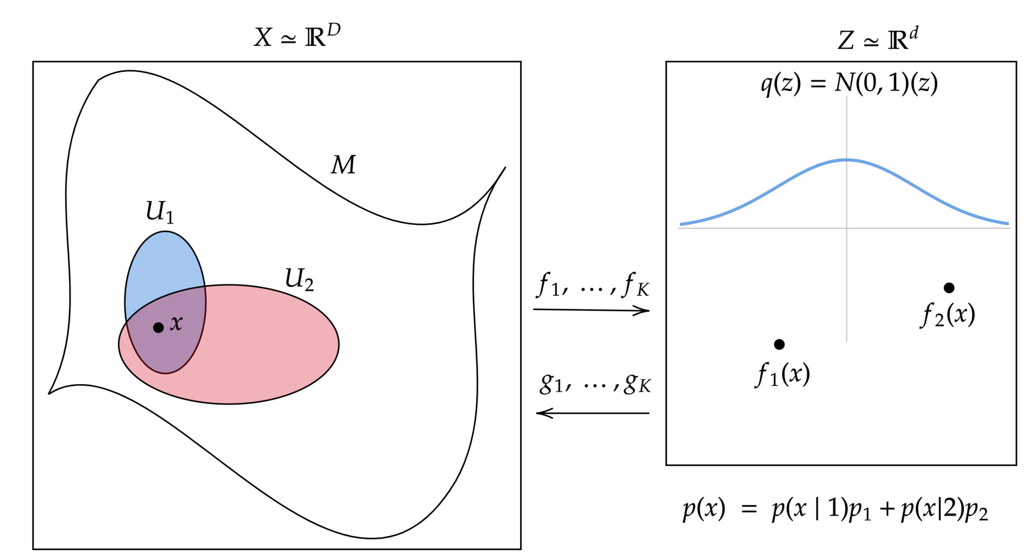

Figure: Normalizing flows as chart maps.

The chart maps $f_1, \dots, f_k$

We model $g_k : \mathcal{Z} \to U_k$ as L layered invertible NFs. To handle dimensionality change, we post-compose with a conformal dimension raising map [2] so that $g_k = c_k \circ g_k^L \circ \dots \circ g_k^1$ and $f_k = f_k^1 \circ \dots \circ f_k^L \circ c_k^\dagger$. In practice, we reduce the number of parameters of our model by restricting each g_k^l (and f_k^l) to depend on k only through the value of the encoded chart center v_k. $g_1, \dots, g_k$ are learned via gradient descent on the objective function (2).

- Sampling: z and k are independent, so sample $z \sim q(z)$ and $k \sim p(k)$ and then compute $x = g_k(z)$.
- Inference: One can perform a stochastic inference via sampling $k \sim p(k|x)$ and computing $z = f_k(x)$, however if deterministic inference is preferred one may instead use $z = \mathbb{E}_{k \sim p(k|x)}[f_k(x)] = \sum_{k:x \in U_k} p(k|x) f_k(x)$.

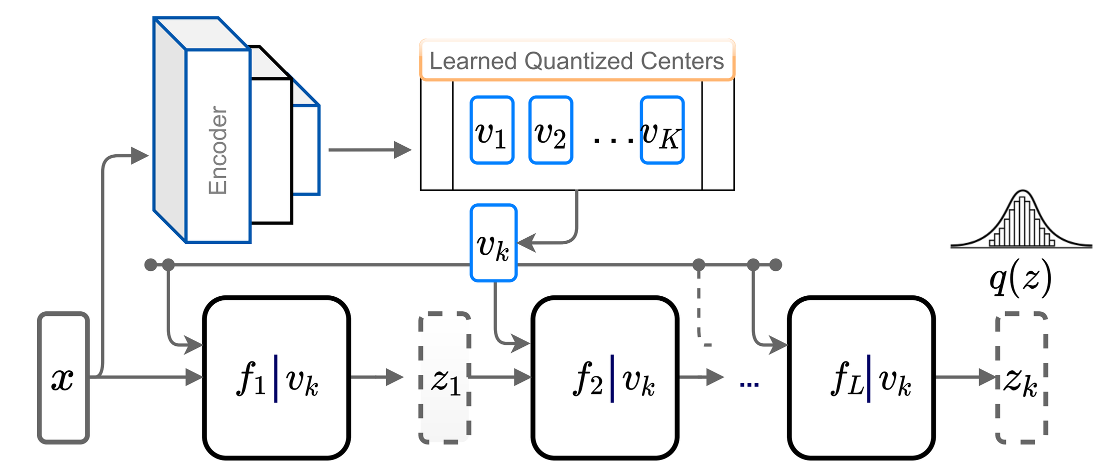

Figure: Learning the data distribution using a family of normalizing flows conditioned on the quantized centers.

Exact Density Evaluation

Denote by $\mathcal{Z}$ the latent space, $\mathcal{X}$ the data space, and $\mathcal{M} \subset \mathcal{X}$ the data manifold. Let $(U_1)_{k=1}^K$ be such that $\mathcal{M} \subset \cup_{k=1}^K U_k$ and let $V_k = U_k \cap \mathcal{M}$. Assume there exists $D_k \subset \mathcal{Z}$ so that $V_k = g_k(D_k)$ for some immersion $g_k : D_k \to U_k$ with inverse $f_k : V_k \to D_k$. If x is a r.v. supported on $\mathcal{M}$, z is a r.v. in $\mathcal{Z}$, k is a discrete random variable and x, z, k have joint distribution

$$p(x, z, k) = \delta(x - g_k(z)) q(z) p_k \tag{1}$$

Then

$$p(x) = \sum_{k: x \in V_k} p_k |\det[Jf_k(x) Jf_k(x)^T]|^{\frac{1}{2}} q(f_k(x)) \tag{2}$$

Results

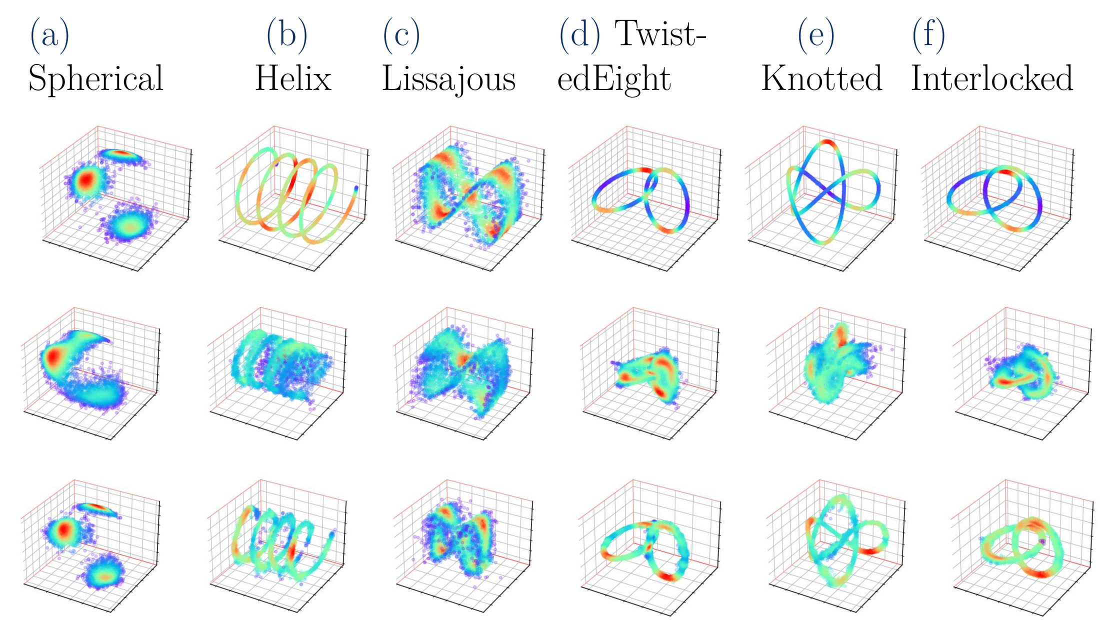

Figure: Qualitative visualization of the samples generated by a classical flow - RealNVP (Middle Row) and its VQ-counterpart (Bottom Row) trained on Toy 3D data distributions (Top Row).

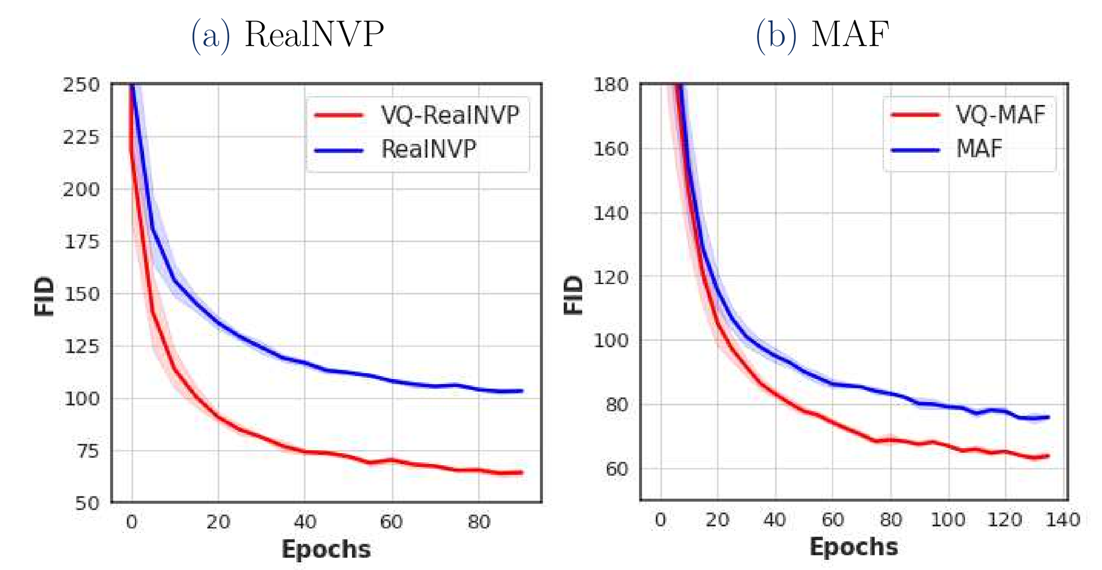

Figure: FID scores (lower the better) across the training of (a) RealNVP and (b) MAF on the MNIST dataset. The shaded region represents the standard deviation over 3 trials.

Future Work

Our framework is well suited to high-dimensional datasets (such as natural images) that obey the manifold hypothesis, an avenue we hope to explore in the sequel.

References

[1] Danilo Rezende and Shakir Mohamed.
Variational inference with normalizing flows.
In *ICML*, 2015.

[2] Brendan Leigh Ross and Jesse C Cresswell.
Tractable density estimation on learned manifolds with conformal embedding flows.
In *NeurIPS*, 2021.