Autonomous Excavation in Granular Materials with Reinforcement Learning

Weitung Chen, Filippos Sotiropoulos, Duane Boning, Harry Asada

***Abstract*—Robotic manipulation task with granular materials is hard to be modeled due to its complicated physics properties. Previous research has tried to incorporate deep reinforcement learning (RL) algorithms to devise robot the optimal policy based on previous experience. However, the training process for these methods often requires a lot of real-world training data. This paper aims to incorporate both data-driven approach and task modeling and achieve high-level control of the manipulation task in granular media. In this paper, a simulation environment was built based on the partial information of the task, i.e., starting scoop location, scoop angle, where the physics of the granular materials were not taken into account. An enormous amount of training data was generated in this "partial model" and then sent to the Soft Actor-Critic, a deep RL algorithm, to perform the training process. To accommodate the difference between this "partial model" and the real-world granular media behavior, we trained a convolution neural network to infer the actual terrain of granular materials from the simulation input. This proposed method has been implemented on a task that manipulates the robot to scoop the granular materials and achieve a specific target pattern. Our system can provide the robot with the optimal policy and achieve an average validation error of XX % [1] for simple patterns, and XX% for complicated patterns. With these efforts, we hope that it can be a starting point towards the development of autonomous excavation.**

***Index Terms*—granular materials, autonomous excavation, reinforcement learning, robotics, manipulation**

I. Introduction

Robotic manipulation of granular materials is essential to the development of autonomous excavation. As operating mining excavators is often an undesirable job (due to location and work conditions) and the operators often require years of training, solutions other than finding skilled workers on the market are needed for matching the increasing demand for excavation works. In this research, the high-level planning of robotic manipulation tasks with granular media is investigated, and a feasible method to provide scooping control planning is introduced. This is the starting point to understand more about autonomous excavation in granular materials.

As technology advances, modern computing power enables us to provide control decisions based on data-driven methods. Therefore, it is feasible to collect a massive amount of sensor data to pre-train a machine learning model and achieve precise control. We will harness the idea of data-driven to explore the application of advanced control and machine learning methods to the area of automated excavation and manipulation of granular materials. In this paper, a technique that incorporates both a data-driven approach and partial task modeling was proposed. It aims to achieve a high-level control of the robot and manipulate the robotic arm to create target patterns in granular materials, such as soil, after a sequence of accurate scooping actions.

Our system composed of two main parts: A simulation model that incorporates the partial information we know to predict the outcome of a specific action (scoop) and a reinforcement learning network to evaluate the best policy. We generated many data using the simulation model and try to train the reinforcement learning network using Soft Actor-Critic (SAC). By giving the system a custom pattern of the granular media terrain that we want the robot to achieve, our model can devise an optimal control policy that manipulates the granular media based on experience, which is the generated data. Then, we designed a predictive network to transfer the learning from the simulation world to the real robot. The result shows that we can control the robot to manipulate the granular media to achieve XX accuracy. This proves our approach can provide acceptable control decisions and perform a high-level path planning for scooping actions. Our proposed system differs from other works in that our model can quickly adapt and learn new tasks in simulation and transfer the learning results to the real robot. The predictive network to aid this transfer process only needs to be trained once. This property makes our system more feasible to be implemented on the real excavator and be used in the real-world scenario.

II. Related Works

There has been some research investigating this similar problem, using deep learning methods to devise policies for the robot when interacting with granular media. Schenck et al. [3] trained a robot to perform a scoop-and-dump action and manipulate the granular materials into the desired shape. The authors trained multiple predictive models and primarily used convolutional neural networks to predict the outcome of a specific action to rate whether the action is good or bad. Clarke [1] investigated similar methods and tried to control a robot to scoop a different desired mass of pellets from a tub. Additionally, Clarke evaluated the possibility of training on the sound signal gathered while performing the scooping action.

The papers above all collected a considerable amount of data. Schenck et al. collected 15000 examples of the scoop and dump action with seven robotic arms working together.

[1]Research works terminated after COVID 19 outbreak. The symbol XX throughout the paper are the placeholders for the results that we didn't obtain.

At the same time, Clarke et al. also collected around 15000 samples of scooping actions using a single robot on five different granular materials. Their methods require a lot of data from the real robot and are often infeasible as the task objective might need to be flexible, and data might be expensive to collect on a real robot.

As the robotic manipulation data is expensive to collect, methods other than the pure data-driven approach must be found. The deterministic method for devising scooping actions is a different way to achieve high-level planning, but soil and granular media, in general, are challenging to model. While the behavior of granular materials is complex to model, we still understand some information about how the robot acts in a real-world environment. This "partial model" of the system will be examined and can be integrated to guide the reinforcement learning agent to find the best scooping control sequence. Therefore, a hybrid of data-driven method and deterministic method is considered for robot manipulation.

Model-free reinforcement learning methods have been used widely in training robots to perform tasks that are hard to model. However, the training process usually takes a long time and a tremendous amount of data[3][1], and it is often expensive to collect the data and train on the actual robot. Our proposed method incorporated the advantage of both a data-driven and deterministic approach and can achieve reasonable control accuracy without much "real-world" robot training data, which is often expensive to collect. In this project, different methods will be investigated to reduce the amount of data needed from the "actual robot" and decrease the training time by exploring the idea of transfer learning. The general concept of "transfer learning" in this context is to allow the robot to learn a specific task in the simulation world and transfer everything it has learned to the real robot. With this concept of transfer learning, our system can adopt various tasks quicker and learn everything from simulation after learning how to map the control in simulation to the real robot.

III. Task Formulation

In this paper, the task that we would like to achieve is to manipulate granular media and reach a target pattern on the terrain. The following explains our task formulation and the reinforcement learning setup in detail.

In episode T, the robot is given an observation of its state O_T before each scoop: a height image of the terrain of the granular materials. The input height images are 80*140 pixels in size and are bounded 0 to 1, with 0 being the lowest point of the terrain and 1 being the highest.

Accordingly, the robot plans a sequence of scooping actions A_T. Each action is a 6D vector that describes a scooping motion. These parameters are described in Fig. 1. In this project, a simple target pattern is designed to be achievable by at most 2 scoops.

The goal G_T describes the target pattern of our task. G_T is a height map with parts eliminated by the scoops, featuring at most two rectangular areas of the terrain with lower height values.

After performing the action sequence, the robot makes an final observation F_T of the terrain, and the robot receives reward $-||F_T - G_T||^2$. We also added some extra reward terms to improve training performance. For instance, we highly reward perfect scoops (by +100) and penalize bad actions (by -100).

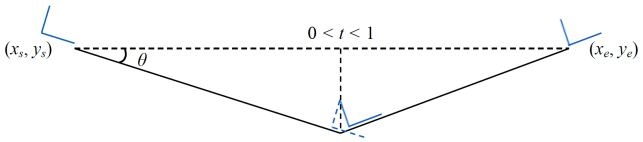

Fig. 1. Parameters for the scooping actions. The scoop starts at a position (x_s, y_s), moves straight down with an angle θ, rotates up at a middle position defined by t in $(0, 1)$, and moves straight up to a position (x_e, y_e). With (x_s, y_s) and (x_e, y_e) forming a line in the xy plane, t defines as fraction the distance from (x_s, y_s) to the rotation position.

IV. System Overview

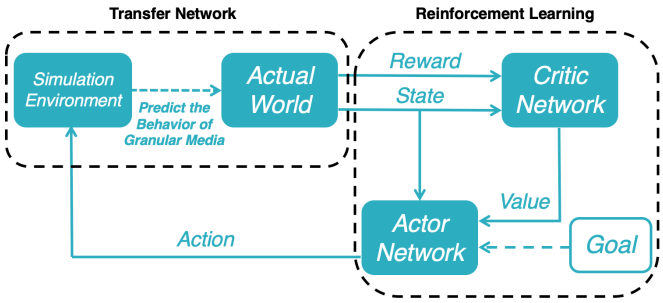

Fig. 2. System Diagram

Our system aims to learn the task of scooping granular media in simulation and transfer its learning to a real robotic arm. The system can be separated into two parts, a simulation model that incorporates the partial information we know to predict the outcome of a specific action (scoop) and a reinforcement learning network to evaluate the best policy. The general strategy is to generate many different scooping trials in the simulation world and to predict the outcome of the same scoop on the real experimental rig via a transfer network. With this transfer network, the model can predict the actual change in the granular media terrain, providing a better policy for achieving a specific goal. In our paper, learning the best policy of scooping is done by training a reinforcement learning agent with the Soft Actor Critic (SAC) model. The SAC model is a robust and sample-efficient actor-critic algorithm based on the maximum entropy RL framework.

The actor optimizes policy to maximize explorations during training. Moreover, the SAC model is also reported to be more stable and with a fast learning speed, which are desirable qualities as we are training a more complex task.

The major benefit of our model is that it can transfer all the training results in the simulation world to the actual robot after learning the behavior of the same type of granular media once. Figure 2 shows the diagram of our system. The workflow of the learning is as follow: it starts from the depth map of a starting terrain gathered from the experimental rig, the actor network devises a policy based on the input goal pattern and the predict value given from the critic network. Next, this action is applied in the simulation environment and generate a simulated height map with this scoop. Then, we applied a pre-trained transfer network to predict the actual height map in the real-world scenario. That way, we are able to calculate the reward and state values for the critic network and starts the second iteration of learning from that terrain map. It is not hard to see that we are able to train the reinforcement learning agent with many different tasks but only need to train the transfer network once for one type of granular media.

V. Simulation Environment

To apply reinforcement learning methods on robotic manipulation in granular materials, we need a lot of sample data. Often, we collect these sample data by building a simulation environment that model the task you try to learn. However, the behavior of the granular media is hard to model. In this paper, we will only model the partially known response that the system will do to a specific input action. More specifically, in the simulated world, we treat each granular particle as a solid substance and assumed that no fluid-like behavior occurs after a scoop.

With regards to the implementation, we built an environment by OpenCV to simulate the height map of the granular media terrain and test the algorithms for manipulating granular materials. The terrain is initialized as a nearly flat surface at the height of 0.6 with random fluctuations within height of ± 0.2. For every step in the RL training, the agent performs an action and scoop a portion of materials out of the terrain. Every action modifies the terrain map by reducing the height values of the pixels in the scooping region according to the action specification (Fig. 1). That is, removing all the materials above the triangular area. The target patterns are similarly generated by 1 or 2 random selected actions. In order to avoid meaningless target patterns, such as scoops that are too short or out of the height range, we restrict the scoop to start from the left half terrain and end at right. The initial scooping angle θ is constrained to be small to avoid negative values in the pattern.

While OpenCV cannot physically account for granular materials, we can simulate the height map of the terrain by using simple geometry and the definition of a scoop. After updating the pixels (height at each locations), we got a height map like the Figure 3 (left). These height maps have sharp edges at the boundaries of the scoop (we called it the voxel scoop), which don't really look like the behavior of the granular materials. Therefore, we tried to apply a simple gaussian blur filter to simulate the fluid-like behavior of the sands. Note that the real granular media height map can be generated simply by applying a gaussian blur filter on a voxel scoop. We believe the real granular material height map can be predicted by a neural network model. For simplicity in this paper, we used the Gaussian blur to simulate the behavior of the granular materials, and the results can be seen in Figure 3 (right).

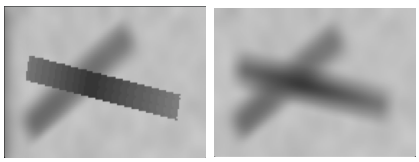

Fig. 3. Height map before (left) and after (right) applying Gaussian blur filter

VI. Network Models

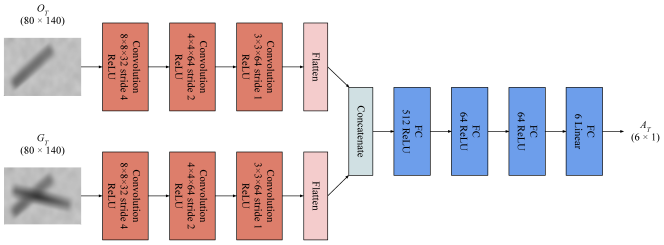

Fig. 4. Architecture of the actor network

There are two primary networks that we need to train for our system: the transfer network, which uses convolutional neural network, and the reinforcement learning model, which is based on the soft actor critic network.

The transfer network aims to predict the effect of a scoop on the terrain of the granular materials. To achieve this, we trained a convolutional neural network with the input from the simulation and the predictive outcome from the real robot for a set of random scoops. After applying a scoop in the simulation, we are able to get a height map of the terrain that doesn't accurately represent the terrain map in the real world. Therefore, we apply the same scoop and control the real robot to perform this policy in the experimental rig. We aim to collect 1000 different terrain map inputs (simulation) and outputs (real terrain) and used them to train the transfer network. The tensorflow library is used for training the predictive model.

Regarding the reinforcement learning model, we expect the model to recognize the goal pattern and devises optimal policies to manipulate granular materials to achieve that

pattern. In this research, we have designed a custom actor network model which is shown in Figure 4. The basic idea is that the observations O_T and target patterns G_T are fed in parallel into the network to generate an action. Multiple convolutional layers are used at the beginning of the model to extract essential features from the given terrain maps. After concatenating the two feature maps, we used multiple fully connected layers and output the six parameters that fully describe an action. The hyper parameters we used for training this model are attached in the Appendix. For the implementation, we use the stable baselines [2] open source library for building our custom actor network and incorporating the SAC model for training the reinforcement learning agent.

Our control method should be evaluated with the mean error between the current terrain and the goal terrain. To understand whether the learning works on the real robot, we will also evaluate the mean error between the real terrain and the goal. Then, the result will be compared with baseline deterministic methods and pure data-driven methods.

VII. Results

A. Devise an optimal scooping sequence to achieve a target pattern

The performance of our system is evaluated based on the mean error between the resulting terrain after a sequence of scoops and the target pattern. There are primarily three different levels of complexity for the given target patterns: single scoop patterns, geometry-shape patterns, and complex patterns. In this section, only the mean error in the simulation is considered. The real-world robot planning can be inferred by the control decision based on our scooping action planning in simulation, which will be evaluated in the next section.

Figure 5 shows the change in mean rewards value during the training process. The orange line indicates that the training converges and get the highest reward of +100 in around 25,000 steps. Also, the results prove that our model, which is based on SAC, performs better than other RL methods, such as DDPG and TD3, in the task of manipulating granular materials specifically. While this is only tested on the simplest task in the simulation world, it implies that our model can learn the single scoop patterns in simulation.

Figure 6 shows the mean error in meters between the terrain after a sequence of scoops and the target pattern. The goal terrain is shown in Figure 7 and the step-by-step scoops is shown in Figure 8. The defined dimension for this target terrain is $3.2(m) \times 4.8(m) \times 20(m)$. After six scoops, the mean error reaches 1.3 meters.

Figure 6 uses geometry shape patterns as the target while Figure XX uses a more complicated target pattern. The X-axis is the number of scoops that our agent perform in our simulated world and Y-axis shows the mean error to the

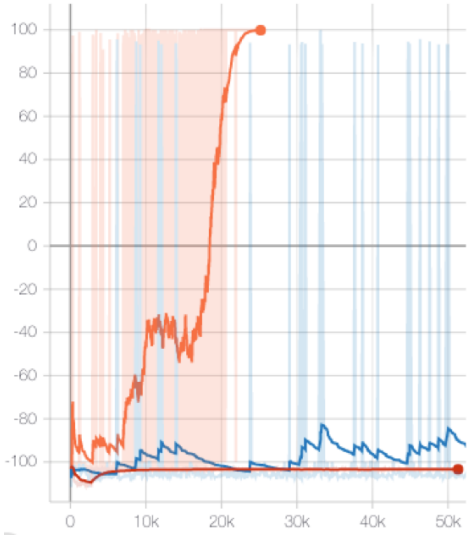

Fig. 5. Reward values during the reinforcement learning training process. The orange line is our RL model based on Soft Actor Critic (SAC). The training performance of our model is compared with RL model using TD3 (blue) and DDPG (red)

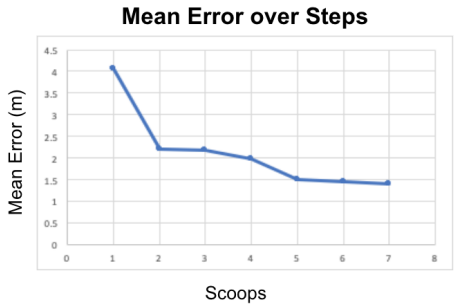

Fig. 6. Mean error values between the terrain after a sequence of devised scooping actions and the target terrain

target pattern. We can see that the mean error will eventually reach an asymptote and have a constant offset to the goal pattern. This error is generally acceptable for applications like large scale excavation.

From Figure XX(c), it is clear that it has a similar trend to that of Figure XX(b), even it has a more complicated task. Noteworthy is that the mean error offset of this type of task is larger and that it takes longer to reach the mean error asymptote due to the increased complexity of the goal pattern.

Figure 9, XX, XX shows examples of the inference of different tasks and example target patterns for each categories of target patterns. Although the resulting terrains don't perfectly match the target terrains, they include most of the features in the goal pattern. From Figure 9, it is clear that the terrain on the left is very similar to that of on the right. While the shapes of the holes are not fully identical, the depth of that area is basically the same.

All of the above experiments are done in simulation and

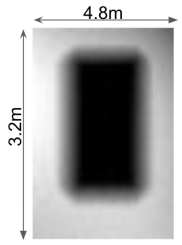

Fig. 7. Goal terrain of the experiment in Figure 6. The dimension of this terrain is $3.2(m) \times 4.8(m) \times 20(m)$.

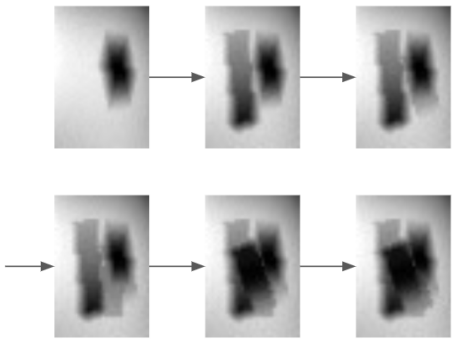

Fig. 8. Terrains of step-by-step scooping actions in the experiment in Figure 6

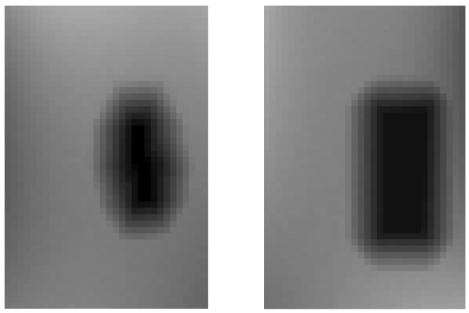

Fig. 9. The comparison between the terrains after around 10 scooping action (on the left) and the goal (on the right)

our study aim to transfer this learning of robotic manipulation scoop task to the real robot, which will be evaluated in the next section.

B. Inferring Real World Terrains from the Simulated Terrains

Convolutional Neural Network is used to predict the real-world terrains from the simulated terrains. It was designed to fill the gap between the terrains of our "partial simulated task model" and the real world granular materials terrains. The setup of our real-world experimental rig can be seen in Figure 10. With this setup, we first record the starting terrain of the granular media before any operations. Then, we sample a random action to control the robotic arm to perform a scoop. Lastly, we record the scoop's parameter and perform the same scoop in the simulation under the same starting terrain collected earlier. That way, we have two terrains, one modified heightmap terrain with the simulated scoop and a real terrain after a scoop.

Fig. 10. Real Robot Experimental Setup

After collecting many sets of data, simulated and real terrain, we trained a convolutional neural network with these data—[See comment 2]. The results show that the training can converge and that the mean error reaches a constant value XX after XX steps of scooping actions (Include graphs in the future). It has proven the concept of "transfer learning" from the simulated scooping environment to the real-world terrain, which enables us to be one step closer to applying to the real excavator.

VIII. Conclusions

High-level planning for robotic manipulation with granular materials that incorporates both a data-driven approach and partial task modeling was proposed. We used Soft Actor-Critic (SAC) to learn the optimal scooping policy in the simulation environment. This simulated model of the manipulation task was built based on the partial information we know about the model, for example, the initial scoop angle and the initial scoop position.

The learning of the robotic manipulation with granular materials is evaluated on manipulation tasks of varying difficulty. The results show that the method can devise an optimal policy for scooping granular media and achieve a target pattern with an error of XX %. This observation is consistent with different target terrains with a similar level of complexity. Also, our study shows that all the learning and planning in the simulation environment can be "transferred" to the real-world robot. The actual robotic arm can manipulate granular materials and achieve a mean error of XX %.

Compared to previous research, our proposed method is able to generate its training data in the simulated environment,

which saves a lot of time and cost to collect these data on the real robot. This benefit enables the robot controller to adopt many different manipulation tasks more quickly. The predictive network to infer real-world terrains from simulated terrain only need to be trained once, which means we only need to collect the real terrain data once. The model can then learn various manipulation tasks in simulation by generating its data in the simulated world. In the future, we hope to investigate more into applying this method to the real robot and contribute to the development of autonomous excavation.

IX. Comments

1) Due to the current situation, I have moved all my research/experiments to simulation environment. Therefore, I'm not able to test my method on the real robot. The big part of my innovation is to prove that all the learning on simulation is able to transfer to the real robot and the planning will still work. However, I'm not able to evaluate that now.
2) Now we have only 200 data sets that was collected last semester so it is not able to train a predictive network with this amount of data.
3) Due to the lack of training machine (I am only using one EC2 instance now due to the pricing), I'm not able to finish all the training that I was planning to do at the beginning of the semester. Leaving some of the results section blank.
4) Last, I just want to say thank you to all the TAs, instructors of this class. Thank you for all the efforts you put into this class and still make it happen despite the special circumstances we are facing.

References

[1] Samuel Clarke. Robot learning for manipulation of granular materials using vision and sound. Master's thesis, 2019.

[2] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, and Yuhuai Wu. Stable baselines, 2018.

[3] Connor Schenck, Jonathan Tompson, Sergey Levine, and Dieter Fox. Learning robotic manipulation of granular media. *ArXiv*, abs/1709.02833, 2017.