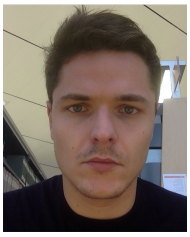
3D Object Recognition with Deep Networks

3D Photography Project Proposal
Supervised by: Martin Oswald, Pablo Speciale
March 7, 2015

GROUP MEMBERS

Tobias Grundmann Adrian Schneuwly Johannes Oswald





I. DESCRIPTION OF THE PROJECT

The goal of this project is to successfully implement 3D object recognition using a Convolutional Neural Network (CNN) approach [3]. Such networks can be trained to recognize complex 3D shapes and determine the corresponding object class given a volumetric representation. This project will mainly use the large-scale 3D CAD model data set called ModelNet [4]. Conceivably, volumetric data can also originate from inexpensive 2.5D depth sensors like Microsoft Kinect, Google Project Tango or Intel RealSense (Fig. 1).

II. WORK PACKAGES AND TIMELINE

A. *Prerequisites in March*

The first goal of the team is to fully understand the approach described in [3]. Therefore everyone in the team will work through the Udacity deep learning course [2] individually in order to gain knowledge and hands-on experience with deep networks using 2D data. Further, we will get familiar with TensorFlow[1] (in Python), which will be used to implement the network architecture. As a first simple exercise to approach the problem we will reproduce the character recognition example[1] using the MNIST data set.
The division of work among the 3 project members will happen after this initial project phase, because at this point in time it is hard to foresee how the the work can be split meaningfully.

[1]open source library for machine intelligence

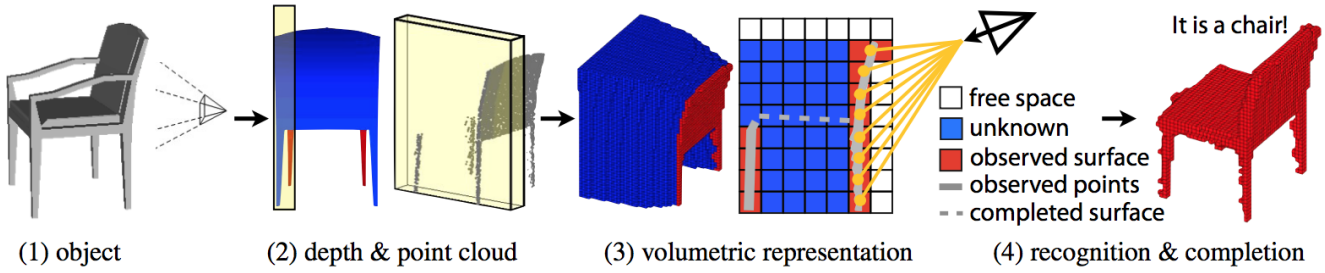


Fig. 1. 2.5D object recognition

B. Data Preparation & Modeling in April

Deep Networks heavily depend on big datasets. Therefore we are going to use the ModelNet dataset from [4], a CAD dataset of 660 objects. The datasets will transformed into 32x32x32 voxel data and split into a training, valuation and test set. If possible we will use data from the project tango tablet to cross test. We begin reimplementing the VoxNet Deep Convolutional Neural Network approaches in Python by starting with the deep learning udacity courses framework. [2] Therefore we will remodel the given 2D object recognition network into a rotation-invariant 3D recognition network which makes heavily use of convolution and pooling. The output of the deep network is the the class of object the network thinks it recognized from the input.

C. Training & Testing in June

After successful implementation and testing, we will train the deep network with a subset of 40 classes called the ModelNet40 (optional: [4] offers an even larger data set).

Due to the computationally expensive training of the Deep Network on the large CAD dataset, we will try to start as soon as possible with this phase of the project.

III. Outcomes and Demonstration

Goal of the project is successfully reimplement the papers approaches for 3D Object Recognition and achieve very similar positive object recognition results. In our live demo, we will snapshot random CAD Models and try to recognise the object through our algorithm.

References

[1] https://www.tensorflow.org/versions/r0.7/tutorials/mnist/beginners/index.html.

[2] https://www.udacity.com/course/deep-learning–ud730.

[3] D. Maturana and S. Scherer. Voxnet: A 3d convolutional neural network for real-time object recognition. *International Conference on Intelligent Robots and Systems (IROS2015)*, 2015.

[4] Z. Wu, S. Song, A. Khosla, L. Zhang F. Yu, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. *Proceedings of 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR2015)*, 2015.