Towards unsupervised object Classification and Prediction in 3D through Semantic Sampling.
R.F.H. Ormesher

Abstract

In this thesis I will investigate the potential use of pre-existing Volumetric Variational Auto-Encoder architectures for object in-filling and de-noising. From the experiments presented here it can be seen that even with relatively simple architectures, complex and varied noises can be repaired by learning generative latent spaces from training with data augmentation. For further improving the VAE's predictive abilities, I propose two novel redefinition of the Variational Bayes Auto-Encoder architecture for management of partial, semantically scaled input samples. The Located-VAE (LVAE) and Prior-VAE (PVAE) are extensions of variational reconstruction networks that attempt to connect real-world sliding window object segments to a latent space of known 3D objects for classification and prediction. Their predictive abilities are shown visually through use of the Classification and Prediction through Auto-Encoder Network (CaPtAEN) application for basic reconstruction tasks, as well as reconstruction with varying noise qualities at input. The classification abilities are demonstrated empirically through comparison of latent space representations of segments taken from the same object. Finally, we argue that although voxel models are visually interesting to work with, the computational complexity and massive sparsity are prohibitive for working with high-resolution models and prevent learning of structured high-level 3D filters. The lack of filter descriptiveness is visually explained using the application presented in this work.