Chapter 9

Conclusions

9.1 Summary

In this dissertation, we have covered convolutional neural networks for various stages of 3D perception. We first introduced 3D reconstruction networks that generate voxelized objects from multi-view images. In Chapter 2, we trained these networks with direct 3D supervision, and with weaker supervision, silhouettes and an unlabeled 3D shape repository in Chapter 3. As we do not have sufficient 3D reconstructions to supervised these networks, using a weaker form of supervision is one of the most critical factors scaling up the system.

In the second part of the dissertation, we present sparse tensor networks, neural networks for spatially sparse tensors in Chapter 4. As we increase the spatial dimension, the sparsity of input data decreases drastically as the volume of the space increases exponentially. Sparse tensor networks exploit such inherent sparsity in the input data and efficiently process them. There are many applications of these sparse tensor networks for 3D representation learning. In Chapter 5, we discussed how to use sparse tensor networks for 3D and 3D spatio-temporal segmentation. We achieved state-of-the-art performance on the large-scale indoor semantic segmentation datasets and show that 4D spatio-temporal convolutional networks can effectively make use of temporal consistency and improve the accuracy of segmentation. In Chapter 6, we apply sparse tensor networks for geometric feature learning. We propose the first fully-convolutional neural networks for 3D correspondences and the hardest contrastive and hardest triplet losses that significanly improve the accuracy of the metric features. Experimentally, sparse tensor networks outperform the state-of-the-art while being a few orders of magnitude faster.

We extend these sparse tensor networks for geometric registration problems in the last part of the dissertation. First, we present high-dimensional convolutional networks for geometric pattern recognition in high-dimensional problems. We validate the sparse tensor networks on high-dimensional linear regression, 6-dimensional hyper-surface detection for 3D registration, and 4-dimensional hyper-conic section detection for image correspondences. Next, we extend the 6D geometric pattern recognition networks to a learning-based framework that aligns real-world 3D scans robustly and accurately. To achieve this, we propose a differentiable Weighted Procrustes analysis for closed-form pose estimation and a gradient-based optimizer for pose refinement.

9.2 Final Remarks

Many problems in computer vision and machine learning involve finding patterns in high-dimensional spaces. Specifically, we covered high-dimensional problems in 3D perception and tackled them with high-dimensional convolutional networks. By converting each stage into a learnable model with inductive biases that capture the inherent geometry of each problem, we explore an end-to-end network for registration by converting each component into a fully differentiable model. However, it is possible to use the 3D reconstruction for the down-stream 3D perception pipeline and train end-to-end. Such models can reconstruct, register, and understand scenes.

We have explored a limited set of problems in 3D perception. However, other types of 3D perception tasks such as object detection and tracking are of great interest in robotics, autonomous driving, and augmented reality applications. Within these applications, there are high-dimensional geometric patterns and the use of high-dimensional convolution will become more vital.

Lastly, high-dimensional data exists in not just in computer vision, but also in machine learning and statistical analysis as well. We will explore the potential application of the sparse tensor network in these areas where spatial geometry plays a key role.