Using B-Splines to Measure Object Representation with Interpolative Quality in Auto-Encoders

William Harrison, Preetham Patlolla, Zeid Kootbally, and Satyandra K. Gupta,

I. Introduction

Models that build low-dimensional representations of the data like autoencoders and Generative Adversarial Networks (GANs) have seen lots of attention over the years. This low-dimensional representation, also called a latent space, has benefits such as creating new instances of the data, noise reduction, and more interpretable data visualizations. Because the latent space may reveal explanatory factors, observations and operations in the latent space can be useful. In the case of data visualization, the dataset may exhibit clustering in the latent space, giving the observer new ways to understand the data. In the case of object pose estimation, the latent space can be used to determine an object's orientation. In the case of data generation, the latent space can be used to generate new and usable data.

The latent space of a trained model contains an implicit representation within it. For the purposes of this paper, the representation is of a physical object. An object's representation in the latent space dictates the model's application and performance. More specifically, applications that utilize a latent space are influenced by the organization of object data in the latent space with respect to factors of variation. Factors of variation refer to observable characteristics that explain the state of an object(in this case). For example in [1] and [2] the relative positioning of latent vectors generated from images can be used to determine the orientation of objects. Additionally, new data may be generated by interpolating between observed data points in the latent space. This can yield novel but possible instances of the data.

Because the latent space is at the heart of generative models, metrics associated with the latent space could serve as a good means for model comparison, determining model confidence, and measuring model performance. Generally, applicable metrics are specific to their downstream task, however, there are metrics that are general and allow for model comparison. These metrics are largely centered around disentanglement. "A disentangled representation is generally described as one which separates the factors of variation, explicitly representing the important attributes of the data" [3]. Factors of variation are expected to remain invariant to one another during interpolation if they are disentangled. A disentangled representation can "improve predictive performance, reduce sample complexity, offer interpretability, improve fairness and have been identified as a way to overcome shortcut learning " [4].

There are a relatively limited number of examples of general measures of disentanglement [4]. The examples that do, face challenges. In [5] the authors "theoretically show that the unsupervised learning of disentangled representations is fundamentally impossible without inductive biases on both the models and the data." In [6] the authors make the point that there is no widely accepted definition of disentanglement, and furthermore show that most disentanglement metrics do not satisfy the desired properties for metrics in general.

While disentanglement is admittedly the most important aspect related to downstream performance, assessing implicit representation by measuring how well factors of variation are organized in the latent space could provide an additional basis for quality measurement and model comparison, leading to increased confidence in deployed models. Furthermore, investigating the organization of factors of variation in the latent space may allow one to verify the existence of abstract representations such as if the model's representation of an object is that of a rigid body, and thus, adding to the confidence in the model if the object is in fact a rigid body. If there is a link between models that exhibit good downstream performance and organization in the latent space with respect to factors of variation, models can better be compared and users will have more information about what a model has learned.

The focus of this paper is on the organization of object data in the latent space as it pertains to a single factor of variation. The application presented here is based on a technique developed in [1] where the orientation of an object is determined by measuring the relative positioning of latent vectors generated by an autoencoder. Our approach is an analysis of interpolation paths of a specific factor of variation through the latent space. The approach is applicable to models trained on data with quantitative factors of variation where the ground truth is known.

To the best of our knowledge, there exists no general metric for measuring the organization of factors of variation in the latent space. The challenges to devising such a measure based on interpolation paths are:

1) There is no clear methodology for establishing latent space organization with respect to factors of variation.
2) A quantitative measure must be developed for measuring interpolative quality.
3) A methodology for entirely and finitely encapsulating the nature of a single factor of variation must be

devised.

4) All methodologies must work in an N-dimensional latent space
5) The approach must account for the precarious nature of distance in the latent space.

We propose an analysis methodology for the quantitative measurement of interpolative quality using a latent space B-Spline analysis. This is accomplished by fitting B-splines to interpolation paths we call trajectories, and comparing them to a quaternion and random vector spaces. Figure 1 shows examples of trajectory B-spines in the quaternion, latent, and random spaces. The analysis includes a set of metrics (torsion and arc length), to investigate the relative positioning and thus the amount of order in a latent space for a single factor of variation. The amount of order in the latent space is measured by analyzing the smoothness (lower torsion and arc length) of interpolation paths transcribed in the three vector spaces. Our analysis is based on an autoencoder pose estimation model but has implications outside of this example. Our main contributions are:

1) A methodology for creating interpolation paths in the latent space.
2) Metrics for measuring the interpolative quality of those paths.

The first contribution is specific to pose estimation models like those used in [1], [7]. The second contribution can be applied to any latent space where the ground truth of factors of variation is known.

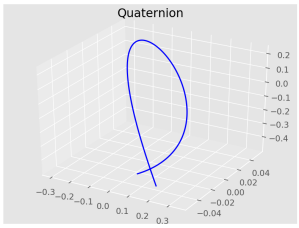
We chose not to use the term interpolation quality but instead interpolative quality because the former implies the quality of the output during interpolation. We, on the other hand, are focused on the relative position of data points in the latent space and the metrics of the interpolation path themselves.

Our approach provides a means of comparison that is not based on disentanglement which is often hard to define, and measure. Contribution 1 also applies outside of just the one application for which it was developed. Additionally, the nature of our approach allows us to compare models despite the added complexities of dealing with distances in the latent space. Finally, our approach shows that, in our test case, the model does demonstrate an ordering of quantifiable factors of variation in the latent space. Figure 2 shows that the arc length and torsion of the autoencoder-generated latent space is lower on average than the random space.

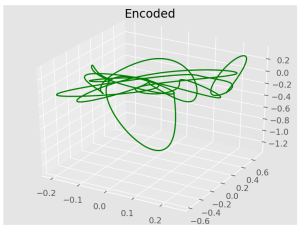
Our analysis shows that the autoencoder-based approach to determining object pose demonstrates spacial organization in the latent space, giving rise to a new method for model quality comparison and assessment for object representations.

II. Disclaimer

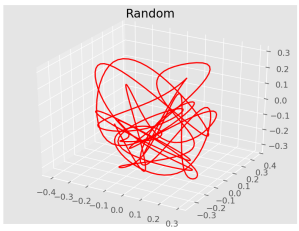
No approval or endorsement of any commercial product by the authors is intended or implied. Certain commercial software systems are identified in this paper to facilitate understanding. Such identification does not imply that these software systems are necessarily the best available for the purpose.



(a) Quaternion Space B-spline



(b) Autoencoder Latent Space B-spline



(c) Random Space B-spline

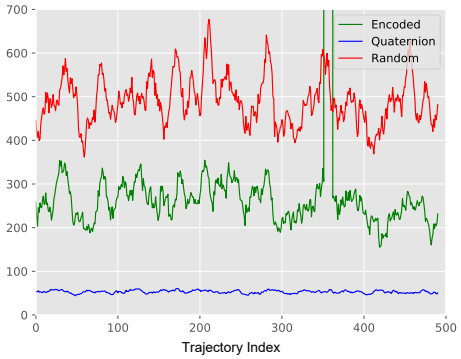
Fig. 1: B-splines are calculated for the quaternion space, the latent space, and the random space. As expected, the random space is the most chaotic with the longest arc length, where the quaternion space is smoother with the shortest arc length, and the autoencoder-generated latent space falls in between.

References

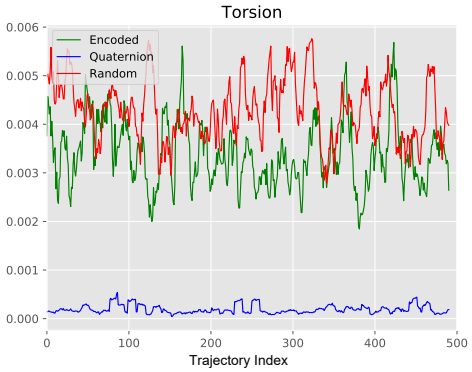
[1] M. Sundermeyer, Z.-C. Marton, M. Durner, M. Brucker, and R. Triebel, "Implicit 3d orientation learning for 6d object detection from rgb images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 699–715.

[2] M. Sundermeyer, M. Durner, E. Y. Puang, Z.-C. Marton, N. Vaskevicius, K. O. Arras, and R. Triebel, "Multi-path learning for object pose estimation across domains," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 13 916–13 925.

[3] C. Eastwood and C. K. Williams, "A framework for the quantitative evaluation of disentangled representations," in *International Conference on Learning Representations*, 2018.



(a) Torsion



(b) Arc Length

Fig. 2: A comparison of trajectories for the single image encoded latent space.

[4] J. Zaidi, J. Boilard, G. Gagnon, and M.-A. Carbonneau, "Measuring disentanglement: A review of metrics," *arXiv preprint arXiv:2012.09276*, 2020.

[5] F. Locatello, S. Bauer, M. Lucic, G. Raetsch, S. Gelly, B. Schölkopf, and O. Bachem, "Challenging common assumptions in the unsupervised learning of disentangled representations," in *international conference on machine learning*. PMLR, 2019, pp. 4114–4124.

[6] A. Sepliarskaia, J. Kiseleva, M. de Rijke *et al.*, "Evaluating disentangled representations," *arXiv preprint arXiv:1910.05587*, 2019.

[7] M. Sundermeyer, Z.-C. Marton, M. Durner, and R. Triebel, "Augmented autoencoders: Implicit 3d orientation learning for 6d object detection," *International Journal of Computer Vision*, vol. 128, no. 3, pp. 714–729, 2020.

Acknowledgment

Dr. Kootbally acknowledges support for this work under grant 70NANB19H009 from the National Institute of Standards and Technology to the University of Southern California.