Computer Vision Term Project: Learning Task-oriented Grasping with Dexterous Hands

Group5 : 310551081 劉安倫　310553017 楊傑祺　311553060 周睦鈞

Contribution

We have equal contribution for

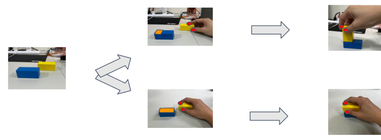
1. coding
2. presentation
3. report

Introduction

Hands are our primary means of manipulation with the physical world. Our hands continuously interact with diverse objects. We are interested in equipping robots with **human-level dexterity** to assist in human-centric environments. In our daily life, tools are usually designed for human hand use. Grasping is the very first step of manipulation. Understanding where or how to grasp an object is the key factor leading to the successful manipulation for humans.

Because of the high degree of freedom (DOF) possessed by a dexterous hand, predicting the grasping pose accurately becomes challenging. Several studies are dedicated to comprehending the geometry of objects in order to achieve stable object grasping.

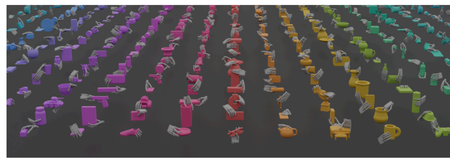
We found the existing works tackle dexterous hand grasping without modeling task information. In our daily life, there are many tasks that contain interaction between objects, e.g. pouring, stacking, and placing. It can provide information about a task. In the following example, we can see that the different object-object interactions can influence how we grasp an object. In this project, we want to prove can the modeling of object-object interaction enables task-oriented dexterous hand grasping.



Data collection

In our specific problem, obtaining a large dataset of grasping poses is crucial. To address this problem, we utilize DexGraspNet, a method published at ICRA 2023.

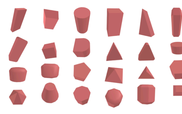
We employ DexGraspNet to collect a comprehensive set of dexterous hand grasping poses.



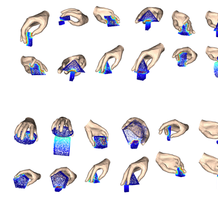
Our dataset comprises both the initial state, representing the process of picking a block from a table, and the goal state, illustrating the stacking of the block onto another one.

We gather information such as position, orientation, hand contact area at the initial state, and object interaction at the goal state. The regions highlighted in red indicate the areas where contact occurs.

24 Object

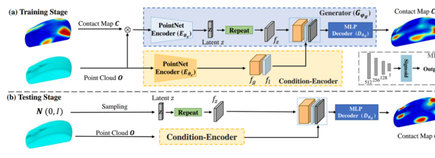


We collect grasping pose for each one of them (see the result in the right picture)

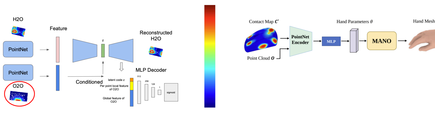


Method

Our method is based on the CVAE(conditional variation auto-encoder) model described in a paper called "Contact2Grasp," where the objective is to predict the hand contact map for a given object.



We have made some modifications to determine if the model can meet our requirements. Specifically, we are interested in exploring whether it can generate the appropriate hand-object contact map for a given task, given object-to-object contact information. Subsequently, we employ the predicted hand-object contact map to estimate mano parameters, which, in turn, enable us to generate a hand mesh using these parameters.



Implement Detail

(a) Parameter (follow the detail mentioned in paper)

- optimizer: Adam
- Batch_size: 32
- Learning rate: 0.0001
- $\alpha : 0.001$
  $\beta : 0.5$
  $\gamma : 0.5$

(b) Loss function

- ContactCVAE Loss: $\alpha \times KLD + \beta \times BCE + \gamma \times DICE$
- GraspNet Loss = Recon + Penetration + Consistency

(c) Data augmentation

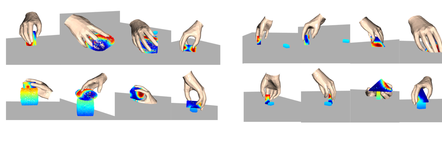
- Translation the point cloud's center of mass to the origin
- Random rotation with [-180, 180] degrees at 3 dimensions (XYZ)
- Random translation center of mass with [-1, 1] cm

Experiment

We design a score to validate if the object-object interaction(o2o) can facilitate generating a better hand-object contact map(h2o). We show the original result and our modified model compared with the dataset.
We can see the modified model can generate a more suitable contact area which is less overlapped with the o2o area.

```
h2o_o2o_consistency = (h2o[(h2o > 0.8)] * o2o[(h2o > 0.8)]).sum() / (h2o > 0.8).sum()
```

	H2O-O2O consistency score
ContactCVAE(MLP decoder) w/o O2O conditioned	0.17672
ContactCVAE(MLP decoder) w O2O conditioned	0.08977
Train_dataset	0.06713
Val_dataset	0.06460



Conclusion

In our work, we show that incorporating the object-object interaction can generate the corresponding hand-object contact map which can generate the task-oriented grasping pose for the designed task.

Package

- Pytorch
- Numpy
- Pytorch3D