

Topic: 3D Deep Learning

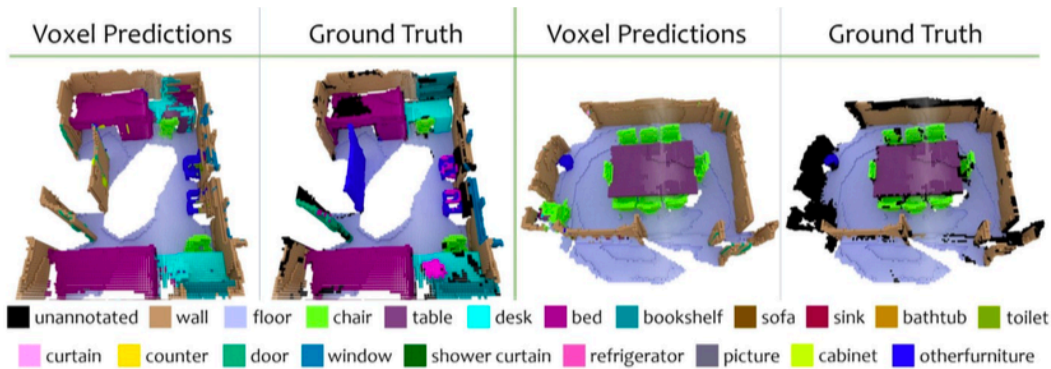


Figure 3: Semantic voxel annotations and predictions on test scenes using the ScanNet dataset [11]. In this proposal, we will extend this line of research by adding a fourth dimension that represents dynamic data, which is important as the real world is constantly in motion; however, the extra dimension also comes with additional memory and compute challenges.

Project Supervisors and affiliations:

Prof. Evgeny Burnaev, CDISE CREI Skoltech

Prof. Denis Zorin, CDISE CREI Skoltech, Courant Institute of Mathematical Sciences at New York University

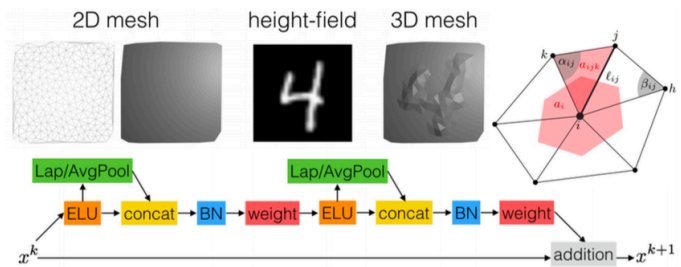


Figure 2: Our Surface Networks [32] operate directly on the manifold mesh by defining ResNet-type blocks using differential geometry operators instead of convolutions. Even though this is a first step towards an important goal, the absence of subsampling limits input mesh resolution and the needs for a well-defined neighborhoods for points limits its applicability for noisy and large-scale real-world scans.

Project Description:

A number of important applied problems in engineering, finance and medicine can be In the last five years, advances in deep learning have led to significant progress in allowing computers to understand the real world from visual input, opening up many new opportunities for dramatic progress in areas ranging from robotics to virtual and augmented reality, as well as medical and industry 4.0 applications.

Most of these machine learning architectures are based on convolutional neural networks (CNNs), which are able to learn powerful features from images, and even generate highly realistic pictures from scratch using generative adversarial networks (GANs). In particular, for images, we have seen tremendous success in both discriminative and generative tasks.

Unfortunately, for 3D data, e.g., data obtained from 3D scans of various types, research remains in its infancy.

Developing machine learning techniques for this type of data has great potential, as our world is inherently three-dimensional, and even four-dimensional when considering the temporal domain.

Performing scene understanding in 3D has significant advantages; for instance, a machine learning approach does not need to learn viewpoint invariance, and as a consequence, requires less training data.

However, this additional third dimension (and fourth for dynamics) comes at significant computational and memory overhead, which has so far been the major bottleneck in applying machine learning to 3D.

In this project, we address these shortcomings by developing efficient machine learning algorithms for 3D and 4D data analysis in the context of several important geometry processing tasks.

In particular, we will develop deep learning architectures and training methods capable of efficiently modeling different types of static and dynamic 3D data representations, including sparse spatial and temporal representations on voxel volumes, RGB-D images, point clouds, multi-view images, and meshes.

We will further construct new datasets designed for our scenarios, captured from the real-world, as well as synthetically generated with simulated renderings, augmented to reduce the reality gap between artificial and real data.

Finally, we will develop new neural network architectures designed for discriminative and generative applications embedded in spatial, and specifically temporal, domains.

In order to showcase our learning methods, we will apply them to static and dynamic 3D reconstruction tasks, as well as semantic scene understanding in 3D and 4D with an emphasis on fusing the spatial and temporal domains.

Key References:

1. Dai, A. and Chang, A. and Savva, M. and Halber, M. and Funkhouser, T. and Nießner, M. ScanNet: Richly-annotated 3D Reconstructions of Indoor Scenes // Proc. Computer Vision and Pattern Recognition 2017 (CVPR), IEEE.
2. Maier, R. and Kim, K. and Cremers, D. and Kautz, J. and Nießner, M. Intrinsic3D: High- Quality 3D Reconstruction by Joint Appearance and Geometry Optimization with Spatially- Varying Lighting // International Conference on Computer Vision 2017 (ICCV).
3. Dai, A. and Qi, Charles R. and Nießner, M. Shape Completion using 3D-Encoder-Predictor CNNs and Shape Synthesis // Proc. Computer Vision and Pattern Recognition 2017 (CVPR), IEEE.
4. Chang, A. and Dai, A. and Funkhouser, T. and and Nießner, M. and Savva, M. and Song, S. and Zeng, A. and Zhang, Y. Matterport3D: Learning from RGB-D Data in Indoor Environ- ments // Proceedings of the International Conference on 3D Vision 2017 (3DV).
5. Dai, A. and Nießner, M. and Zolotov, M. and Izadi, S. and Theobalt, C. Bundle Fusion: Real-time Globally Consistent 3D Reconstruction using On-the-fly Surface Re-integration // ACM Transactions on Graphics 2017 (TOG).
6. Notchenko A., Kapushev E., Burnaev E. Large Scale Shape Retrieval with Sparse 3D Con- volutional Neural Networks // Proceedings of the 6th International Conference on Analysis of Images, Social Networks and Texts (AIST-2017), July 27-29 2017, W.M.P. van der Aalst et al. (Eds.): AIST 2017, LNCS 10716, pp. 236-245, 2018. arXiv preprint arXiv:1611.09159.
7. Kuleshov A., Bernstein A., Burnaev E., Yanovich Yu. Machine Learning in Appearance- based Robot Self-localization // 16th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE Conference Publications, 2017.
8. Ilya Kostrikov, Joan Bruna, Daniele Panozzo, Denis Zorin. Surface Networks. <https://arxiv.org/abs/1705.10819>.
9. R. Klotz and V. Lempitsky. Escape from Cells: Deep Kd-Networks for the Recognition of 3D Point Cloud Models. IEEE International Conference on Computer Vision (ICCV), 2017.
10. A. Pumarola and A. Vakhitov and A. Agudo and A. Sanfeliu and F. Moreno-Noguer . PL- SLAM: Real-Time Monocular Visual SLAM with Points and Lines //International Conference in Robotics and Automation, 2017.

Ratio of effort:

Theoretical: 20%

Computational: 40%
Programming: 40%

Recommended Classes/Pre-requisites: Machine Learning, Deep Learning, Computer Vision