Manifold Learning in Quotient Spaces: Supplementary Material

Éloi Mehr[1], André Lieutier , Fernando Sanchez Bermudez , Vincent Guitteny , Nicolas Thome[1], and Matthieu Cord[1]

[1]LIP6, UPMC Sorbonne Universités

1. Mathematical background

In this section, we provide the reader with a short mathematical background in relation with our work, which falls within geometry and topology applied to machine learning.

In order to forget the unneeded information about 3D, we represent an intrinsic geometry as an element of $X/\mathcal{G}$, the quotient of the original input space X by the action of the rotations group. The quotient space $X/\mathcal{G}$ is the set of all the orbits of the group action: $X/\mathcal{G} = \{\overline{x} \mid x \in X\}$, where $\overline{x} = \{h.x \mid h \in \mathcal{G}\}$, and $(h, x) \mapsto h.x$ is the group action. If (X, T) is a topological space, then $X/\mathcal{G}$ is naturally equipped with the so-called quotient topology $T_{X/\mathcal{G}} = \left\{U \subset X/\mathcal{G} \mid q^{-1}(U) \in T\right\}$, where q is the canonical surjection which maps x onto its orbit. When X is a metric space, proposition 1 makes explicit the quotient topology in the case where the action is an isometry. Especially, under mild assumptions the quotient loss (Equation (2) in the paper) also defines a distance in the quotient space $X/\mathcal{G}$.

A Lie group G is a group that is also a differentiable manifold. More formally it is group with a differentiable manifold structure such that both group multiplication and inverse map are differentiable. Since the multiplication by an element $h \in G$ defines a diffeomorphism from G to G that sends the identity element to h, the local behavior near the identity reflects the local behavior near any element $h \in G$. Therefore the tangent space at the identity together with its linear structure, *i.e.* the first order behavior of elements infinitesimally close to the identity, called Lie Algebra, is a central object in the study of Lie groups. Among the most common and intuitive Lie groups are rotation groups in n-dimensional Euclidean space $SO(n)$. In our paper, we use the Lie structure to parameterize an infinite group with a finite-dimensional vector h representing the transformation.

2. Implementation details

To discretize an infinite group, we "evenly" pick transformations according to a Riemannian metric. For instance, we parameterize $SO(2)$ by the rotation angle and evenly split $[0, 2\pi]$ to get 36 rotations. Depending on the considered group, it might be better to deal with the infinite group itself without any discretization, as done in section 5 of the paper. The sampling in the orbit pooling is theoretically chosen such that the orbit distribution is invariant by the group action. To simplify, in section 5 we bounded the deformations set to a scaled hypercube of $\mathbb{R}^{10}$ and used a uniform distribution.

Our deep architectures are detailed in Figure 1. They are inspired by [8]. These architectures are the same for both the QAE and the vanilla autoencoder, except that the vanilla autoencoder has no orbit pooling while the QAE has an orbit pooling layer. We use ReLU activation functions with a slope of 0.2 for the non-linearities, except for the activation of the last layer which is the hyperbolic tangent function. We use ADAM optimization with mini-batch size of 32 samples for the optimization of the networks' parameters.

We experiment the QAE on version 1 of the core ShapeNet dataset [2], called ShapeNetCore v1. This release contains 7497 chairs, 6778 cars, and 4045 airplanes.

3. Quotient loss

We prove here that our quotient loss $\overline{d}$ benefits from nice properties under weak conditions, fulfilled for instance with rotations acting on shapes.

Proposition 1. *Let (X, d) be a metric space, $\mathcal{G}$ a group, and $(h, x) \mapsto h.x$ a group action of $\mathcal{G}$ on X. We note $\overline{x} = \{h.x \mid h \in \mathcal{G}\}$ the orbit of x. We suppose that the action is an isometry, that is for every $x, y \in X$, and any $h \in \mathcal{G}$, $d(h.x, h.y) = d(x, y)$. Then,*

$$\overline{d} \colon (\overline{x}, \overline{y}) \mapsto \inf_{h \in \mathcal{G}} d(h.x, y) \tag{1}$$

well defines a pseudometric on the quotient space $X/\mathcal{G} = \{\overline{x} \mid x \in X\}$.

Moreover, if $\mathcal{G}$ is a compact space and its action is continuous, then $\overline{d}$ is also a metric on $X/\mathcal{G}$.

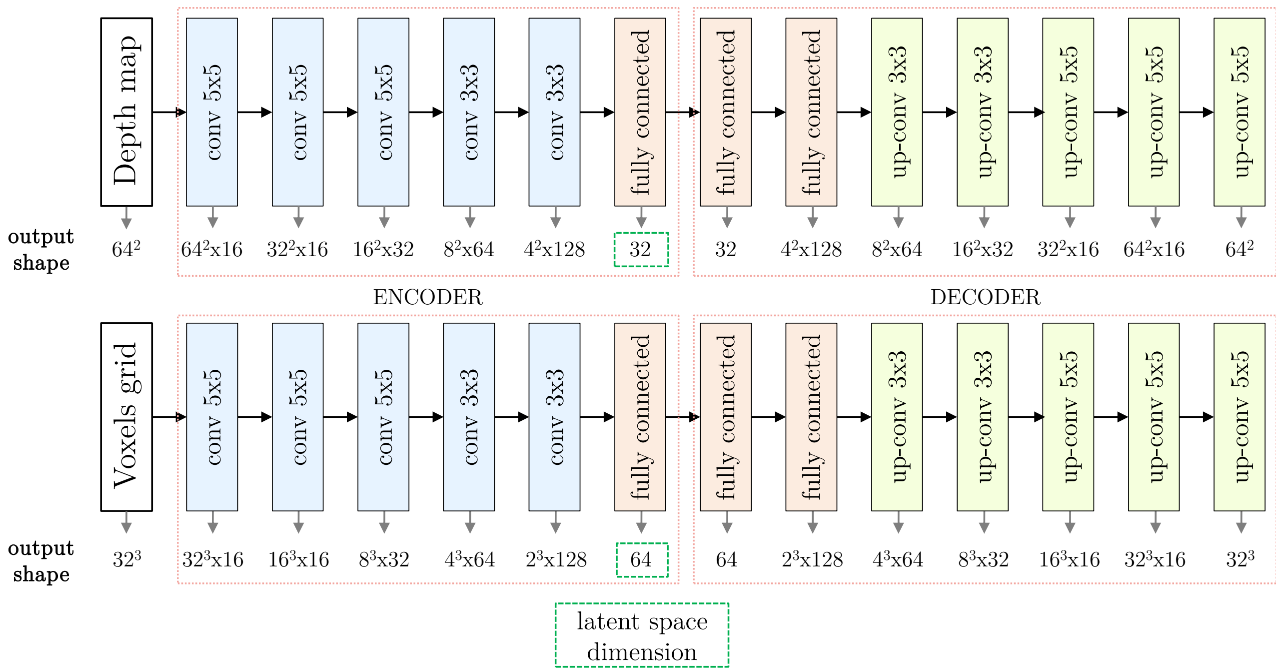

Figure 1: Description of our 2D and 3D deep convolutional architectures.

Proof. $\overline{d}$ is well defined on $X/\mathcal{G}$. Indeed, if x_1 and x_2 are two representatives of the same orbit $\overline{x}$, and similarly y_1, y_2 for $\overline{y}$, then there exists $h_x, h_y \in \mathcal{G}$ such that $x_1 = h_x.x_2$ and $y_1 = h_y.y_2$. So, for any $h \in \mathcal{G}$,

$$d(h.x_1, y_1) = d(hh_x.x_2, h_y.y_2) \quad (2)$$
$$= d(h_y^{-1}hh_x.x_2, y_2)\,. \quad (3)$$

As $h \mapsto h_y^{-1}hh_x$ is a bijection of $\mathcal{G}$, taking the infimum in Equation (3) gives

$$\inf_{h\in\mathcal{G}} d(h.x_1, y_1) = \inf_{h\in\mathcal{G}} d(h.x_2, y_2)\,. \quad (4)$$

Obviously $\overline{d}$ is non-negative and $\overline{d}(\overline{x}, \overline{x}) = 0$ for any $x \in \mathcal{G}$.

Let $x, y \in X$. Then, for any $h \in \mathcal{G}$, we have

$$d(h.x, y) = d(x, h^{-1}.y) = d(h^{-1}.y, x)\,. \quad (5)$$

Moreover, $h \mapsto h^{-1}$ is a bijection of $\mathcal{G}$, so by taking the infimum in Equation (5), we get

$$\overline{d}(\overline{x}, \overline{y}) = \overline{d}(\overline{y}, \overline{x})\,, \quad (6)$$

so $\overline{d}$ is symmetric.

Let $x, y, z \in X$, and $\epsilon > 0$. Then there exists $h_1 \in \mathcal{G}$ such that

$$d(h_1.x, y) \le \overline{d}(\overline{x}, \overline{y}) + \epsilon/2 \quad (7)$$

and $h_2 \in \mathcal{G}$ such that

$$d(h_2.z, y) \le \overline{d}(\overline{z}, \overline{y}) + \epsilon/2\,. \quad (8)$$

Thus,

$$d(h_1.x, h_2.z) \le d(h_1.x, y) + d(h_2.z, y) \quad (9)$$
$$\le \overline{d}(\overline{x}, \overline{y}) + \overline{d}(\overline{z}, \overline{y}) + \epsilon\,. \quad (10)$$

We also have

$$\overline{d}(\overline{x}, \overline{y}) \le d(h_2^{-1}h_1.x, z) = d(h_1.x, h_2.z)\,. \quad (11)$$

Taking the limit as $\epsilon \to 0$ concludes the proof of the triangular inequality.

Finally, suppose that $\mathcal{G}$ is compact and its action is continuous. Let $x, y \in \mathcal{G}$ such that $\overline{d}(\overline{x}, \overline{y}) = 0$. The application $h \mapsto d(h.x, y)$ from the compact $\mathcal{G}$ to $\mathbb{R}_+$ is continuous, so the extreme value theorem implies that there exists $\widehat{h} \in \mathcal{G}$ such that

$$\inf_{h\in\mathcal{G}} d(h.x, y) = d(\widehat{h}.x, y)\,. \quad (12)$$

By unicity of the limit we have

$$d(\widehat{h}.x, y) = 0\,. \quad (13)$$

So $\widehat{h}.x = y$, which exactly means that $\overline{x} = \overline{y}$ and concludes our proof. ∎

4. Further work

The extension of the QAE to a variational autoencoder is straightforward, as it simply adds a regularization term (and a genuine sampling step during the optimization) to the regular loss of the autoencoder. It is more challenging to wonder how to use a quotient framework in a generative adversatial network [4, 10], as generative adversarial networks do not use an explicit reconstruction error or likelihood.

QAE could be applied to recent multi-resolution voxels grids learning methods [6, 9], but we are also interested in applying our QAE to richer 3D representations, such as point clouds [1, 3] or meshes [5].

Further work could also study the connections between the QAE and spatial transformer networks [7], although spatial transformer networks are not able to deal with discrete transformations.

References

[1] P. Achlioptas, O. Diamanti, I. Mitliagkas, and L. J. Guibas. Representation learning and adversarial generation of 3d point clouds. *arXiv*, 2017. 3

[2] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. ShapeNet: An Information-Rich 3D Model Repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015. 1

[3] H. Fan, H. Su, and L. J. Guibas. A point set generation network for 3d object reconstruction from a single image. In *Computer Vision and Pattern Regognition*, 2017. 3

[4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, pages 2672–2680. 2014. 3

[5] T. Groueix, M. Fisher, V. G. Kim, B. Russell, and M. Aubry. AtlasNet: A Papier-Mâché Approach to Learning 3D Surface Generation. In *Computer Vision and Pattern Regognition*, 2018. 3

[6] C. Häne, S. Tulsiani, and J. Malik. Hierarchical surface prediction for 3d object reconstruction. In *arXiv*. 2017. 3

[7] M. Jaderberg, K. Simonyan, A. Zisserman, and k. kavukcuoglu. Spatial transformer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2017–2025. Curran Associates, Inc., 2015. 3

[8] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Multi-view 3d models from single images with a convolutional network. In *European Conference on Computer Vision*, pages 322–337, 2016. 1

[9] M. Tatarchenko, A. Dosovitskiy, and T. Brox. Octree generating networks: Efficient convolutional architectures for high-resolution 3d outputs. In *International Conference on Computer Vision*, 2017. 3

[10] J. Wu, C. Zhang, T. Xue, W. T. Freeman, and J. B. Tenenbaum. Learning a Probabilistic Latent Space of Object Shapes via 3D Generative-Adversarial Modeling. In *Advances In Neural Information Processing Systems*, volume 29, pages 82–90, 2016. 3