

## A geometric-information-enhanced crystal graph network for predicting properties of materials

Jiucheng Cheng<sup>1,4</sup>, Chunkai Zhang <sup>1,4</sup>✉ & Lifeng Dong <sup>2,3</sup>✉

Graph neural networks (GNNs) have been used previously for identifying new crystalline materials. However, geometric structure is not usually taken into consideration, or only partially. Here, we develop a geometric-information-enhanced crystal graph neural network (GeoCGNN) to predict the properties of crystalline materials. By considering the distance vector between each node and its neighbors, our model can learn full topological and spatial geometric structure information. Furthermore, we incorporate an effective method based on the mixed basis functions to encode the geometric information into our model, which outperforms other GNN methods in a variety of databases. For example, for predicting formation energy our model is 25.6%, 14.3% and 35.7% more accurate than CGCNN, MEGNet and iCGCNN models, respectively. For band gap, our model outperforms CGCNN by 27.6% and MEGNet by 12.4%.

<sup>1</sup>School of Computer Science and Technology, Harbin Institute of Technology (Shenzhen), Shenzhen, PR China. <sup>2</sup>College of Materials Science and Engineering, Qingdao University of Science and Technology, Qingdao, PR China. <sup>3</sup>Department of Physics, Hamline University, St. Paul, MN, USA. <sup>4</sup>These authors contributed equally: Jiucheng Cheng, Chunkai Zhang. ✉email: [ckzhang@hit.edu.cn](mailto:ckzhang@hit.edu.cn); [donglifeng@qust.edu.cn](mailto:donglifeng@qust.edu.cn)

Machine learning is a widely used tool to predict material properties, which is not only much faster (multiple orders of magnitude) than AB initio calculations, but also with close prediction accuracy based on large material databases<sup>1–4</sup>. Traditional machine learning methods<sup>5–13</sup> have been developed for various property prediction tasks, such as formation energy, band gap, thermal conductivity, and so on. Compared to traditional machine learning, many deep networks like equivariant 3D convolution networks<sup>14–17</sup> and graph neural network (GNN) have been proposed recently. As a novel machine learning model, GNN has advantages in representing material's topological structures by graph<sup>18,19</sup>. GNN has been designed to let networks automatically learn complex features from nearly raw structure data which could just contain atomic numbers and the position of each atom. Up to now, GNN has already achieved remarkably high accuracy for predicting material properties that traditional machine learning has never achieved before<sup>19–22</sup>.

The topological structure can be well embedded as the adjacency matrix  $A$  in the graph for material property predictions. For example, the Crystal Graph Neural Network (CGNN)<sup>21</sup> achieved a mean absolute error (MAE) of formation energy ( $E_f$ ) = 0.0346 eV/atom with a single model that only used topological information in the aggregation process. In material prediction domain, the geometrical structure information like spatial distance and direction is also important because the relative spatial position of atoms in the microstructure is closely related to the charge interaction between them and thus affects macroscopic properties of the material.<sup>23</sup> Many works have already introduced geometric structure information into their models. The Crystal Graph Convolutional Neural Network (CGCNN)<sup>19</sup> chose the distance between atoms to represent the edges in the crystal graph. The Materials Graph Network (MEGNet)<sup>24</sup> introduced manual features that included topological distance and spatial distance. The directional message passing neural network (DimeNet)<sup>25</sup> encoded the directional information into GNN models for molecular materials<sup>26,27</sup>. However, previous GNN based works have provided incomplete spatial geometrical information, such as just the distance<sup>19,22,24</sup>, or one additional angle<sup>25</sup>, which makes the models unable to learn the complete local geometrical relationship between atoms.

There are two different ways to embed the geometrical information into a GNN model. One way is to directly embed the geometrical information into node or edge features, and the other one is to encode the geometrical information first and then use the encoded information to act on the message passing process. A proper encoder could transform discrete and unnormalized geometrical information into a set of normalized data which is better to be learned. In addition, specific formation of the encoder could introduce the physical meanings related to the specific task. For example, PhysNet<sup>28</sup> introduced gaussian attention mask to filter the message between nodes according to the physical knowledge that bound state wave functions in two-body systems decay exponentially. DimeNet<sup>25</sup> introduced the 2D Fourier-Bessel basis as an attention mask under the hypothesis that each atom exists in an infinite deep spherical well. The utilization of attention masks is an excellent way to encode the geometrical structure among atoms. Since it is just beginning of the use in the field of crystal material prediction, there is still room for development. In this work, we propose a GNN model to accurately predict properties for any crystalline materials, which is invariant to global 3D rotations, translations, and node permutations. Our model achieves unprecedented prediction accuracy by introducing complete local spatial geometrical information. On the one hand, we construct a directed multi-graph and define the edge feature as the distance vector between atoms and their neighbors to incorporate the complete geometrical information in the crystal lattice coordinate system. Unlike distance and angles used in previous works<sup>19,25</sup>, distance

vector registers the complete information about local spatial geometrical structure. In addition, to validate discrete and unnormalized distance vector data better learned by the model, we propose an encoder as an attention mask to transform the discrete distance vector to two sets of orthogonal basis functions inspired by mixed basis<sup>29,30</sup> in the solution space of Schrödinger's equation. As the Bloch theorem utilizes plane waves to describe periodic structure of crystals, we introduce the plane waves in the mask function for the crystal material predictions. Experimental results prove that they can help our model better learn crystal structures. Finally, without compromising accuracy, to ensure the universality, the initial features of nodes and edges in our model only contain atomic numbers and the position of atoms.

Our main contributions can be summarized as below: (1) We propose a message passing neural network (MPNN)<sup>31</sup>-based GNN architecture with high prediction accuracy for the formation energy and band gap; and (2) we provide an effective way to encode the local geometrical information in the process of aggregation, that is, an attention mask composed by Gaussian radial basis and plane waves.

## Results

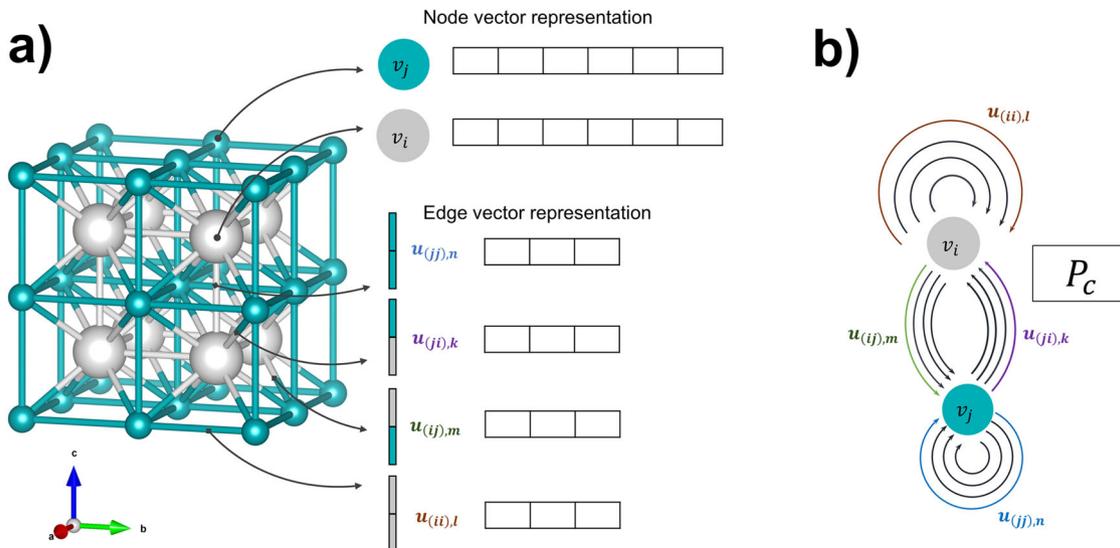
**Crystal graph definition and the introduction of geometric information.** In this section, we construct a crystal graph representation suitable for any stoichiometric crystalline material. Such a graph retains the information of the topological and geometric structure of crystals. It also records the periodicity and the key crystal information, such as the crystal lattice vector and the cell volume. Besides, representations in the graph meet translation invariance and node permutation invariance.

In a molecule graph representation, nodes usually represent atoms in the molecule and edges represent the chemical bonds between atoms<sup>18,31,32</sup>. But in crystals, there are no clearly defined chemical bonds among atoms. Hence, it is necessary to define the adjacency relationship among atoms first. Similar to CGCNN<sup>19</sup>, we define the neighbors of each atom to be the nearest  $k$  atoms in the cutoff radius  $c$ . In this work,  $k = 12$  and  $c = 8\text{Å}$ . Note that this is a multi-graph due to the periodicity,  $u_{(ij),k}$  means the  $k_{\text{th}}$  edge between node  $i$  and  $j$ . We implement the idea using the open python library pymatgen<sup>33</sup>. By defining the adjacency relationship in this way, our model embeds the topological structure and the periodicity of crystals. Then we define the node and edge representations on the graph. To construct a general crystal graph appropriate for all kinds of crystals, we should introduce as few manual features as possible. Node features  $v_i^0$  are defined as a one-hot encoding that depends on atomic number, as shown in Eq. (1). The matrix  $W$  is to resize the feature's dimension. To introduce the geometric information, we keep the distance vector between atoms as the edge features, that is,  $u_{(ij),k} = \mathbf{r}_{(ij),k}$  (We will omit  $k$  in the rest of the paper if there is no ambiguity). Note that the graph in this paper is also a directed graph because the edge  $u_{(ij)}$  needs to record the distance vector from node  $i$  to node  $j$  and it is obviously different from  $u_{(ji)}$ , or even the  $u_{(ji)}$  doesn't exist. In addition, we record the lattice vector  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  and cell volume  $\Omega$  as  $P_c$  which means crystal parameters and they will be used in section C. The full picture of the crystal graph is shown in Fig. 1.

$$v_i^0 = \text{Embedding}(z_i) = W(\text{Onehot}(z_i)); u_{ij} = \vec{r}_{ij} \quad (1)$$

## A MPNN-based GNN for crystal property prediction

As described in the “Methods” part, the forward propagation process of GNN can be explained as two steps: the node updating and the target outputting. Note that in this work the edges will not update with iteration. The node updating process in our



**Fig. 1 The crystal graph.** **a** NaCl crystal structure. **b** crystal graph with num of neighbors=8, each arrow represents a directed edge from one atom node to another. In this graph  $G = (P_c, \{v\}, u)$ . The global descriptor  $P_c = (\Omega, \mathbf{a}, \mathbf{b}, \mathbf{c})$ , node set  $v = (v_i, v_j)$ , edge set  $u = (u_{(ii),l}, u_{(ij),m}, u_{(ji),k}, u_{(jj),n}, m, n, k \in [1, 4])$ .

model can be written as Eq. (2).

$$v_i^t = f_{\text{update}}\left(v_i^{t-1}, f_{\text{agg}}(v_j^{t-1}, \mathbf{r}_{ij}, P_c) | j \in N_i\right) \quad (2)$$

Similar to CGCNN<sup>19</sup>, we utilize the architecture of Message Passing Neural Network<sup>31</sup> (MPNN) and Gated Convolution (Gated Conv) to implement Eq. (2). In MPNN, the message being passed from node  $i$  to node  $j$  is a function of  $v_j \oplus v_i$ , where  $\oplus$  denotes the concatenate operation. In this work, we add additional information into the message (Eq. 3): the difference quotient of node features  $\nabla v_{ij} = (v_j - v_i)/|\mathbf{r}_{ij}|$ , which means the change of node features with distance between nodes. Experiments have proven that it improves the performance of the model by 5%. As the distance vector  $\mathbf{r}_{ij}$  plays a role to represent the relationship between atoms and it is hard to be directly learned by the model due to its discreteness and non-normality, we embed  $\mathbf{r}_{ij}$  and  $P_c$  into a learnable attention mask<sup>34</sup>  $M$  to filter the message being passed from the neighbors, where  $M$  will be carefully discussed in the next section. The mathematical expression is shown in Eqs. (3–5), where  $W$  denotes the learnable weight matrix and  $\odot$  denotes the element-wise product operation.  $\sigma$  denotes any nonlinear function which is the Elu function in this work and  $g$  denotes a Sigmoid function to filter the message passing from node  $i$  to  $j$ . Note that there is a significant difference between the two functions  $g$  and  $M$ . Although both are used to filter messages, the former is just based on topological information, but the latter is based on geometric information. The concatenated node feature  $\tilde{v}_{ij}$  first passes into an aggregation function  $f_{\text{agg}}$  with an attention mask  $M$  to get the aggregated feature  $\omega_i^t$  for each node where  $i$  denotes the  $i$ th node and  $t$  denotes the  $t$ th layer. Then  $\omega_i^t$  passes into an update function  $f_{\text{update}}$  to get the updated node feature  $v_i^{t+1}$ . One node updating process finishes and the next layer of updating starts.

$$\tilde{v}_{ij} = v_j \oplus v_i \oplus \nabla v_{ij} \quad (3)$$

$$\omega_i^t = f_{\text{agg}} = v_i^t + \sum_{j \in N\{i\}} \sigma(W_{\sigma}^t \tilde{v}_{ij}) \odot g(W_g^t \tilde{v}_{ij}) \odot M(\mathbf{r}_{ij}, P_c) \quad (4)$$

$$v_i^{t+1} = f_{\text{update}} = \text{MLP}(\omega_i^t) \quad (5)$$

To enable the model to learn different scales of information, we use a basic strategy to set the gated pooling (Eq. 6) layer after each gated Conv layer to get a layer vector  $\gamma^t$  at each layer  $t$  where  $g'$  is

a Tanh function. Then get a graph vector by a simple summation  $\sum_t \gamma^t$ . Finally, input the graph vector  $\sum_t \gamma^t$  into a Multilayer Perceptron to get a real number  $P$  (Eq. 7).

$$\gamma^t = \frac{1}{N} \sum_i g'(W_g^t \omega_i^t) \odot W_{\text{omega}}^t \omega_i^t \quad (6)$$

$$P = \text{MLP}(\sum_t \gamma^t) \quad (7)$$

The training process of the model can be seen as an optimization problem. As in Eq. (8), the target is to optimize all the following weight matrices  $W$  and to minimize the loss function which is Mean Square Error (MSE) between model predictions and DFT calculations in this work.

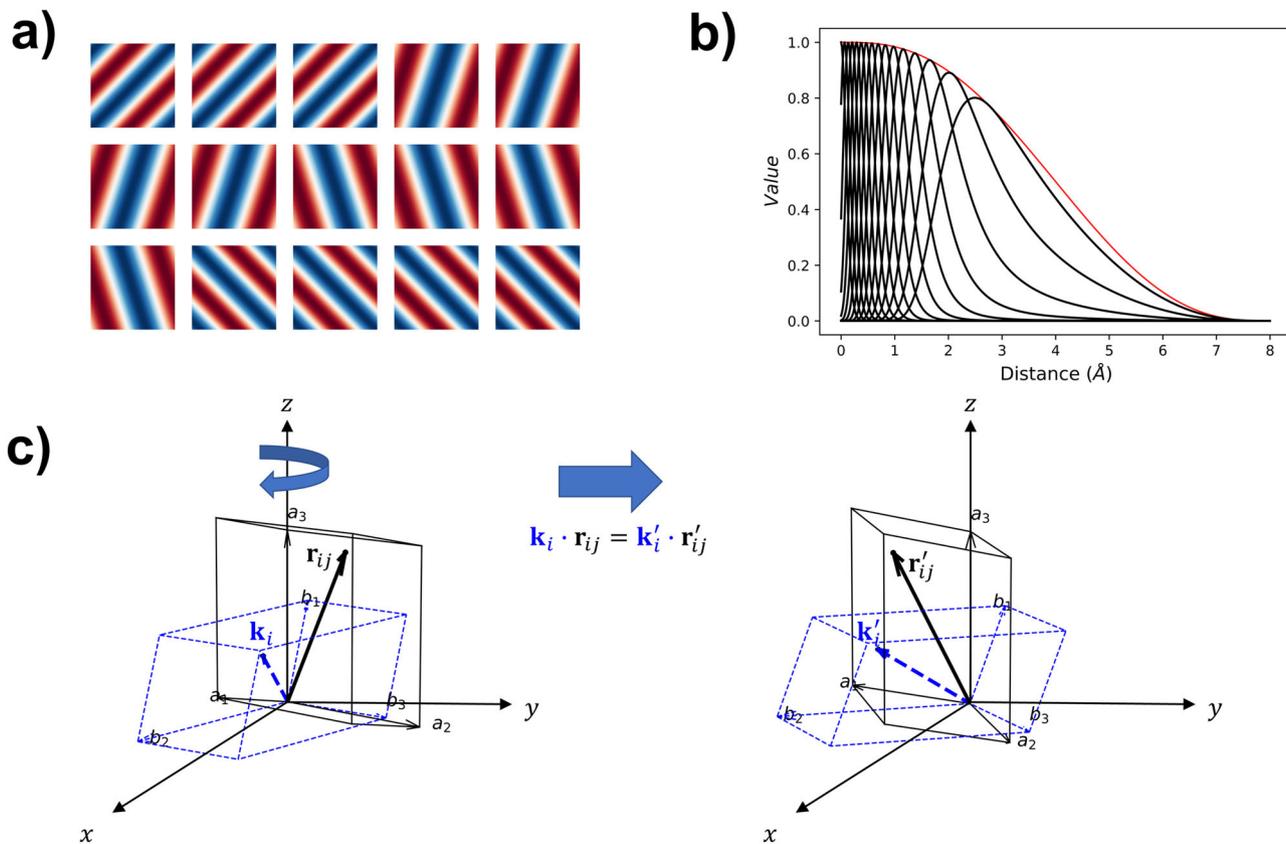
$$\min \text{MSE}(P(W), y) \quad (8)$$

where  $y$  is the DFT calculated data and  $P$  is the model predictions.  $W$  denotes all the learnable matrices in the model. This optimization problem can be solved by back propagation and gradient descent.

**Utilize an attention mask to encode the crystal geometric structure.** In 2017, DTNN<sup>35</sup> introduced quantum chemical insights into GNN and encoded the distance between atoms into the node updating function of GNN. Recently, other works like PhysNet<sup>28</sup> and DimeNet<sup>25</sup> refined this idea and proposed attention masks respectively based on Gaussian radial function and Fourier-Bessel basis function to encode spatial and chemical information. In the latest work DimeNet<sup>25</sup>, it proved that even utilizing the simplest wave function (the solution of Schrödinger equation under infinite sphere potential) as an attention mask can significantly improve the performance of the GNN model. In this part, we introduce an attention mask  $M$  with a more precise physical meaning. It effectively encodes the geometric structure information, which is given by

$$M = W_R \{a_{\text{RBF}}(|\mathbf{r}_{ij}|)\} + W_P \{a_{\text{PW}}(\mathbf{r}_{ij}, \Omega, \mathbf{a}, \mathbf{b}, \mathbf{c}) \oplus G\} \quad (9)$$

Since plane waves are eigenfunctions of the Schrödinger equation with constant potential, they are the natural basis in the nearly-free-electron approximation. On the other hand, local orbitals can be used as a basis to carry out a full self-consistent solution of independent particle equations. And analytic forms, especially gaussians, are extensively used in chemistry<sup>23,30</sup>. Based on these theories, the mixed basis<sup>29,30</sup> has been developed and



**Fig. 2** Several features of the attention mask function  $M$ . **a**  $\{a_{\text{PW}}\}$  versus  $\mathbf{r}_{ij}$  (Å). 15 plane waves  $\{a_{\text{PW}}\}$  in a plane of  $z = 0$  determined by different  $\mathbf{k}$ -points.  $q = 4$ . (crystal structure used here is  $\text{CaTiO}_3$ ). **b**  $\{a_{\text{RBF}}\}$  versus  $|\mathbf{r}_{ij}|$  (Å). The smooth cutoff function  $\phi$  (red line) with a set of radial bases  $\{a_{\text{RBF}}\}$  (black line,  $n \in [1, 16]$ , cutoff = 8Å). **c** A simple example to show the rotational invariance of  $\{a_{\text{PW}}\}$ .  $a_1, a_2, a_3$  define a crystal lattice and  $b_1, b_2, b_3$  define the corresponding reciprocal space. The distance vector  $\mathbf{r}_{ij}$  and the sampled  $\mathbf{k}$ -point  $\mathbf{k}_i$  change with the rotation of the whole crystal but their dot product keeps the same.

utilized to expand the traditional approaches, which provides a convenient way to describe some complex electronic systems like transition-metals<sup>29</sup>. Equation (9) takes the form of mixed basis: the Gaussian radial basis  $\{a_{\text{RBF}}\}$  and the plane wave  $\{a_{\text{PW}}\}$  with gate function  $G$ .  $W$  denotes learnable weight matrixes. Note that in Eq. (9), two tensor product operations are set here to make the dimension of two basis set match so that the element-wise plus operation can be implemented. We neither utilize any nonlinear function except for the gate function  $G$ , nor add any bias item after linear combination of the basis sets.

We choose the Gaussian basis set proposed by PhysNet<sup>28</sup> as  $\{a_{\text{RBF}}\}$  in Eq. (9), which is given by

$$\{a_{\text{RBF}}\} = \phi(|\mathbf{r}_{ij}|) \cdot \exp(-\beta_n(\exp(-|\mathbf{r}_{ij}|) - \mu_n)^2) \quad (10)$$

where  $|\mathbf{r}_{ij}|$  denotes the distance between atoms;  $\phi(|\mathbf{r}_{ij}|)$  denotes a smooth cutoff function<sup>36</sup> to ensure continuous behavior when an atom enters or leaves the cutoff sphere;  $\beta_n, \mu_n$  denote the constant parameters of  $n^{\text{th}}$  order.

$\{a_{\text{PW}}\}$  in Eq. (9) is the plane wave basis set where  $\Omega$  denotes the volume of the crystal cell.  $\mathbf{k}$  denotes the points in reciprocal space ( $k$  space). A  $k$ -point mesh with  $q \times q \times q$  Monkhorst Pack special points<sup>37</sup> is employed, which is a widely used sampling method in the first Brillouin zone. The mathematical definition of  $\{a_{\text{PW}}\}$  and the  $k$ -point sampling method are given by

$$\{a_{\text{PW}}\} = \text{real}\left(\frac{1}{\sqrt{\Omega}} e^{i\mathbf{k} \cdot \mathbf{r}_{ij}}\right) \quad (11)$$

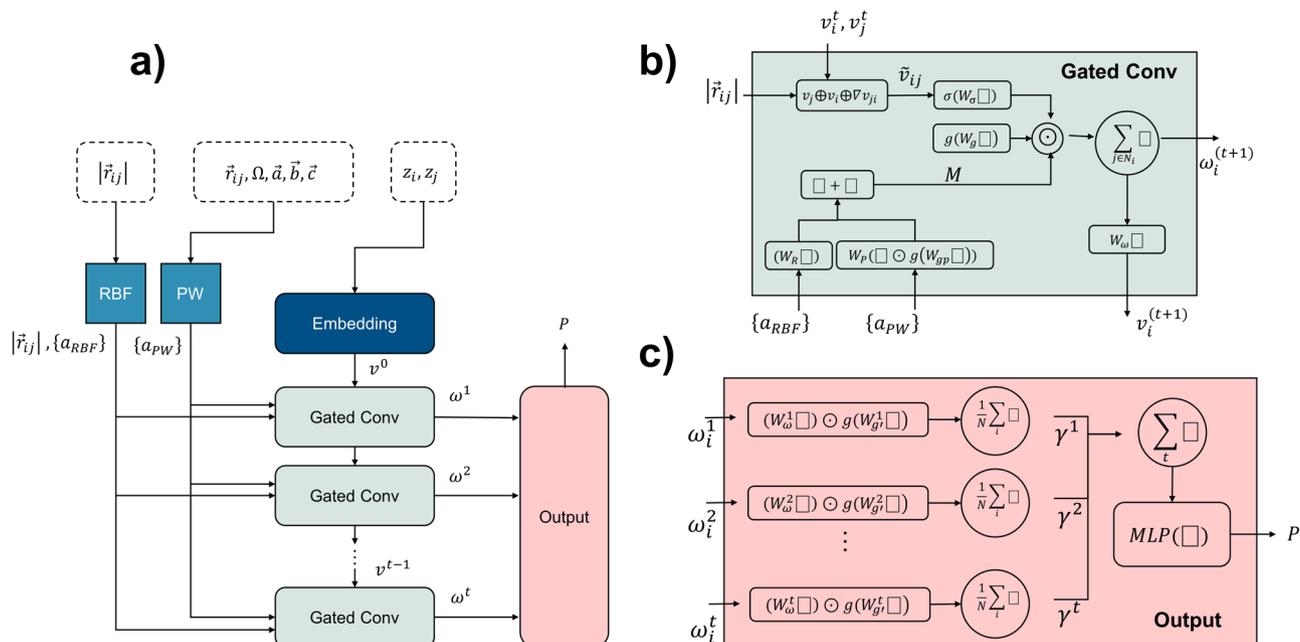
$$u = \frac{2r-q-1}{2q} (r = 1, 2, 3, \dots, q) \quad (12)$$

$$\mathbf{k}_{prs} = u_p \mathbf{b}_1 + u_r \mathbf{b}_2 + u_s \mathbf{b}_3; u_p, u_r, u_s \in u \quad (13)$$

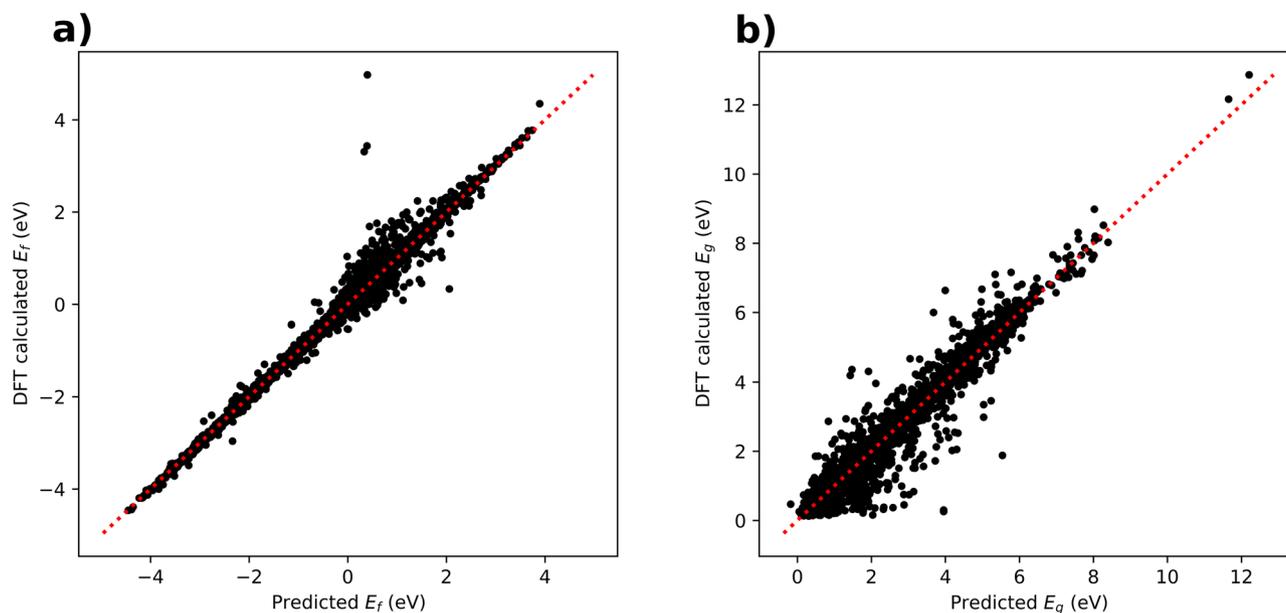
where  $q$  denotes an integer to determine the number of sampling grids and  $u_r$  denotes a set of real numbers for the weight at each basis vector.  $\mathbf{b}_i$  denotes the basis vector in reciprocal space corresponding to the crystal lattice vector  $\mathbf{a}_i$ . To be specific,  $\mathbf{b}_1 = 2\pi i \frac{(\mathbf{a}_2 \times \mathbf{a}_3)}{\Omega}$ ;  $\mathbf{b}_2 = 2\pi i \frac{(\mathbf{a}_3 \times \mathbf{a}_1)}{\Omega}$ ;  $\mathbf{b}_3 = 2\pi i \frac{(\mathbf{a}_1 \times \mathbf{a}_2)}{\Omega}$ . Note that in some previous works<sup>37,38</sup>, the special points in the  $k$ -space were chosen according to the specific characteristics of the given crystal system.

In this work, instead of manually selecting special  $\mathbf{k}$ -points for various crystal systems, a learnable gate  $G$  is utilized to filter the  $\{a_{\text{PW}}\}$  automatically.  $G$  has the same dimension  $q^3$  with  $\{a_{\text{PW}}\}$  and the value of each dimension ranges from 0 to 1. We tried two different formats of  $G$ ,  $g(W\{\text{SG}\})$  and  $g(W\{a_{\text{PW}}\})$ , where  $g$  is the Sigmoid function,  $\{\text{SG}\}$  denotes the one-hot vector of current space group and  $W$  is a learnable matrix. Experiments have proved that the latter is better, which means that the self-adapting gate is better than the space-group based gate. The reason is probably that the value of plane wave  $\{a_{\text{PW}}\}$  at each  $\mathbf{k}$ -point already reflects the space information of a given crystal structure. All the results in this paper are based on the format of  $G = g(W\{a_{\text{PW}}\})$ .

Because of the rotational invariance of mask function  $M$ , our model is invariant w.r.t. rotations of the input crystal. The rotational invariance of the  $\{a_{\text{RBF}}\}$  in  $M$  is obvious. Figure 2(c) presents a simple example to show why  $\{a_{\text{PW}}\}$  is also rotation invariant. The sampled  $\mathbf{k}$ -points  $\mathbf{k}_i$  and the distance vector between atoms  $\mathbf{r}_{ij}$  change with the rotation operation, but their dot product will keep the same because the reciprocal space also rotates with the coordinate space.



**Fig. 3 Model framework.** **a** Model architecture,  $z$  denotes the atomic number of each atom; Embedding block generates one-hot encoding and implements a linear transformation to adjust the embedding's dimension;  $\{a_{RBF}\}$ ,  $\{a_{PW}\}$  denote respectively the Gaussian radial basis (Equation 10) and the plane wave basis (Eq. 11),  $t$  is the layer index. **b** The detailed Gated Graph Convolution block includes the aggregation and updating functions (Eqs. 4 and 5). **c** The detailed Output block (Eqs. 6 and 7). MLP means the multi-layers perceptron where dimension = 1 in the last layer.



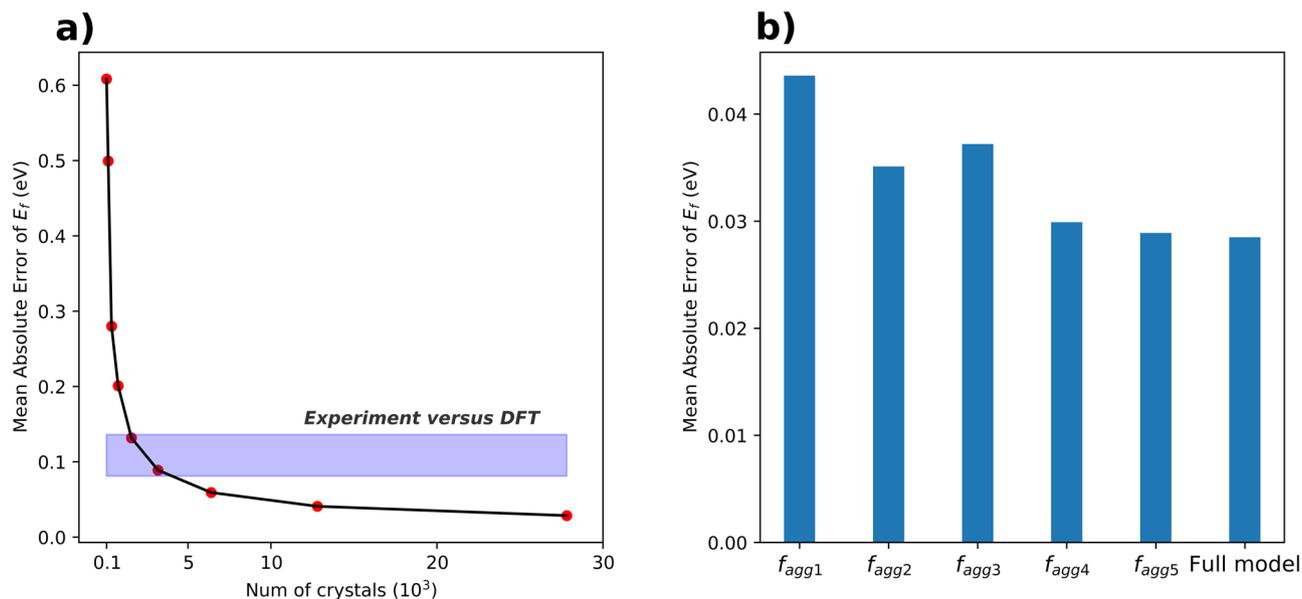
**Fig. 4 Comparison between DFT calculations and predictions.** **a** 56K predicted  $E_f$  versus the corresponding DFT calculated values. The model used is trained by 449K DFT data in the OQMD database. **b** 3K predicted  $E_g$  versus the corresponding DFT calculated values. The model used is trained by 24K DFT data in the OQMD database.

As shown in Fig. 3, our model contains three main blocks: embedding, gated convolution, and output blocks. The embedding block generates an initial one-hot vector of each atom according to the atomic number and passes the initialized vector to the first gated convolution block. The gated convolution block updates the node embedding via gated graph convolution among the neighbors of each node and passes each layer's node embedding to the output block. The output block aggregates on each crystal graph's node features to generate each layer vector, and then passes them into the multilayer perceptron to get the final target.

**Model evaluation.** Different datasets cause different performances of machine learning. However, many of previous works ignore this point when evaluating their models. To evaluate our model sufficiently, we compared our model against three other models with the same datasets (training set, validation set, and testing set) which comprised by different amounts of data in Material Project (MP)<sup>2</sup> and Open Quantum Material Database (OQMD)<sup>1</sup>. MP and OQMD contain over 130K and 560K computational crystal structures, respectively. The targets we choose are formation energy per atom ( $E_f$ ), band gap ( $E_g$ ) and two elastic metrics: bulk modulus  $K_{VRH}$

**Table 1 Comparison with other approaches recently reported in the literature.**

Database	Target(train/validate/test)	Units	CGCNN	MEGNet	iCGCNN	This work
Material Project	$E_f$ (27,824/9274/9274)	eV	0.039	/	/	$0.029 \pm 0.001$
	$E_g$ (16,266/5422/5,422)	eV	0.388	/	/	$0.281 \pm 0.003$
	$E_f$ (60,000/4619/4619)	eV	/	0.028	/	$0.024 \pm 0.000$
	$E_f$ (106,736/13,342/13,342)	eV	/	0.026	/	$0.023 \pm 0.001$
	$E_g$ (36,720/4590/4590)	eV	/	0.330	/	$0.289 \pm 0.004$
	$K_{VRH}$ (4664/583/583)	$\log_{10}(\text{GPa})$		0.050		$0.057 \pm 0.001$
OQMD	$G_{VRH}$ (4664/583/583)	$\log_{10}(\text{GPa})$		0.079		$0.077 \pm 0.000$
	$E_f$ (180,000/20,000/230,000)	eV	/	/	0.0305	$0.0196 \pm 0.000$
	$E_f$ (449,506/56,188/56,188)	eV	/	/	/	$0.015 \pm 0.000$
	$E_g$ (24,296/3,036/3,036)	eV	/	/	/	$0.254 \pm 0.002$

**Fig. 5 Experiments on 9274 DFT data in Material Project. a** Different volume of training data versus the MAE of  $E_f$ . **b** The MAE of  $E_f$  on 6 different aggregation functions.

and shear modulus  $G_{VRH}$ . Note that there are a few data mismatches with CGCNN because of database updating of MP. In addition, we trained our model on the largest datasets we can find (over 560K of OQMD) to get the best performance we can get now.

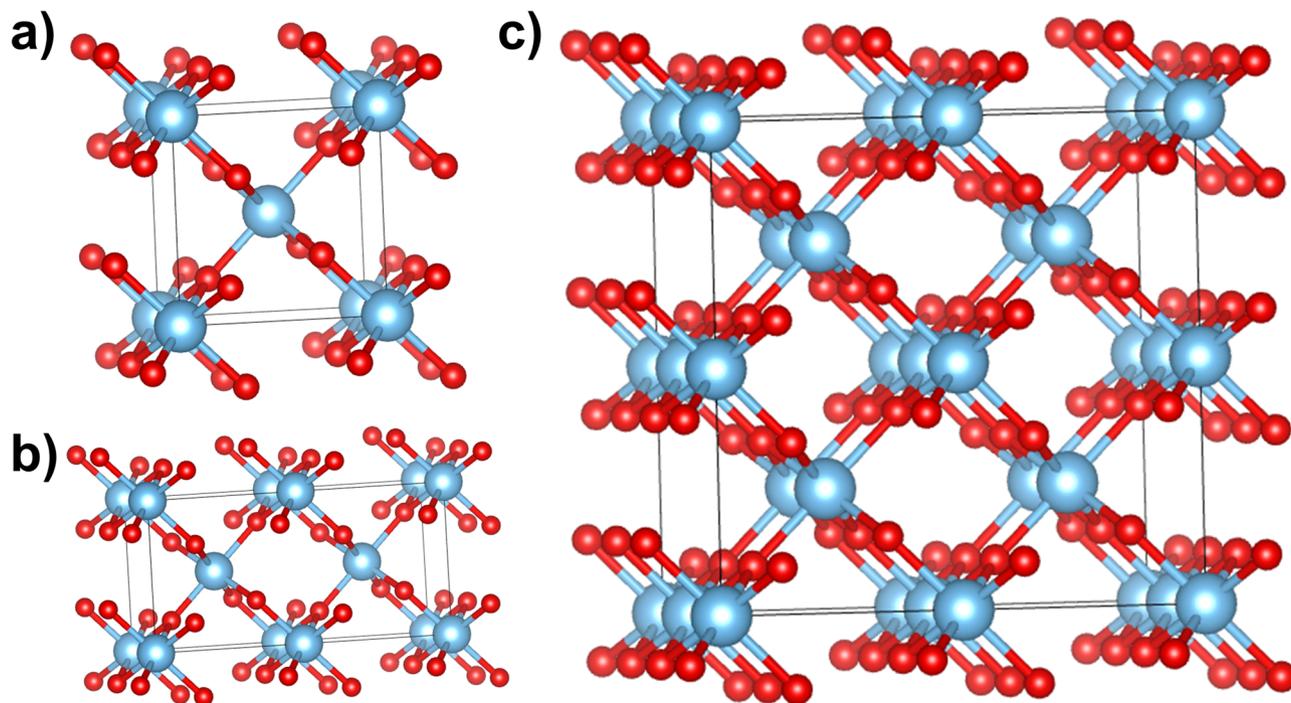
Figure 4 shows that our model performs well over the entire range of  $E_f$  and  $E_g$ . The MAE of  $E_f$  and  $E_g$  of DFT calculations against experimental measurements are 0.81–0.136 eV/atom and 0.6 eV, respectively<sup>1,39</sup>. Also, Table 1 summarizes the MAE of four models (CGCNN<sup>19</sup>, MEGNet<sup>24</sup>, iCGCNN<sup>22</sup> and this work) on different datasets. In comparison, our model can achieve much better performance. For example, for  $E_f$ , our model outperforms CGCNN, MEGNet, and iCGCNN by 25.6%, 14.3% and 35.7% respectively. For  $E_g$ , our model outperforms CGCNN and MEGNet by 27.6% and 12.4% respectively. For elastic metrics  $K_{VRH}$  and  $G_{VRH}$ , our model reaches the same-level error of the MEGNet model. The good performance on diverse datasets shows that our model has both high precision and good generalization.

We note that the volume of dataset is closely related to the performance. To better understand this relationship, we implemented several experiments on different volumes of training sets. The testing sets were the same 9274 DFT data from Material Project. Figure 5(a) shows that the precision improves significantly with the increase of the number of training data. Our model reaches a comparable precision with that of experiment vs. DFT calculation when the number of training data reaches  $\sim 10^3$ .

To validate the contribution of our model, we implemented the ablation experiments on random 27,824 training data in the MP database. To test the validity of the attention mask  $M$ , we run the model with a set of different aggregate functions (Eq. 4). First, we validate the model without  $M$  and the aggregate function changes to  $f_{agg1}$  in Eq. (14). Compared with our full model, the MAE is increased by 53%. The results demonstrate that the attention mask  $M$  is the main contributor to the high scores of our model. To further validate the two parts of attention mask  $M$ , we implemented additional experiments on the same database as above. We compared 2 different aggregate functions  $f_{agg2}$  in Eq. (15) and  $f_{agg3}$  in Eq. (16) to prove that both the Gaussian radial basis and plane waves are valuable. Figure 5(b) shows that two single basis function (Gaussian radial basis or Plane Waves) offers limited improvements to the model, but when combined they can significantly improve the model's performance. We also evaluate alternatives to implement  $\tilde{v}_{ij}$  and  $G$ . Based on our full model, we respectively change  $\tilde{v}_{ij}$  to  $\tilde{v}'_{ij}$  where  $\tilde{v}'_{ij} = v_j \oplus v_i$ , and  $G$  to  $G'$  where  $G' = g(W\{SG\})$  as in Eq. (17) and Eq. (18).  $\{SG\}$  denotes the one-hot vector of space group.

$$f_{agg1} = \sum_{i \in N\{j\}} \sigma(W_{\sigma}^t \tilde{v}_{ij}) \odot g(W_g^t \tilde{v}_{ij}) \quad (14)$$

$$f_{agg2} = \sum_{i \in N\{j\}} \sigma(W_{\sigma}^t \tilde{v}_{ij}) \odot g(W_g^t \tilde{v}_{ij}) \odot W_R \{a_{RBF}\} \quad (15)$$



**Fig. 6 Supercell transformation invariance.** **a** Crystal cell of TiO<sub>2</sub>(mp-2657<sub>111</sub>). **b** Crystal supercell of TiO<sub>2</sub>(mp-2657<sub>112</sub>). **c** Crystal supercell of TiO<sub>2</sub>(mp-2657<sub>222</sub>).

$$f_{\text{agg}3} = \sum_{i \in N\{j\}} \sigma(W_{\sigma}^t \tilde{v}_{ij}) \odot g(W_{\text{g}}^t \tilde{v}_{ij}) \odot W_{\text{R}}\{a_{\text{PW}} \otimes G\} \quad (16)$$

$$f_{\text{agg}4} = \sum_{i \in N\{j\}} \sigma(W_{\sigma}^t \tilde{v}'_{ij}) \odot g(W_{\text{g}}^t \tilde{v}'_{ij}) \odot M, \tilde{v}'_{ij} = v_j \oplus v_i \quad (17)$$

$$\begin{aligned} f_{\text{agg}5} &= \sum_{i \in N\{j\}} \sigma(W_{\sigma}^t \tilde{v}_{ij}) \odot g(W_{\text{g}}^t \tilde{v}_{ij}) \odot M(G'), G' \\ &= g(W\{SG\}) \end{aligned} \quad (18)$$

**Self-learned supercell invariance.** In theory, static properties like  $E_{\text{g}}$  and  $E_{\text{f}}$  would not change with supercell transformation (present multi primitive cells in a bigger supercell), so a machine learning model should satisfy this characteristic. The output of our model is mainly based on the correlative relationship of atoms and their local structures of input crystals. As given in Eq. (11), the plane wave mask function imports the information of the crystal structure which alters with supercell transformation, which means the absence of inherent a priori supercell invariance in our model.

But this does not mean that our model cannot capture the supercell invariance by learning multi scales of crystal structure data. There are many supercell data in MP and OQMD database and the good test results on these databases already demonstrate the ability of our model to learn the supercell invariance. From the point of view of Machine Learning, our model can learn similar representations of the Mask function (showed in Eq. 9) of similar crystal structures. The definition of “similarity” is automatically determined by the model according to the prediction target. For the properties like  $E_{\text{g}}$  or  $E_{\text{f}}$ , the model can learn this kind of “similarity” as well. In other words, the supercell transformation invariance is one kind of similarities that our model should learn from massive different scales of training data. In detail,  $W_{\text{p}}$  and  $G$  of Eq. (9) enable our model to learn the similarities among crystals. For different supercells of one

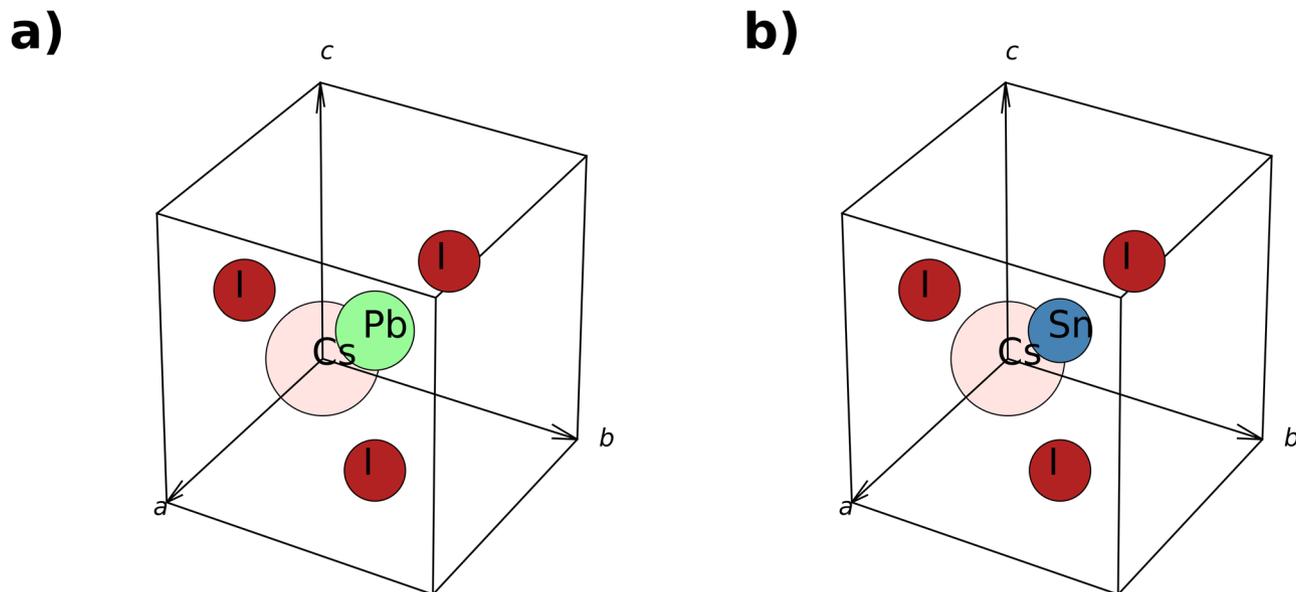
**Table 2 Predictions ( $E_{\text{f}}$ ) of eight different supercells for three crystal structures.**

Material id	Prediction (eV)	Target (eV)	Error (%)
mp-5827 <sub>111</sub>	-3.4884093	-3.5091639	0.59
mp-5827 <sub>112</sub>	-3.501649	-3.5091639	0.21
mp-5827 <sub>121</sub>	-3.490446	-3.5091639	0.53
mp-5827 <sub>122</sub>	-3.5019953	-3.5091639	0.20
mp-5827 <sub>211</sub>	-3.4550805	-3.5091639	1.54
mp-5827 <sub>212</sub>	-3.4929543	-3.5091639	0.46
mp-5827 <sub>221</sub>	-3.492474	-3.5091639	0.48
mp-5827 <sub>222</sub>	-3.496754	-3.5091639	0.35
mp-2657 <sub>111</sub>	-3.4740384	-3.481761	0.22
mp-2657 <sub>112</sub>	-3.4811313	-3.481761	0.02
mp-2657 <sub>121</sub>	-3.4829652	-3.481761	0.03
mp-2657 <sub>122</sub>	-3.4868681	-3.481761	0.15
mp-2657 <sub>211</sub>	-3.455293	-3.481761	0.76
mp-2657 <sub>212</sub>	-3.470849	-3.481761	0.31
mp-2657 <sub>221</sub>	-3.473609	-3.481761	0.23
mp-2657 <sub>222</sub>	-3.4764066	-3.481761	0.15
mp-5229 <sub>111</sub>	-3.5618715	-3.5686102	0.19
mp-5229 <sub>112</sub>	-3.5654297	-3.5686102	0.09
mp-5229 <sub>121</sub>	-3.5705163	-3.5686102	0.05
mp-5229 <sub>122</sub>	-3.5727057	-3.5686102	0.11
mp-5229 <sub>211</sub>	-3.553746	-3.5686102	0.42
mp-5229 <sub>212</sub>	-3.5589035	-3.5686102	0.27
mp-5229 <sub>221</sub>	-3.5634522	-3.5686102	0.14
mp-5229 <sub>222</sub>	-3.5494866	-3.5686102	0.54

mp-5827: perovskite CaTiO<sub>3</sub>; mp-2657: rutile TiO<sub>2</sub>; mp-5229: perovskite SrTiO<sub>3</sub>.

primitive cell,  $a_{\text{PW}}$  is different on each dimension but the value of  $M$  function should be very close.

To demonstrate the supercell transformation invariance clearer, we selected 3 crystals: perovskite CaTiO<sub>3</sub>, rutile TiO<sub>2</sub>, and perovskite SrTiO<sub>3</sub>, each of them is with 8 different supercell configurations. Figure 6 illustrate three supercell configurations of TiO<sub>2</sub>. Note that  $mp - XXX_{ijk}$  means the lengths of the first,



**Fig. 7** Crystal structures of halide perovskites in cubic system. **a** crystal structure of CsPbI<sub>3</sub>. **b** crystal structure of CsSnI<sub>3</sub>.

second, and third basis are magnified by  $i$ ,  $j$ , and  $k$  times, respectively; and  $mp-XXX$  is the material ID in the Material Project database. Take formation energy as an example: Table 2 shows the DFT calculations (Target value) and the predictions of our model for three crystal structures with eight different supercell configurations. The results demonstrate the supercell transformation invariance where the errors of prediction of each supercell are comparable. The model used here is trained on 69K DFT calculation data (formation energy) which not includes the three selected crystals.

## Discussion

Deep neural networks including GNN nearly always outperform traditional machine learning, but the interpretability is the main drawback of deep neural networks due to their multiple nonlinear transformation layers and a large number of neurons. Like all other prediction models, our model is evaluated by the prediction error. Although the prediction accuracy is high, the features learned by the model have nothing to do with physical properties but merely fit the results based on a limited dataset. For some crystalline materials, even just changing one atom can make big differences in their physical properties. It is understandable for experts, but it is difficult for a machine learning model to recognize it because the input is just a set of one-hot vectors with similar distances between any two atoms. If our model can still accurately predict properties even if the input data is only fine-tuned, then we can be more confident that the model does learn features with physical meanings. In this section, we present a specific case of band gap predictions of perovskite materials to show the plausibility of the model.

The data we use in this section is a group of halide perovskites with the formular AMX<sub>3</sub> (A = CH<sub>3</sub>NH<sub>3</sub>, Cs; M = Pb, Sn, X = I, Br, Cl) in cubic or orthorhombic system. As the demonstration in Fig. 7, the structures of AMX<sub>3</sub> with the same crystal system are similar, especially relative positions among atoms, which could be a big challenge for the machine learning model to make a satisfying prediction. In detail, we selected five different perovskites, including 3 cubic crystals and 2 orthorhombic crystals, to evaluate the sensitivity of our model to slight differences in atomic types and crystal structures. The prediction target is band gap, an important parameter for solar cell materials.

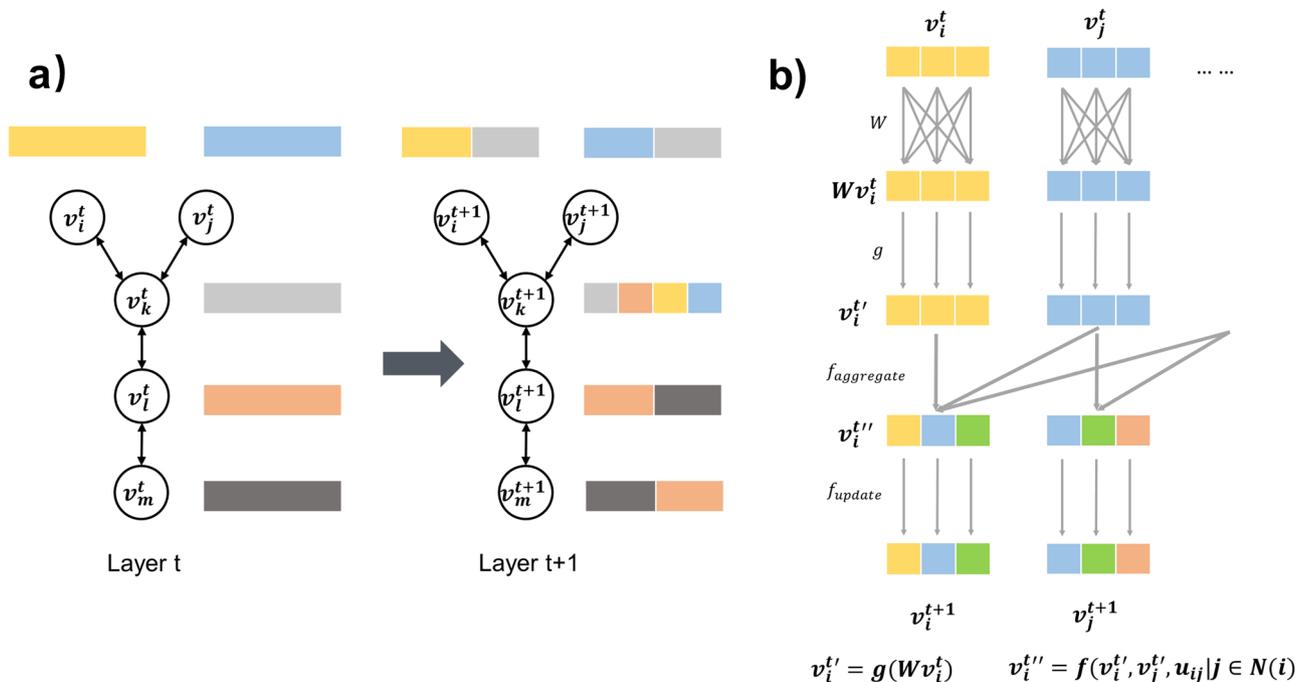
**Table 3** Band gap predictions vs. experimental results vs. DFT calculations (eV).  $\alpha$ ,  $\gamma$  denotes cubic system and orthorhombic system, respectively.

Formular	Prediction	Expt.	DFT <sup>40</sup>
$\alpha$ - CH <sub>3</sub> NH <sub>3</sub> PbCl <sub>3</sub>	3.133	2.88–3.13 <sup>41–43</sup>	3.090
$\alpha$ - CH <sub>3</sub> NH <sub>3</sub> PbBr <sub>3</sub>	2.007	2.33–2.35 <sup>41,43</sup>	2.400
$\alpha$ - CH <sub>3</sub> NH <sub>3</sub> PbI <sub>3</sub>	1.662	1.60–1.61 <sup>44–48</sup>	1.810
$\gamma$ - CsPbI <sub>3</sub>	2.180	1.74	2.000
$\gamma$ - CsSnI <sub>3</sub>	1.381	1.30	1.340

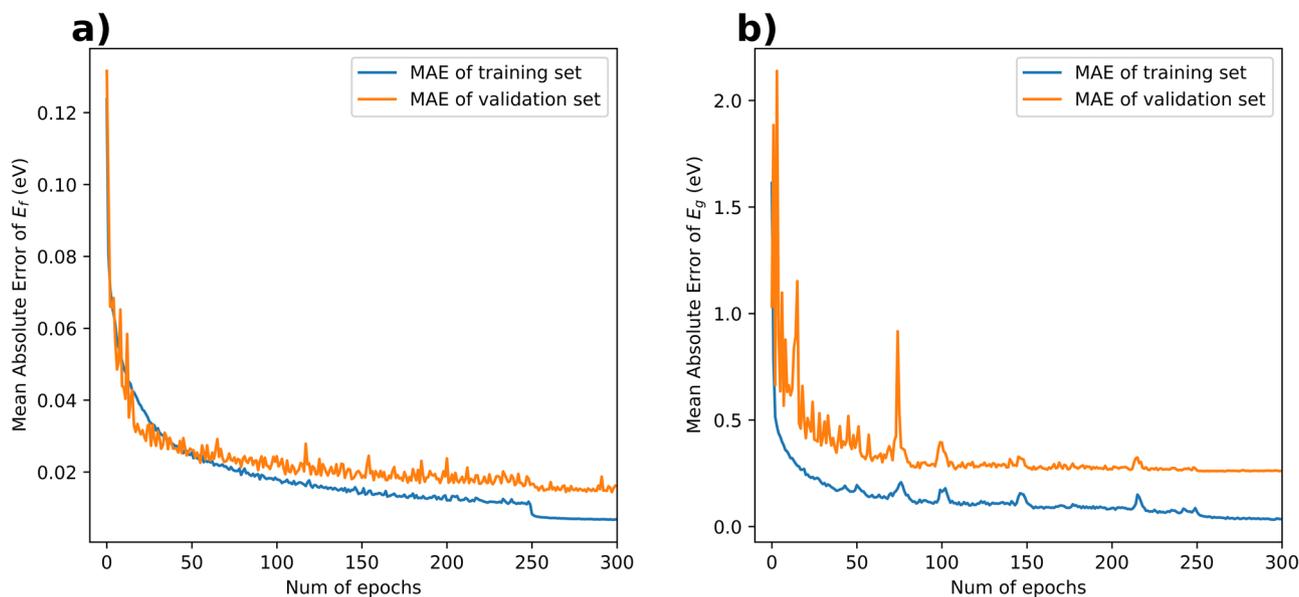
The model used in this section is trained on 69,000 DFT calculated band gaps  $E_g$  in the open database Material Project. As shown in Table 3, the errors between model predictions and experimental results are comparable with those between DFT and experimental results. In details, we can find that the band gap predictions of CH<sub>3</sub>NH<sub>3</sub>PbCl<sub>3</sub>, CH<sub>3</sub>NH<sub>3</sub>PbBr<sub>3</sub> and CH<sub>3</sub>NH<sub>3</sub>PbI<sub>3</sub> descend in turn due to different intrinsic properties of Cl, Br and I. Our predictions are consistent with both DFT calculations and experimental results. As from CsPbI<sub>3</sub> to CsSnI<sub>3</sub>, the band gap prediction sharply decreases from 2.180 eV to 1.381 eV, which also matches with the DFT calculations and experimental results. Note that the DFT calculations in Table 3 are conducted using the latest approach of approximate quasiparticle DFT-1/2 method<sup>40</sup>. We believe that our model can get better performance along with the iterating of more available open databases.

## Conclusion

In summary, GeoCGNN presents a machine learning approach for crystal property prediction. With an attention mask containing Gaussian basis and plane wave basis, we demonstrate that encoding the spatial geometrical information in a proper way can significantly improve the prediction accuracy. The predicted targets of our model can be very close to the DFT calculated results with a training set of  $\sim 10^4$ . It indicates that our model can be a potential substitute for the DFT calculation, especially in a high-throughput screening scenario. Finally, introducing distance vector between atoms can help our model to learn the complete geometrical information with the rotational invariance. Hence our model can be used and transformed in the databases with unified defined crystal coordinate systems.



**Fig. 8 A demonstration of the node aggregation and updating in GNN. a** The full picture of one step of forward propagation in GNN. **b** The detailed process in two nodes.  $v_i^t$  means the  $i$ th node in the  $t$ th layer,  $u_{ij}$  means the edge between node  $i$  and  $j$  in the  $t$ th layer (in this work,  $u_{ij}$  doesn't vary with  $t$ ),  $W$  is a learnable parameter matrix,  $f$  and  $g$  are nonlinear functions.



**Fig. 9 The plot of learning history on full OQMD database. a** Red line shows the MAE of training set in the learning process of 449,506  $E_f$  training data. Blue line shows the MAE of validation set in the training process of 56,188  $E_f$  validation data. **b** The learning history of 24,296  $E_g$  training data and 3,036  $E_g$  validation data.

**Methods**

**Graph Neural Networks.** This section demonstrates how GNN learns structural information during each step of forward propagation. Figure 8 uses different colors to represent different nodes' information, and we can note that each node gathers the information of its neighbors after one step of node updating (from layer  $t$  to later  $t+1$ ). Equation (19) and the right part of Fig. 8 give a simple demo of details in each step. Equation (20) shows the simplest output function where  $P$  is the target. The output function usually contains several pooling layers which compress all nodes and edges into a single vector and a final full connected layer to transform the vector to a real number  $P$ . The graph  $G$  is defined by Eq. (21), where  $v$ ,  $u$  and  $A$  denote the node, the edge, and the adjacency matrix, respectively. It is clear that the nodes with different positions in this topological graph can gather different

information from each other. Thus, each node in the  $t$ th layer encodes the  $t$ -order topological information around itself. The nodes can also encode the geometric information in a similar way just by adding the geometric information into the edge representation  $u$ .

$$v_i^t = f_{\text{update}}(v_i^{t-1}, f_{\text{aggregate}}(v_j^{t-1}, u_{ij}^{t-1} | j \in N_i)) \tag{19}$$

$$P = f_{\text{output}}(\{v_i^t\}, \{u_{ij}^t\} | i, j \in G) \tag{20}$$

$$G = (v, u, A) \tag{21}$$

**Hyperparameters.** In this part, we will discuss several main hyperparameters in our model. The results in this part are based on the  $E_f$  data in MP dataset. Data splitting: train/val/test = 60K/4.6K/4.6K. For all of the models presented in this work, 300 training epochs and the Adam optimizer<sup>49</sup> were used. Figure 9 shows that 300 epochs are enough for the convergence of our model.

**Learning rate.** Learning rate is one of the most important hyperparameters. We tried  $1e^{-4}$ ,  $5e^{-4}$ ,  $1e^{-3}$ ,  $5e^{-3}$ , and we found the model is sensitive to the learning rate. The performance of the model fluctuates within 23%. The best performance is got at  $1e^{-3}$ . In addition, we implemented a basic approach to reduce the learning rate to  $1e^{-4}$  during the later 50 epochs for tighter convergence.

**Batch size.** Experiments show that our model is not sensitive to batch size. We tried 128, 256, 300, 396, 512, and the performance fluctuates within 5%. Finally, we got the best point at batch size 300.

**The number of convolution layers  $N_{\text{layer}}$ .** Because of the issue of *oversmoothing*, we just tried 2–7  $N_{\text{layer}}$ . We can find that the model's performance is stable if  $N_{\text{layer}} \geq 3$ . And the best performance was obtained when  $N_{\text{layer}} = 5$ .

**Dimension of node  $\text{Dim}_{\text{node}}$  and graph  $\text{Dim}_{\text{graph}}$ .** We tried 64, 128, 192, and 256 for  $\text{Dim}_{\text{node}}$  and  $\text{Dim}_{\text{graph}}$ , respectively, and found that  $\text{Dim}_{\text{node}} = 192$  and  $\text{Dim}_{\text{graph}} = 192$  are the best. The performance has no obvious change when both dimensions are higher than 128.

**The number of sampled  $k$ -points  $N_k$  and the number of Gaussian radial basis  $N_{\text{Gaussian}}$ .** We tried 27, 64, 125 for  $N_k$  and 32, 64, 128 for  $N_{\text{Gaussian}}$ , respectively and found that  $N_k = 64$  and  $N_{\text{Gaussian}} = 64$  are the best.

## Data availability

All the ab initio data presented in this paper can be found in two open material databases: Material Project<sup>2</sup> and the Open Quantum Material Database<sup>1</sup>. For the convenience of reproducing this work, we provide the list of material id we used in: [https://drive.google.com/drive/folders/1kdNpbjcsyJ-Hfn0DjvpgTODfEsdMCn\\_?usp=sharing](https://drive.google.com/drive/folders/1kdNpbjcsyJ-Hfn0DjvpgTODfEsdMCn_?usp=sharing).

## Code availability

Code of this work is available at: <https://github.com/Tinystormjojo/geo-CGNN>.

Received: 30 November 2020; Accepted: 12 July 2021;

Published online: 06 September 2021

## References

- Kirklin, S. et al. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
- Jain, A. et al. Commentary: The Materials Project: a materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
- Curtarolo, S. et al. AFLOWLIB.ORG: a distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
- Belsky, A., Hellenbrandt, M., Karen, V. L. & Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. B* **58**, 364–369 (2002).
- Lu, S. et al. Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nat. Commun.* **9**, 3405 (2018).
- Isayev, O. et al. Universal fragment descriptors for predicting properties of inorganic crystals. *Nat. Commun.* **8**, 15679 (2017).
- Im, J. et al. Identifying Pb-free perovskites for solar cells by machine learning. *npj Comput. Mater.* **5**, 37 (2019).
- Ward, L. et al. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B* **96**, 024104 (2017).
- de Jong, M. et al. A statistical learning framework for materials science: application to elastic moduli of  $k$ -nary inorganic polycrystalline compounds. *Sci. Rep.* **6**, 34256 (2016).
- Seko, A. et al. Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and Bayesian optimization. *Phys. Rev. Lett.* **115**, 205901 (2015).
- Pilania, G., Gubernatis, J. E. & Lookman, T. Structure classification and melting temperature prediction in octet AB solids via machine learning. *Phys. Rev. B* **91**, 214302 (2015).
- Seko, A., Hayashi, H., Nakayama, K., Takahashi, A. & Tanaka, I. Representation of compounds for machine-learning prediction of physical properties. *Phys. Rev. B* **95**, 144110 (2017).
- Lee, J., Seko, A., Shitara, K., Nakayama, K. & Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B* **93**, 115104 (2016).
- Esteves, C., Allen-Blanchette, C., Makadia, A. & Danilidis, K. Learning SO(3) Equivariant Representations with Spherical CNNs. In *Computer Vision – ECCV 2018. Lecture Notes in Computer Science*, Vol 11217 (Eds Ferrari, V., Hebert, M., Sminchisescu, C. & Weiss, Y.) (Springer, Cham., 2018) [https://doi.org/10.1007/978-3-030-01261-8\\_4](https://doi.org/10.1007/978-3-030-01261-8_4).
- Poulenard, A., Rakotosaona, M.-J., Ponty, Y. & Ovsjanikov, M. Effective rotation-invariant point cnn with spherical harmonics kernels. In *2019 International Conference on 3D Vision (3DV)* 47–56. (Springer, IEEE). <https://doi.org/10.1109/3DV.2019.00015>.
- Jiang, C. M., Huang, J., Kashinath, K., Prabhat, Marcus, P. & Niessner, M. (2019). Spherical CNNs on Unstructured Grids. In *ICLR*, (2019).
- Thomas, N. et al. Tensor field networks: rotation- and translation-equivariant neural networks for 3D point clouds. Preprint at <https://ui.adsabs.harvard.edu/abs/2018arXiv180208219T> (2018).
- Duvenaud, D. K. et al. Convolutional Networks on Graphs for Learning Molecular Fingerprints. In *Advances in Neural Information Processing Systems*. Vol. 28 (Eds Cortes, C., Lawrence, N., Lee, D., Sugiyama, M. & Garnett, R.) (Curran Associates, Inc., 2015)
- Xie, T. & Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Phys. Rev. Lett.* **120**, 145301 (2018).
- Zeng, S. et al. Atom table convolutional neural networks for an accurate prediction of compounds properties. *npj Comput. Mater.* **5**, 84 (2019).
- Yamamoto, T. *Crystal Graph Neural Networks for Data Mining in Materials Science* (Research Institute for Mathematical and Computational Sciences, LLC, 2019).
- Park, C. W. & Wolverton, C. Developing an improved crystal graph convolutional neural network framework for accelerated materials discovery. *Phys. Rev. Mater.* **4**, 063801 (2020).
- Griffiths, D. J. & Schroeter, D. F. *Introduction to Quantum Mechanics* (Cambridge University Press, 2018).
- Chen, C., Ye, W., Zuo, Y., Zheng, C. & Ong, S. P. Graph networks as a universal machine learning framework for molecules and crystals. *Chem. Mater.* **31**, 3564–3572 (2019).
- Klicpera, J., Groß, J. & Günnemann, S. Directional Message Passing for Molecular Graphs. In *ICLR*, (2020).
- Ramakrishnan, R., Dral, P. O., Rupp, M. & von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules. *Sci. Data* **1**, 140022 (2014).
- Chmiela, S. et al. Machine learning of accurate energy-conserving molecular force fields. *Sci. Adv.* **3**, e1603015 (2017).
- Unke, O. T. & Meuwly, M. PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges. *J. Chem. Theory. Comput.* **15**, 3678–3693 (2019).
- Louie, S. G., Ho, K.-M. & Cohen, M. L. Self-consistent mixed-basis approach to the electronic structure of solids. *Phys. Rev. B* **19**, 1774–1782 (1979).
- Martin, R. M. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, 2020).
- Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O. & Dahl, G. E. Neural message passing for quantum chemistry. In *International conference on machine learning*. 1263–1272 (PMLR, 2017).
- Kearnes, S., McCloskey, K., Berndl, M., Pande, V. & Riley, P. Molecular graph convolutions: moving beyond fingerprints. *J. Comput.-Aided Mol. Des.* **30**, 595–608 (2016).
- Ong, S. P. et al. Python Materials Genomics (pymatgen): a robust, open-source python library for materials analysis. *Comput. Mater. Sci.* **68**, 314–319 (2013).
- Mnih, V., Heess, N., Graves, A. & Kavukcuoglu, K. Recurrent Models of Visual Attention. In *Advances in Neural Information Processing Systems*. (Eds Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N. & Weinberger, K. Q.) Vol. 27 (Curran Associates, Inc., 2014).
- Schütt, K. T., Arbabzadah, F., Chmiela, S., Müller, K. R. & Tkatchenko, A. Quantum-chemical insights from deep tensor neural networks. *Nat. Commun.* **8**, 13890 (2017).
- Ebert, D. S., Musgrave, F. K., Peachey, D., Perlin, K. & Worley, S. *Texturing & Modeling: a Procedural Approach* (Morgan Kaufmann, 2003).
- Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* **13**, 5188–5192 (1976).
- Chadi, D. J. & Cohen, M. L. Special points in the Brillouin zone. *Phys. Rev. B* **8**, 5747–5753 (1973).
- Jain, A. et al. A high-throughput infrastructure for density functional theory calculations. *Comput. Mater. Sci.* **50**, 2295–2310 (2011).
- Tao, S. X., Cao, X. & Bobbert, P. A. Accurate and efficient band gap predictions of metal halide perovskites using the DFT-1/2 method: GW accuracy with DFT expense. *Sci. Rep.* **7**, 14386 (2017).

41. Papavassiliou, G. C. & Koutselas, I. J. S. M. Structural, optical and related properties of some natural three-and lower-dimensional semiconductor systems. *Synth. Met.* **71**, 1713–1714 (1995).
42. Ryu, S. et al. Voltage output of efficient perovskite solar cells with high open-circuit voltage and fill factor. *Energy Environ. Sci.* **7**, 2614–2618 (2014).
43. Kitazawa, N., Watanabe, Y. & Nakamura, Y. J. Optical properties of CH<sub>3</sub>NH<sub>3</sub>PbX<sub>3</sub> (X = halogen) and their mixed-halide crystals. *J. Mater. Sci.* **37**, 3585–3587 (2002).
44. Phuong, L. Q. et al. Free carriers versus excitons in CH<sub>3</sub>NH<sub>3</sub>PbI<sub>3</sub> perovskite thin films at low temperatures: charge transfer from the orthorhombic phase to the tetragonal phase. *Phys. Chem. Lett.* **7**, 2316–2321 (2016).
45. Chen, L.-C. & Weng, C.-Y. J. N r l. Optoelectronic properties of MAPbI<sub>3</sub> perovskite/titanium dioxide heterostructures on porous silicon substrates for cyan sensor applications. *Nanoscale Res. Lett.* **10**, 1–5 (2015).
46. Yamada, Y., Nakamura, T., Endo, M., Wakamiya, A. & Kanemitsu, Y. J. A. P. E. Near-band-edge optical responses of solution-processed organic–inorganic hybrid perovskite CH<sub>3</sub>NH<sub>3</sub>PbI<sub>3</sub> on mesoporous TiO<sub>2</sub> electrodes. *Appl. Phys. Express* **7**, 032302 (2014).
47. Zhumekenov, A. A. et al. Formamidinium lead halide perovskite crystals with unprecedented long carrier dynamics and diffusion length. *ACS Energy Lett.* **1**, 32–37 (2016).
48. Protesescu, L. et al. Nanocrystals of cesium lead halide perovskites (CsPbX<sub>3</sub>, X = Cl, Br, and I): novel optoelectronic materials showing bright emission with wide color gamut. *Nano Lett.* **15**, 3692–3696 (2015).
49. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. In *ICLR*, (2015).

## Acknowledgements

This work was partially supported by Shenzhen Research Council (JCYJ20200109113427092, GJHZ20180928155209705). J. Cheng acknowledges the support from Huiping Li of the University of Science and Technology of China. Also, L. Dong thanks financial support from Qingdao University of Science and Technology as well as the Malmstrom Endowed Fund at Hamline University. C. Zhang acknowledges the support from Harbin Institute of Technology (Shenzhen) and Shenzhen Research Council.

## Author contributions

J.C., C.Z. and L.D. conceived the idea, developed, and implemented the method. J.C. prepared the data and performed the analysis. All authors discussed the results and contributed to the writing of the article.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to C.Z. or L.D.

**Peer review information** *Communications Materials* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Milica Todorović and Aldo Isidori.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021