Predicting Dynamics of Deformable Objects using Deep Generative Models

Zehang Weng, Hang Yin, Anastasiia Varava, and Danica Kragic

Robotics, Perception and Learning Lab, EECS at KTH Royal Institute of Technology
zehang@kth.se, hyin@kth.se, varava@kth.se, dani@kth.se

Deformable object manipulation is an important sub-area in current robotics research that is still in its infancy. One key challenge is the lack of physical simulators that are stable and reliable when modeling deformable objects. Modern machine learning approaches can provide powerful tools for representing and predicting objects' state and dynamics. Most of them focus on rigid objects[2][6] or human bodies [3]. The recent DPI-Net[4] learn a simulator from data based on particles instead of unordered point clouds. For deformable objects, point cloud data is often used as an explicit representation of the large configuration space. Point cloud classification problem has been addressed[1][5]. However, there is little research on highly deformable objects represented as point clouds, especially for dynamics prediction, like clothes and bags.

In this project, we focus on studying and predicting the dynamics of deformable objects. We aim to learn a deep generative model to capture the system dynamics and simulate the future of unseen frames, based on a dataset of point cloud trajectories of complex clothing items. We aim to design efficient representations capturing the invariant information about the system and modeled objects, incorporating some prior knowledge.

The challenge comes from three main perspectives. First, there is no existing public point clouds dataset for doing such a study. Second, there is no existing properly designed structure to capture the dynamics of deformable unordered point cloud. Third, there is no standard criterion to evaluate model performance. We address these challenges separately.

Point clouds are collections of unordered 3D points in the space. For further study of the dynamics of cloth-like objects, it is crucial to have a clean point clouds dataset, which is hard to obtain. It is inconvenient to evaluate some approaches in the real world that the raw point clouds data from acquisition devices contain noise and unexpected outliers. At the current stage, we choose to use the game engine Unity, together with a recent particle-base cloth simulator, to generate a synthetic dataset. It allows us to perform certain actions on the cloth and easily collect the point clouds observations for further analysis.

To ensure that it is feasible to predict the cloth dynamics by using a deep generative model, we simplify the research question as a one-step prediction problem as depicted in Fig.1(A). The proposed deep model should be able to forecast the state of the next frame given an unseen point cloud set. In our problem settings, the performing action and internal cloth parameters remain the same among different trajectories in the dataset. Under this assumption, we expect historical frames to store enough information to forecast the next frame.

Unlike rigid objects, the trajectory of a non-rigid object always consists of two parts: global movement and internal deformation. We propose a simple model Fig.1(B) to decouple them for the dynamics prediction task.
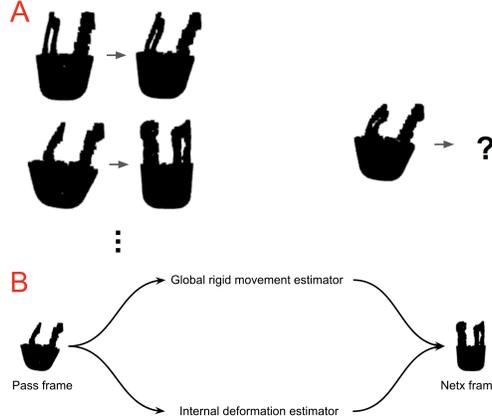

Fig. 1. (A) Example of our point clouds dataset and the task illustration. Given pairs of point cloud frames in left hand side, the task is to predict the future of an unseen data frame. (B) Overview of the proposed method.

According to the evaluation criterion, there is no standard way to measure the model performance. In this project, we will evaluate the models from two perspectives: with respect to the Chamfer Distance between the predicted frame and the ground truth, and by visually inspecting the predicted trajectories.

References

1. Ben-Shabat, Y., Lindenbaum, M., Fischer, A.: 3dmfv: Three-dimensional point cloud classification in real-time using convolutional neural networks. IEEE Robotics and Automation Letters **3**(4), 3145–3152 (2018)
2. Finn, C., Levine, S.: Deep visual foresight for planning robot motion. In: 2017 IEEE International Conference on Robotics and Automation (ICRA) (2017)
3. Kanazawa, A., Zhang, J.Y., Felsen, P., Malik, J.: Learning 3d human dynamics from video. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5614–5623 (2019)
4. Li, Y., Wu, J., Tedrake, R., Tenenbaum, J.B., Torralba, A.: Learning particle dynamics for manipulating rigid bodies, deformable objects, and fluids. arXiv preprint arXiv:1810.01566 (2018)
5. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 652–660 (2017)
6. Vijayanarasimhan, S., Ricco, S., Schmid, C., Sukthankar, R., Fragkiadaki, K.: Sfmnet: Learning of structure and motion from video. arXiv preprint :1704.07804 (2017)