Computer Methods in Applied Mechanics and Engineering 432 (2024) 117445

Contents lists available at ScienceDirect

Comput. Methods Appl. Mech. Engrg.

journal homepage: www.elsevier.com/locate/cma

Machine-learning-enabled discrete element method: The extension to three dimensions and computational issues

Shuai Huang [a,b], Pei Wang [a,b,*], Zhengshou Lai [c,d,e,**], Zhen-Yu Yin [b], Linchong Huang [c,d,e], Changjie Xu [a]

[a] Institute of Geotechnical Engineering, School of Civil Engineering and Architecture, East China Jiaotong University, Nanchang, Jiangxi, China
[b] Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, China
[c] School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China
[d] State Key Laboratory for Tunnel Engineering, Guangzhou, 510275, China
[e] Guangdong Key Laboratory of Marine Civil Engineering, School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China

ARTICLE INFO

Keywords:
Discrete element method
Machine learning
Irregular-shaped particle
Contact detection and resolution

ABSTRACT

The detection and resolution of contacts among irregular-shaped particles pose significant challenges in the discrete element method (DEM) and recent advancements have introduced a machine learning-enabled approach specifically tailored for contact detection and resolution in two dimensions. Building upon this progress, this paper extends the application of machine learning-enabled DEM to encompass the more complex and realistic three-dimensional (3D) scenario. Particles are modeled using a polyhedral representation with arbitrary shapes, and contact behavior is governed by an energy-conserving contact model based on contact volumes. The efficacy of the machine learning-enabled 3D DEM is evaluated through comparative analyses of computational time and simulation results across individual contact as well as whole DEM simulations against those obtained from the conventional DEM. The findings indicate that the machine learning-enabled approach adeptly identifies and resolves contacts among 3D irregular-shaped particles while accurately reproducing the mechanical characteristics of densely contacting particle assemblies. The computational issues and challenges associated with the machine learning-enabled DEM are also discussed. The study highlights that the machine learning-enabled approach significantly enhances computational efficiency, showcasing its potential to advance complex DEM simulations in a more efficient manner.

1. Introduction

The discrete element method (DEM) [1] is a widely used discontinuous method that has proven an effective tool for modeling granular materials. In recent years, an increasing number of researchers have applied DEM to model irregular-shaped particles and investigate the impact of particle shape on the mechanical and flow behavior of particulate systems [2–7]. Significant efforts have been devoted to the development of DEM algorithms tailored specifically for irregular-shaped particles, including advancements in particle models [8–11] and the corresponding contact algorithms [12–15]. These advancements in irregular-shaped particle modeling have greatly enhanced the fidelity of modeling particulate systems.

* Corresponding author at: Institute of Geotechnical Engineering, School of Civil Engineering and Architecture, East China Jiaotong University, Nanchang, Jiangxi, China.
** Corresponding author at: School of Civil Engineering, Sun Yat-sen University, Guangzhou 510275, China.
E-mail addresses: peiwang@ecjtu.edu.cn (P. Wang), laizhengsh@mail.sysu.edu.cn (Z. Lai).

https://doi.org/10.1016/j.cma.2024.117445
Received 22 April 2024; Received in revised form 23 September 2024; Accepted 3 October 2024

Contact detection and resolution are the core components of DEM and constitute the most time-consuming steps in DEM calculation [16]. The conventional DEM usually utilizes geometry-based methods for contact detection and resolution, which come with significant computational costs. Prevalent contact algorithms for irregular-shaped particles includes clump-based approach [17,18], Gilbert–Johnson–Keerthi approach [13,19], node-to-surface approach [20,21], mesh Boolean approach [22], polygon-based contact approach [23], etc. Contact algorithms essentially determine the contact status and geometric characteristics of particles based on known particle geometry and positions. For irregular-shaped particles, the definition of their contact geometric features is inconsistent and ad-hoc in previous studies. For instance, the contact normal orientation has been defined in various forms, such as the spatial derivatives of the level set function [9], the mean of the surface normal orientations [24], and the vertical direction of the line at the intersection of two contact particles [14]. This inconsistency may lead to a decoupling of different contact geometric features, potentially resulting in numerical instability. In light of this, Feng [22,25,26,27] developed a contact theory applicable to arbitrary shaped particles and proposed the contact volume-based contact model. This contact model strictly derives and defines the contact geometric features based on energy conservation. The core challenge of contact volume-based model lies in computing these geometric features, with a common strategy involving the utilization of polyhedral particle models to represent irregular shapes and their resolution through the mesh Boolean approach. It is worth noting that Feng [25] has proposed an enhanced strategy for resolving the contact geometry of polygonal particles, which proves to be more efficient than the mesh Boolean approach. These geometric computations demand substantial computational resources, particularly when faced with a high quantity of surface vertexes. In this context, machine learning can serve as a surrogate model of the traditional geometry-based contact algorithms. Machine learning models, primarily relying on linear operations, tend to exhibit high computational efficiency and wide adaptability.

Inspired by the work of Lai et al. [28], which introduced a machine learning-enabled DEM for addressing contact-related challenges in 2D, this study extends their approach to three dimensions. The results show that the machine learning-enabled contact detection and resolution approach can accurately and efficiently detect the contacts and solve the contact geometry features of particles. It provides a pioneering attempt to apply machine learning to DEM. Nonetheless, there remain certain aspects within the machine learning-enabled DEM to be further explored. Firstly, the current machine learning-enabled DEM has solely showcased their proficiency in handling 2D particles, necessitating further investigation into their applicability to 3D particles. Secondly, the existing study has predominantly focused on particles with simplified shapes, such as elliptical and convex particles, making it worthwhile to explore their effectiveness in dealing with general irregular-shaped particles, including concave geometries. Contact detection and resolution of 3D arbitrarily shaped particles is significantly more complex than 2D convex particles. This increased complexity stems from the presence of additional contact geometric features and the inclusion of multiple contact points. Consequently, it is essential to conduct further investigation on the definition of input–output parameters and generation of datasets.

It is worth noting that a recent study [29] has developed a machine learning-based particle–particle collision model for irregular-shaped particles. One limitation of this study is that it solely focuses on the particle–particle scale and lacks a comprehensively validated DEM model. Similar to the methodology developed in this work as well as the earlier original work by Lai et al. [28], the machine learning approach takes particle relative positions as inputs and outputs the contact flag and properties. Although the concept of machine learning for contact detection and properties is straightforward, there are some challenges regarding its integration into DEM simulation and addressing its computational accuracy, efficiency and stability. Special considerations need to be taken for the selection of data inputs and outputs, preparing the training samples, regularizing the data inputs and outputs from a DEM simulation, etc. In a comprehensive DEM simulation, it is imperative to evaluate the prediction accuracy of contact geometry characteristics, the predictive capabilities of the macro- and micro-mechanical properties of the particle assembly, and the computational efficiency to thoroughly ascertain the efficacy of machine learning algorithms. Addressing the insufficient discourse in the literature [29], this work delves into and substantiates these crucial aspects.

In this work, we focus on extending the previous machine learning-enabled DEM into 3D general irregular-shaped particles. The performance of the proposed method is demonstrated in terms of the contact between two particles as well as the particle medium in dense contacts. Furthermore, we adopt the energy-conserving contact theory to resolve the contact behavior of general irregular-shaped particles, which is beneficial to numerical stability. The rest of the paper is organized as follows. Section 2 shows the methodologies of integrating machine learning for contact detection and resolution. Section 3 presents the design, training, and evaluation of machine learning models. Section 4 shows three numerical examples to demonstrate the performance of the 3D machine learning-enabled DEM. Section 5 presents some further discussions and Section 6 summarizes the concluding remarks.

2. Methodologies

There are several algorithms for dealing with generally irregular-shaped particles in DEM, including the contact volume-based approach [22,24], the signed distance field-based approach [30], and their variants [31], etc. In this work, the contact volume-based approach is adopted as the reference and to create the database for training the machine learning models. The ingredients of particle representation, contact detection and resolution, integration of machine learning and numerical implementation are presented in the following.

2.1. Particle representation

In the contact volume-based approach, particles are represented by polyhedron. Specifically, a triangular mesh is employed to represent the surface of each particle. For particle surface meshes initially obtained from X-ray computed tomography [32], they may contain large amount of surface vertexes and be of low quality. In this work, we reconstructed the particle surface meshes with a sampling scheme proposed by Lai et al. [30], which utilizes a weighted spherical centroidal Voronoi tessellation. This approach allows the generation or reconstruction of a uniform triangular mesh, as illustrated in Fig. 1(a).

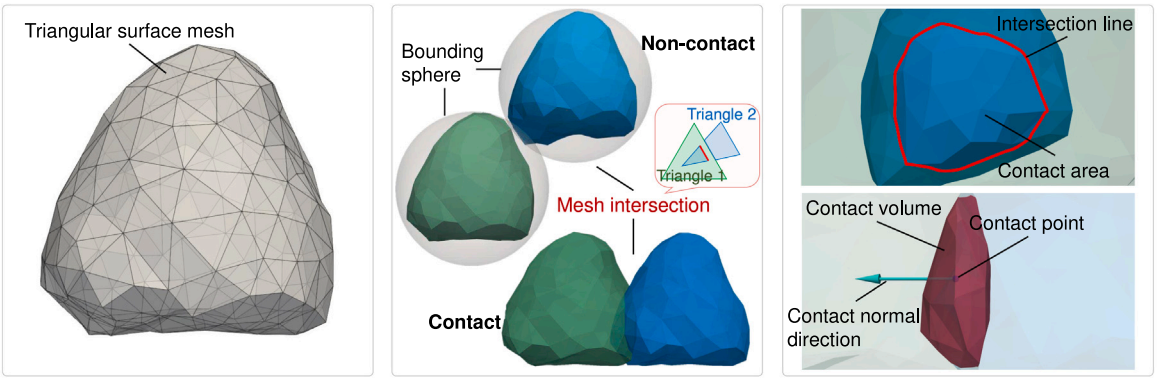


Fig. 1. Illustration of the polyhedral particle model and the contact volume-based approach for contact detection and resolution. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

2.2. Contact detection and resolution

Contact detection and resolution are the key components of DEM. In the contact volume-based approach, a contact is identified as two particles have an overlap in their volume, and the contact behavior is defined as the negative gradient of contact potential. Typically, as proposed in Feng [22], the contact potential is defined as a function of the contact volume. The detailed formulation of the contact forces are presented in Appendix. Accordingly, in this work the mesh Boolean approach is used to test whether two particles are in contact and then to obtain the contact geometries, namely the overlap volumes [33]. As shown in Fig. 1 (b), contact detection is achieved by performing the mesh intersection operation on the triangular mesh of two particles. The basic operation is to find the intersection line segment between two given triangles from each particle surface. If the intersection line exits, two particles are in contact, and vice versa. Following the mesh intersection operation, we obtain closed line segments of intersection, which are used to reconstruct the triangular mesh of the contact region (see the red polyhedron in Fig. 1(c)). In this way, the contact geometry features, involved contact volume, contact area, contact normal direction, and contact point can be resolved from the triangular mesh of the contact region. Fig. 1(b) and (c) illustrate the contact detection and resolution procedure for particle–particle contacts. The handling of particle-boundary contacts follows a similar procedure. In this case, the plane boundary is converted into the triangular mesh for the contact detection and resolution.

2.3. Integration of machine learning

The machine learning-enabled DEM follows the same calculation steps as the conventional DEM, i.e., the detection and resolution of contacts, the evaluation of contact behavior, the calculation of particle motion, and the updating of particle geometric descriptions, as shown in Fig. 2. It uses the machine learning method to replace traditional geometry-based methods for contact detection and resolution. The contact detection and resolution problem is of two stages. The first stage (contact detection) is to detect whether two particles are in contact, and the second stage (contact resolution) is to resolve the contact geometry features. These two stages correspond exactly to the classification and regression problems in machine learning, respectively.

In this work, two types of artificial neural networks (ANNs) are adopted for contact detection and resolution. Fig. 3 shows the inputs and outputs of ANNs. Specifically, the geometry descriptors of particle and boundary are taken as their inputs. Generally, there are two types of geometric descriptors for particles: shape descriptors and position descriptors. The shape descriptors will not be considered in this work. As will be discussed in Section 5, there are extremely large number of shape descriptor variables for the 3D general irregular-shaped particles, making it difficult to be precisely quantified and used as the suitable inputs. Including shape descriptors into the inputs would require deeper ANNs to achieve high-accuracy contact detection and resolution results, which thus could impact the computational efficiency of the machine-learning-DEM. As such, in the present work we will use one set of ANN for each shape. For a selected irregular-shaped particle, the position descriptors are the coordinates of the particle centroid and the rotation of the particle with respect to its initial alignment. For two particles (namely object particle and cue particle) in potential contact, we translate and rotate the coordinate system to make the centroid of the object particle located at the origin and the principal axes align with the coordinate axes. Then, the coordinate system is scaled to ensure that the object particle have a unit size. For the particle-boundary contacts, the inputs of ANNs are the normal direction of boundary and its distance to the centroid of object particle.

The outputs of the classification and the regression networks are the contact flag and contact geometric features, respectively. A true contact flag means the two particles are in-contact while a false one indicates no contact. The contact geometric features are related to the contact model used. In this work, we use the contact volume-based model. The key contact geometric features of this contact model are contact volume, contact area, contact normal direction and contact position. These geometric features are

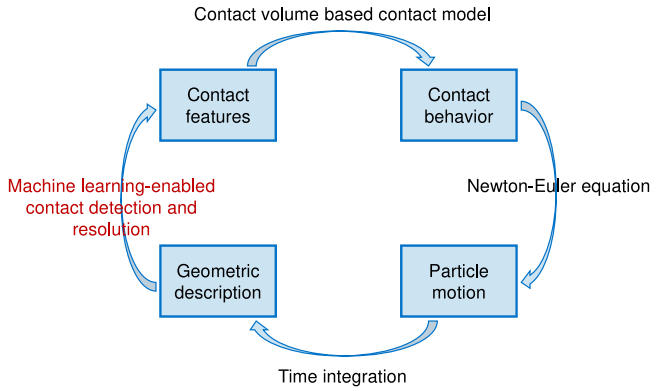
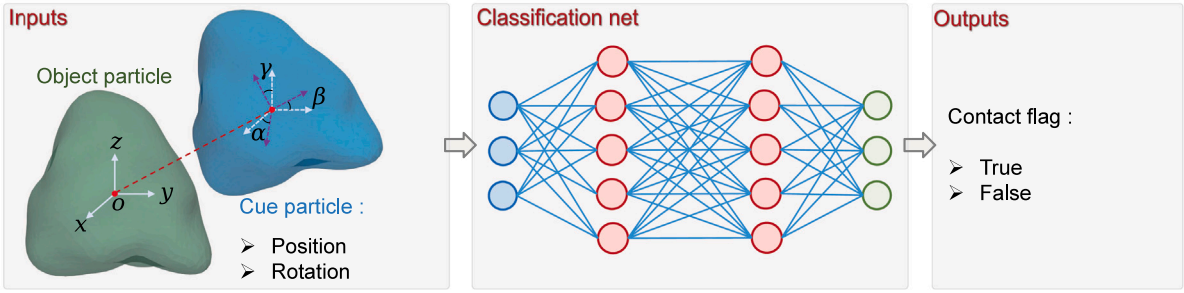
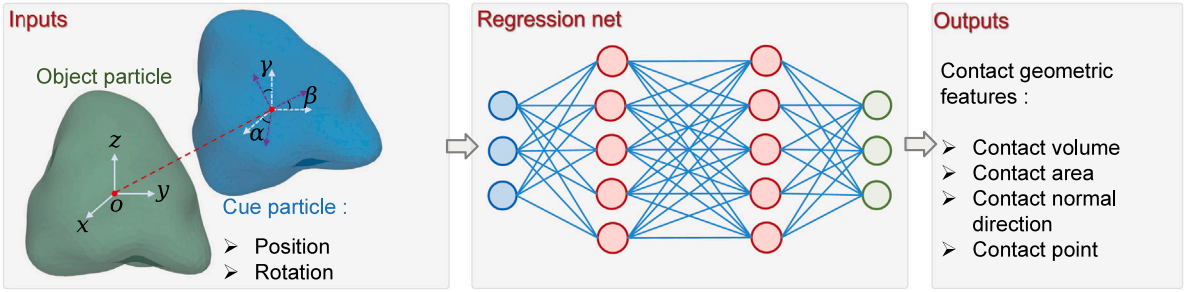


Fig. 2. Workflow of a typical calculation cycle of the machine learning-enabled DEM.



(a)



(b)

Fig. 3. Illustration of machine learning approach for contact detection and resolution: (a) classification network for contact detection, and (b) regression network for contact resolution. The object particle is colored in green, with its centroid located at the origin and its principal axes aligned with the coordinate axes. The cue particle is colored in blue, and its position is described using the coordinates of the particle centroid and rotation. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the outputs of the regression networks. There are also some other contact geometry features, such as branch vectors and tangential direction. These features can be deduced from the known key features above and thus are not included into the outputs.

2.4. Numerical implementation

Fig. 4 illustrates the implementation of machine learning-based contact detection and resolution approach in DEM. The algorithms are implemented on the top of a general-purpose in-house DEM code, *NetDEM*, developed and maintained by the third author (Lai). The name of *NetDEM* derives from the integration of machine learning capabilities within our in-house DEM code. *NetDEM* offers a variety of contact detection and resolution approaches, including machine learning-enabled, geometry-based [22], Gilbert–Johnson–Keerthi [34], and signed distance field [30]. Both NetDEM (machine learning-enabled DEM) and GeoDEM (the conventional geometric algorithm-based DEM) used in Section 4 of this work are implemented in our in-house code, sharing

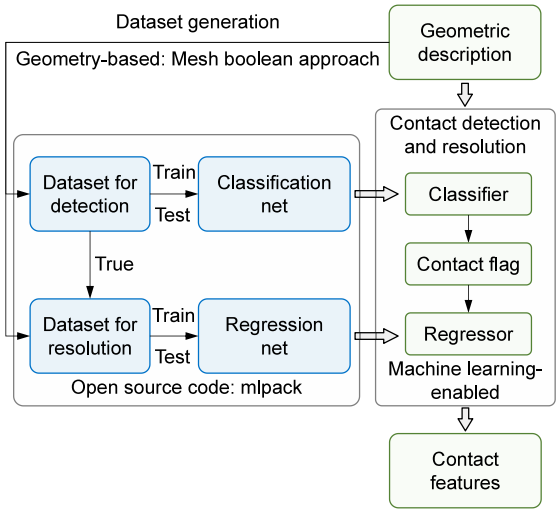


Fig. 4. Implementation of the machine learning-enabled contact detection and resolution approach in DEM.

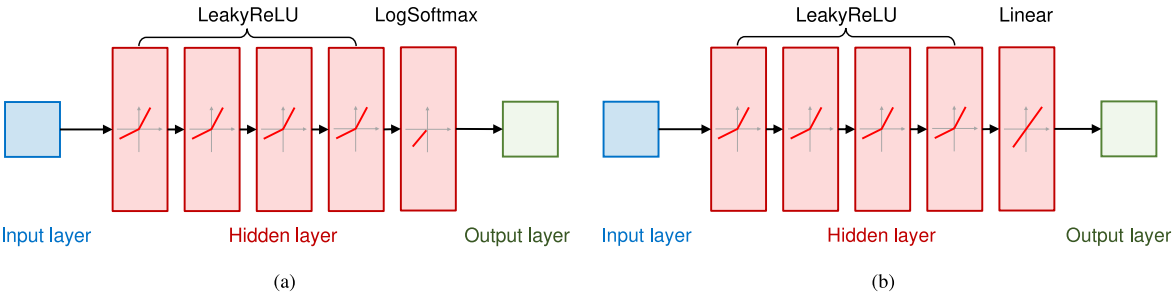


Fig. 5. Architecture of (a) the classification network for contact detection, and (b) the regression network for contact resolution.

identical architectures. The primary divergence lies in the contact detection and resolution approach: GeoDEM employs the mesh Boolean approach, while NetDEM incorporates the machine learning-enabled contact algorithm introduced in this work. The code implements the shared memory parallelism with OpenMP (Open Multi-Processing) and the distributed memory parallelism with MPI (Message Passage Interface). For the simulations in this work, only the OpenMP parallelism is used. The calculations, including contact detection and resolution, contact force evaluation and particle position, run in parallel in each DEM cycle. For the machine learning part, an open-source machine learning library, *mlpack*, is adopted to implement the classification and regression networks. *mlpack* is a general-purpose machine learning library that aims to provide fast, lightweight implementations of machine learning algorithms [35]. As shown in Fig. 5, the architectures of the classification and regression networks are similar. The ANNs are composed of an input layer, several hidden layers and an output layer. The activation function of the last hidden layer in the classification network differs from that in regression networks. The LogSoftmax activation function is often used in multi-class classification tasks, and the goal is to assign an input to one of several possible classes. The Linear activation function is often used in regression tasks, and the goal is to predict continuous outputs. On this basis, LogSoftmax and Linear activation functions are used in classification and regression networks, respectively.

2.5. Preparation of training data

Once the architectures of ANNs are determined, the datasets need to be prepared for model training. Because training data directly affects the performance of machine learning models, we introduce the following procedures to generate reliable datasets for the particle–particle and the particle-boundary contacts. The procedures of dataset generation for the particle–particle contacts is illustrated in Fig. 6. The detailed steps are as follows.

(1) Insert the object particle into the contact scene. The centroid of the object particle is located at the origin and the principal axes align with the coordinate axes. The object particle are scaled to be unit size. This work uses the polyhedral particle model for shape characterization.

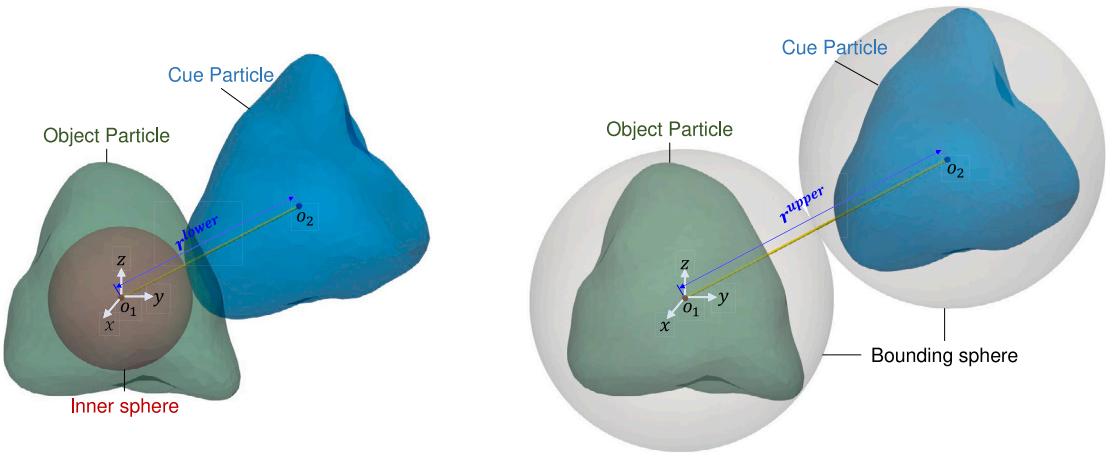


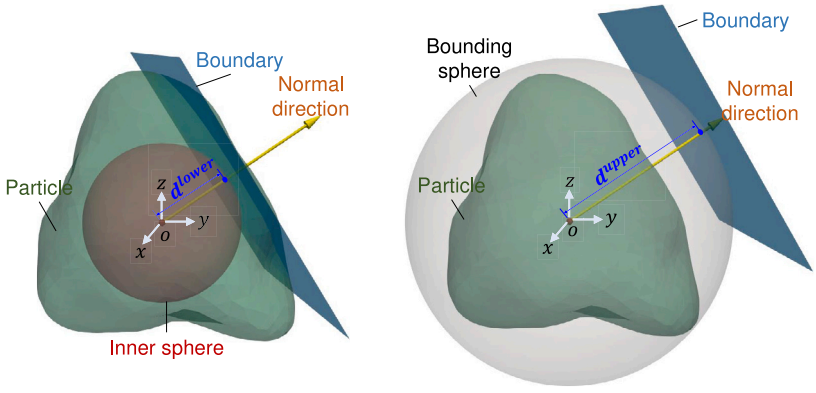
Fig. 6. Illustration of the procedure of preparing training and testing data for particle–particle contacts: (a) the minimum distance case, and (b) the maximum distance case. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

(2) Generate the centroid coordinates of the cue particle. In DEM, the contact detection is usually implemented with two phases, namely, a broad phase and a narrow phase. The broad phase is used to quickly identify and filter out pairs of elements that are not in contact based on fast algorithms such as bounding sphere algorithm, whereas the narrow phase performs the more expensive and exact contact detection calculations. It should be noted that the machine learning-enabled approach is implemented in narrow phase. Therefore, the centroid coordinates of the cue particle are generated such that the two particles fall into the range of narrow-phase contact detection, and the centroid positions of the cue particle are spatially uniformly distributed.
In the implementation, the centroid coordinates of the cue particle is obtained with a sub-routine described as following.

   (a) The centroid coordinates o_2 (see Fig. 6) of the cue particle can be determined using the vector $\overrightarrow{o_2o_1}$. This vector can be decomposed into two components: the distance between the centroid coordinates of the cue particle and the object particle, denoted as $r = |o_2o_1|$, and the unit direction vector $\boldsymbol{n} = \overrightarrow{o_2o_1}/|o_2o_1|$. Figs. 6(a) and 6(b) illustrate the scenarios for the minimum distance r^{lower} and the maximum distance r^{upper} between object particle and cue particle, respectively.
   (b) The minimum distance r^{lower} occurs when the inner spheres of the two particles are tangent to each other. The inner sphere, represented by the red sphere in Fig. 6(a), has a radius equal to 90% of the distance from the nearest surface point to the centroid of the object particle. The maximum distance r^{upper} occurs when the bounding spheres of the two particles are tangent to each other. For irregularly shaped particles, the radius of the bounding sphere is the distance from the centroid to the point that is farthest from the centroid on the particle surface. The distance between the object particle and the cue particle lies within the range from r^{lower} to r^{upper}. To achieve uniform sampling of the centroid distance, we generate a uniformly distributed random number u in the range [0, 1], and calculate $r = r^{lower} + u(r^{upper} - r^{lower})$.
   (c) The direction vectors $\boldsymbol{n}$ are generated through uniform sampling in space. For instance, we can employ the spherical centroidal Voronoi-based approach to sample points on a unit-size sphere, resulting in a set of surface points that are uniformly distributed. The vectors pointing from the center of the sphere to these surface points serve as direction vectors with a uniform spatial distribution.

(3) Determine the rotation of the cue particle. Commonly used metrics for representing 3D rotation include quaternion, Euler angles and rotation matrix. Rotation calculations using quaternions are more efficient than rotation matrix and do not suffer from the problem of gimbal locking in Euler angles [36,37]. Thus, we use the quaternions for particle rotation in this work. In this way, the random quaternions with uniform distributions are generated as inputs.
(4) Repeat steps 2 and 3 until the required number of input data samples are generated. Then, take the inputs into the geometry-based contact algorithm to obtain the outputs. Lastly, the inputs and outputs together comprise the complete datasets.

The procedures of datasets generation for the particle-boundary contacts are illustrated in Fig. 7. The position of the boundary is determined by the distance from the boundary to the particle centroid d, and the unit normal direction vector of the boundary plane $\boldsymbol{n}$. The minimum distance d^{lower} and the maximum distance d^{upper} between the particle and boundary correspond to the case where the boundary is tangent to the inner sphere and the bounding sphere, respectively. The unit normal direction vector of the boundary plane $\boldsymbol{n}$ can be characterized by a series of spatially uniformly distributed direction vectors from the origin. Subsequently,



(a) (b)

Fig. 7. Illustration of the procedure of preparing training and testing data for particle–boundary contacts: (a) the minimum distance case, and (b) the maximum distance case.

Table 1
Parameter settings of the ANN models.

Parameter	Hidden layers	Nodes	Epochs	Batch size
Classification nets for particle–particle contacts	5	32	100	128
Regression nets for particle–particle contacts	5	128	100	128
Classification nets for particle-boundary contacts	5	32	100	128
Regression nets for particle-boundary contacts	5	64	100	128

d and $\boldsymbol{n}$ are imported into the geometry-based contact algorithm to solve the contact flags and contact geometry features, which form the datasets.

3. Example contact detection and resolution

To validate the effectiveness of machine learning-based contact detection and resolution algorithm, the results of model training and evaluation are presented. Additionally, we investigate the effects of training parameters, ANN structure and particle shape on the performance of ANN models.

3.1. Model training and evaluation

Based on the procedure described in Section 2, we generate 500,000 samples of particle-boundary contacts and 1,000,000 samples of particle–particle contacts. Then, we split the samples into a training and a testing sets, in which 10% of the data is held out for testing. In the machine learning-enabled DEM, a total of four networks need to be developed, i.e., two classification networks for particle–particle and particle-boundary contact detection, and two regression networks for particle–particle and particle-boundary contact resolution. Table 1 presents the training parameters of the ANN models, which is selected on the basis of trial calculations and prior experience.

The performance of classification networks is evaluated based on the accuracy of the predictions, which is defined as the percentage of the correctly predicted contact flags. For the particle-boundary contacts, the accuracies of the trained classification networks are 99.62% and 99.48% for the training and testing datasets, respectively. For the particle–particle contacts, the accuracies are 96.57% for the training dataset and 96.29% for the testing dataset. The accuracy of the particle–particle contacts is slightly lower than that of particle-boundary contacts, due to the higher complexity. The results show that the classification networks perform very well and can effectively detect particle–particle contacts and particle-boundary contacts.

Next, we investigate the performance of regression networks. In this work, we use the mean absolute errors (MAEs) to evaluate their performance. Fig. 8 presents the predicted values and the ground truths of contact geometry features for the particle-boundary contacts. The predicted values of the contact geometry features match well with the ground truths, and the data points distributed closely around the 1:1 line. The MAEs in the predicted contact volume and contact area are in the order of 10^{-5} and 10^{-4}, respectively, which are smaller than those of the predicted contact points. Overall, the MAEs for all contact geometry features are in the order of 10^{-3}. It is noted that only the results of the test datasets are plotted in Fig. 8, and the results of the training datasets are basically consistent with the test datasets and thus are not exhaustively presented.

Fig. 9 shows the predicted values and ground truths of contact geometric features for particle–particle contacts. The MAEs of contact volume, contact area and contact point for particle–particle contacts are an order of magnitude larger than those for the particle-boundary contacts. The data points in Fig. 9 are more scattered comparing with those in Fig. 8. This means that the ANNs

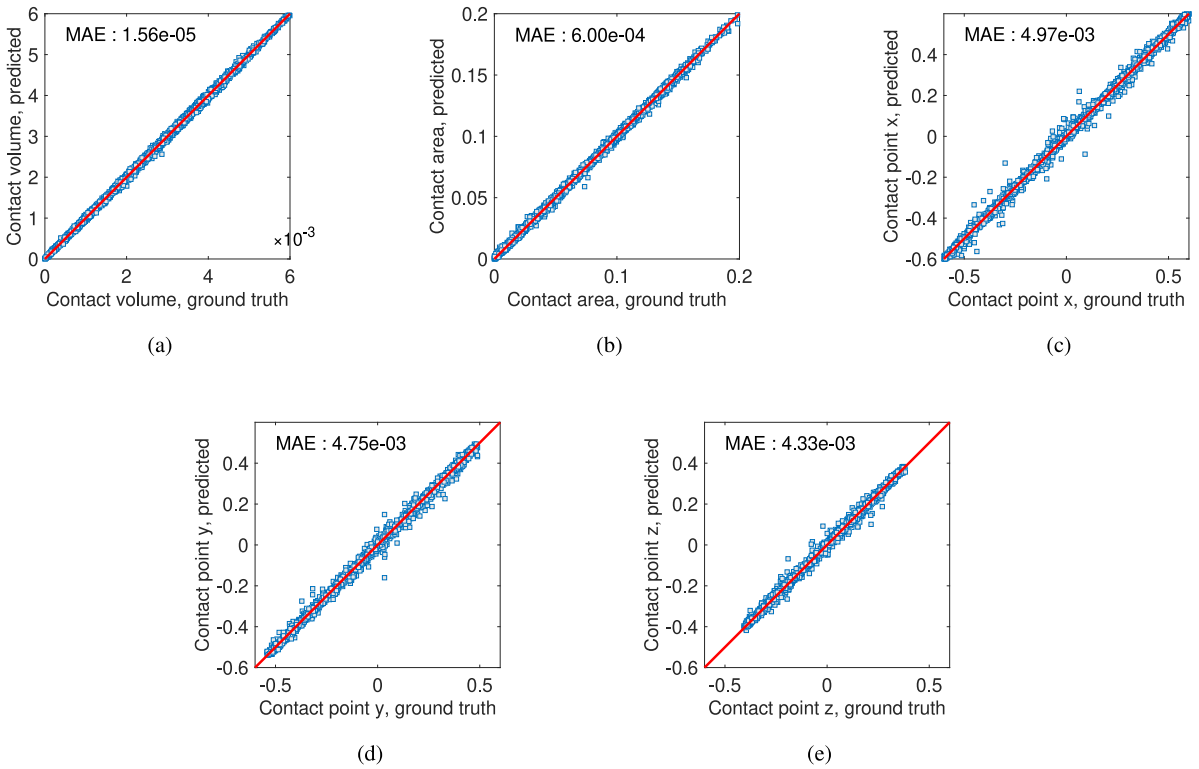


Fig. 8. Performance of the regression network for contact resolution of the particle-boundary contacts: (a) contact volume, (b) contact area, (c) x of contact point, (d) y of contact point, and (e) z of contact point.

perform better in the particle-boundary contacts than in the particle–particle contacts. The reason for this phenomenon is that the resolution of contact between two irregular-shaped particles is much more complex than that between a particle and a boundary. Even so, the predicted contact geometry features are still in good agreement with the ground truths for particle–particle contacts.

3.2. Effect of training parameter

The selection of appropriate training parameters plays a pivotal role in the predictive performance of an ANN model. In this work, we focus on the effect of two training parameters, namely the batch size and epochs, on the performance of ANN models. The batch size refers to the number of training examples processed together in one iteration, while the epoch represents one complete pass through the entire training dataset. In general, larger batch sizes led to improved memory utilization and parallel processing capabilities, resulting in faster training times. However, excessively large batch sizes could hinder the ability of the network to escape local optima and find better global optima. The results of the MAEs of each contact geometric feature with different batch sizes and epochs for particle–particle contacts are plotted in Figs. 10 and 11. According to results of the four evaluated cases of batch sizes, including 32, 64, 128 and 256, the batch size of 128 is a good option as a balance of the model accuracy and computational cost. On the other hand, the number of epochs, which represents the complete traversal of the training dataset, influenced the convergence and generalization abilities of the networks. As shown in Fig. 11, the performance of ANN model is better with the number of epochs being 100, whereas further training epochs may degrade the performance.

3.3. Effect of ANN structure

We further investigate the effect of ANN structure (i.e., the number of hidden layers and nodes in a hidden layer) on the performance of the ANNs. The results of regression network for particle–particle contacts are presented in Figs. 12 and 13. In this work, the number of nodes in each hidden layer is set to be the same, and six cases with node numbers of 8, 16, 32, 64, 128, and, 256 are considered. As shown in Fig. 12, the MAEs decrease as the number of nodes increases. The curve tends to be stable when the number of nodes reaches 128. Fig. 13 shows the effect of the number of hidden layers on the performance of regression networks. With the number of hidden layers increases from 3 to 9, MAEs decrease rapidly. However, hidden layers being too large (e.g., 11) leads to a slight increase in MAEs, which is likely resulted from the increased difficulty in training. The performance of ANNs is more sensitive to the number of hidden layers than the number of nodes. In general, increasing the number of hidden layers or nodes allows the network to capture more complex relationships in the data. This increased capacity can lead to improved

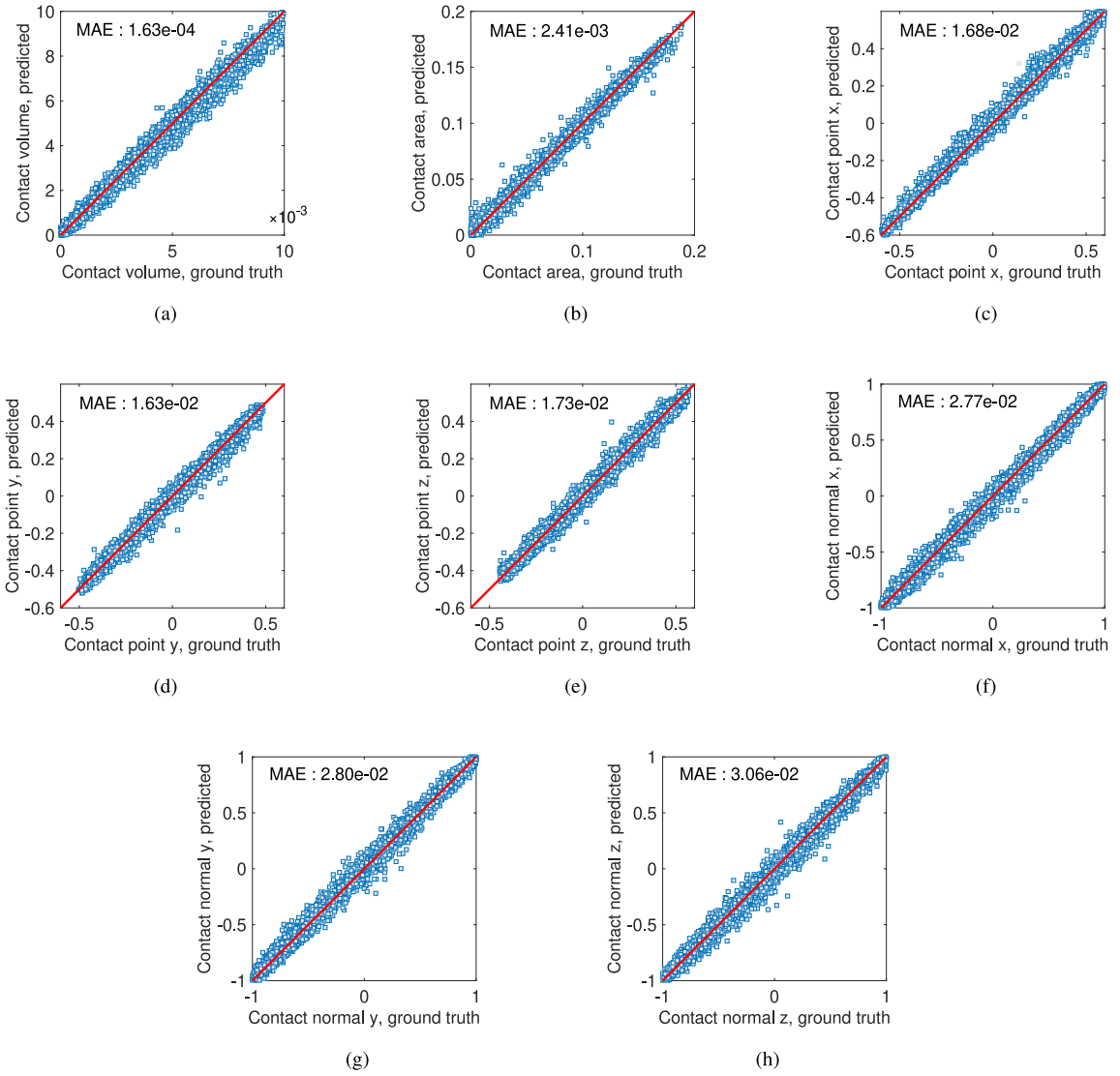


Fig. 9. Performance of the regression network for contact resolution of the particle–particle contacts: (a) contact volume, (b) contact area, (c) x of contact point, (d) y of contact point, (e) z of contact point, (f) x of contact normal, (g) y of contact normal, and (h) z of contact normal.

performance of ANNs. However, having too many nodes or hidden layers can also lead to over-fitting or increase computational complexity, ultimately resulting in a poor performance.

3.4. Effect of particle shape

It should be pointed out that the aforementioned discussions are all based on the same particle shape of a sand, which is shown in Fig. 6. Therefore, we also investigate the effect of particle shape on the performance of ANN models. Three types of particles with different morphological characteristics are evaluated and the parameters of their ANN models are the same, as listed in Table 2. Fig. 14 presents the results of the performance of the ANN models for particle–particle contacts and particle-boundary contacts. As shown in Fig. 14, the second green particle is the one that has been used in this work. A sphere particle and a bolt particle are added for comparison. The sphere particle has a completely smooth surface, while bolt particle exhibits sharp edges. The findings indicate that the performance of the ANN models is best for sphere particle and worst for bolt particle. The main differences arise from the prediction of particle–particle contacts. For instance, the MAE of the sphere particle an order of magnitude smaller than that of the bolt particle for particle–particle contacts. This means that complex shapes may require a deeper ANN models for accurate contact prediction.

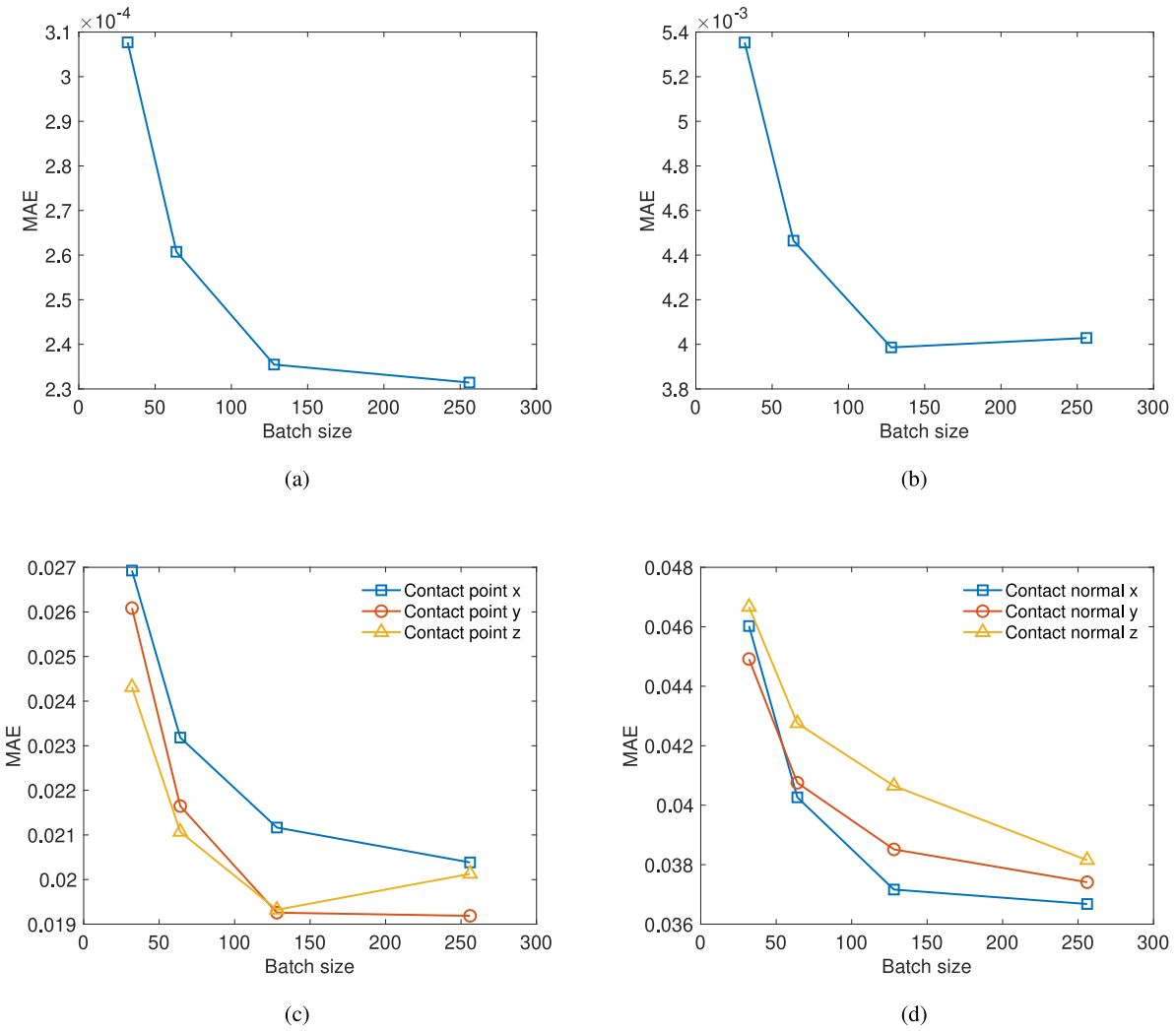


Fig. 10. Effect of batch size on the performance of regression networks: (a) contact volume, (b) contact area, (c) contact point and (d) contact normal.

4. Example DEM simulations

The 3D machine learning-enabled DEM is further validated by three numerical examples in typical quasi-static and dynamic conditions, including a random packing test, a column collapse test, and a triaxial compression test. For each numerical example, a reference simulation is conducted using the conventional DEM to examine the performance of the machine learning-enabled DEM. For simplicity, the 3D machine learning-enabled DEM and the conventional DEM are denoted as NetDEM and GeoDEM, respectively.

4.1. Random packing

The random packing test involves 2400 irregular-shaped particles with an equivalent size of 0.01 m. In this context, the equivalent size of an irregular-shaped particle is defined as the diameter of a sphere that has the same volume as the particle. Particles are inserted into a box container with a base of 0.1-by-0.1 m at once, and are allowed to settle down under gravity. The initial particle configurations of NetDEM and GeoDEM are identical, as shown in Fig. 15(a). The polyhedral particle model with 400 surface vertexes is used. The parameters of the DEM simulations are listed as follow: normal stiffness k_n is 1.0×10^{14} N/m, tangential stiffness k_t is 1.0×10^5 N/m^3 (note that contact volume-based contact model is adopted), damping coefficient c is 0.7, contact friction μ is 0.3, particle density is 2650 kg/m^3, and timestep is 1.0×10^{-5} s. Fig. 15(b) displays the snapshots of particle configurations at the end of random packing based on NetDEM and GeoDEM. It can be observed that the packing heights of two particle assemblies are basically the same at approximately 0.23 m, and particles are in close contact without unrealistic particle overlaps or penetrations. The porosities of final packing of NetDEM and GeoDEM are approximately 0.452 and 0.445, respectively.

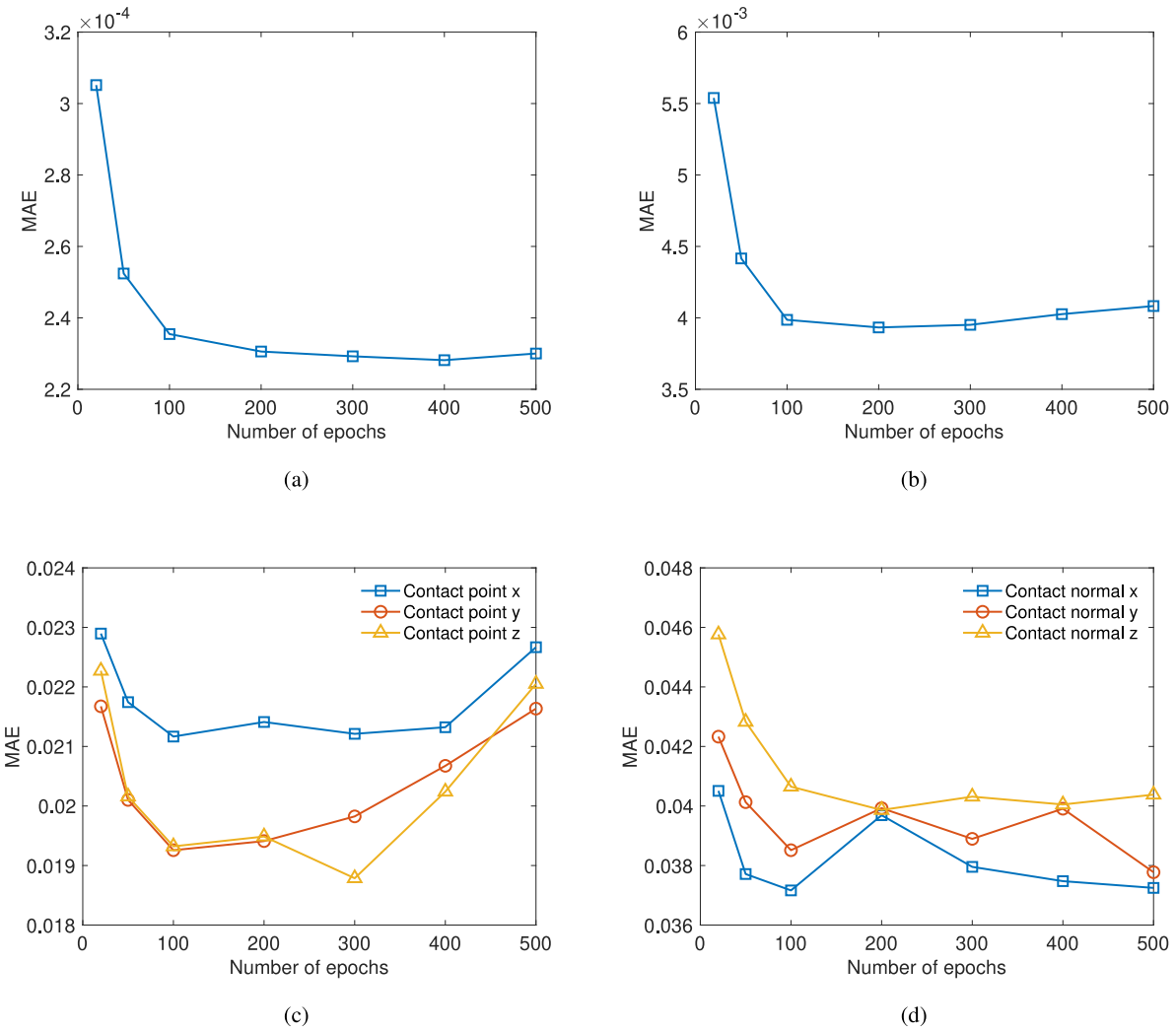


Fig. 11. Effect of the number of epochs on the performance of regression networks: (a) contact volume, (b) contact area, (c) contact point and (d) contact normal.

To compare the microscopic behaviors of particles in the NetDEM model and the GeoDEM model, the particle velocities, unbalanced force, and contact forces are extracted. As shown in Fig. 16(a), the average velocity of the particles increases at the beginning, followed by a rapid decrease, and become stable at the time around 0.7 s. At the end of random packing, the particle average velocity for both NetDEM and GeoDEM reaches a stable value, approximately 2.0×10^{-3} m/s and 1.0×10^{-3} m/s, respectively. It indicates that the particle assembly reaches a quasi-static equilibrium state. The velocity band, depicting the minimum and maximum velocities of particles, is also illustrated in Fig. 16(a), showcasing the range of particle velocities across various times. The velocity bands of NetDEM and GeoDEM are largely overlapping before 0.4 s, but afterward, the velocity distribution of NetDEM tends to skew higher overall. This trend aligns with the results of particle average velocities. Upon further analysis, the discrepancy between NetDEM and GeoDEM in Fig. 16(a) can be attributed to the fact that particle velocities are influenced by the force states of the particles. In the contact volume-based contact model, the normal contact force is determined by the product of contact stiffness, contact volume, and contact area. Machine learning models may introduce errors when estimating contact volume and contact area. Given the typically substantial values of contact stiffness, such inaccuracies can amplify the errors in calculating contact forces within NetDEM, thereby causing variations in velocity distributions. Nonetheless, the discrepancy in velocity is considerably small. The visually big difference between NetDEM and GeoDEM in Fig. 16(a) is due to the adoption of a logarithmic scale for velocity.

Fig. 16(b) illustrates the evolution of particle unbalanced force ratio (i.e., ratio of the mean unbalanced force to the mean contact force [38]) during the random packing test. During the initial stage of packing (particularly before 0.1 s), a discernible discrepancy in the unbalanced force ratio between NetDEM and GeoDEM is observed. This disparity arises from the limited number of contacts between particles at the initial stage, leading to discrepancies in mean contact force caused by predictive errors in the machine

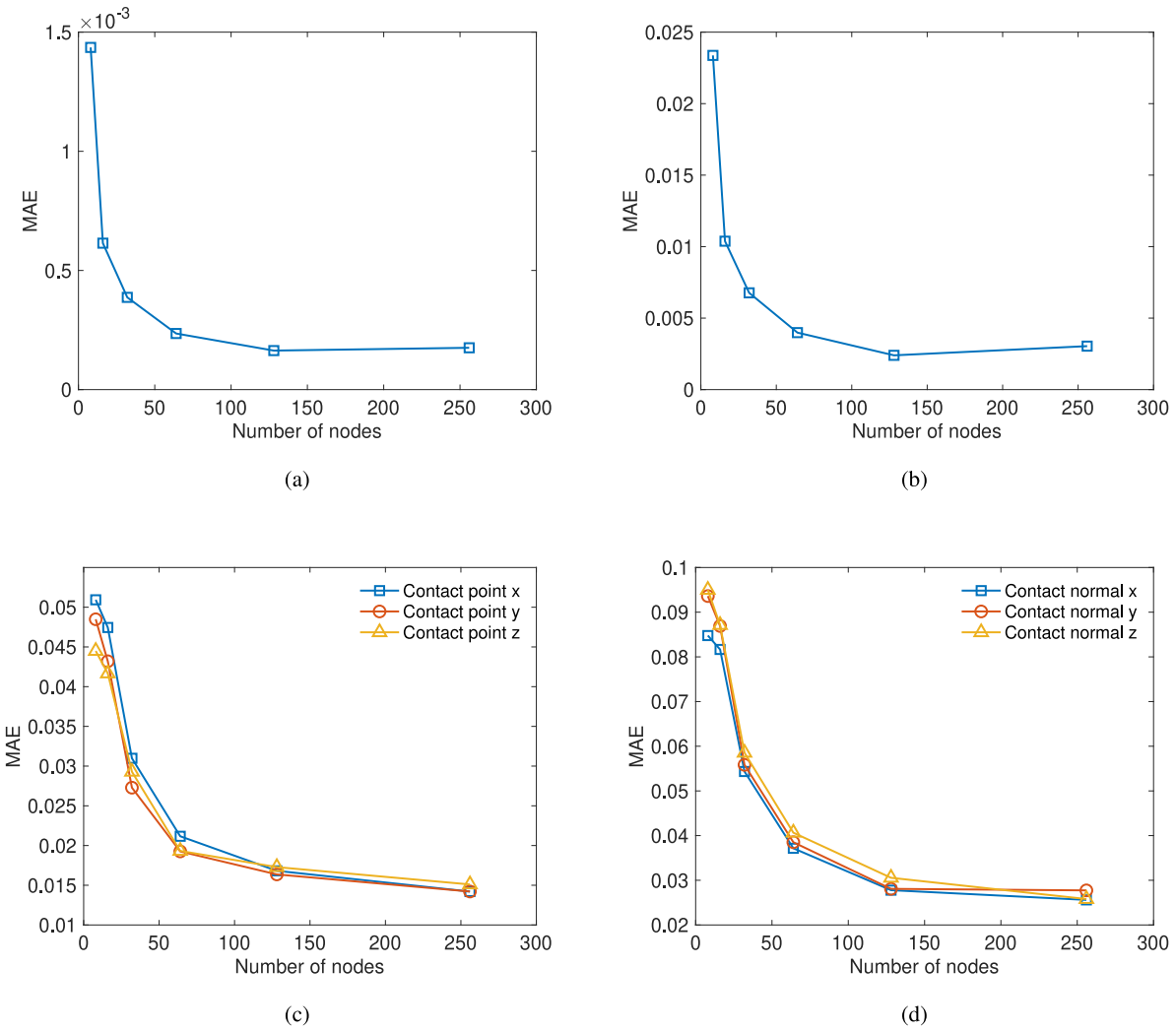


Fig. 12. Effect of the number of nodes in a hidden layer on the performance of regression networks: (a) contact volume, (b) contact area, (c) contact point and (d) contact normal.

learning model. Subsequently, these differences extend to the unbalanced force ratio. Notably, as the number of contacts increases, this discrepancy diminishes significantly. After 0.4 s, the unbalanced force ratios of NetDEM and GeoDEM are essentially converged to very minor values, reaffirming that the particles have reached a quasi-static state. Furthermore, Fig. 16(c) displays the variation in maximum packing height over time. It is noteworthy that the maximum height of the NetDEM remains lower than that of the GeoDEM until 0.4 s. In the initial stages of a complex collision involving multiple particles, inaccuracies in contact force predictions by the NetDEM are propagated to both the velocity and position of the particles, leading to disparities in the maximum packing height compared to the results of GeoDEM. These heights stabilize after 0.4 s, with NetDEM and GeoDEM reaching final maximum packing heights of 0.227 m and 0.231 m, respectively, indicating a close similarity.

The contact forces are normalized by the average of contact force and their distributions are presented in Fig. 17. The distributions of contact force from two DEM models are in good agreement, following an exponential distribution [39,40]. The majority of contact forces are very low in magnitudes. For instance, about 35.35% in NetDEM and 34.74% in GeoDEM of contact forces less than $0.3\overline{f}_n$, forming the weak chain of contact forces in the packing progress. The probability of contact forces between $0.3\overline{f}_n$ and $7\overline{f}_n$ (forming the strong contacts) accounts for 63.78% in NetDEM and 64.51% in GeoDEM. The slight differences between two DEM models could be the result of prediction errors of machine learning.

As a more quantitative analysis, we examine the fabric anisotropy of the particles in both the NetDEM model and the GeoDEM model. The details of the formulation and definition of the anisotropy can be found in Guo and Zhao [41]. Fig. 18 visualizes the anisotropy in contact normal orientation, contact normal force, and contact shear force of the particles after random packing. The spherical histograms of the contact normal direction and contact normal force are more uniformly distributed and only slightly elongated in the vertical direction, which is attributed to the aspect ratio of the particles being close to 1. The difference of anisotropy

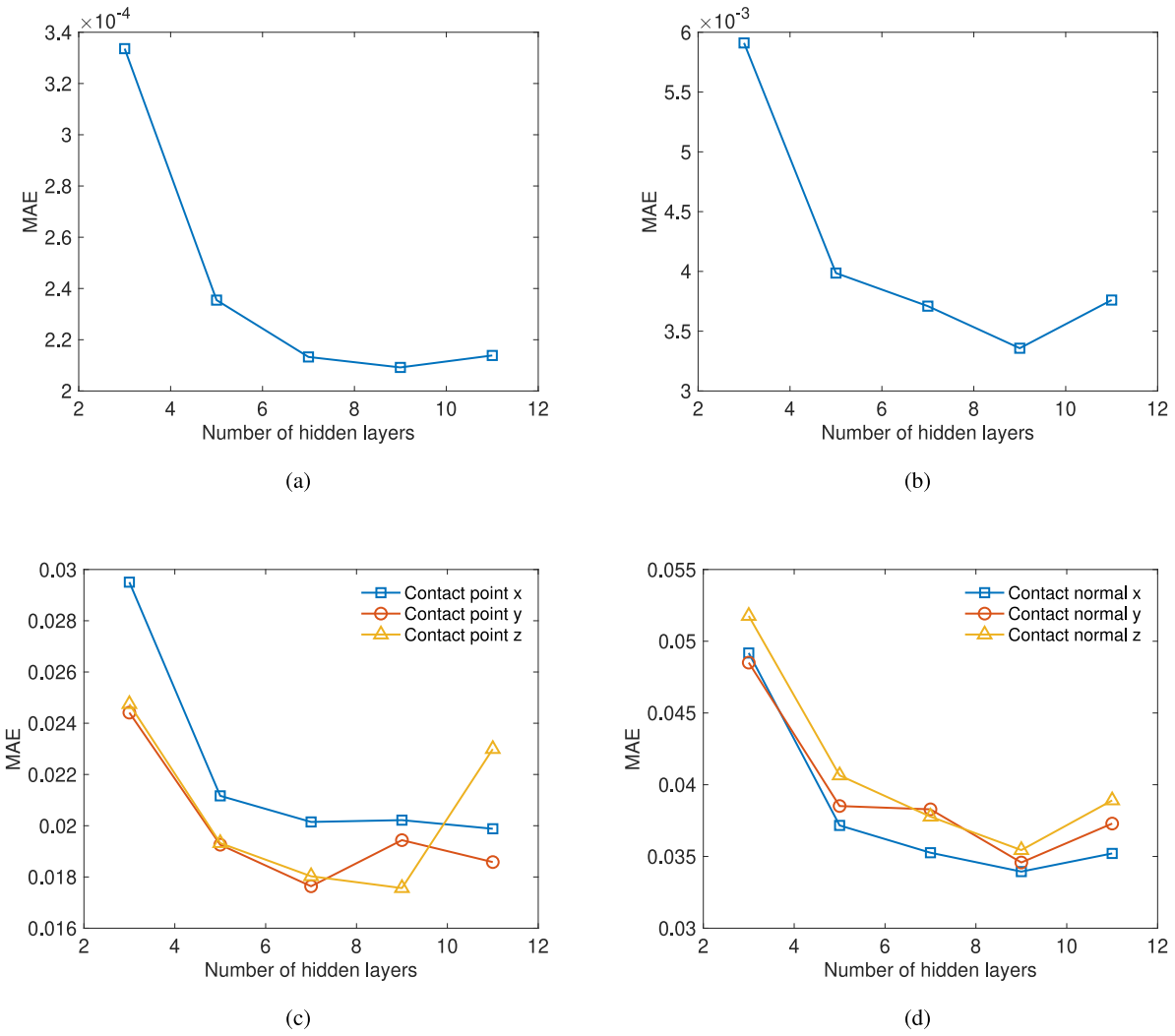


Fig. 13. The effect of the number of hidden layers on the performance of regression networks: (a) contact volume, (b) contact area, (c) contact point and (d) contact normal.

results between NetDEM and GeoDEM model is relatively small yet clearly visible, likely stemming from the limited particle count and predictive errors.

4.2. Column collapse

In the column collapse test, the initial particle assembly is created by random packing and has a dimension of 0.1 m × 0.1 m × 0.2 m. The column collapse test is proceeded by removing the lateral wall of the box container. The simulation lasts for 5.0 s. The parameters of DEM model are the same as those in the random packing test. To eliminate the effect of packing randomness, both the NetDEM and GeoDEM simulations utilize the same initial particle assembly created by GeoDEM. The porosity of the initial packing is approximately 0.451. Fig. 19 presents several snapshots of particle configurations during the reposing process. Particles suffer a severe rearrangement as time lapses, and gradually reach equilibrium in about 5.0 s. The run-out profiles of NetDEM simulation match well with those of GeoDEM simulation in the first 1.0 s. At the end of column collapse test, the repose angle from NetDEM is slightly lower than that from GeoDEM. As shown in Fig. 20, the final angle of repose of NetDEM and GeoDEM are calculated to be approximately 22.8° and 24.3°, respectively. To investigate this aspect, the contact forces from NetDEM and GeoDEM are extracted during the repose process, as shown in Fig. 21. The average contact force increases at the beginning, followed by a rapid decrease, and becomes stable at the time around 1.0 s. The average contact force simulated by NetDEM is overall a fairly good match with that simulated by GeoDEM. However, Fig. 19 shows that the particle configurations of the NetDEM model changed slightly within 1.0 s ∼ 5.0 s, whereas the particle conformations of the GeoDEM model were almost the same. In view of this, we locally zoomed

Particle shapes		Sphere 400 surface vertexes	Sand 400 surface vertexes	Bolt 1000 surface vertexes
Particle – particle contacts	Classification nets	99.25%	96.29%	94.87%
	Regression nets	3.10×10^{-3}	1.73×10^{-2}	4.79×10^{-2}
Particle – boundary contacts	Classification nets	99.73%	99.48%	99.12%
	Regression nets	2.41×10^{-3}	4.66×10^{-3}	1.33×10^{-2}

Fig. 14. The effect of particle shape on the performance of ANN models. The MAE for regression nets is the mean value of all contact geometry features. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

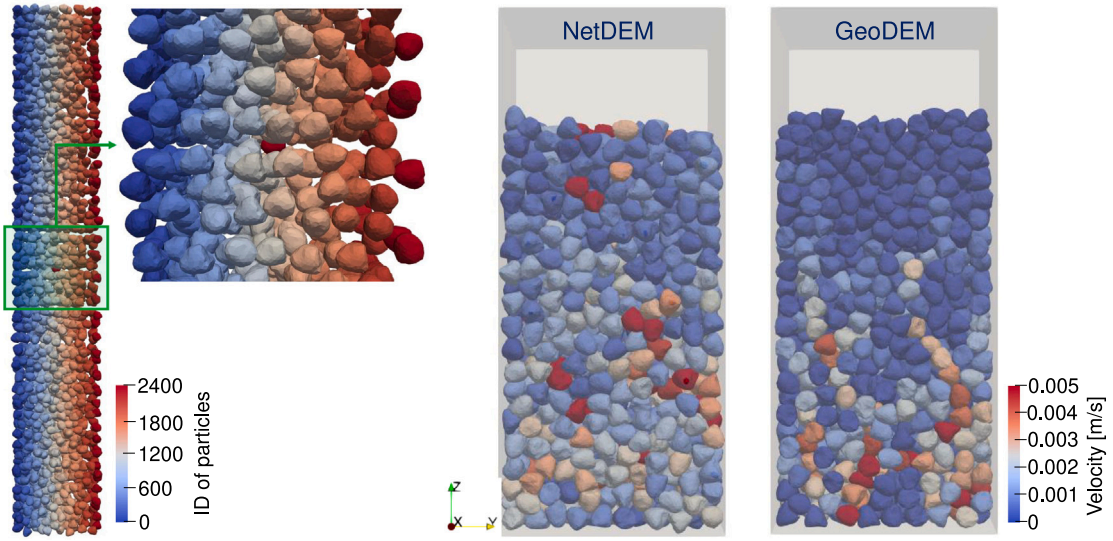
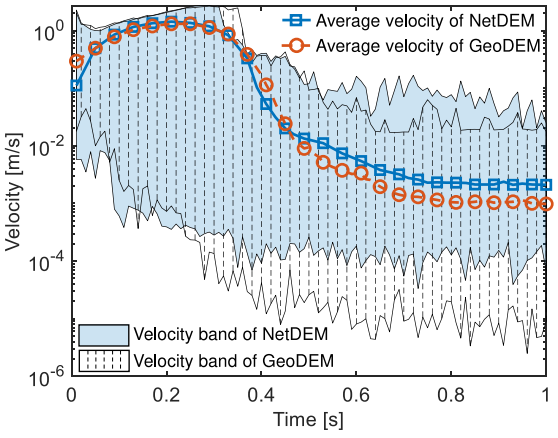


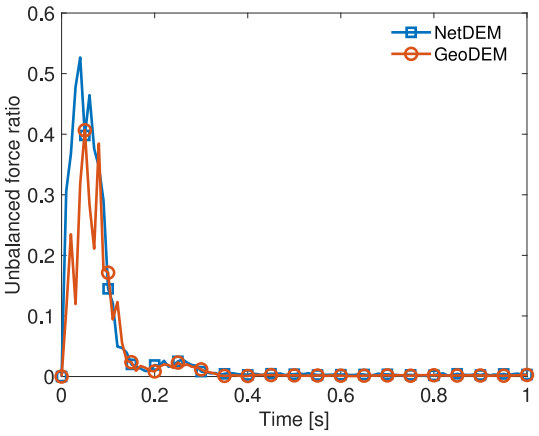
Fig. 15. Snapshots of the packing simulation: (a) initial stage, and (b) final packing. The colorbar in (a) indicates the IDs of the particles, and the colorbar in (b) indicates the magnitude of velocity of particles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in on the average contact force from 1.0 s to 5.0 s, as sketched in the embedded plot in Fig. 21. It is observed that the fluctuation of average contact force from NetDEM is significantly larger than that from GeoDEM. The ratio of their coefficients of variation (COV) is close to 20. The frequent fluctuations of contact force can cause minor vibrations in particles, resulting in movement of particles. Consequently, the NetDEM model exhibits a slightly smaller angle of repose in the column collapse test. The fluctuation of contact force in NetDEM is intrinsically resulted from the prediction errors of the machine learning models, as illustrated in Fig. 9(a).

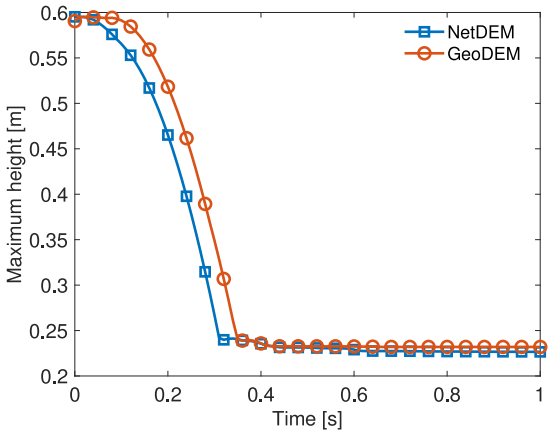
Fig. 22(a) plots the evolution of particle average velocity during the column collapse test. The evolutions of average particle velocity from NetDEM and GeoDEM show the same trend, i.e., a rapid increase followed by a gradual decrease to a stable value. At the end of column collapse test, the average velocity from NetDEM and GeoDEM are approximately 1.08×10^{-3} m/s and 2.21×10^{-4} m/s, respectively. The difference in average velocity between NetDEM and GeoDEM mainly comes from the prediction errors of the machine learning models. The average particle velocity in NetDEM is consistently larger than that of GeoDEM during the collapsing processes. Furthermore, the velocity bands of NetDEM and GeoDEM exhibit noticeable differences, particularly in the distribution of minimum particle velocities. Notably, the velocities in Fig. 22(a) are depicted on a logarithmic scale, with the



(a)



(b)



(c)

Fig. 16. Evolutions of (a) particle velocity, (b) unbalanced force ratio, and (c) maximum packing height during the random packing test.

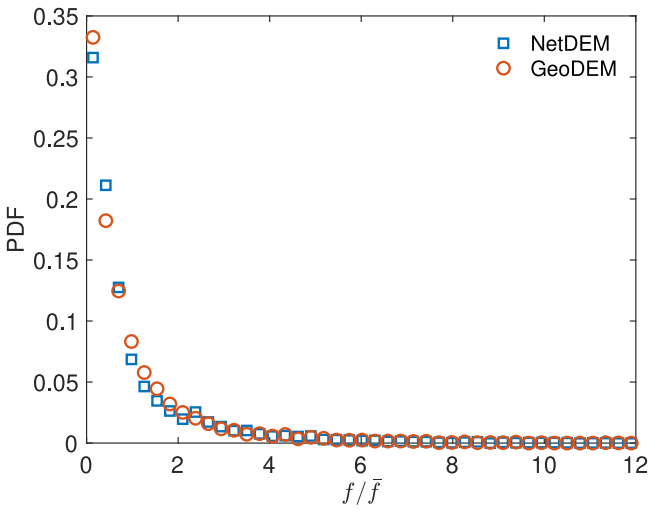
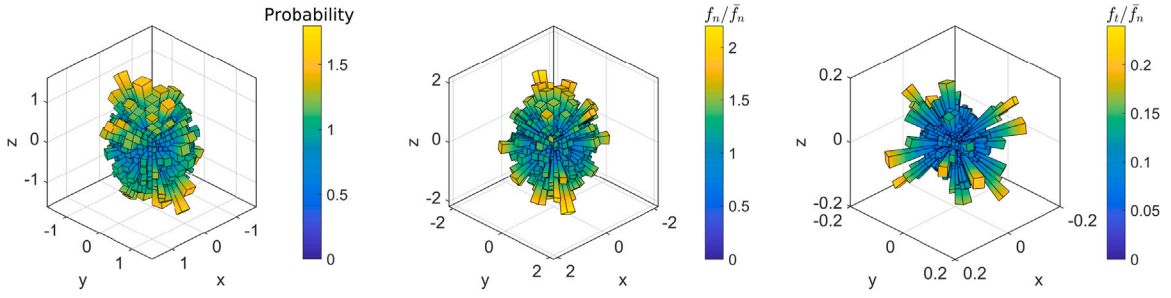
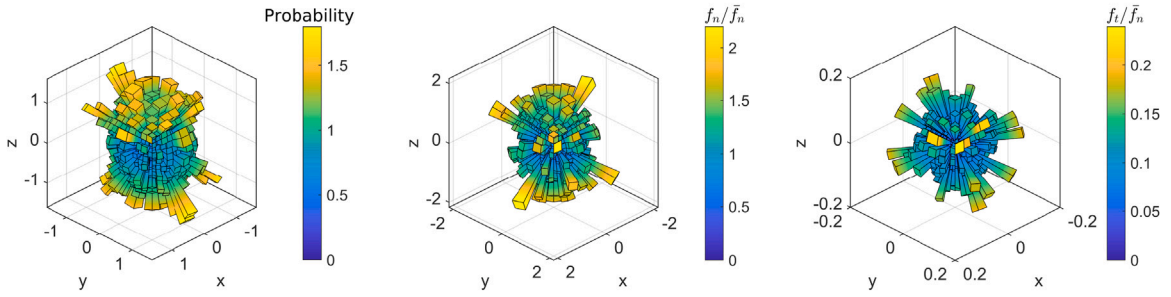


Fig. 17. Distribution of contact forces. The contact forces are normalized by the average contact force $\bar{f}$.



(a) NetDEM: contact normal orientation (b) NetDEM: contact normal force (c) NetDEM: contact shear force



(d) GeoDEM: contact normal orientation (e) GeoDEM: contact normal force (f) GeoDEM: contact shear force

Fig. 18. Fabric anisotropy at the end of packing. The top row represents the results of NetDEM model, and the bottom row represents the results of GeoDEM model.

minimum velocity stabilizing at around 10^{-5} m/s in NetDEM and around 10^{-9} m/s in GeoDEM, implying some variance between the two. Additionally, it is noted that the velocity range in NetDEM exceeds that of GeoDEM. The larger velocity explains the larger run-out distance of the particles as well as the smaller response angle in NetDEM.

Fig. 22(b) depicts the evolution of unbalanced force ratio during the column collapse test. Both NetDEM and GeoDEM exhibit similar trends in unbalanced forces, initially experiencing a sudden increase followed by a rapid decrease and eventual stabilization. It can be observed that the stability of the unbalanced force ratio is slightly lower for NetDEM than that for GeoDEM from 1.0 s to 5.0 s. This variation primarily arises from the larger fluctuation in the mean contact force for NetDEM as opposed to GeoDEM, as illustrated in Fig. 21. Furthermore, the variation of the column height over time is recorded, as shown in Fig. 22(c). As discussed in Section 3, one intrinsic limitation of the NetDEM approach is that there exist avoidable prediction errors in the contact forces. Such prediction errors would then result in an overall greater value and also a little bit unstable unbalanced force for the NetDEM simulations, comparing with the conventional GeoDEM simulations. This phenomenon is also evidenced by the greater averaged velocity of the particles, as depicted in Fig. 22(a). As a result, it looks like that the NetDEM simulation exhibits some form of "creeping". Remedies to mitigate the prediction errors in contact forces and thereby the "creeping" merits further exploration. The final stabilized maximum heights for NetDEM and GeoDEM are 0.119 m and 0.126 m, respectively.

4.3. Triaxial compression

In the triaxial compression test, as illustrated in Fig. 23, 2150 irregular-shaped particles are packed into the box container with a base of 0.1-by-0.1 m, and the final packing height is approximately 0.2 m. The parameters of DEM model are the same as those in the random packing test. After packing, the particles are allowed to relax for 1.0 s to reach the equilibrium state. The porosity of the particle assembly is approximately 0.453. It should be noted that the aforementioned packing and equilibrium processes are both simulated using GeoDEM. Subsequent compression processes are simulated using NetDEM and GeoDEM separately, ensuring the consistency of the initial particle configurations. To proceed with the compression test, the particles first undergo an initial compression with an isotropic pressure of 100 kPa. Then, the particles are sheared by continuously moving the top wall downwards at a speed of 0.005 m/s, meanwhile maintaining the lateral stress at 100 kPa through a servo control [42]. The shearing process lasts for 10.0 s with a final axial strain of approximately 30%.

To begin with, we investigated the macroscopic mechanical response of the granular packing in terms of deviatoric stress ratio and volumetric strain. Herein, the deviatoric stress ratio is defined as p/q, where p is the deviatoric stress, and q is the mean stress. The volumetric strain is defined as $\epsilon_V = \ln(V_0/V_p)$, where V_0 and V_p, respectively, are the initial and current volume of the packing. $\epsilon_V < 0$ means dilatation of the packing. More details of the definitions are referred to Zhao and Zhao [43] and Lai and Huang [44]. The results of stress–strain evaluation during the shearing process are shown in Fig. 24. The deviatoric stress ratio gradually

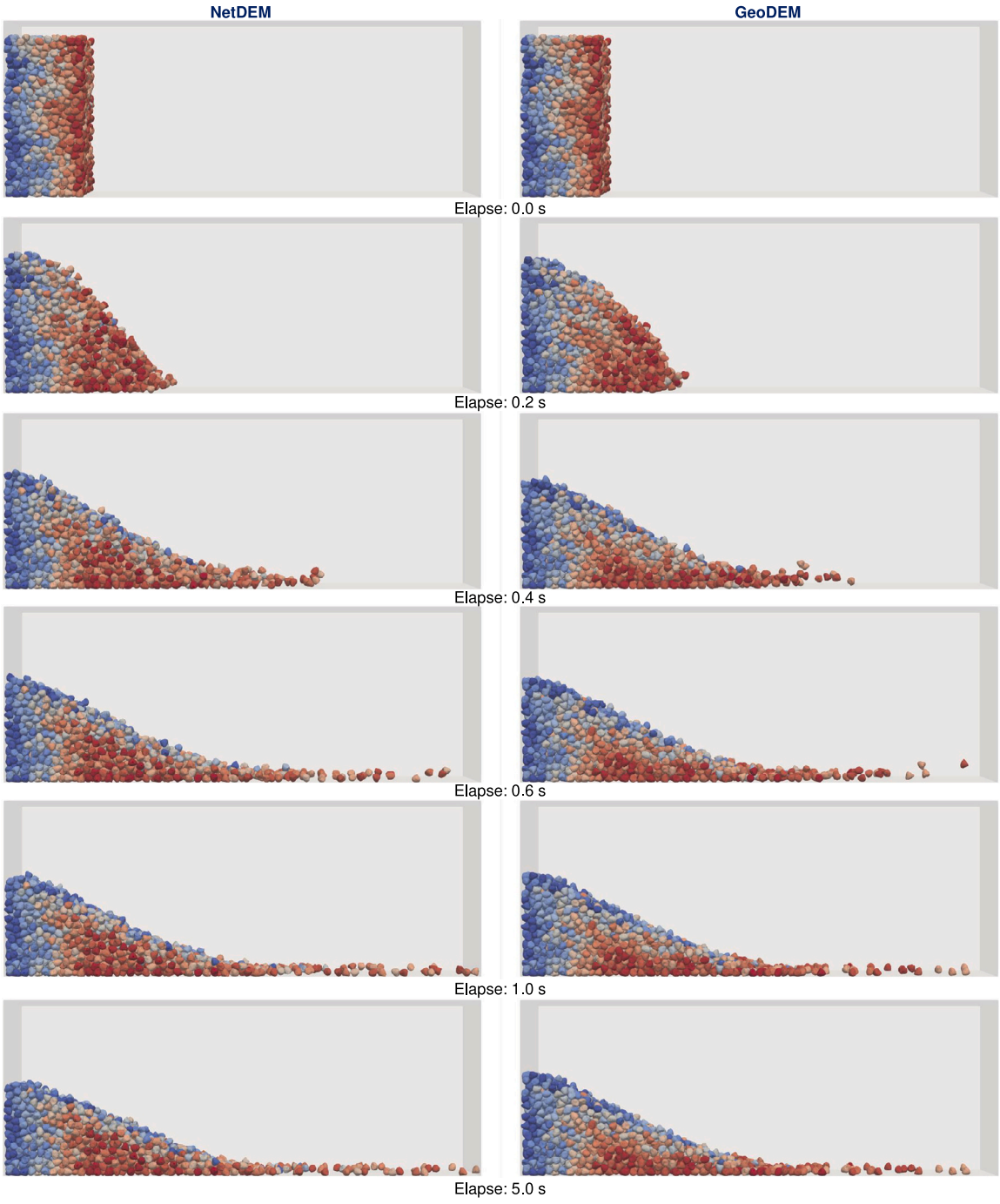


Fig. 19. Snapshots of particle configurations during the column collapse test. The color represents different particle indices for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

increases as the axial strain approaches approximately 5%, after which it becomes stable and exhibits a nearly constant value (about 1.2 for NetDEM and GeoDEM). The volumetric strain increases with increasing axial strain, showing typical shear behaviors of loose soils. When the axial strain reaches approximately 15%, the volumetric strain almost plateaus. It can be observed that the results simulated by NetDEM and GeoDEM match well with each other, indicating that the NetDEM can accurately reproduce the macroscopic mechanical behavior in triaxial shear tests. It should be mentioned that some fluctuations in the volumetric strain of the

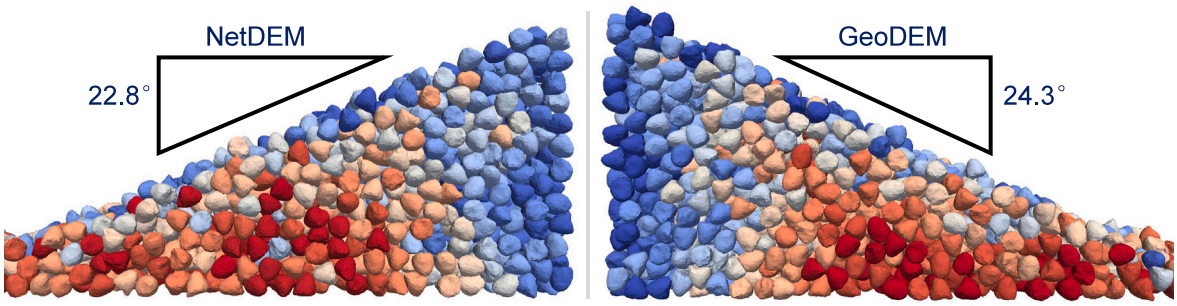


Fig. 20. Zoom-in snapshots of the particle configurations at the end of collapsing and the corresponding repose angle. The color represents different particle indices for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

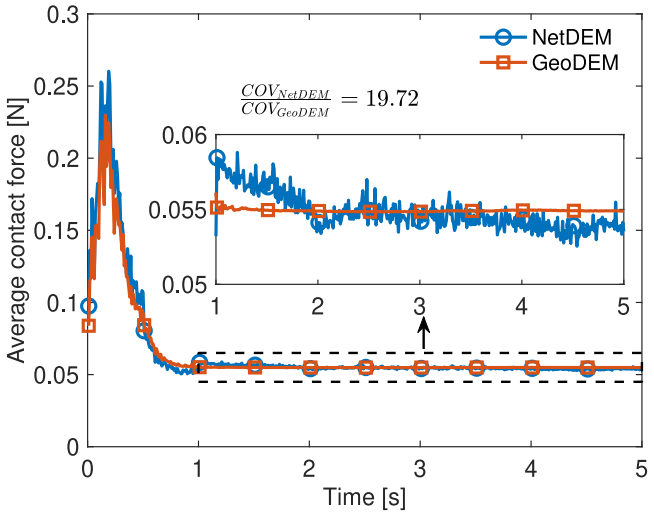


Fig. 21. Evolution of average contact force during the column collapse test. The average contact force represents the sum of the contact forces divided by the number of contacts.

NetDEM model, as compared to the GeoDEM model, were observed when the axial strain is in the range of 5% ∼ 15%. The inertia number is utilized to verify whether the particles are in quasi-static state during the shearing process. The inertia number is defined as $I = \gamma d/\sqrt{\rho/P}$, where γ is the shear rate, d is particle size, ρ is particle density and P is the confining stress [45]. The inertial number of the particles in the triaxial compression test stands at 3.1×10^{-4}, falling below 10^{-3}, indicating that the particles are in a quasi-static state. Therefore, the fluctuation phenomenon can be caused by the prediction errors of machine learning, which could be mitigated by employing a deeper neural network. This would increase the computational expenses of NetDEM, necessitating a balance between computational efficiency and accuracy according to the actual situation. Nevertheless, the final volumetric strain from NetDEM and GeoDEM are fairly close, at approximately 3.5%.

Furthermore, the microscopic mechanical behaviors of the granular packing are examined. The distribution of particle contact forces is presented in Fig. 25. The contact force distributions from the two DEM models exhibit a good match, with only a slight discrepancy observed in the portion of small contact forces.

The aforementioned compression tests demonstrate the capability of NetDEM to precise capture of the mechanical behaviors of loose soils. Given the distinct mechanical behaviors between loose and dense soils, we extended our investigation by conducting triaxial compression tests on the dense particle assemblies to enhance the comprehensiveness of the validation. A dense particle assembly with a porosity of 0.362 is prepared, and the steps and parameters of this simulation are identical to the previous triaxial compression test. Fig. 26 illustrates the results of stress–strain evolution during the triaxial compression process. The deviatoric stress ratio exhibits an initial rapid increase followed by a gradual decrease. The peak deviatoric stress ratio is approximately 1.38 at an axial strain of 5%. Correspondingly, the volumetric strain initially shows a slight increase, signifying a contraction phase, which is succeeded by a decrease indicating a dilation phase. These macroscopic responses of the particles closely resemble the behavior commonly observed in laboratory settings for dense soils. The distribution of particle contact forces is also plotted in Fig. 27. The results indicate a close alignment between the simulations of NetDEM and GeoDEM.

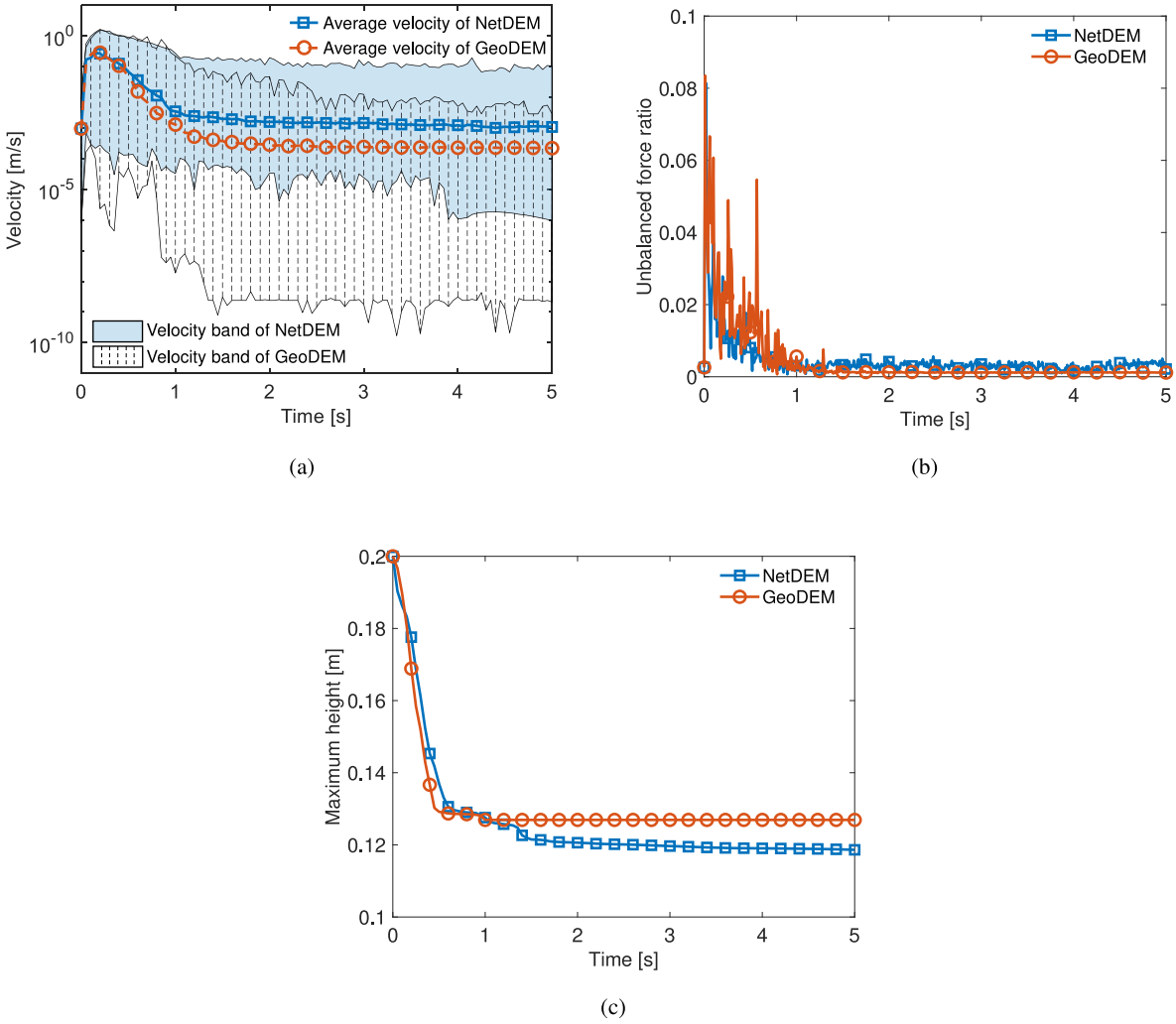


Fig. 22. Evolutions of (a) the particle velocity, (b) unbalanced force ratio, and (c) maximum height during the column collapse test.

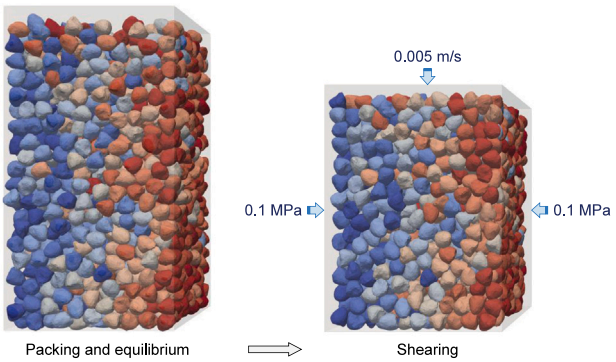
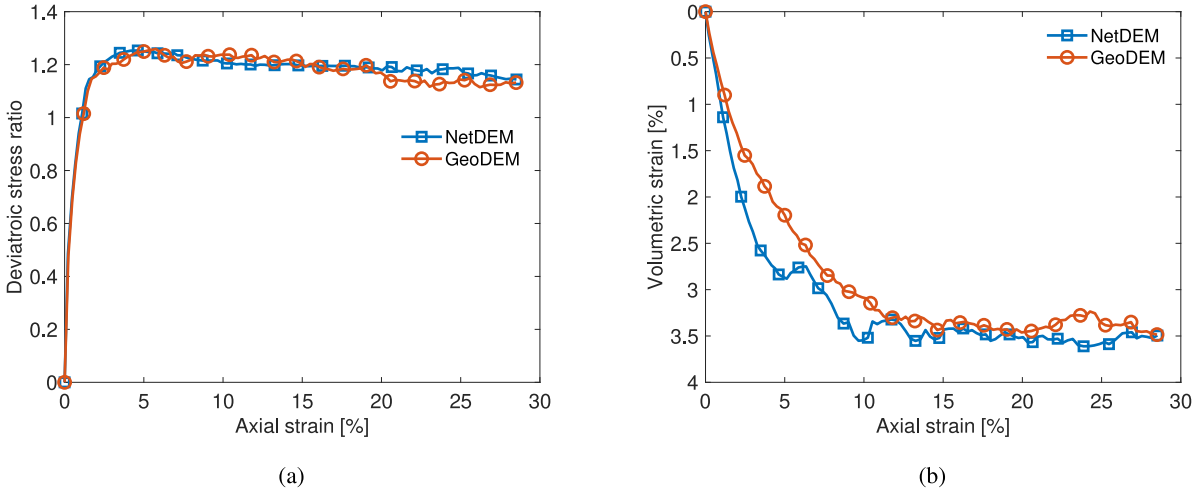
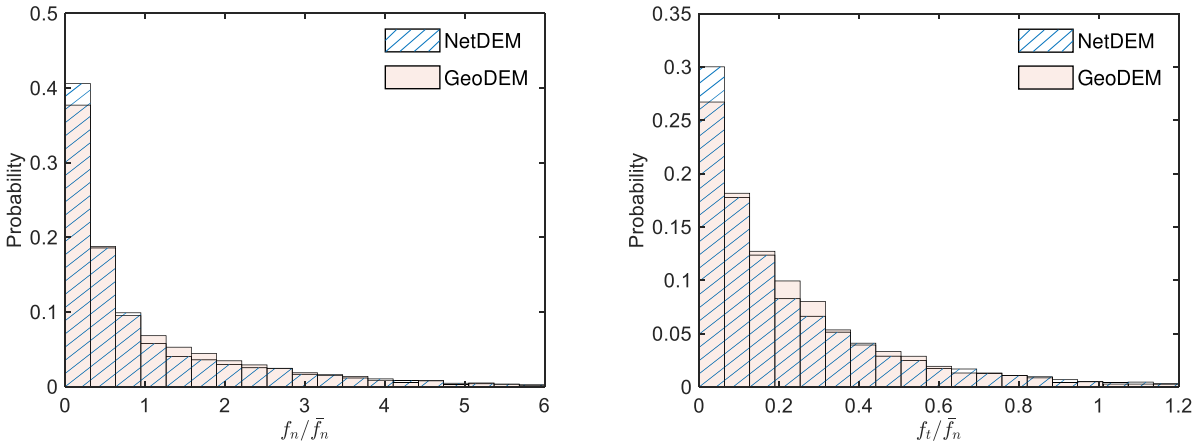


Fig. 23. Illustration of the main steps of the triaxial compression test. The color represents different particle indices for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



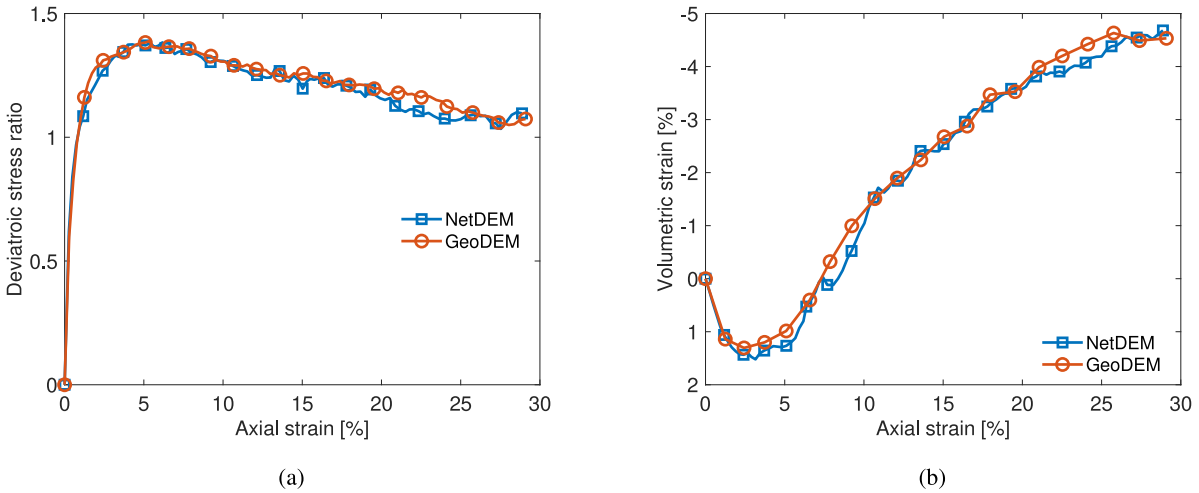
(a) (b)

Fig. 24. Evolution of (a) deviatoric stress ratio, and (b) volumetric strain of the loose soils during the triaxial compression process.



(a) (b)

Fig. 25. Distribution of (a) contact normal forces, and (b) contact shear forces of the particles of the loose soils. Both the contact normal and shear forces are normalized by the average contact normal force.



(a) (b)

Fig. 26. Evolution of (a) deviatoric stress ratio, and (b) volumetric strain of the dense soils during the triaxial compression process.

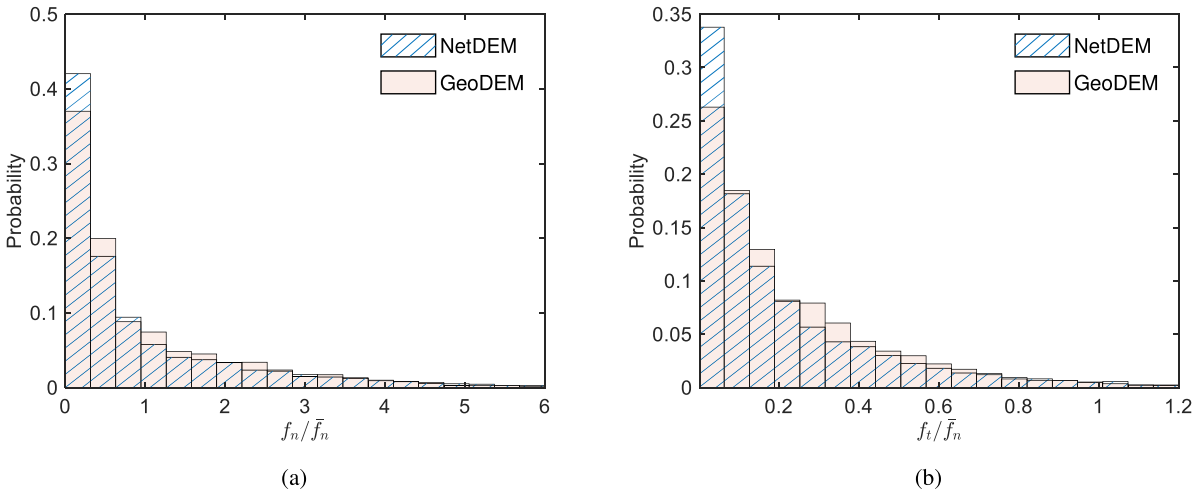


Fig. 27. Distribution of (a) contact normal forces, and (b) contact shear forces of the particles of the dense soils. Both the contact normal and shear forces are normalized by the average contact normal force.

5. Discussions

5.1. Acquisition of reference contact datasets

The performance of the machine learning-enabled contact detection and resolution depends heavily on the reference contact datasets generated by the reference contact approach. In this work, the contact volume-based contact approach is selected for the generation of training datasets. The advantages of the contact volume-based approach are mainly in the following three aspects: (1) The contact geometry features are defined explicitly and analytically. For instance, the contact volume is obtained by the mesh intersection operation on the polyhedral particle. On this basis, other contact geometry features, i.e., contact area, contact normal direction, and contact point can be calculated without additional approximations. (2) The contact volume-based approach is derived based on the energy-conserving contact theory, which is beneficial to the numerical stability. (3) The contact geometric features in the contact volume-based approach are additive. This implies that when multiple contacts exist between two particles, we can compute the contact geometry features separately in each subdomain and then sum them up to obtain the overall contact geometry features. In other words, we can aggregate the multiple contact problems into a single contact by accumulation. This property ensures consistency in the dimensionality of the output variables for machine learning models.

It is worth noting that there are some other studies for irregular-shaped particles DEM, e.g., level set-DEM [9,46], spherical harmonics-DEM [11,24], and signed distance field-DEM [30]. The common features of these methods are node-based and grid-dependent. Specifically, for two particles in a contact pair, the surface of one particle is discretized into a series of nodes, and the contact flag is determined by checking for the existence of nodes intruding into the other particle. Subsequently, the contact force is calculated for each intruding node and the total contact force is synthesized from the contact forces on all intruding nodes. The integration of machine learning with these irregular-shaped particles DEM also merits further exploration.

5.2. Computational efficiency

One of the most important advantage of NetDEM lies in its high computational efficiency. To investigate this aspect, we calculated the Cundall number of the aforementioned three numerical experiments, as shown in Table 2. The Cundall number is defined as $C = N_T N / T_{CPU}$, where N_T is the number of timesteps, N is the number of particles, and T_{CPU} is the CPU time of the simulation. Note that the DEM simulations in this work are performed on an Intel i910980XE CPU with 32 logic cores. The significantly higher Cundall numbers observed in the NetDEM simulations compared to GeoDEM indicate a notable increase in computational efficiency. The Cundall number for random packing, column collapse, and triaxial compression in NetDEM simulations are 4.25, 4.57, and 5.73 times, respective, higher than those in GeoDEM. This implies that the computational efficiency of NetDEM is enhanced by approximately fivefold compared to GeoDEM. The Cundall number of NetDEM compared to GeoDEM is largest in triaxial compression simulation and smallest in random packing simulation. This phenomenon may result from the larger number of contacts in the triaxial compression. As a result, the computational efficiency of the machine learning approach can be fully utilized.

To gain an insight into computational efficiency of NetDEM, we further investigate the impact of the number of polyhedron vertexes. The computational efficiency of the polyhedral particle model used in this work is sensitive to the number of surface vertexes. It is difficult to accurately describe the shape of real particles when using the polyhedral particle model with too few surface vertexes. However, too many surface vertexes again significantly increase the computation time. Hence, GeoDEM usually

Table 2
Comparison of the Cundall number of NetDEM and GeoDEM. (Simulations are conducted with 32 cores.)

Cundall number	Random packing	Column collapse	Triaxial compression
NetDEM	5.25×10^4	4.46×10^4	6.05×10^4
GeoDEM	1.23×10^4	9.8×10^3	1.05×10^4
Ratio	4.26	4.57	5.73

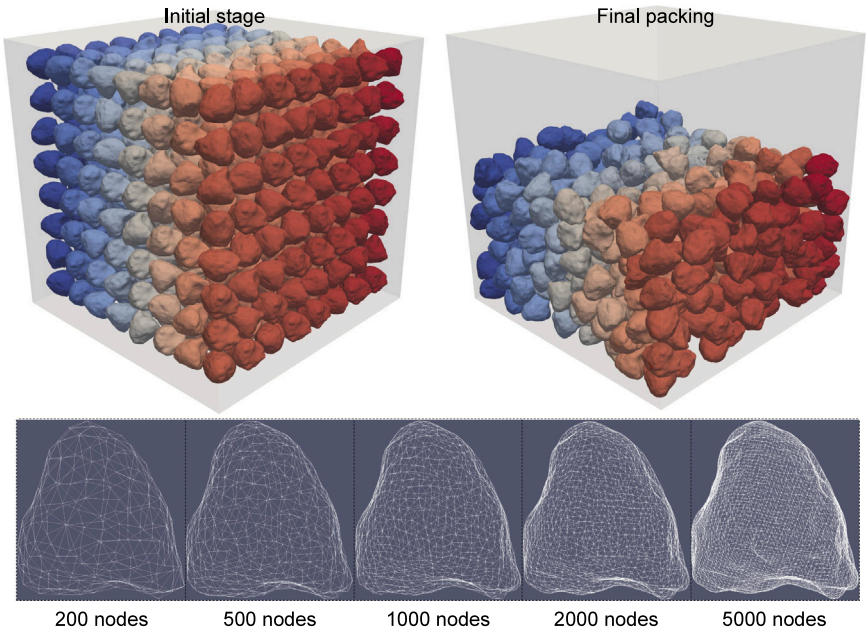


Fig. 28. Snapshots of the packing simulation with dense surface vertexes. The color represents different particle indices for better visualization. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

needs to weigh a balance between the accuracy of particle shape characterization and the computational efficiency to obtain the optimal number of surface vertexes [26]. As a more quantitative analysis, we simulate the random packing test involved 512 particles with an equivalent size of 0.1 m. Particles are insert into a box container with size of 1 m by 1 m by 1 m, and are allowed to settle down under gravity. The packing progress lasts one second, and timestep is set as 1.0×10^{-4} s, thus the number of DEM cycle is 10,000. Fig. 28 shows the particle configurations at initial and final stage. Five sets of random packing tests with different numbers of particle surface vertexes are performed. Particles with the number of 200, 500, 1000, 2000, and 5000 surface vertexes are plotted in Fig. 28. It is important to note that, different from the numerical tests in Section 4 which use 32 cores, the additional random packing tests in this subsection (i.e., Section 5.2) are simulated on a single core with the intention of ensuring a more equitable comparison.

Fig. 29 demonstrates the results of Cundall number simulated by NetDEM and GeoDEM. As expected, the Cundall number for GeoDEM decreases with the number of particle surface vertexes, while the Cundall number for NetDEM remains a stable value. It is because the surface vertexes of the particles in GeoDEM are directly involved in the contact computations. However, the contact detection and resolution in NetDEM are handled by ANNs, whose computational efficiency is relatively independent from particle surface meshes and is mainly correlated to the architecture of the ANNs. The computational efficiency of NetDEM against GeoDEM is more pronounced as the number of particle surface vertexes increases. For instance, the Cundall number ratio of NetDEM to GeoDEM is 72.17 when the particle with 5000 surface vertexes. The results indicate that NetDEM offers great advantages in terms of computational efficiency for DEM simulations involving particles with complex shapes. It is worth mentioning that there are also some studies on improving the computational efficiency of polyhedral particles, e.g., GPU-based method [47], global direct search method [48], and so on. However, these methods are still essentially the geometry-based contact algorithms, and their computational efficiency is affected by the number of vertexes of the polyhedral particles.

Fig. 30 provides a deeper evaluation of the influence of the number of logical cores on the Cundall number. The results indicate that the Cundall numbers for both NetDEM and GeoDEM increase with the number of logical cores.

After discussing the impact of the number of particle surface vertexes on computational efficiency, a subsequent question arises regarding the influence of the number of surface vertexes on the predictive performance of the ANN models. Datasets for particle–particle contacts are generated using particles characterized by varying quantities of surface vertexes. The ANN models corresponding to distinct surface vertexes are individually trained with the parameters delineated in Table 1. Subsequently, the

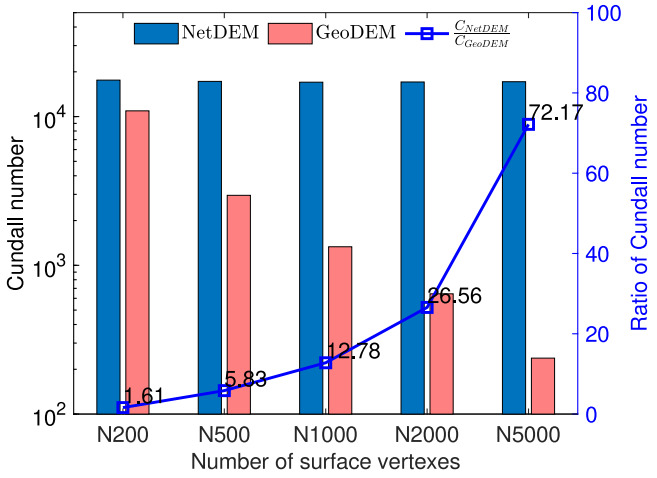


Fig. 29. Cundall number for random packing simulations with varying surface vertex number on a single core. The label N5000 indicates that the polyhedral particle has 5000 surface vertexes.

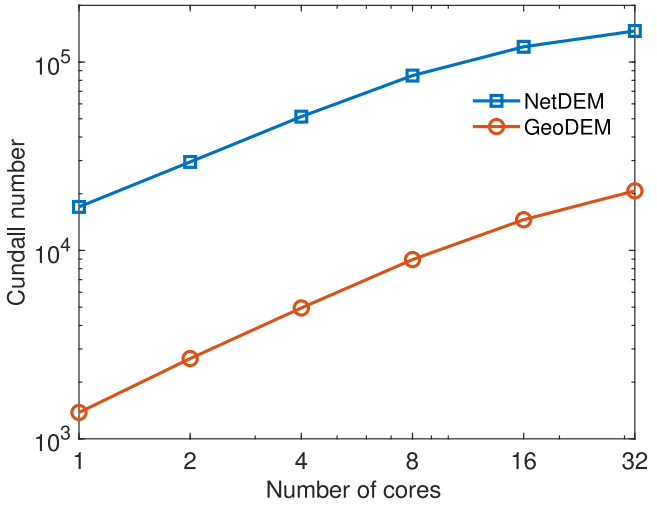


Fig. 30. Cundall number versus the number of logic cores. Note that the number of surface vertexes is 1000.

predictive performance of the ANN models is assessed on testing datasets, as depicted in Fig. 31. The results reveal that both the classification and regression networks exhibit minor sensitivity to the number of surface vertexes. In scenarios involving intricate geometries, the machine learning-enabled approach may exhibit more.

5.3. One model for various shapes

One limitation of the present machine learning-enabled DEM is that each particle shape requires a set of ANNs. In general, a granular specimen often consists of a large number of particles with different shapes and thus would require multiple sets of ANNs. The development of one set of ANNs accommodative to all shapes merits further exploration. The great challenge of this issue is how to choose reasonable particle shape descriptors as the inputs of the ANN model. There are some particle shapes using closed forms with geometric equations, such as poly-ellipsoid [49], poly-super-ellipsoid [43], super-quadrics [10,50]. As illustrated in Fig. 32, the shape of these three particles is completely controlled by the corresponding shape descriptors.

Taking the poly-super-quadrics as an example, we briefly describe the process of integrating poly-super-quadrics particles with different shapes into NetDEM. The framework for the integration is essentially the same as those described in Section 2. The only discrepancy is the inclusion of the shape descriptors as inputs to ANNs in addition to the position descriptors. The complete inputs of ANNs for poly-super-quadrics particles could be written as: (1) The shape descriptors of the object particle, i.e., $[r_{x+}, r_{x-}, r_{y+}, r_{y-}, r_{z+}, r_{z-}, \epsilon_{x+}, \epsilon_{x-}, \epsilon_{y+}, \epsilon_{y-}, \epsilon_{z+}, \epsilon_{z-}]_{Object}$; (2) The shape descriptors of the cue particle, i.e., $[r_{x+}, r_{x-}, r_{y+}, r_{y-}, r_{z+}, r_{z-}, \epsilon_{x+}, \epsilon_{x-}, \epsilon_{y+}, \epsilon_{y-}, \epsilon_{z+}, \epsilon_{z-}]_{Cue}$; (3) The position descriptors of the cue particle, i.e., $[x_c, y_c, z_c, q_0, q_1, q_2, q_3]$.

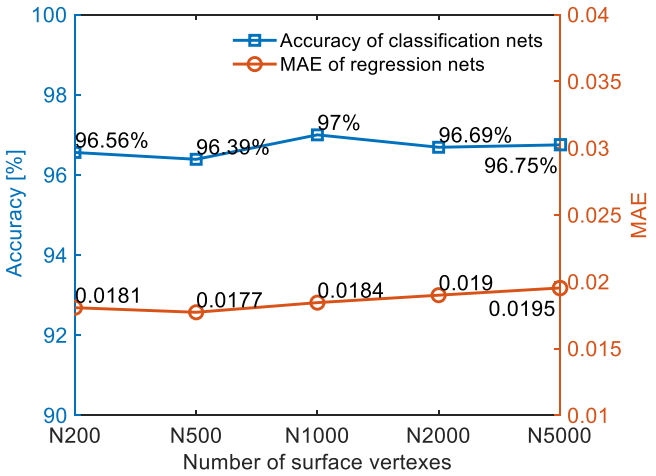


Fig. 31. The performance of ANN models with varying surface vertex number.

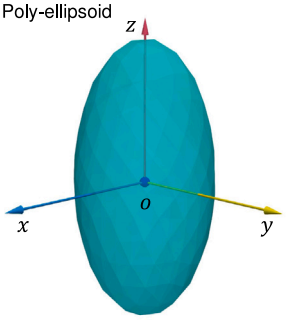
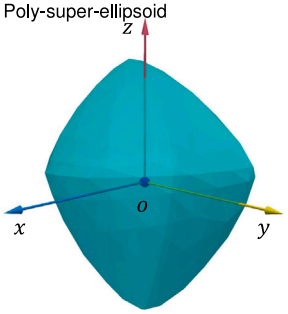
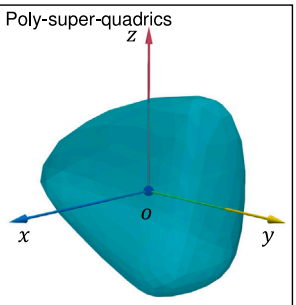
	Poly-ellipsoid	Poly-super-ellipsoid	Poly-super-quadrics
Paericle surface			
Formulation	$\left(\frac{x}{r_x}\right)^2+\left(\frac{y}{r_y}\right)^2+\left(\frac{z}{r_z}\right)^2=1$	$\left(\left\lvert\frac{x}{r_x}\right\rvert^{\frac{2}{\epsilon_1}}+\left\lvert\frac{y}{r_y}\right\rvert^{\frac{2}{\epsilon_1}}\right)^{\frac{\epsilon_1}{\epsilon_2}}+\left\lvert\frac{z}{r_z}\right\rvert^{\frac{2}{\epsilon_2}}=1$	$\left\lvert\frac{x}{r_x}\right\rvert^{\frac{2}{\epsilon_x}}+\left\lvert\frac{x}{r_y}\right\rvert^{\frac{2}{\epsilon_y}}+\left\lvert\frac{z}{r_z}\right\rvert^{\frac{2}{\epsilon_z}}=1$
Shape descriptors	Semi-major axis lengths: r_x, r_y, r_z	Semi-major axis lengths: r_x, r_y, r_z Shape parameters: ϵ_1, ϵ_2	Semi-major axis lengths: r_x, r_y, r_z Shape parameters: $\epsilon_x, \epsilon_y, \epsilon_z$

Fig. 32. Illustration of the particle surface, formulation, and shape descriptors for poly-ellipsoids, poly-super-ellipsoids and super-quadrics.

The number of input variables totals 31, which may increase the difficulty of ANNs in training and impact the computational efficiency.

There also exist many more complex shapes, such as the particle shape used in this work, that are difficult to be characterized by mathematical functions. In such cases, it may require large number of combined shape descriptors, making it difficult to use them all as inputs. One feasible approach is to utilize spherical harmonics to describe irregular-shaped particles, adopting the spherical harmonics coefficients as particle descriptors. The spherical harmonics coefficients depend on the exact particle orientation in the prescribed coordinate system [51]. Consequently, one can use rotation-invariant shape descriptor, i.e., the amplitude of each spherical harmonics frequency, as the inputs of the ANNs [52,53].

As mentioned in the introduction, although the concept of machine learning for contact detection and properties is straightforward, there are some challenges regarding its integration into DEM simulation and addressing its computational accuracy, efficiency and stability. The numerical aspects and computational issues, including the selection of data inputs and outputs, preparing the training samples, regularizing the data inputs and outputs from a DEM simulation, the design of machine learning architecture and hyper-parameters, etc., have been discussed in this work. Additionally, this work discusses and explores preliminary directions for advancing machine learning-enabled DEM, including the acquisition of reference contact datasets, improving computational efficiency, and developing a single model adaptable to various shapes. While DEM has successfully been applied to various natural and engineering problems, its high computational expense remains a significant limitation for large-scale particulate systems. Machine learning techniques have shown promising performance in other computational methods, such as the finite element

method and computational fluid dynamics. This work addresses critical aspects of integrating machine learning into DEM, and it is anticipated that the insights presented here will inspire further research in this area.

6. Conclusion

This study has extended the applicability of the machine learning-enabled DEM to the intricate 3D realm involving irregular-shaped particles. The approach integrates two neural network architectures into the DEM framework, namely a classification network for contact detection and a regression network for contact resolution. The input parameters for these networks are geometric descriptors such as particle positions and rotations, while the outputs comprise contact flags and contact geometric features, respectively. The polyhedral particle model is employed to represent general irregular shapes, and a contact volume-based energy-conserving model is utilized to assess contact behavior and generate the requisite training dataset of contact geometry features.

The performance of the machine learning-enabled approach has been evaluated by first concerning individual contacts, aiming to reveal the effect of various ANN model parameters on the prediction accuracy and consequently to determine the optimal model architecture. The trained classification networks demonstrate good accuracy, surpassing 99% for particle-boundary contacts and 96% for particle–particle contacts. The errors in contact geometric features are within the order of 10^{-3} for particle-boundary contacts and 10^{-2} for particle–particle contacts. These results underscore the predictive capabilities of the developed ANNs in accurately detecting contact flags and resolving contact features among irregular particles. Subsequently, the overall performance of the machine learning-enabled DEM has been assessed through comprehensive simulations of random packing, column collapse, and triaxial compression. The results indicate that the machine learning-enabled DEM adeptly simulates 3D irregular-shaped particles with dense contacts, accurately capturing their mechanical behavior while maintaining good numerical stability. Furthermore, the computational efficiency of the machine learning-enabled DEM has been also discussed. Comparative analyses between NetDEM and GeoDEM in the three numerical experiments reveal that the Cundall number of NetDEM is approximately five times greater than that of GeoDEM, signifying a substantial improvement in computational efficiency. This advantage in computational efficiency becomes particularly pronounced when handling complex shaped particles with a large number of surface vertex. The study demonstrates the efficacy and potential of machine learning-enabled DEM in handling complex 3D scenarios involving irregular-shaped particles, offering enhanced predictive accuracy, simulation fidelity, and computational efficiency compared to the conventional DEM.

CRediT authorship contribution statement

Shuai Huang: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation. **Pei Wang:** Writing – review & editing, Supervision, Resources. **Zhengshou Lai:** Writing – review & editing, Software, Methodology, Conceptualization. **Zhen-Yu Yin:** Writing – review & editing, Supervision, Project administration. **Linchong Huang:** Writing – review & editing, Supervision. **Changjie Xu:** Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work has been financially supported by the National Key R and D Program of China (2023YFC3009400), the National Natural Science Foundation of China (42207210), the Research Grants Council of Hong Kong (15226322, 15220221, 15226322, 15229223, 15227923), the Natural Science Foundation of Jiangxi Province, China (20242BAB26079), and the Shenzhen Science and Technology Project for Sustainable Development, China (KCXFZ202002011008532). We gratefully acknowledge the insightful comments and constructive suggestions provided by the reviewers, which significantly enhanced the quality of this manuscript.

Appendix. Contact volume-based energy conserving contact theory

The energy-conserving contact (ECC) theory originally proposed by Feng [25] is adopted to resolve the contact behavior. According to the ECC theory, two particles in contact are placed in configuration space ($\boldsymbol{x}$-o-$\boldsymbol{\theta}$), taking the centroid of one particle as

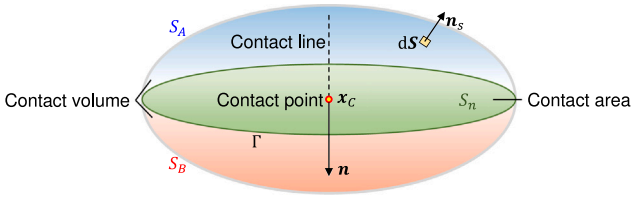


Fig. A.33. Illustration of contact geometric features in contact volume-based contact resolution approach.

the origin, and the configuration of the other particle can be characterized by $(\boldsymbol{x}, \boldsymbol{\theta})$. The key assumption of ECC is that the contact behavior can be characterized by a contact energy potential, denoted as w. The contact energy potential is a function of position $\boldsymbol{x}$ and rotation $\boldsymbol{\theta}$ of two particles. Then, the inter-particle normal contact force $\boldsymbol{F_n}$ and moment $\boldsymbol{M_n}$ can be characterized as

$$\boldsymbol{F}_n = -\frac{\partial w(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{x}} \tag{A.1}$$

$$\boldsymbol{M}_n = -\frac{\partial w(\boldsymbol{x}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \tag{A.2}$$

The form of contact energy potential $w(\boldsymbol{x}, \boldsymbol{\theta})$ can be defined on demand, whereas it is required to satisfy both translational and rotational invariants. The contact potential is usually expressed as a function of the contact geometry features, such as overlap and contact volume [22,27,30,54,55]. In this study, the contact volume-based energy-conserving contact model is adopted. In this approach, the contact potential is denoted as $w(V_c)$, where V_c is the volume of contact region between two particles in contact. With the polyhedral particle model, the contact volume can be resolved by the mesh Boolean method, and the contact detection can be achieved by checking whether the contact volume is greater than zero. The normal contact force $\boldsymbol{F_n}$ is written as:

$$\boldsymbol{F_n} = -w'(V_c)\boldsymbol{S}_n \tag{A.3}$$

where the contact energy potential $w(V_c) = k_n V_c^m$, $(')$ represents derivative operation, k_n is the normal stiffness. It should be noted that the normal stiffness k_n in the contact volume based model is different from that in the conventional linear spring model. The conversion relationship between them is $k_n = k_n^{linear}/\pi r_c$, and r_c is the radius of curvature at contact region. $\boldsymbol{S}_n$ is the vector area of contact surface S_A, and the scalar value of vector area is the contact area, denoted as $S_n = |\boldsymbol{S}_n|$. The magnitude of the normal contact force is $F_n = w'(V_c)S_n = mk_n V_c^{m-1} S_n$. When $m = 2$, F_n becomes $F_n = 2k_n V_c S_n$. Then, the contact normal direction is defined by

$$\boldsymbol{n} = \boldsymbol{F_n}/F_n \tag{A.4}$$

The contact point $\boldsymbol{x}_c$ is given by

$$\boldsymbol{x}_C = \frac{\boldsymbol{n} \times \boldsymbol{G}_n}{S_n} + \lambda \boldsymbol{n} \tag{A.5}$$

where $\boldsymbol{G}_n$ is the surface integral over S_A, $\boldsymbol{G}_n = \int_{S_A} r \times d\boldsymbol{S}$. The contact geometric features mentioned above are schematically illustrated in Fig. A.33. Interested readers are referred to Feng [22] for more details about the evaluation schemes for $\boldsymbol{S}_n$ and $\boldsymbol{G}_n$.

References

[1] P.A. Cundall, O.D.L. Strack, A discrete numerical model for granular assemblies, Géotechnique 29 (1) (1979) 47–65.
[2] P. Wang, Z.-Y. Yin, Z.-Y. Wang, Micromechanical investigation of particle-size effect of granular materials in biaxial test with the role of particle breakage, J. Eng. Mech. 148 (1) (2022) 04021133.
[3] P. Wang, Z.-Y. Yin, P.-Y. Hicher, Y.-J. Cui, Micro-mechanical analysis of one-dimensional compression of clay with DEM, Int. J. Numer. Anal. Methods Geomech. (2023).
[4] P. Wang, C. Xu, Z.-Y. Yin, S.-x. Song, C. Xu, S. Dai, A DEM-based generic modeling framework for hydrate-bearing sediments, Comput. Geotech. 171 (2024) 106287.
[5] Q. Huang, W. Zhou, G. Ma, J. Mei, K. Xu, Investigation of the grain breakage behaviour of 2D granular materials with disordered pore distribution, Comput. Part. Mech. 8 (2021) 1033–1045.
[6] Y. Xia, F. Chen, J.L. Klinger, J.J. Kane, T. Bhattacharjee, R. Seifert, O.O. Ajayi, Q. Chen, Assessment of a tomography-informed polyhedral discrete element modelling approach for complex-shaped granular woody biomass in stress consolidation, Biosyst. Eng. 205 (2021) 187–211.
[7] D.-Y. Liu, M.-Z. Lyu, Capturing the random mechanical behavior of granular materials: A comprehensive stochastic discrete element method study, Géotechnique (2024) 1–35.
[8] K.-W. Lim, J.E. Andrade, Granular element method for three-dimensional discrete element calculations, Int. J. Numer. Anal. Methods Geomech. 38 (2) (2014) 167–188.
[9] R. Kawamoto, E. Andò, G. Viggiani, J.E. Andrade, Level set discrete element method for three-dimensional computations with triaxial case study, J. Mech. Phys. Solids 91 (2016) 1–13.
[10] A. Podlozhnyuk, S. Pirker, C. Kloss, Efficient implementation of superquadric particles in discrete element method within an open-source framework, Comput. Part. Mech. 4 (1) (2017) 101–118.
[11] R. Capozza, K.J. Hanley, A hierarchical, spherical harmonic-based approach to simulate abradable, irregularly shaped particles in DEM, Powder Technol. 378 (2021) 528–537.

[12] J.E. Andrade, K.-W. Lim, C.F. Avila, I. Vlahinić, Granular element method for computational particle mechanics, Comput. Methods Appl. Mech. Engrg. 241 (2012) 262–274.
[13] A. Wachs, L. Girolami, G. Vinay, G. Ferrer, Grains3D, a flexible DEM approach for particles of arbitrary convex shape—Part I: Numerical model and validations, Powder Technol. 224 (2012) 374–389.
[14] Z. Lai, Q. Chen, L. Huang, Fourier series-based discrete element method for computational mechanics of irregular-shaped particles, Comput. Methods Appl. Mech. Engrg. 362 (2020) 112873.
[15] D. Su, X. Wang, Fourier series-based discrete element method for two-dimensional concave irregular particles, Comput. Geotech. 132 (2021) 103991.
[16] J.R. Williams, R. O'Connor, Discrete element simulation and the contact problem, Arch. Comput. Methods Eng. 6 (4) (1999) 279–304.
[17] C. Shi, D.-j. Li, W.-y. Xu, R. Wang, Discrete element cluster modeling of complex mesoscopic particles for use with the particle flow code method, Granul. Matter 17 (2015) 377–387.
[18] J. Zheng, R.D. Hryciw, An image based clump library for DEM simulations, Granul. Matter 19 (2) (2017) 26.
[19] A.D. Rakotonirina, J.Y. Delenne, F. Radjai, A. Wachs, Grains3D, a flexible DEM approach for particles of arbitrary convex shape, part III: extension to non-convex particles modelled as glued convex particles, Comput. Part. Mech. 6 (1) (2019) 55–84.
[20] A.R. Khoei, S. Biabanaki, S. Parvaneh, 3D dynamic modeling of powder forming processes via a simple and efficient node-to-surface contact algorithm, Appl. Math. Model. 37 (1–2) (2013) 443–462.
[21] M. Paggi, P. Wriggers, Node-to-segment and node-to-surface interface finite elements for fracture mechanics, Comput. Methods Appl. Mech. Engrg. 300 (2016) 540–560.
[22] Y. Feng, An energy-conserving contact theory for discrete element modelling of arbitrarily shaped particles: Contact volume based model and computational issues, Comput. Methods Appl. Mech. Engrg. 373 (2021) 113493.
[23] B. Smeets, T. Odenthal, S. Vanmaercke, H. Ramon, Polygon-based contact description for modeling arbitrary polyhedra in the discrete element method, Comput. Methods Appl. Mech. Engrg. 290 (2015) 277–289.
[24] X. Wang, Z.-Y. Yin, H. Xiong, D. Su, Y.-T. Feng, A spherical-harmonic-based approach to discrete element modeling of 3D irregular particles, Internat. J. Numer. Methods Engrg. 122 (20) (2021) 5626–5655.
[25] Y.T. Feng, An energy-conserving contact theory for discrete element modelling of arbitrarily shaped particles: Basic framework and general contact model, Comput. Methods Appl. Mech. Engrg. 373 (2021) 113454.
[26] Y.T. Feng, An effective energy-conserving contact modelling strategy for spherical harmonic particles represented by surface triangular meshes with automatic simplification, Comput. Methods Appl. Mech. Engrg. 379 (2021) 113750.
[27] Y.T. Feng, Thirty years of developments in contact modelling of non-spherical particles in DEM: a selective review, Acta Mech. Sin. 39 (2023) 722343.
[28] Z. Lai, Q. Chen, L. Huang, Machine-learning-enabled discrete element method: Contact detection and resolution of irregular-shaped particles, Int. J. Numer. Anal. Methods Geomech. 46 (1) (2022) 113–140.
[29] S. Hwang, J. Pan, A.A. Sunny, L.-S. Fan, A machine learning-based particle-particle collision model for non-spherical particles with arbitrary shape, Chem. Eng. Sci. 251 (2022) 117439.
[30] Z. Lai, S. Zhao, J. Zhao, L. Huang, Signed distance field framework for unified DEM modeling of granular media with arbitrary particle shapes, Comput. Mech. 70 (4) (2022) 763–783.
[31] D.L. van der Haven, I.S. Fragkopoulos, J.A. Elliott, A physically consistent discrete element method for arbitrary shapes using volume-interacting level sets, Comput. Methods Appl. Mech. Engrg. 414 (2023) 116165.
[32] Z. Cheng, J. Wang, W. Xiong, A machine learning-based strategy for experimentally estimating force chains of granular materials using X-ray micro-tomography, Géotechnique (2023) 1–13.
[33] Á. Orosz, K. Bagi, Comparison of contact treatment methods for rigid polyhedral discrete element models, Int. J. Rock Mech. Min. Sci. 170 (2023) 105550.
[34] Z. Lai, S. Zhao, J. Zhao, L. Huang, Revisiting the GJK and shape erosion method for contact resolution in DEM, Powder Technol. 394 (2021) 363–371.
[35] R.R. Curtin, J.R. Cline, N.P. Slagle, W.B. March, P. Ram, N.A. Mehta, A.G. Gray, MLPACK: A scalable C++ machine learning library, J. Mach. Learn. Res. 14 (1) (2013) 801–805.
[36] Q. Bai, M. Shehata, A. Nada, Review study of using Euler angles and Euler parameters in multibody modeling of spatial holonomic and non-holonomic systems, Int. J. Dyn. Control 10 (5) (2022) 1707–1725.
[37] S.R. Nekoo, J.Á. Acosta, A. Ollero, Quaternion-based state-dependent differential riccati equation for quadrotor drones: Regulation control problem in aerobatic flight, Robotica 40 (9) (2022) 3120–3135.
[38] B. Farhang, A.A. Mirghasemi, A study of principle stress rotation on granular soils using DEM simulation of hollow cylinder test, Adv. Powder Technol. 28 (9) (2017) 2052–2064.
[39] T.S. Majmudar, R.P. Behringer, Contact force measurements and stress-induced anisotropy in granular materials, Nature 435 (7045) (2005) 1079–1082.
[40] S. Zhao, N. Zhang, X. Zhou, L. Zhang, Particle shape effects on fabric of granular random packing, Powder Technol. 310 (2017) 175–186.
[41] N. Guo, J. Zhao, The signature of shear-induced anisotropy in granular media, Comput. Geotech. 47 (2013) 1–15.
[42] A. Abbas, E. Masad, T. Papagiannakis, A. Shenoy, Modelling asphalt mastic stiffness using discrete element analysis and micromechanics-based models, Int. J. Pavement Eng. 6 (2) (2005) 137–146.
[43] S. Zhao, J. Zhao, A poly-superellipsoid-based approach on particle morphology for DEM modeling of granular media, Int. J. Numer. Anal. Methods Geomech. 43 (13) (2019) 2147–2169.
[44] Z. Lai, L. Huang, A polybézier-based particle model for the DEM modeling of granular media, Comput. Geotech. 134 (2021) 104052.
[45] T. Shire, K.J. Hanley, K. Stratford, DEM simulations of polydisperse media: efficient contact detection applied to investigate the quasi-static limit, Comput. Part. Mech. 8 (4) (2021) 653–663.
[46] R. Kawamoto, E. Andò, G. Viggiani, J.E. Andrade, All you need is shape: Predicting shear banding in sand with LS-DEM, J. Mech. Phys. Solids 111 (2018) 375–392.
[47] G.-Y. Liu, W.-J. Xu, A GPU-based DEM framework for simulation of polyhedral particulate system, Granul. Matter 25 (2) (2023) 27.
[48] X. Wang, W. Wu, H. Zhu, H. Zhang, J.-S. Lin, A. Bobet, A global direct search method for high-fidelity contact detection between arbitrarily shaped three-dimensional convex polyhedral blocks, Comput. Geotech. 150 (2022) 104891.
[49] F. John, Peters, A. Mark, Hopkins, Raju, Kala, E. Ronald, A poly-ellipsoid particle for non-spherical discrete element method, Eng. Comput. (2009).
[50] S. Wang, S. Ji, Poly-superquadric model for DEM simulations of asymmetrically shaped particles, Comput. Part. Mech. 9 (2) (2022) 299–313.
[51] J.Y. Nie, D.Q. Li, Z.J. Cao, B. Zhou, A.J. Zhang, Probabilistic characterization and simulation of realistic particle shape based on sphere harmonic representation and nataf transformation, Powder Technol. 360 (2020) 209–220.
[52] M. Kazhdan, T. Funkhouser, S. Rusinkiewicz, Rotation invariant spherical harmonic representation of 3D shape descriptors, in: Symposium on Geometry Processing, Vol. 6, 2003, pp. 156–164.
[53] B. Zhao, D. Wei, J. Wang, Particle shape quantification using rotation-invariant spherical harmonic analysis, Géotech. Lett. 7 (2) (2017) 190–196.
[54] G.-Y. Liu, W.-J. Xu, Q. Zhou, X.-L. Zhang, Contact overlap calculation algorithms and benchmarks based on blocky discrete-element method, Int. J. Geomech. 22 (12) (2022) 04022227.
[55] G.-Y. Liu, W.-J. Xu, Q. Zhou, DEM contact model for spherical and polyhedral particles based on energy conservation, Comput. Geotech. 153 (2023) 105072.