MAE4GM: Visuo-Tactile Learning for Property Estimation of Granular Material using Multimodal Autoencoder

Zeqing Zhang[1,2], Guangze Zheng[1], Xuebo Ji[1], Guanqi Chen[1], Ruixing Jia[1], Wentao Chen[1]
Guanhua Chen[3], Liangjun Zhang[4] and Jia Pan[1,2]

I. Introduction

Granular materials (GMs) refer to a collection of solid particles, that are formable to be flowing [1], or jamming [2] under external forces. Some attributes of GMs, e.g., particle size, can be easily determined by vision. However, other properties, e.g., particle density, require specialized tools, like the force sensor, to be acquired. Estimating these properties presents a greater challenge for visual-based estimators alone. However, in everyday life, humans could estimate attributes that typically require other sensory modalities (e.g., taste) solely based on visual perception, which is attributed to humans' extensive multi-perception training in their daily lives. Inspired by this, we propose a multimodal training framework utilizing visuo-tactile data, called **MAE4GM** (Multimodal AutoEncoder for Granular Material). By training the MAE with visual and tactile signals, the model is ultimately capable of estimating particle properties and mechanical signals solely based on visual cues. The overview is summarised in Fig. 1.

Main contributions:

- We propose an MAE framework for GM property estimation as well as force inference using visuo-tactile signals from GM-probe interaction.
- We propose a particle tracking algorithm that extracts granule motion features from visual streams, facilitating multimodal learning with MAE.
- We extensively validate the generalization capability of the proposed model for unseen GMs, different data collection equipment, and scenarios.

To the best of our knowledge, it is the first work to study property estimation for GMs using visuo-tactile learning. Videos are available at `https://sites.google.com/view/mae4gm/site`.

II. Methodology

A. Physical Principle

This paper is based on the physical principles of GM-probe interaction [4]–[6], as explained in Fig. 1-(a). The advancing probe will create an area in front of it, that is the failure wedge zone [3]. Particles in this area block the probe's movement and generate force F_d, which can be

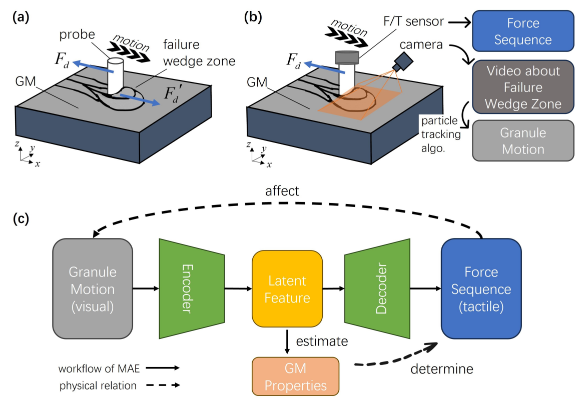

Fig. 1. Overview. (a) GM-probe interaction. The advancement of the probe creates a failure wedge zone [3], where the force F_d exerted on the probe originates from the resistance of particles within this zone. The reaction force F_d' will disturb ambient GMs.. (b) Visuo-tactile data. Force sequence F_d is measured by the F/T sensor, and the granule motion is extracted by the proposed particle tracking algorithm from a video clip. (c) Workflow of **MAE4GM** and inherent physical relationship. The input of MAE is the extracted visual information of granule motion and the output is time-series force sequence. Then GM properties are estimated based on its latent features. From (1), the force F_d received by the probe is determined by particle size and density, and its reactive force F_d' propels them into motion.

experimentally expressed as [7]

$$F_d = \eta \rho g d_c H^2, \tag{1}$$

where ρ and d_c refer to the density and particle size of GMs. So, we try to estimate GM density and size from latent embeddings using force F_d and resulting granule motions.

B. Dataset and Framework

In this study, we employ the UR5 robot arm to collect visuo-tactile data (see Fig. 1-(b)) in 15 common GMs, as depicted in Fig. 2. Here, we designate 10 GMs as "seen materials", while the remaining ones are considered as "unseen materials" absent from training, whose IDs are displayed on the green background in Fig. 2.

The framework of proposed **MAE4GM** is an encoder-decoder structure, as shown in Fig. 1-(c). Instead of directly inputting the video into the MAE, we preprocess the data by applying the proposed particle tracking algorithm to further extract the motion of particles. Since both the input and output are temporal signals, we employ two 3D convolutional layers to extract features from granule motion in the encoder and four 1D deconvolution layers are used to predict the force sequence in the decoder.

III. Results

We will show the trained model provides accurate estimations of GM properties and force sequences, as well

[1]The University of Hong Kong.
[2]Center for Transformative Garment Production.
[3]Southern University of Science and Technology.
[4]Robotics and Autonomous Driving Lab, Baidu Research, USA.

as its generalization ability to unseen GMs, different data acquisition devices, and scenarios.

A. Property Estimation and Force Inference

In Fig. 2-(a), we observe a distribution pattern where particle sizes range from small to large, exhibiting a progression from the bottom-left to the top-right. Red circles cluster in the bottom-left (smaller GMs like millet, cassia seed, cat litter). Moderate-sized GMs (coffee bean, shelled peanut) are concentrated in the middle. Larger particles (broad bean, large macaroni) with non-spherical shapes are found in the top-right. In Fig. 2-(b), we discover a distribution pattern of particle density ranging from small to large, exhibiting a clustering feature from the two ends towards the center.

Fig. 2-(c) shows the contact force estimation predicted by the trained model for GMs from the test set, e.g., cassia seeds and millet. The estimated forces (red dashed line) closely align with the real forces (blue solid line). Furthermore, we also test the model on unseen GMs. E.g., the model excels with sands but struggles with sunflower seed's force fluctuations despite generally aligned mean values.

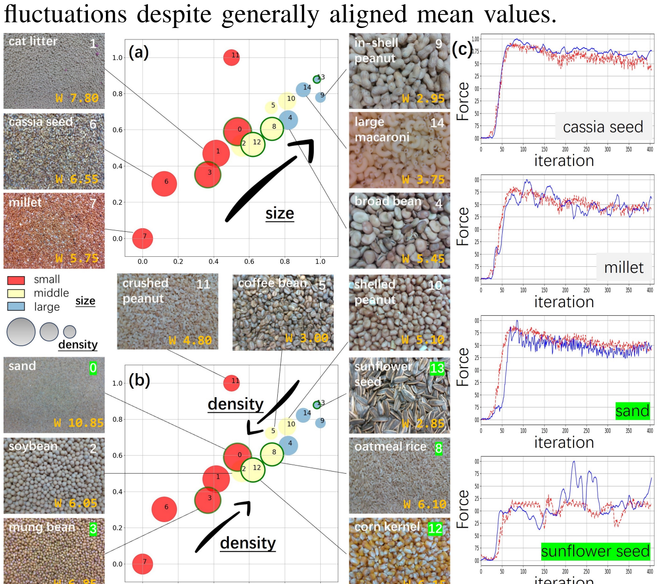

Fig. 2. Granule property estimation and force inference. The GM name (top-left), particle size category (bottom-left), and measured weight (unit: kg) (bottom-right) are presented for each GM. IDs of unseen GMs are displayed on the green background. (a) GMs are arranged diagonally in order of particle size in the low-dimension space from latent features. (b) GMs with large densities tend to be more concentrated in the latent embedding. Here, circle size refers to the value of particle density, and its color indicates particle size. (c) Force inference on seen and unseen GMs.

B. Generalization Validation

Fig. 2 already gives the model's property estimates and force predictions for unseen GMs. In Fig. 3, we use handheld devices instead of a robot arm to collect videos in two GMs. The trained model accurately estimates properties for seen GM (coffee bean) and shows reasonable estimates for unseen GM (sunflower seed) in the latent embedding.

Furthermore, we extend the data collection scenario to the natural environment rather than in the confined laboratory. Similarly, we use handheld devices to capture the GM-probe interaction on a beach (top left of Fig. 4-(a)) and take samples from 3 locations depending on the water content in sands (bottom right of Fig. 4-(a)). In Fig. 4-(a), we can see that GM 15 and GM 16 are located near the center of the latent space, indicating that the model considers these 2 GMs to have relatively high density, which aligns with the observed property distribution of sands in Fig. 2. Additionally, the model assigns GM 15 and GM 16 with medium (or slightly large) particle sizes, which corresponds well to the aggregation of sand particles in the presence of water, as observed in the videos. GM 17 exhibits large-scale cracking due to significant moisture content, which deviates from conventional GM motion, making the model unable to generalize. However, GM 15 and GM 16 with lower moisture content exhibit traditional granule movement. Despite the limitations of manual video collection, we observe increased probe resistance with higher water content, validated by inferred force sequence in Fig. 4-(c).

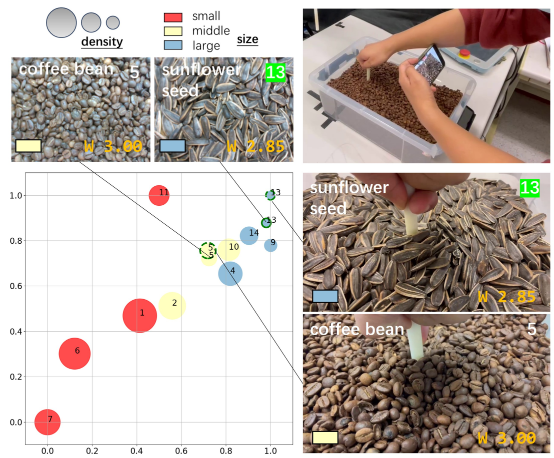

Fig. 3. Using a handheld device, we collected videos of a vertical probe dragged through GMs. We tested the model's generalization on seen (coffee bean) and unseen (sunflower seed) GMs, displaying their estimated properties with dashed green edges.

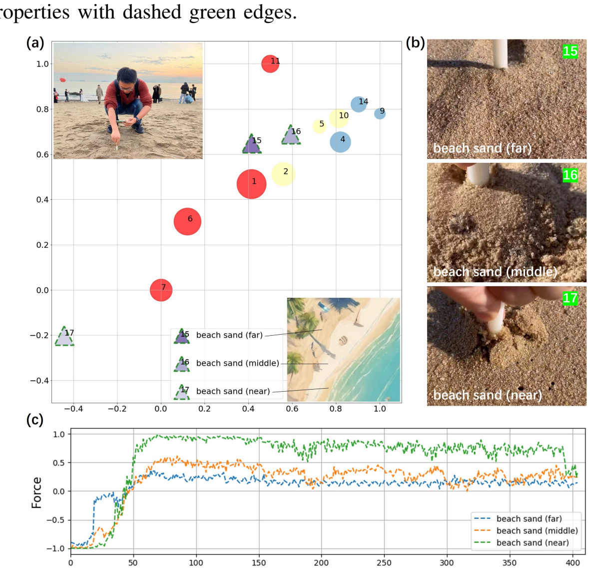

Fig. 4. Using a handheld device on a beach, we take videos from 3 sampling sites with varying water content in beach sands (see (b)). (a) The model estimates the size and density of GM15 and GM16, aligning with video observations. However, the model fails to generalize to GM17 due to its high water content, resulting in different granule motions. (c) The estimated probe force correlates with the force we felt in the sampling, increasing with water content.

REFERENCES

[1] O. Zik, J. Stavans, and Y. Rabin, "Mobility of a sphere in vibrated granular media," *EPL (Europhysics Letters)*, vol. 17, no. 4, p. 315, 1992.

[2] P. Richard, M. Nicodemi, R. Delannay, P. Ribiere, and D. Bideau, "Slow relaxation and compaction of granular systems," *Nature materials*, vol. 4, no. 2, pp. 121–128, 2005.

[3] W. Swick and J. Perumpral, "A model for predicting soil-tool interaction," *Journal of Terramechanics*, vol. 25, no. 1, pp. 43–56, 1988.

[4] R. D. Maladen, Y. Ding, C. Li, and D. I. Goldman, "Undulatory swimming in sand: subsurface locomotion of the sandfish lizard," *science*, vol. 325, no. 5938, pp. 314–318, 2009.

[5] Y. Zhu, L. Abdulmajeid, and K. Hauser, "A data-driven approach for fast simulation of robot locomotion on granular media," in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 7653–7659.

[6] M. Kobayakawa, S. Miyai, T. Tsuji, and T. Tanaka, "Local dilation and compaction of granular materials induced by plate drag," *Physical Review E*, vol. 98, no. 5, p. 052907, 2018.

[7] R. Albert, M. Pfeifer, A.-L. Barabási, and P. Schiffer, "Slow drag in a granular medium," *Physical review letters*, vol. 82, no. 1, p. 205, 1999.