Learning a Regularizer by Approximating the Patch Manifold

Giovanni S. Alberti* and Johannes Hertrich†

The goal of this project is to derive a learned regularizer for image reconstruction by approximating the manifold of patches by a mixture of generative models. Below, we describe the project in detail, structuring it in three consecutive steps.

1 Manifold Approximation with Mixtures of VAEs

Background We are given data points $x_1, ..., x_N \in \mathbb{R}^n$ and assume that the elements of the datasets are located on an embedded d-dimensional submanifold $\mathcal{M} \subseteq \mathbb{R}^n$ with $d \ll n$, which is often true in practice (for image datasets this assumption is known as the "manifold hypothesis", see, e.g., [3]). In this project, we aim to learn this manifold and consider related applications. A common approach for manifold learning is to train an injective neural network $D \colon \mathbb{R}^d \to \mathbb{R}^n$ and represent $\mathcal{M} \approx \mathrm{Range}(D)$. To derive a loss function, we can use an adaption of the change-of-variables formula from normalizing flows [7, 6, 8], or we interpret D as the decoder of a variational autoencoder [5]. However, this approach can only approximate manifolds that admit a global parametrization, which is not true for disconnected manifolds or manifolds with holes. As a remedy, we can approximate $\mathcal{M}$ by a mixture model of VAEs, where each decoder represents (the inverse of) a chart from the manifold [1].

Objective Replace the mixture of VAEs in [1] by a mixture of injective free form flows [8], which directly approximate the likelihood of each generator and do not have such tight architectural constraints. Try a state-of-the-art architecture for image generation and examine whether this alternative architecture improves generation quality.

2 Approximating the Patch-Manifolds

Background We consider the dataset of all $p \times p$ (possibly overlapping) patches from some image dataset. In [4] the authors observe that for 3×3 patches of gray-valued images, this dataset is located on a two-dimensional manifold with non-trivial topology.

Objective Approximate this manifold by a mixture of VAEs (or other architectures, see previous task) and find out the topology of this manifold by considering which charts overlap at the boundary. Can the claims from [4] be reproduced numerically? What happens for $p \geq 4$? Is the topology the same as for 3×3 patches?

*Machine Learning Genoa Center (MaLGa), Department of Mathematics, University of Genoa
†Department of Computer Science, University College London

3 Patch-based Regularizers

Background Once we have approximated the manifold of all patches in a certain image dataset, we want to apply this approach to define a regularizer which can be used for image reconstruction in inverse problems. Our goal is to reconstruct an unknown image x from an observation y generated as

$$y = \mathrm{noisy}(F(x)),$$

where F is a non-invertible or ill-posed forward operator. A common approach for reconstructing x is to minimize the

$$\hat{x} \in \arg\min_x d(F(x), y) + \lambda R(x),$$

where the first term measures the consistency of x with the observed data y, and the second term incorporates some prior knowledge. Following the approaches of the expected patch log-likelihood regularizer (EPLL [9]) and patch normalizing flow regularizer (patchNR [2]), we aim to define a regularizer

$$R(x) = -\sum_{i=1}^{N} \log(p(P_i(x))),$$

where P_i extracts the i-th patch from x and p is the density function of the patch distribution. Of course, this requires that we have learned the patch-distribution a-priori.

Objective While EPLL [9] learns the patch distribution p as a Gaussian mixture model and the patchNR [2] as a normalizing flow, we now want to insert our manifold approximation from the previous task as p. Can we improve the results of EPLL/patchNR?

References

[1] G. S. Alberti, J. Hertrich, M. Santacesaria, and S. Sciutto. Manifold Learning by Mixture Models of VAEs for Inverse Problems. *Journal of Machine Learning Research*, 25(202):1–35, 2024.

[2] F. Altekrüger, A. Denker, P. Hagemann, J. Hertrich, P. Maass, and G. Steidl. PatchNR: learning from very few images by patch normalizing flow regularization. *Inverse Problems*, 39(6):064006, 2023.

[3] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[4] G. Carlsson, T. Ishkhanov, V. De Silva, and A. Zomorodian. On the local behavior of spaces of natural images. *International journal of computer vision*, 76:1–12, 2008.

[5] D. P. Kingma. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[6] K. Kothari, A. Khorashadizadeh, M. de Hoop, and I. Dokmanić. Trumpets: Injective flows for inference and inverse problems. In *Uncertainty in Artificial Intelligence*, pages 1269–1278. PMLR, 2021.

[7] S. G. Krantz and H. R. Parks. *Geometric integration theory*. Springer Science & Business Media, 2008.

[8] P. Sorrenson, F. Draxler, A. Rousselot, S. Hummerich, L. Zimmermann, and U. Koethe. Lifting architectural constraints of injective flows. In *The Twelfth International Conference on Learning Representations*, 2024.

[9] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *2011 international conference on computer vision*, pages 479–486. IEEE, 2011.