Supplementary Material for "AVT: Unsupervised Learning of Transformation Equivariant Representations by Autoencoding Variational Transformations"

Guo-Jun Qi[1,2,*], Liheng Zhang[1], Chang Wen Chen[3], Qi Tian[4]
[1]Laboratory for MAchine Perception and LEarning (MAPLE)
http://maple-lab.net/
[2]Huawei Cloud, [4]Huawei Noah's Ark Lab
[3] The Chinese University of Hong Kong at Shenzhen and Peng Cheng Laboratory
guojun.qi@huawei.com
http://maple-lab.net/projects/AVT.htm

In the conventional definition of transformation equivariance, there should exist an automorphism $\boldsymbol{\rho}(\mathbf{t}) \in \mathrm{Aut}(\mathcal{Z}) : \mathcal{Z} \rightarrow \mathcal{Z}$ in the representation space, such that $\mathbf{z} = [\boldsymbol{\rho}(\mathbf{t})](\tilde{\mathbf{z}})$, where $\tilde{\mathbf{z}}$ is the representation of the original image without transformation. Here, the essence is the representation $\mathbf{z}$ of a transformed sample can be completely determined by the original representation $\mathbf{z}$ and the applied transformation $\mathbf{t}$ without accessing the original sample $\mathbf{x}$, which is called "steerability" in literature [1].

This property can be generalized beyond the linear automorphism $\boldsymbol{\rho}(\mathbf{t})$. Instead of sticking with a linear transformation. From an information theoretical point of view, this requires $\{\tilde{\mathbf{z}}, \mathbf{t}\}$ should contain all necessary information about $\mathbf{z}$ so that $\mathbf{z}$ can be best estimated from them without accessing $\mathbf{x}$.

This leads us to maximizing the mutual information $I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t})$ to learn a generalized transformation equivariant representation. Indeed, by the chain rule and nonnegativity of mutual information, we have

$$I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t}) = I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t}, \mathbf{x}) - I_\theta(\mathbf{z}; \mathbf{x}|\tilde{\mathbf{z}}, \mathbf{t}) \leq I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t}, \mathbf{x}),$$

where $I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t})$ attains the upper bound $I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t}, \mathbf{x})$ as its maximum value, when $I_\theta(\mathbf{z}; \mathbf{x}|\tilde{\mathbf{z}}, \mathbf{t}) = 0$, i.e., $\mathbf{x}$ provides no additional information about $\mathbf{z}$ with $(\tilde{\mathbf{z}}, \mathbf{t})$ given. This implies that one can estimate $\mathbf{z}$ from $(\tilde{\mathbf{z}}, \mathbf{t})$ directly, satisfying the "steerability" property.

In the proposed variational approach, however, we maximize the following lower bound of $I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t})$

$$I_\theta(\mathbf{z}; \mathbf{t}|\tilde{\mathbf{z}}) = I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t}) - I_\theta(\mathbf{z}; \tilde{\mathbf{z}}) \leq I_\theta(\mathbf{z}; \tilde{\mathbf{z}}, \mathbf{t})$$

between the representation and the transformation as presented in Section 3 to pursue the generalized form of transformation equivariant representation. This will be elaborated in the long version of this paper [2].

*Corresponding author: G.-J. Qi. Email: guojunq@gmail.com. The idea was conceived and formulated by G.-J. Qi, and L. Zhang performed experiments while interning at Huawei Cloud.

References

[1] Taco S Cohen and Max Welling. Steerable cnns. *arXiv preprint arXiv:1612.08498*, 2016. 1

[2] Guo-Jun Qi. Learning generalized transformation equivariant representations via autoencoding transformations. *arXiv preprint arXiv:1906.08628*, 2019. 1