Bachelor Project Proposal: Auto-encoder-based Representation Learning for 3D Object Recognition in Open-Ended Domains

Vlad Iftime (S3426394)
Dept. of Artificial Intelligence
University of Groningen, the Netherlands.
`v.c.iftime@student.rug.nl`

Hamidreza Kasaei*
Dept. of Artificial Intelligence
University of Groningen, the Netherlands.
`hamidreza.kasaei@rug.nl`

This research is aimed at recognizing and categorizing 3D objects. In this project, we will try to categorize different objects and also learn new categories, when a new type of object is introduced. This type of recognition and learning is vital in robotics applications since there is never enough training data that will prepare the robot for all types of objects that it might encounter in a real-life scenario. In this work, we will try to look at how an autoencoder-based model performs in categorizing 3D objects in open-ended domains. Thus, the research question that we will try to answer is: *Which autoencoder models perform best for learning and recognizing 3D objects in open-ended domains?*. The performance of the agent will be assessed based on four metrics: *how much does it learn?*, *how well does it learn?*, *how fast does it learn?*, *how much memory does it take?*.

The requirements of the relevant Bachelor project include a literature review within the discipline, the implementation of relevant code, experimentation with benchmarks as well as writing a thesis. We mainly use C++ based ROS as the main programming language. In addition to these, we will also use python and Keras framework for deep learning. In this study, we assume that an object has already been segmented from the point cloud of the scene, and we will mainly focus on detailing 3D object category learning and recognition.

To give a brief overview of the stages that make up the research, we begin by using the orthographic projection to obtain the views of the object.We then feed these orthographic views to an Autoencoder model to create a compact representation for the given object. The obtained representation is finally used for both learning and recognition processes. Before the system can categorize, the autoencoder needs to be trained to provide unique object representation. In particular, we implement a representation learning approach. Below we present a representation of a autoencoder model.

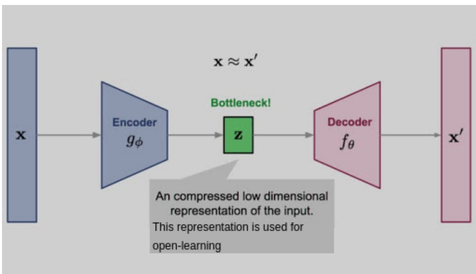

Figure 1: Autoencoder model

X represents one view of an object feed into the network. *Z* is the low dimension representation of the object. From this representation the network will then try to rebuild the image, thus obtaining *X'*. The network training is done by minimizing the error between *X* and *X'*. The network is trained when we are confident enough that the representation it provides is similar with the input(i.e. the difference between *X* and *X'* is small enough). After the network is trained the representation learning is done by categorizing the different *Z*'s that we obtain. Finally, in order to show the applicability of the system, we will test it on a real-life robot.

*H. Kasaei is a Faculty of Science and Engineering at the University of Groningen, the Netherlands (www.ai.rug.nl/hkasaei)