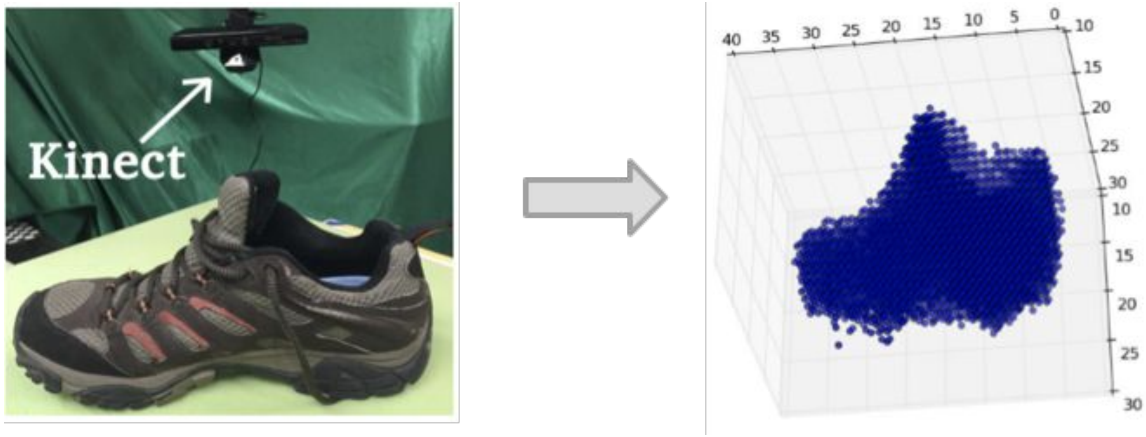
3D Object Reconstruction via Latent Space Recovery

Given a partial depth view of an object (e.g., from Kinect), reconstruct a prediction of the full object.

- Representation: Voxel - discretized 3D space.
- Approach: Neural Networks + Optimization.

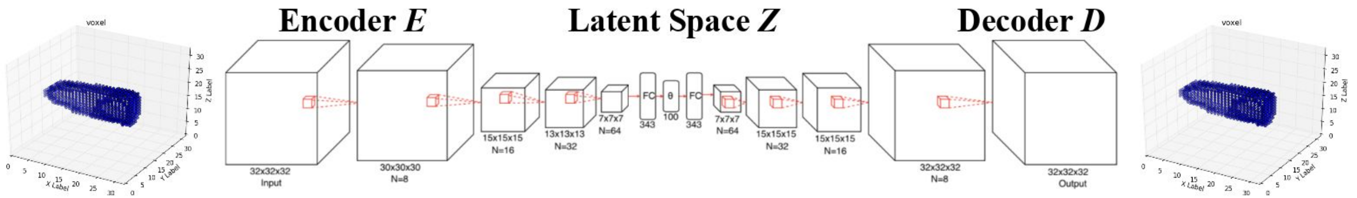


Jiahui Chen, Mark Van der Merwe

Related Work

- Representation: Voxel [1,5,6], Point-Cloud, Mesh, and Implicit [3,4] 3D Representations.
- Approach: Direct Regression [5,6], Learning Embedding[1,3,4], and Optimization over Generative Network [2].
- Inputs to reconstruction problem: 2.5D [3,4,5], RGB [3,6].
- Applications: robotic grasping [5], reconstruction for reconstruction's sake [3,4,6].
- Structure of the autoencoder we used in approaches 1 & 2 was adapted from [1].
- Latent space recovery and latent vector optimization in approach 3 was inspired by [2].

Overview of Approaches



Let x be the full 3D object and x^* be the partially viewed 3D object. We would like to recover the true latent z:

$$z = E(x)$$

1. Partial to latent via Encoder trained on Full Views

$$z \approx E(x^*)$$

2. Partial to latent via Encoder trained on Partial Views

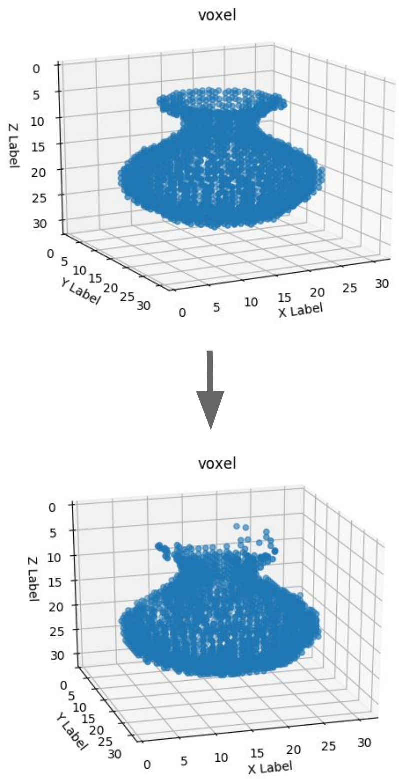
$$z \approx F(x^*)$$

3. Optimization of latent space using Decoder and latent vectors of Full Views.

$$z \approx \mathrm{argmin}_z \mathcal{L}(D(z), x^*)$$

Results: Reconstruction Error on Full (F1)

	F1
Training Set (MN40)	0.764661640947
Validation Set (MN40)	0.758852964618
Test Set (MN40)	0.757912622477
YCB	0.705109139338



Results: Reconstruction Error on Partial (F1):

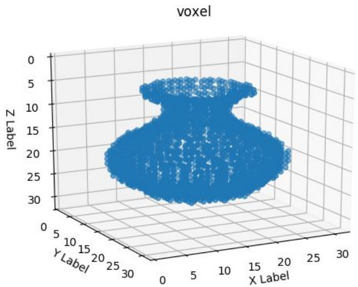
	$z \approx E(x^*)$	$z \approx \text{opt}(E(x^*))$	$z \approx F(x^*)$	$z \approx \text{opt}(F(x^*))$
Training Set (MN40)	0.447567994181	0.470707842207	0.671892934301	0.685943800753
Validation Set (MN40)	0.441867692329	0.471083651243	0.661604040081	0.662987213014
Test Set (MN40)	0.445513516677	0.47933072912	0.656182251038	0.666070905061
YCB	0.404187364693	0.44343301744	0.567149598346	0.637112220782

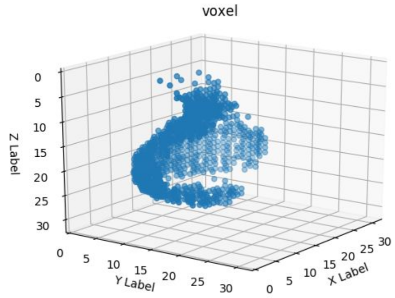
Results: Reconstruction (Qualitative):

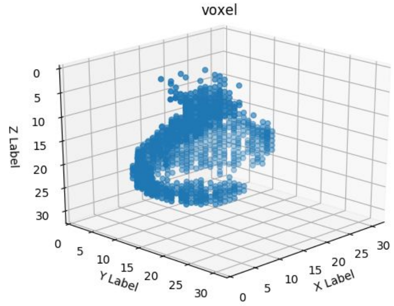
Full Voxel:

$z \approx E(x^*)$

$z \approx \mathrm{opt}(E(x^*))$



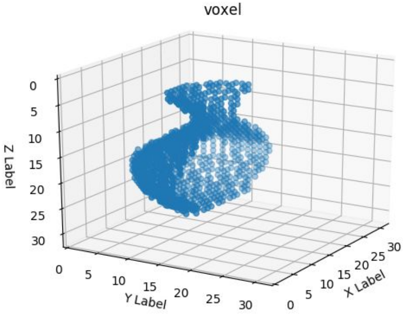


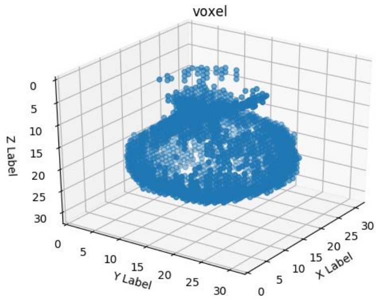


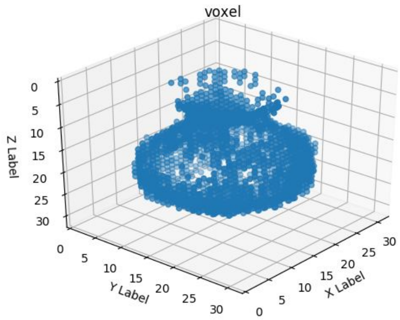
Partial Voxel (Input):

$z \approx F(x^*)$

$z \approx \mathrm{opt}(F(x^*))$







Conclusions & Future Work

- What Did Work:
  - Autoencoder able to learn complex embedding for objects.
  - Training a new encoder for partial to latent allowed for good reconstruction.
  - Gradient Optimization over Latent Space had marginal improvement.

- What Didn’t Work:
  - Aligning to PCA principal axis removed valuable context.
  - Autoencoder on its own was not able to reconstruct - learned “too well.”
  - Slicing did not generalize to true partial views.

- Future Work:
  - Train on larger datasets + true partial views.
  - Incorporate more context into the inference problem.

References

[1] Andrew Brock, Theodore Lim, James M Ritchie, and Nick Weston. Generative and discriminative voxel modeling with convolutional neural networks. *arXiv preprint arXiv:1608.04236*, 2016.

[2] Fangchang Ma, Ulas Ayaz, and Sertac Karaman. Invertibility of convolutional generative networks from partial measurements. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 9628–9637. 2018.

[3] Lars M. Mescheder, Michael Oechsle, Michael Niemeyer, Sebastian Nowozin, and Andreas Geiger. Occupancy networks: Learning 3d reconstruction in function space. *CoRR*, abs/1812.03828, 2018.

[4] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. *arXiv preprint arXiv:1901.05103*, 2019.

[5] Jacob Varley, Chad DeChant, Adam Richardson, Joaquín Ruales, and Peter Allen. Shape completion enabled robotic grasping. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2442–2447. IEEE, 2017.

[6] Hanqing Wang, Jiaolong Yang, Wei Liang, and Xin Tong. Deep single-view 3d object reconstruction with visual hull embedding. *arXiv preprint arXiv:1809.03451*, 2018.

Dataset

Full Embedding Round 1:

- Full Dexnet Model Set - 13,252 objects.
- Data augmentation - align by principal axis (PCA). Rotate.
- Issue - embedding too exact, training slow.

Full Embedding Round 2:

- ModelNet40 - 2,539 objects. Tested generalization on YCB - 80.
- Data augmentation - only rotate around initial z axis.

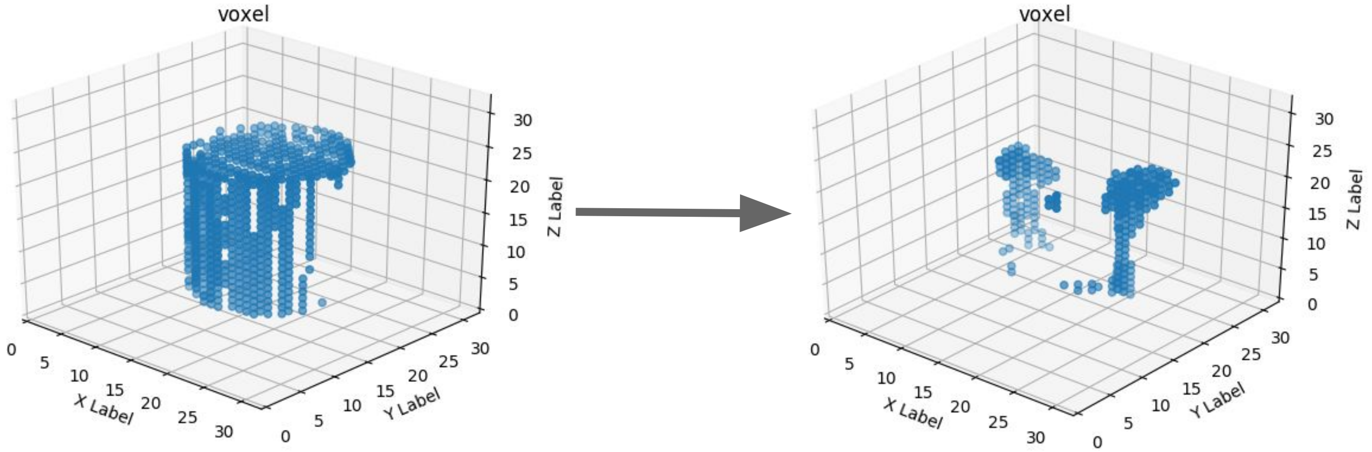
Partial View:

- Take each object from above sets, remove all voxels where y>16.

Results: Latent Space Recovery (MSE):

	$z \approx E(x^*)$	$z \approx \text{opt}(E(x^*))$	$z \approx F(x^*)$	$z \approx \text{opt}(F(x^*))$
Training Set (MN40)	15.214705	14.25329	13.025468	12.448365
Validation Set (MN40)	15.408493	15.402458	13.606855	13.664517
Test Set (MN40)	15.305848	15.144078	13.298943	13.398145
YCB	27.77371	27.648096	21.398005	21.443647

Results: Partial View (Gazebo) Reconstruction:



Partial view contains additional voxels which our slicing method did not generalize to.