Staying in Shape: Learning Invariant Shape Representations using Contrastive Learning (Supplementary Material)

Jeffrey Gu[1] **Serena Yeung**[2]

[1]Institute for Computational & Mathematical Eng., Stanford University, Stanford, California, USA
[2]Depts. of Biomedical Data Science and Computer Science, Stanford University, Stanford, California, USA

A EUCLIDEAN ISOMETRIES ARE ORTHOGONAL MATRICES

The isometries of n-dimensional Euclidean space are described by the Euclidean group $E(n)$, the elements of which are arbitrary combinations of rotations, reflections, and translations. One way to describe this structure mathematically is that the group $E(n) = O(n) \rtimes T(n)$ is the semi-direct product of the group of n-dimensional orthogonal matrices $O(n)$ by the group of n-dimensional translations $T(n)$. For the purpose of learning representations from point clouds, it suffices to only consider the non-translation components of $E(n)$ since we can always normalize input point clouds, which has the effect of centering all point clouds at the origin. Mathematically, this is achieved by taking the quotient of $E(n)$ by the translation group $T(n)$, so it suffices to work only with the orthogonal group $O(n) \cong E(n)/T(n)$.

B RIP MATRICES

Here we provide additional characterizations of RIP matrices in terms of the spectral norm and 2-norm. We will find it easier to work with the following definition of RIP matrices:

Definition B.1 (Adapted from Zhao et al. [2020])**.** For all s-sparse vectors $x \in \mathbb{R}^n$, that is vectors x with at most s nonzero coordinates, matrix A satisfies s-restricted isometry with constant δ if

$$(1-\delta)\|x\|^2 \leq \|Ax\|^2 \leq (1+\delta)\|x\|^2 \qquad (1)$$

To see why it makes sense to describe matrices satisfying the RIP condition as almost-orthogonal, we will follow the argument of Zhao et al. [2020]. In our case, our vectors will not be sparse, so we will have s equal to the size of the vector n. Then we can rewrite this condition as

$$\left| \frac{\|Ax\|^2}{\|x\|^2} - 1 \right| \leq \delta, \forall x \in \mathbb{R}^n \qquad (2)$$

Since $\|A\|_2 = \sigma(A)$, where $\sigma(A)$ is the spectral norm of A; that is, the largest singular value of A. Using the min-max characterization of singular values, we know that

$$\sigma(A^TA - I) = \max_{x \neq 0} \frac{x^T(A^TA - I)x}{\|x\|^2} \qquad (3)$$

and simplifying we get

$$\sigma(A^TA - I) = \max_{x \neq 0} \frac{\|Ax\|^2}{\|x\|^2} - 1 \qquad (4)$$

Plugging this in to Equation 2, we get

$$\sigma(A^TA - I) \leq \delta \qquad (5)$$

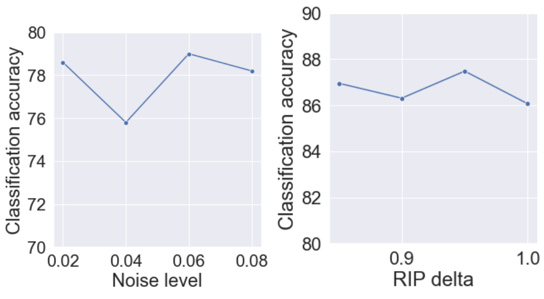
From this equation, we can see that RIP matrices are almost-orthogonal, and therefore almost-isometric, with respect to the spectral norm.

C HYPERPARAMETER SENSITIVITY

We investigate the sensitivity of our model to the Gaussian noise parameter (standard deviation) σ for Gaussian perturbations and the stretching parameter δ for RIP matrices. Results can be found in Figure and Figure, respectively. We find that the performance of our model is not heavily effected by the choice of either parameter.

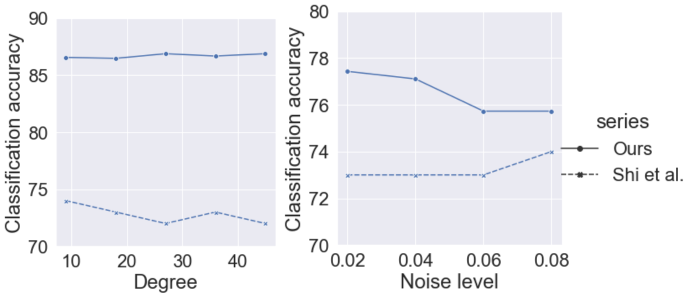
D ROBUSTNESS COMPARISON TO BASELINE

Results for the rotation and Gaussian perturbation robustness experiments on ModelNet40 of Section 4.2 using the baseline method [Shi et al., 2020] can be found in Figure 2. An identical experiment was carried out in their paper, except the classification part (see Section 4.1) was carried out on ShapeNet instead of ModelNet40. The experiments were carried out using their publicly available implementation here: `https://github.com/WordBearerY`



(a) (b)

Figure 1: Hyperparameter sensitivity plots for (a) σ, the standard deviation of Gaussian noise in the random Gaussian perturbation augmentation, and (b) δ, the deviation from isometry for our random RIP augmentation. We find that our model is not particularly sensitive to either hyperparameter.



(a) (b)

Figure 2: Plots of accuracy vs variation strength for (a) rotations by a fixed angle, (b) Gaussian noise of varying standard deviations for the baseline Shi et al. [2020]. We see that the method is fairly robust but less accurate than our method. One caveat is that we were unable to fully reproduce their results using their publicly available code.

`I/Unsupervised-Deep-Shape-Descriptor-with-Point-Distribution-Learning`. We find that differing amounts of Gaussian noise do not affect the classification accuracy, contrary to their results on ShapeNet where as increasing rotations have a slight negative effect on classification accuracy, which reflects their ShapeNet results. We note that we were unable to reproduce their result in Table 3 with their code. With the results we were able to produce, we find that our model has similar robustness but much better accuracy than Shi et al. [2020]. We will also make our code publicly available.

E POINTNET ENCODER ARCHITECTURE

A exact specification of our PointNet Qi et al. [2017] encdoer architecture can be found in Table 1.

F EXAMPLES OF TRANSFORMATIONS

In Figure 4 we provide additional examples of randomly sampled transformations from each of our proposed data augmentation methods, which are the uniform orthogonal transformation, random RIP transformation, and smooth perturbation transformation.

G FAILURE CASES

In Figure 3 we show examples from ModelNet40 that were misclassified by our method, and similar examples from the class it was misclassified as. The highest error rate ModelNet40 class is the flower pot class, which has an error rate much higher than any other class. Our method frequently mistakes the examples from the flower pot class for the plant class, which is much larger, and more rarely as other classes. As shown in Figure 3, examples from one class can be very similar visually to an example from another class, and we believe that this similarity is challenging for contrastive learning algorithms.

References

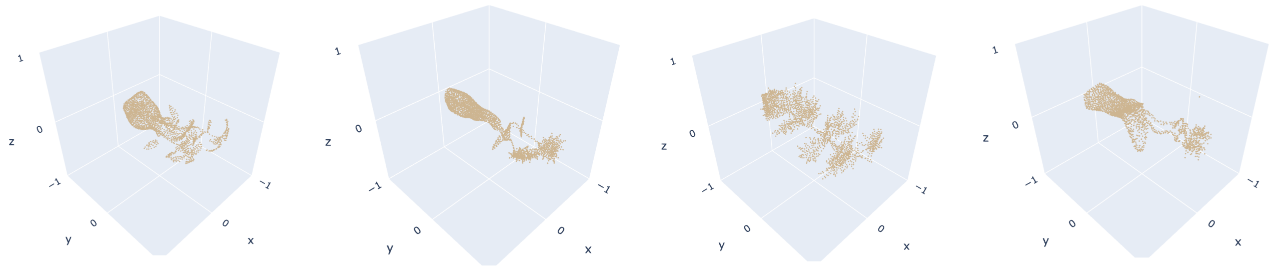
Charles R Qi, Hao Su, Kaichun Mo, and Leonidas J Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 652–660, 2017.

Yi Shi, Mengchen Xu, Shuaihang Yuan, and Yi Fang. Unsupervised deep shape descriptor with point distribution learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9353–9362, 2020.

Yue Zhao, Yuwei Wu, Caihua Chen, and Andrew Lim. On isometry robustness of deep 3d point cloud models under adversarial attacks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1201–1210, 2020.

Table 1: The PointNet encoder architecture used for all versions of our model. Each layers is followed by a batch normalization layer and a ReLU layer except for the last two linear layers. The identity is added to the third linear layer as in Qi et al. [2017], and the output is reshaped at the before the second block of 1D convolutions. C is the number of classes for classification.

Layer Type	In channels	Kernel size	Stride	Out channels
Conv1D	3	1	1	64
Conv1D	64	1	1	128
Conv1D	128	1	1	1024
Linear	1024	–	–	512
Linear	512	–	–	256
Linear	256	–	–	9
Conv1D	3	1	1	64
Conv1D	64	1	1	128
Conv1D	128	1	1	1024
Linear	1024	–	–	C



(a)

(b)

(c)

(d)

Figure 3: (a) and (b) are examples of the flower pot class that are misclassified by our method as the plant class, and (c) and (d) are similar looking examples from the plant class.

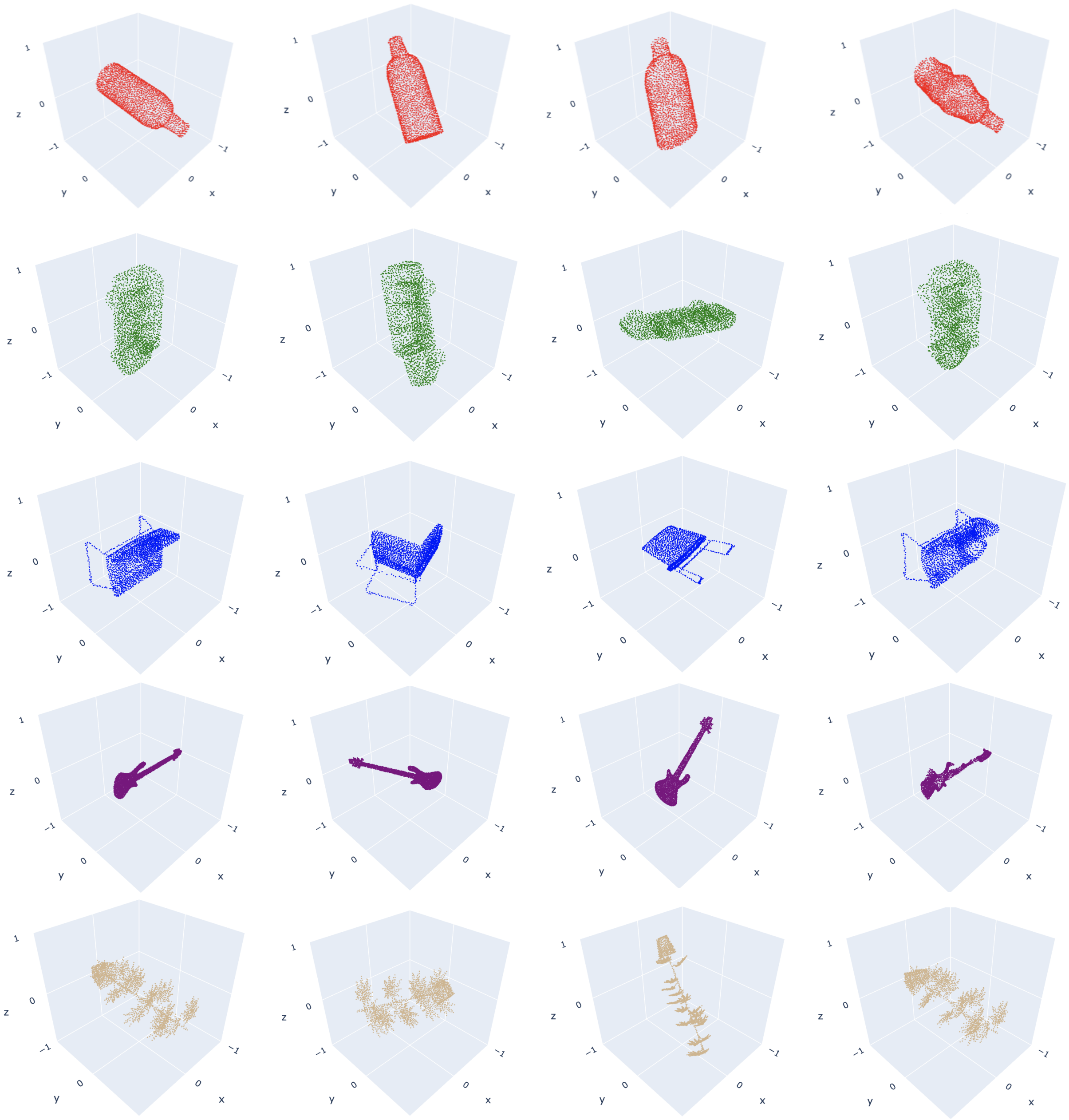


Figure 4: Additional examples of randomly sampled uniform orthogonal, random RIP, and smooth perturbation transformation using our methods. In the first column from the left is the original image. In the second, third, and fourth columns from the right, we apply a randomly sampled orthogonal, RIP, and smooth perturbation transformation, respectively. We see that in general that the orthogonal transform rotates and possibly reflects the object, that the RIP transform generally rotations and slightly elongates the object, and that the smooth noise smoothly deforms the objects.