Supplementary Material:
Interpretable Transformations with Encoder–Decoder Networks

Daniel E. Worrall
d.worrall@cs.ucl.ac.uk

Stephan J. Garbin
s.garbin@cs.ucl.ac.uk

Daniyar Turmukhambetov
d.turmukhambetov@cs.ucl.ac.uk

Gabriel J. Brostow
g.brostow@cs.ucl.ac.uk

Computer Science Department
University College London

Abstract

Here we present classification performance on the ModelNet10 dataset and the mathematical definition of the homomorphism property *from the main paper.*

1. ShapeNets (ModelNet10) classification accuracy

The ModelNet10 classification task is evaluated on 908 models from the test set. For this task we trained the Modelnet architecture autoencoder with a 2-layer MLP (256-128-10) on the relative phase between all subvectors of the codes.

We minimize the sum of two losses: cross-entropy loss for classification and the reconstruction loss. We follow [3] for the binary cross-entropy reconstruction loss:

$$\mathcal{L}_{\text{recon}} = \sum_{i \in \text{voxels}} -\gamma t_i \log(o_i) - (1-\gamma)(1-t_i)\log(1-o_i), \tag{1}$$

where t_i are the target values rescaled to $[-1,2]$, o_i is the output of the autoencoder rescaled to [0.1,0.9999] and γ is set to 0.98 to compensate for the sparseness of volumetric data. Thus, the loss is:

$$\mathcal{L} = \mathcal{L}_{\text{recon}} + 10\mathcal{L}_{\text{classification}} \tag{2}$$

We optimize the loss using Adam and minibatch size 16, and learning rate of 10^{-4}. We use the augmentation strategy of Maturana *et al*. [6].

We accurately classify 821 models out of 908, with an accuracy of 90.4%.

2. The Homomorphism Property

The homomorphism property (Equation 6) is

$$\mathbf{F}_{\theta_2\theta_1} = \mathbf{F}_{\theta_2}\mathbf{F}_{\theta_1}. \tag{3}$$

Thus if $I \in \Theta$ is the identity transformation, then

$$\mathbf{F}_\theta = \mathbf{F}_{I\theta} = \mathbf{F}_{\theta I} = \mathbf{F}_I\mathbf{F}_\theta = \mathbf{F}_\theta\mathbf{F}_I \tag{4}$$

$$\implies \mathbf{F}_I = \mathbf{I}, \tag{5}$$

where $\mathbf{I}$ is the identity matrix. This in turn implies the invertability property $\mathbf{F}_{\theta^{-1}} = \mathbf{F}_\theta^{-1}$, since

$$\mathbf{I} = \mathbf{F}_I = \mathbf{F}_{\theta\theta^{-1}} = \mathbf{F}_\theta\mathbf{F}_{\theta^{-1}} \tag{6}$$

$$\implies \mathbf{F}_{\theta^{-1}} = \mathbf{F}_\theta^{-1}. \tag{7}$$

Method	Accuracy
VRN Ensemble [3]	97.14%
ORION [7]	93.8%
LightNet [1]	93.39%
FusionNet [4]	93.11%
Pairwise [5]	92.8%
GIFT [2]	92.35%
VoxNet [6]	92%
3D-GAN [8]	91.00%
Ours	90.4%

Table 1. State of the Art methods and their classification accuracy on ModelNet10 benchmark.

References

[1] A lightweight 3d convolutional neural network for real-time 3d object recognition. *http://modelnet.cs.princeton.edu/*. 2

[2] S. Bai, X. Bai, Z. Zhou, Z. Zhang, and L. Jan Latecki. Gift: A real-time and scalable 3d shape search engine. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5023–5032, 2016. 2

[3] A. Brock, T. Lim, J. Ritchie, and N. Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016. 1, 2

[4] V. Hegde and R. Zadeh. Fusionnet: 3d object classification using multiple data representations. *arXiv preprint arXiv:1607.05695*, 2016. 2

[5] E. Johns, S. Leutenegger, and A. J. Davison. Pairwise decomposition of image sequences for active multi-view recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3813–3822, 2016. 2

[6] D. Maturana and S. Scherer. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition. In *IROS*, 2015. 1, 2

[7] N. Sedaghat, M. Zolfaghari, and T. Brox. Orientation-boosted voxel nets for 3d object recognition. *arXiv preprint arXiv:1604.03351*, 2016. 2

[8] J. Wu, C. Zhang, T. Xue, B. Freeman, and J. Tenenbaum. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling. In *Advances in Neural Information Processing Systems*, pages 82–90, 2016. 2