Classifying 3D objects as a whole

JunYoung Gwak, Christopher Choy (not in class)
This project is part of a research at CVGL, Stanford

ABSTRACT

In this project, we approached **classification of a 3D objects by using the whole voxel representation as an input to the neural network**. There has been a lot of works that tries to classify a 2D image to correct class but there was a little effort to classify 3D model as a whole. This is partially due to lack of training data and difficulty of building intuitive hand-engineered features on 3D space. We have access to enough data with advent of **ShapeNet** and **random data augmentation**. With enough data, we could approach this problem without engineering features using **3D convolutional neural network**.

NETWORK

We convoluted on 3D filters to use a whole voxel as an input to the network. Our network is configured as following:

(3dconv + ReLu + pool) x 4

+ (FC +ReLu + Dropout) x 2 + SoftMax

Input: (32, 1, 32, 32, 32)
Conv1 filter: (32, 1, 3, 3, 3), Pool1 filter: (2,2,2)
Conv2 filter: (64, 32, 3, 3, 3), Pool2 filter : (2,2,2)
Conv3 filter: (128, 64, 3, 3, 3), Pool3 filter : (2, 2, 2)
Conv4 filter: (256, 128, 3, 3, 3), Pool4 filter : (2, 2, 2)
FC5 output: 1024, FC6 output: 1024

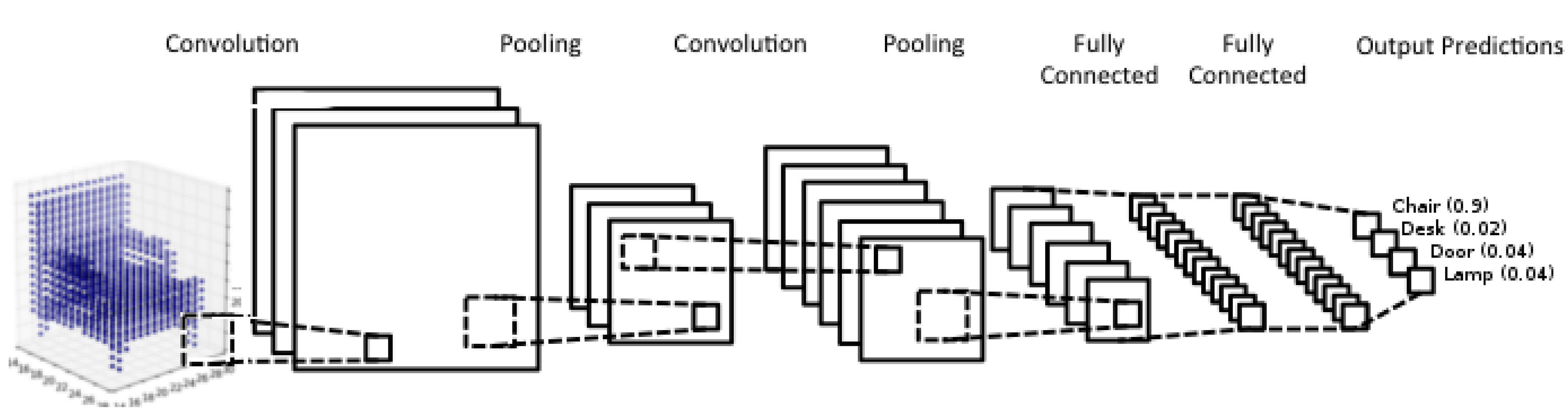

Fig 1. Abstract overview of our 3D convolutional neural network making decision in classification

ANALYSIS

Through projection of activation of a channel with respect to the input voxel, we could analyze what the network has learned.

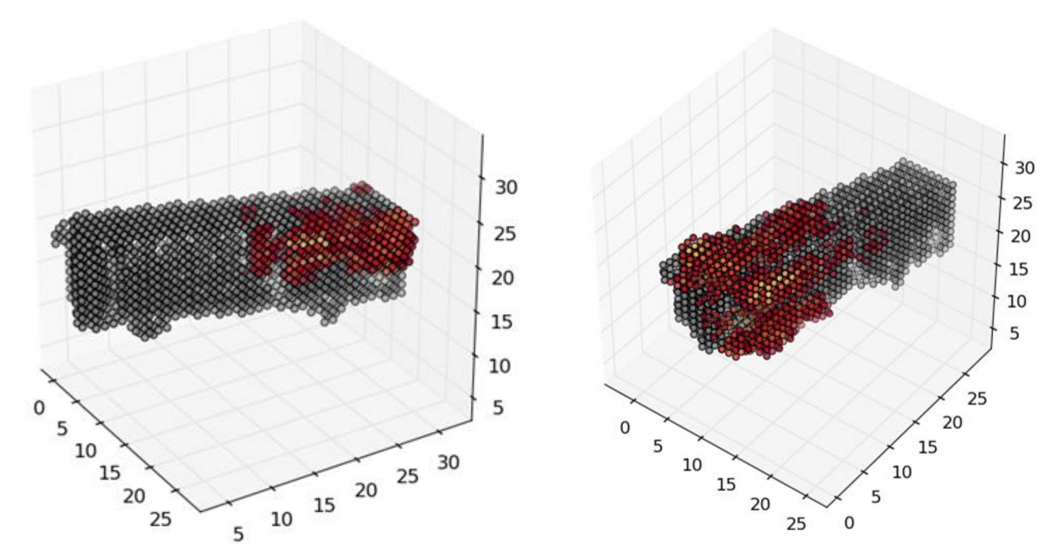

Fig 5. Projection of activation of different neurons with respect to the same voxel of bus.

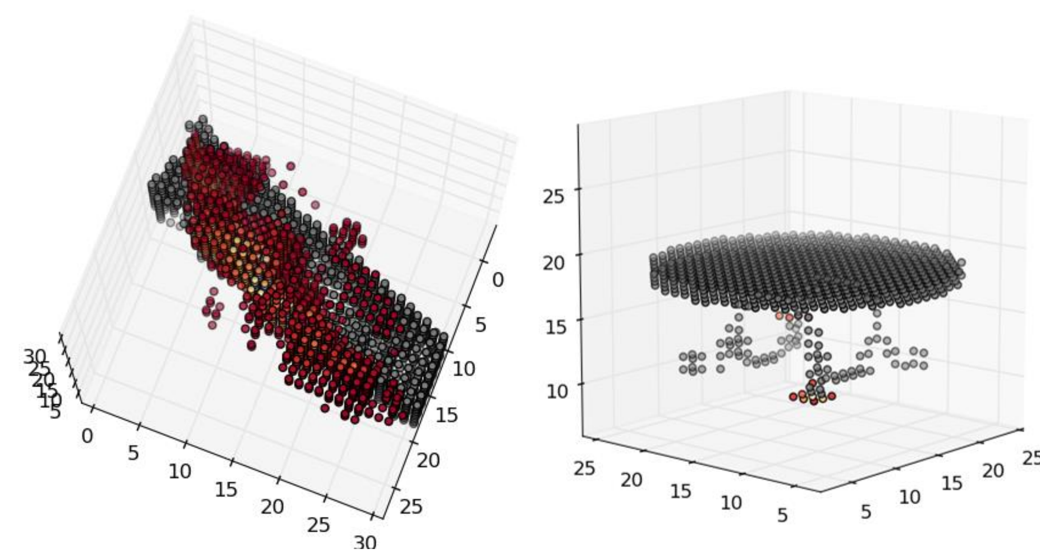

Fig 6. Projection of activation of same neuron with respect to different input voxels.

RELATED WORK

Few works has been made to classify 3D object

- Atmosukarto et al. mapped 3D into 2D depth map and learned salient points to classify 3d models.
- Kassimi et al. utilized high-level semantic annotation and low-level voxel clusters to classify 3D models.
- Wu et al. approached this problem with generative model of class prior and convolutional deep belief net. Ours is discriminative model thus expected to perform better than this approach.

INPUT DATA

We use **ShapeNet**, a richly-annotated, large-scale dataset of 3D shapes. With ShapeNetCore, we have access to 57k models in 57 classes.

We **augment data** by applying random scale, yaw rotation, and translation transformation to the model on the fly.

The scale of random yaw rotation(i.e., the pose of the model) is utilized to estimate the pose of the model as an extension of this work.

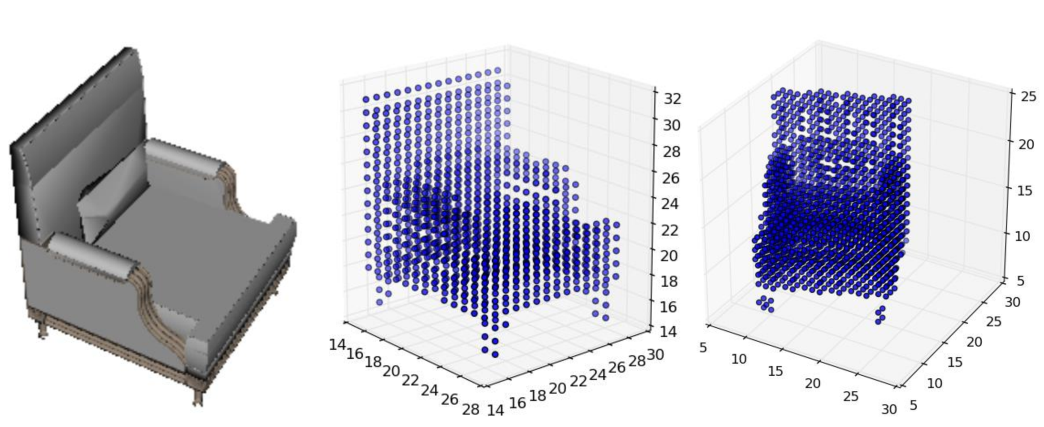

Fig 2. 3D model from ShapeNet(left), Voxelized(center), Data augmented with random transformation(right)

RESULT

57 classes classification result:

Test error: **0.2046**

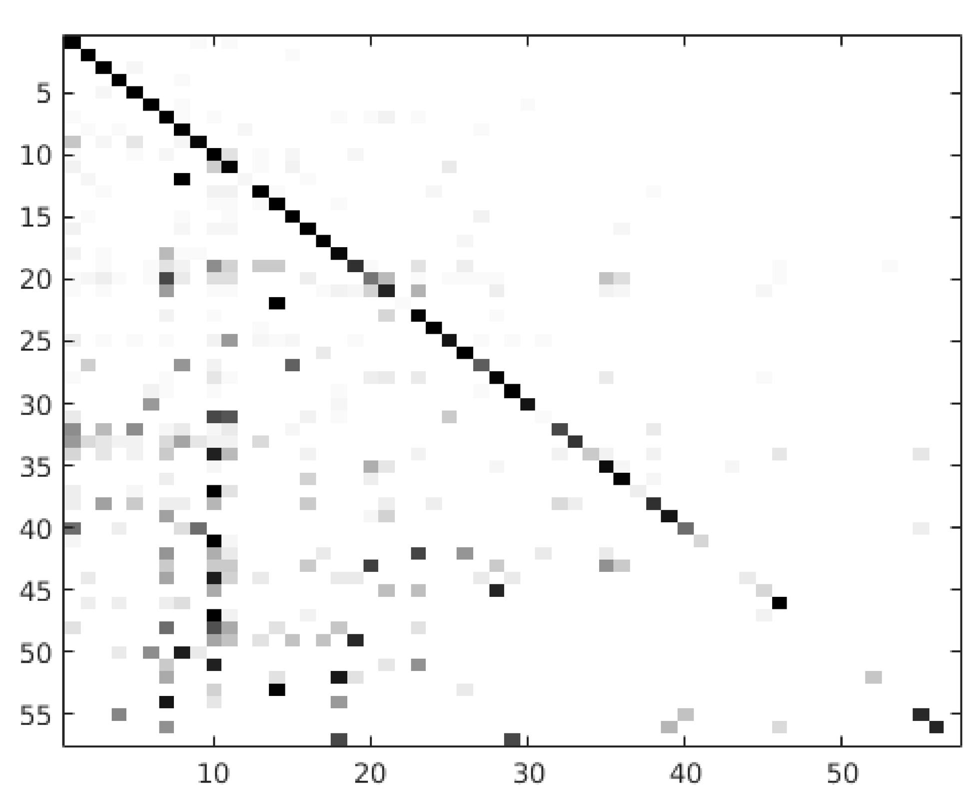

Fig 3. Confusion Matrix of classification result. Index sorted by number of models in each class.

As shown in Fig 3, there exists a strong preference in classification on classes with more models. This is due to a bias in number of models in ShapeNet, varying from 44 to 6807. Also, there exists some redundant classes in ShapeNet such as “vessel” and “boat”. We are currently running the test again with cleaned-up data.

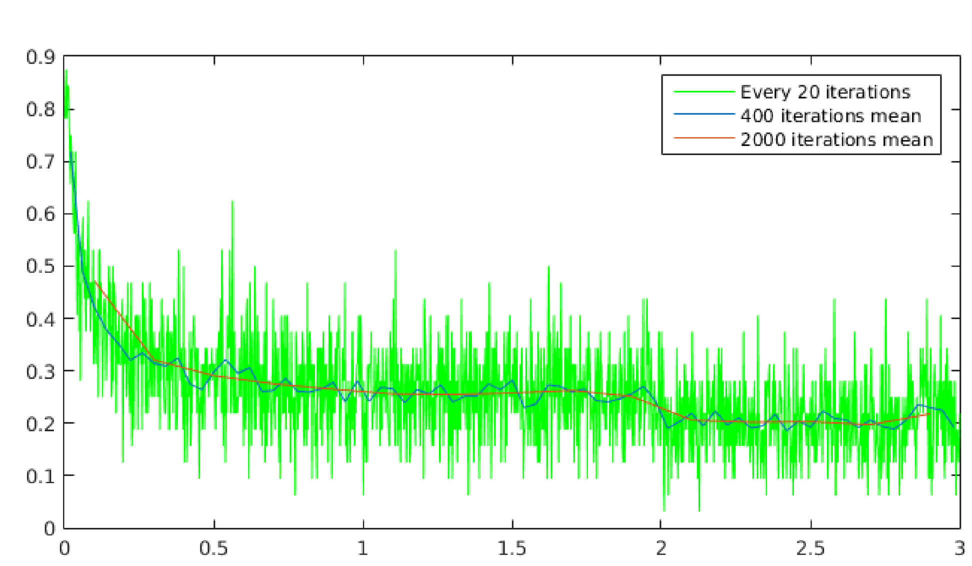

Fig 4. Training loss vs # iteration.

POSE ESTIMATION

By adding another SoftMax layer classifying model pose into 8 bins of $[-\pi, \pi]$ and defining cost as **total_cost = classification_cost + λ pose_cost**, we enforced the network to learn model pose along with its class. We observed that pose errors come from bins nearby or 180° apart.

Test error: **0.1264**

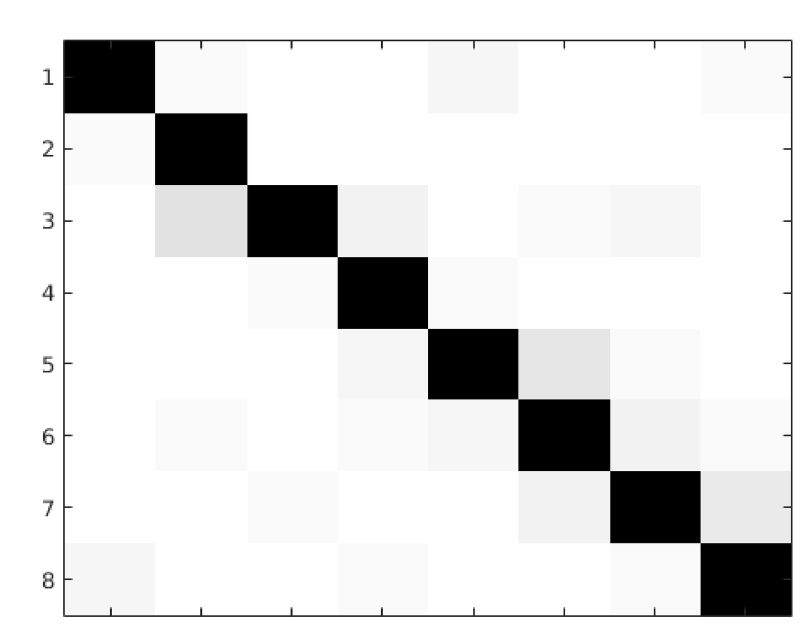

Fig 7. Confusion Matrix of pose estimation result

CONCLUSION

- We demonstrated that classification of 3D objects as a whole can be achieved with low error using 3d convolutional neural network.
- We analyzed that the network has learned simple gradients on layers closer to input and class-specific data on layers closer to output.
- As an extension of this work, we additionally enforced the network to learn the pose of the model by adding another softmax with weighted cost.