

A Additional Details of GRAIN

A.1 Algorithm for Image Representation of Robot Excavation Action

To highlight the relationship among the robot excavation action, avalanche behavior and obstacle movement, we introduce our image representation of the excavation actions in Sec. 4.1. The pseudocode in Alg. 1 shows how we obtain this representation. We show several examples of the image representation in Fig. 9, for different locations of the robot leg.

Algorithm 1 Image Representation of Robot Excavation Action

Require: White square size A , coordinates of robot excavation action location $\mathbf{a}_t = (x_t, y_t)$

Ensure: Output image O of size (H, W)

```

1:  $O \leftarrow$  Zero matrix of size  $(H, W)$ 
2:  $x_{\text{start}} \leftarrow \max(0, x_t - \frac{A}{2})$ 
3:  $y_{\text{start}} \leftarrow \max(0, y_t - \frac{A}{2})$ 
4:  $x_{\text{end}} \leftarrow \min(H, x_t + \frac{A}{2})$ 
5:  $y_{\text{end}} \leftarrow \min(W, y_t + \frac{A}{2})$ 
6: for  $i = x_{\text{start}}$  to  $x_{\text{end}}$  do
7:   for  $j = y_{\text{start}}$  to  $y_{\text{end}}$  do
8:      $O[i, j] \leftarrow 255$ 
9:   end for
10: end for

```

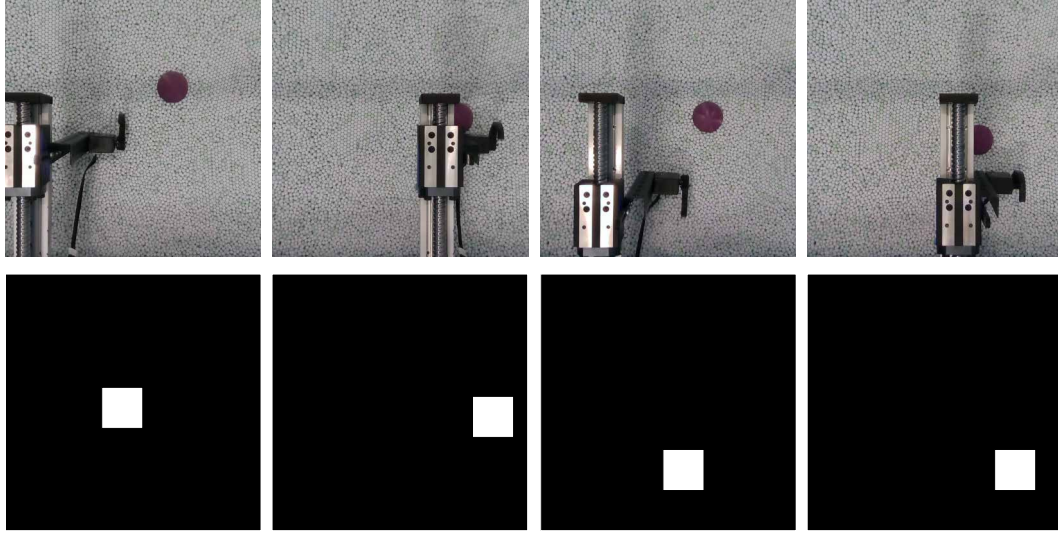


Figure 9: Additional examples of image representations of excavation actions. **Top:** RGB images of robot excavation actions. **Bottom:** The corresponding image representations of robot excavation actions.

A.2 Algorithm for Masking Obstacles in \mathbf{x}_t

Sec. 4.3 discusses how we handle multiple obstacles. In Alg. 2, we formalize our method to mask other unselected obstacles. This lets the masked images look similar to images from the single obstacle case, which we use for training the ViT. See Fig. 10 for an example of an image and its corresponding masked version (obtained via Alg. 2). During training, we use `cv2.colormap`, a package in `opencv`, to convert depth to RGB, as shown in the figure. For visual clarity, we overlay “Obstacles” and “Masked Obstacles.”

Algorithm 2 Masking Unselected Obstacles in Multiple Obstacles Manipulation

Require: Input image I of size (H, W) , window size B , list of obstacle center coordinates $\{(x_k, y_k)\}_{k=1}^{K-1}$, list of pixels sets (obstacles) $\{P_k\}_{k=1}^{K-1}$ for each coordinate

Ensure: Output image O of size (H, W)

```
1:  $O \leftarrow I$ 
2: for each  $(x, y) \in \{(x_k, y_k)\}_{k=1}^{K-1}$  with corresponding pixels set  $P$  do
3:    $x_{\text{start}} \leftarrow \max(0, x - \frac{B}{2})$ 
4:    $y_{\text{start}} \leftarrow \max(0, y - \frac{B}{2})$ 
5:    $x_{\text{end}} \leftarrow \min(H, x + \frac{B}{2})$ 
6:    $y_{\text{end}} \leftarrow \min(W, y + \frac{B}{2})$ 
7:    $S \leftarrow \{I[i, j] \mid x_{\text{start}} \leq i < x_{\text{end}}, y_{\text{start}} \leq j < y_{\text{end}}\}$ 
8:    $\text{avg} \leftarrow \frac{1}{(x_{\text{end}} - x_{\text{start}})(y_{\text{end}} - y_{\text{start}})} \sum_{(i, j) \in S} I[i, j]$ 
9:   for each  $(i, j) \in P$  do
10:    if  $x_{\text{start}} \leq i < x_{\text{end}}$  and  $y_{\text{start}} \leq j < y_{\text{end}}$  then
11:       $O[i, j] \leftarrow \text{avg}$ 
12:    end if
13:  end for
14: end for
```

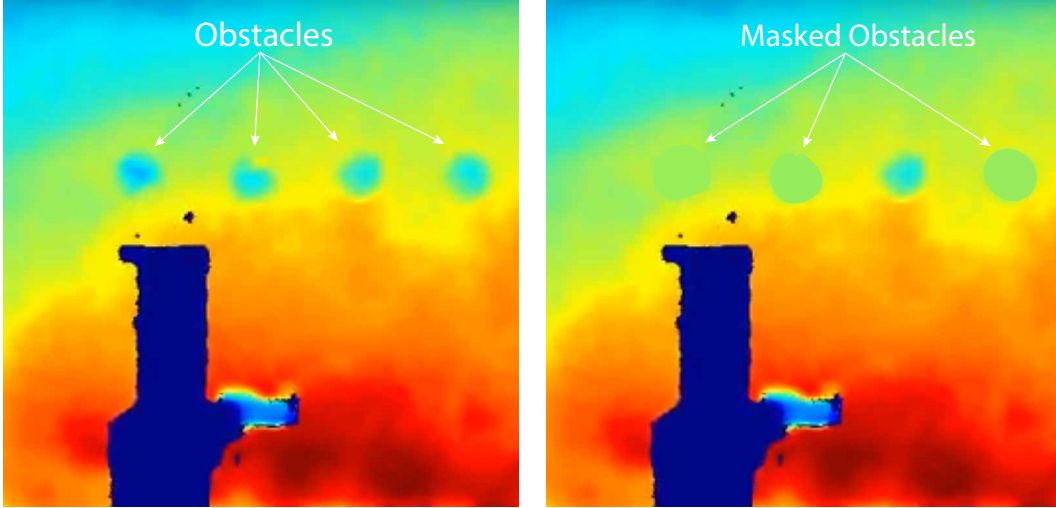


Figure 10: Example of masked image \tilde{x}_t . The left image is an example of x_t and the right image is the corresponding masked image \tilde{x}_t . Specifically, the second to the right most obstacle is the selected obstacle and other 3 obstacles are masked.

B Additional Experiment Details

B.1 Single Leg Manipulation Results

Fig. 11 shows one experiment trial for two manipulation tasks: “Single obstacle with single task,” and “Unseen obstacle.” We refer the reader to Section 6.2 for a description of what the tasks mean, and for example trials from other tasks. The statistics of all manipulation trials are shown in Tab. 1.

B.2 Multiple Unseen Obstacles Manipulation

To test our trained model’s generalization ability, we use obstacles with different shapes and weights in this task. Specifically, we use a star shape obstacle, a cuboid obstacle that has half of the weight of the obstacles used in the training dataset, and a hemisphere obstacle the same size but 4 times the weight of the obstacles used in the training dataset. All obstacles are 3D-printed. We placed these unseen obstacles on the granular slope plus the obstacle we used in the training dataset as a total of

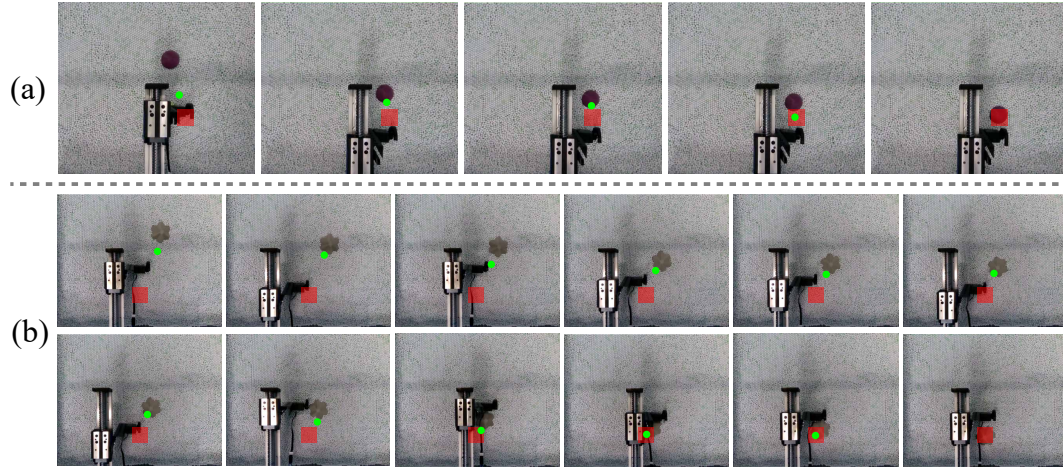


Figure 11: Single leg manipulation experiment results: (a) is “Single obstacle with single tasks,” and (b) is “Unseen obstacle.” In the above, red shaded areas are target areas. The green dots are the prediction of post-excavation locations of obstacles after the leg performs the action.

452 $K = 4$ obstacles with a random distribution and executed the manipulation policy. We show one
 453 manipulation trial in Fig. 12. Our system succeeds in 3 out of 5 trials.

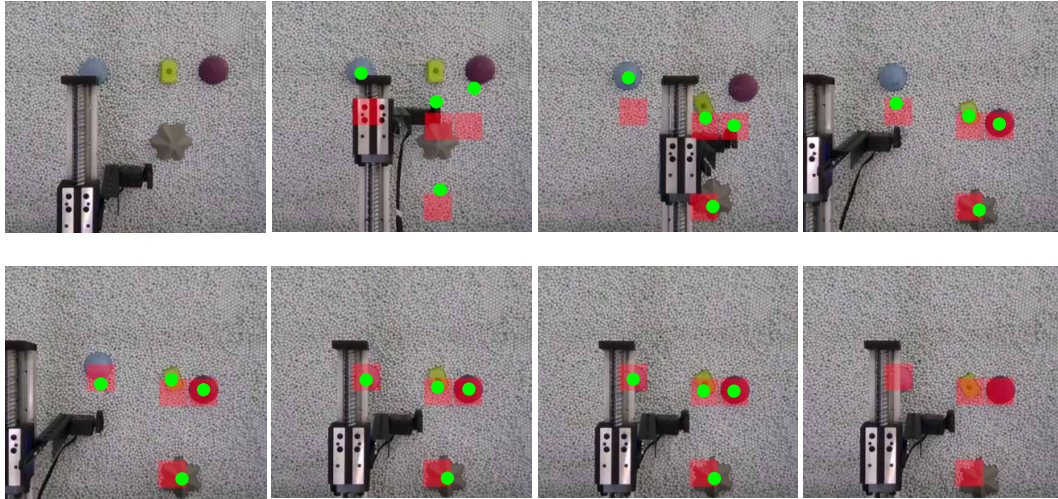


Figure 12: Multiple Unseen Obstacles Manipulation. Red shaded areas are target areas. The green dots are the prediction of post-excavation locations of obstacles after the leg performs an action at its current location.