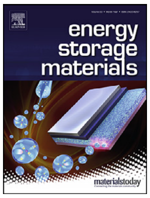
Energy Storage Materials 39 (2021) 45–53

Contents lists available at ScienceDirect

Energy Storage Materials

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



Deep learning for ultra-fast and high precision screening of energy materials

Zhilong Wang[a], Qingxun Wang[a], Yanqiang Han[a], Yan Ma[b], Hua Zhao[b], Andrzej Nowak[c], Jinjin Li[a,*]

[a] *Key Laboratory for Thin Film and Microfabrication of Ministry of Education, Department of Micro/Nano-electronics, Shanghai Jiao Tong University, Shanghai, 200240, China*

[b] *Beijing Institute of Tracking and Telecommunications Technology, Beijing, 100094, China*

[c] *Faculty of Chemistry, Gdańsk University of Technology, Narutowicza 11/12, 80-233 Gdańsk, Poland*

ARTICLE INFO

Keywords:
Multifunctional energy storage materials
Nanostructured materials
band gap
Transfer learning
ab initio calculations

ABSTRACT

Semiconductor materials for energy storage are the core and foundation of modern information society and play important roles in photovoltaic system, integrated circuit, spacecraft technology, lighting applications, and other fields. Unfortunately, due to the long experiment period and high calculation cost, the high-precision band gap (the basic characteristic parameter) of semiconductor is difficult to obtain, which hinders the development of new semiconductor materials. Since the traditional Perdew–Burke–Ernzerhof (PBE) functional not only requires a long calculation time, but also significantly underestimates the band gap, we developed a deep learning model that can predict the more precise Heyd–Scuseria–Ernzerhof (HSE06) band gaps in milliseconds for 1,503 binary metallic oxides, nitrides, and sulfides, with a mean absolute error (MAE) of 0.35 eV, a mean squared error (MSE) of 0.21 eV, and a coefficient of determination (R^2) of 0.98. Based on transfer learning, only $<$ 5% of the data set (64 structures) was required to train the model and predict the band gaps of the remaining 1,439 structures. From the 1,503 candidate materials, we quickly identified 75 carrier transport materials, 33 electrode and electrocatalytic materials, 299 power switching materials, and 114 sensing materials. This work is the first to demonstrate the feasibility of transfer learning in band gap prediction, from the low-level PBE to the high-level HSE06 calculation, with a computation speed at least 10^4 times faster than the *ab initio* calculation. The proposed method could be further expanded to incorporate entire organic/inorganic crystal materials databases ($> 10^6$ crystals), which is of great significance for the screening and discovery of new semiconductor energy storage materials.

1. Introduction

Semiconductor energy storage materials are crucial for various electronic devices, such as p-n junctions, ultraviolet (UV) luminescence devices, photocatalysts, and thin-film transistors[1–5], which support the modern electronic information society. Thus, the characteristic parameters of semiconductor materials must be meticulously and comprehensively understood, particularly the band gap (E_g)[6–9]. For example, in perovskite solar cells (PSCs), the hole transport layer (HTL) and the electron transport layer (ETL) should have appropriate E_g that match the band gaps of perovskite to ensure the efficient transmission of holes and electrons[10,11], while their E_g are within the range of optimal optical conversion efficiency (0.9–1.6 eV). E_g is also an essential parameter for determining electronic conductivity and helps to identify more efficient conductive materials in batteries for a wide range of applications[12,13]. Electrode materials generally have high conductivity, i.e. narrow E_g, while solid electrolyte materials require extremely low electronic conductivity, i.e. wide E_g[14–16]. In addition, one criterion for identifying suitable sensor materials includes having a band gap with an empirical value of 2–4 eV. Semiconductors with larger E_g have low carrier concentrations and are less able to absorb gas molecules, while the resistance change for semiconductors with a smaller band gap is imperceptible due to the high carrier concentration[17].

Functional inorganic semiconductor materials have a wide range of applications in energy systems. Specifically, binary metallic oxides, nitrides, and sulfides are often used as HTL/ETL in PSCs[18–21], electrode materials[14,15,22], power switch materials[23–25], and sensing materials[26–29], because of their excellent semiconductor characteristics. Zhang *et al.* have used Co_3O_4-based oxides with an E_g of 1.51 eV (in the visible wavelength range of 0.9–1.6 eV) as HTL in organometallic halide $MAPbI_3$ and $(FASnI_3)_{0.6}(MAPbI_3)_{0.4}$ PSCs, achieving approximately 14% and 7% power conversion efficiencies, respectively[11]. Furthermore, III-nitride semiconductors such as InN and AlN have been

---

[*] Corresponding author.
E-mail address: lijinjin@sjtu.edu.cn (J. Li).

https://doi.org/10.1016/j.ensm.2021.04.006
Received 5 January 2021; Received in revised form 17 March 2021; Accepted 1 April 2021

intensively investigated for optoelectronics applications owing to their suitable band gaps[30,31]. Ga_2O_3 is emerging as an interesting semiconductor for high-power electronics and solar-blind ultraviolet photodetectors, because of its ultrawide band gap (∼5.0 eV)[24,32]. In addition, the narrow band gap porous Mn_2O_3 and α-MnO_2 have been used as electrode materials[16]. At 3.0 eV, the wide E_g of MoO_3 is a significant reason for the widespread application of MoO_3 in the field of sensors, because of its ability to adsorb gas molecules[33]. Thus, accurately determining the E_g is the key to selecting and accelerating the development of new semiconductor materials.

On the experimental side, through a complex and costly process, E_g can be measured by diffuse reflectance, surface photovoltage, UV-vis measurements, etc., but it is impractical to synthesize and measure all the candidate materials experimentally. Density functional theory (DFT) computations have been widely used to determine the electronic structures of thousands of solids[34–36]. However, employing standard semi-local exchange-correlation functionals such as the Perdew–Burke–Ernzerhof (PBE) functional is well known to underestimate the E_g substantially[3,37]. Beyond conventional DFT, more expensive methods such as hybrid functional calculations (Heyd–Scuseria–Ernzerhof, HSE06)[38–40] and GW-type methods[41] demonstrably improve the calculation accuracy. However, these methods are not currently amenable to high-throughput computation, because of their excessive computational cost. Machine learning (ML), a research trend and a high-efficiency technique used in material data engineering[42–49], can circumvent many obstacles in using DFT for estimating E_g. For example, Zhuo *et al.* applied support vector regression (SVR) to predict the experimental E_g based on chemical compositions, resulting in a root-mean-square error (RMSE) of 0.45 eV[37]. Grossman *et al.* developed a crystal graph convolutional neural network (CGCNN) to predict the E_g at PBE level based on crystal materials from the Materials Project[50]. Moreover, Shi *et al.* reviewed the typical mode and basic procedures for applying ML in more material property prediction[51,52], providing guidance for the discovery and design of energy materials. In short, some of these previous ML models did not consider the crystal structure, and some were trained based on the PBE functional. The scarcity of high-precision band gap data set (i.e., at HSE06 or GW level) limits the ability to accurately predict the E_g for a wide range of materials. Recently, transfer learning (TL) has been intensively explored by scientists in the field of chemical engineering and materials science, which has good applications in overcoming the problem of data scarcity[53,54].

Here, we reported a deep learning strategy that can predict the band gaps of binary metallic oxides, nitrides, and sulfides at the HSE06 level in milliseconds. The chosen data set included 1,503 semiconductor materials taken from the Materials Project[55]. Based on the CGCNN model, a prediction model of the PBE band gap was obtained by training from scratch (CGCNN-FS), resulting in a mean absolute error (MAE) of 0.37 eV, a mean squared error (MSE) of 0.35 eV, and a coefficient of determination (R^2) of 0.89. From the 1,503 candidate materials, we calculated the band gaps of 64 (less than 5% of the data set) randomly selected structures using the high-precision HSE06 functional. Then we established a transfer learning model for rapid prediction of HSE06 band gap values based on these 64 data points, with a MAE of 0.35 eV, a MSE of 0.21 eV, and a R^2 of 0.98. Finally, the band gaps of 1,503 semiconductors were predicted at the HSE06 level by the CGCNN-TL model with a computation speed at least 10^4 times faster than the traditional *ab initio* method. This work is the first to demonstrate the feasibility of transfer learning in band gap prediction, from the low-level PBE functional to

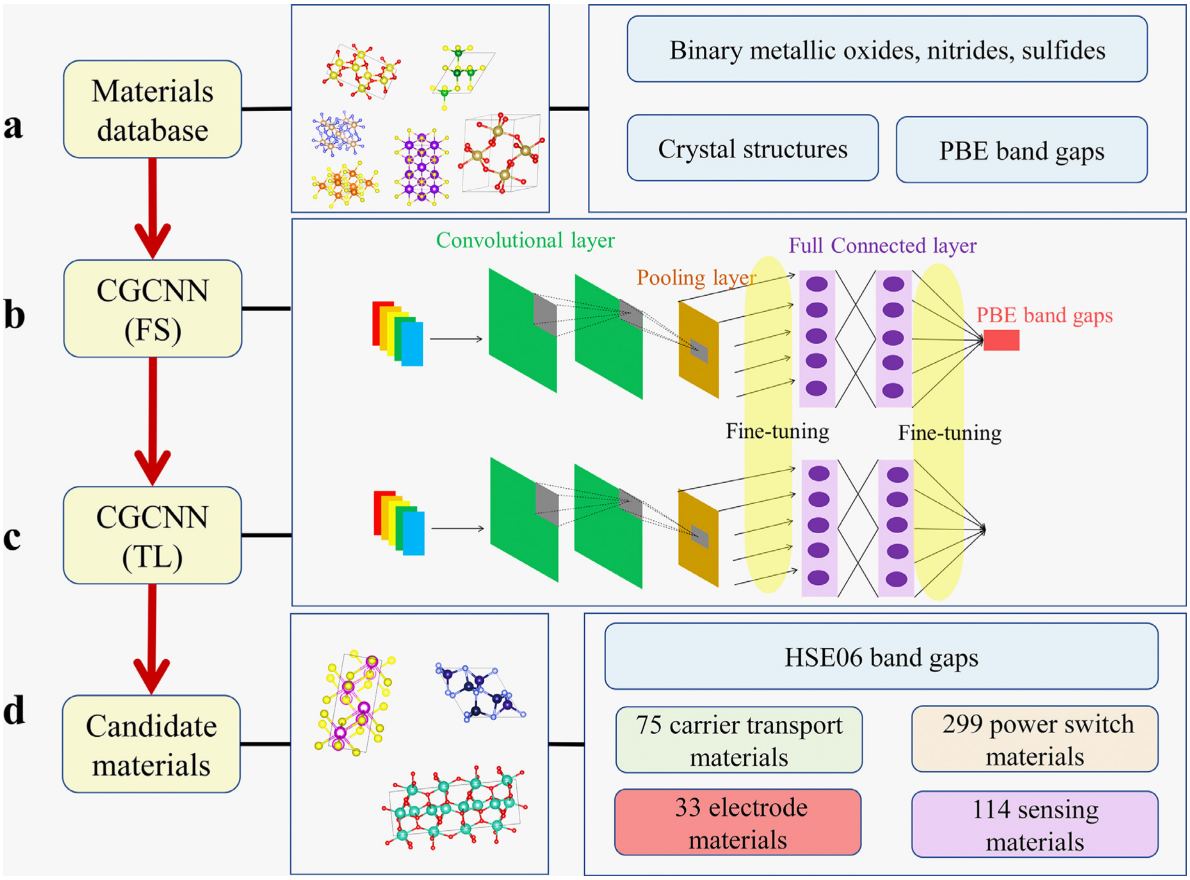


Fig. 1. Framework of DL model. **a** Collection of materials, where a total of 1,503 crystal structures make up the data set. **b** The PBE band gap prediction model created by training the CGCNN model from scratch (CGCNN-FS). **C** The HSE06 band gap prediction model created by training the CGCNN model with transfer learning (CGCNN-TL). **d** Screening of candidate energy materials for further applications using the more accurate, high-level HSE06 band gap predictions.

the high-level HSE06 functional, and does not require large amounts of data. Importantly, according to the HSE06 band gap predictions, we identified 75 materials with band gaps in the visible wavelength region of 0.9–1.6 eV, which may be potential transport layer materials for PSCs; 33 materials with band gaps less than 0.9 eV, as the promising electrode and electrocatalytic materials; 299 materials with band gaps more than 4.0 eV, providing new bipolar oxide switch materials capable of managing large amounts of power or electrolyte materials; and 114 materials with band gaps around 3.0 eV, which may be explored as superior sensing materials. The method proposed in this paper is not only applicable to the screening of semiconductor materials, but also can be expanded to include more organic/inorganic databases for qualitative screening of the same type of materials (including band gap prediction, structural stability prediction, etc.), which is of great significance for the discovery of new materials.

2. Results and discussion

2.1. Deep learning (DL) framework

The deep learning process for this work is presented in Fig. 1, including the collection of material datasets (Fig. 1**a**), the establishment of the CGCNN model based on training from scratch (CGCNN-FS, Fig. 1**b**), fine-tuning the CGCNN model based on transfer learning (CGCNN-TL, Fig. 1**c**), and selecting candidate materials for potential applications (Fig. 1**d**). In this process, a total of 1,503 crystal materials, making up of all the binary metal oxides, nitrides, and sulfides, were collected from the Materials Project database. The crystal structures and the corresponding PBE band gaps of these 1,503 materials were described in the Methods section. In Fig. 1**b**, the crystal materials were converted into graph vectors based on their atomic information and crystal structures. The prediction model for the PBE band gaps was developed using the CGCNN-FS model, which included input layer, convolutional layers, pooling layers, fully connected layers, and output layer). The CGCNN-FS was trained from scratch; that is, the parameters of the network were randomly initialized. From the 1,503 crystal materials, 64 structures (a single batch size, accounting for $< 5\%$ of the total data set) were randomly selected and their band gap values were calculated at the HSE06 level, creating the data set for the CGCNN-TL model. In Fig. 1**c**, the network parameters of the CGCNN-FS were used as the starting point for the CGCNN-TL model. After fixing some of the network parameters, the prediction model for the HSE06 band gap values was obtained after fine-tuning the remaining network parameters. More details about CGCNN architecture and hyper-parameters were provided in the section of Methods. The well-trained CGCNN-TL model was then employed to predict the HSE06 band gap values of all 1,503 materials with high precision, providing a reference for further screening and searching for advanced semiconductor energy storage materials, as shown in Fig. 1**d**.

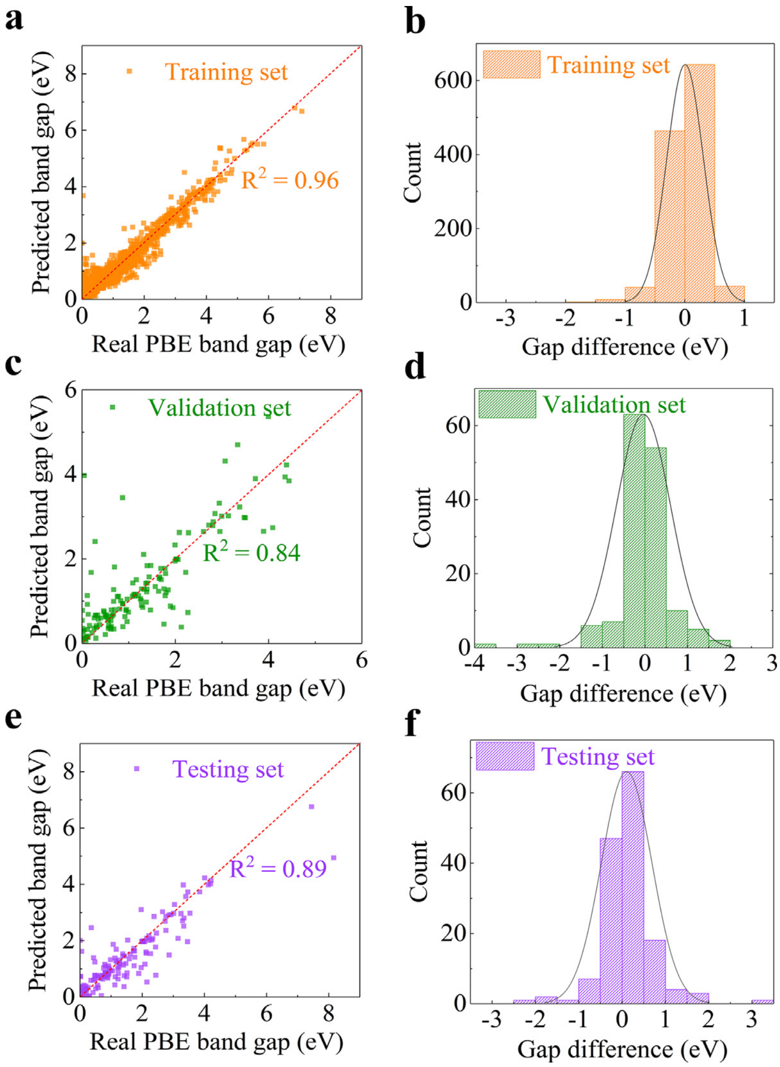


Fig. 2. Performance of the CGCNN-FS model with the training set, validation set, and testing set. **a, c, e** Correlation plots between the established PBE band gap values (taken from the Materials Project) and the band gap values predicted by the CGCNN-FS model with the training set, validation set, and testing set, respectively. **b, d, f** Statistical histograms of the difference in band gap values between the CGCNN-FS predictions and each data set.

Table 1

The data sets, MAEs, MSEs and R^2 of the CGCNN-FS model and CGCNN-TL model, including the training, validation, and testing stages.

	Training				Validation				Testing			
	Data	MAE (eV)	MSE (eV)	R^2	Data	MAE (eV)	MSE (eV)	R^2	Data	MAE (eV)	MSE (eV)	R^2
CGCNN-FS	1203	0.20	0.09	0.96	150	0.37	0.40	0.84	150	0.37	0.35	0.89
CGCNN-TL	52	0.44	0.30	0.94	6	0.32	0.22	0.72	6	0.32	0.13	0.99
	52	0.38	0.20	0.96	6	0.34	0.17	0.97	6	0.37	0.21	0.98
	52	0.44	0.27	0.92	6	0.35	0.19	0.95	6	0.35	0.28	0.98
Average[a]		0.42	0.26	0.94		0.34	0.19	0.88		0.35	0.21	0.98

[a] The average of training, validation and testing of CGCNN-TL model

2.2. PBE band gap prediction based on CGCNN-FS

As a popular deep learning method, CGCNN has been increasingly applied to structure-property prediction for periodic crystal systems[50]. In this work, 1,503 crystal materials were selected according to their atomic information and crystal structures and converted into graph vectors. These vectors were used as the input of CGCNN with the PBE band gap values as the output. Before training, the hyper-parameters of the CGCNN were randomly initialized, which was trained from scratch, or CGCNN-FS for short. In the CGCNN-FS model, as shown in Fig. 1**b**, there were three types of network structures: the convolutional layers, the pooling layers, and the fully connected layers. The convolutional layers were used to learn the atomic feature vectors from the input data by iteratively incorporating the surrounding environment of each atom. The pooling layers were used to produce the overall feature vectors for the crystal structures that satisfied the permutational invariance of the atomic indices and the size invariance of the unit cell. In addition to the convolutional and pooling layers, two hidden, fully connected layers were added to capture the complex mapping between the crystal structures and band gaps. Finally, the output layer, which was connected to the second hidden layer, predicted the band gaps.

In order to evaluate the generalization ability of the model, especially to avoid overfitting phenomenon, the dataset was divided into a training set, a validation set, and a testing set at a ratio of 1203:150:150 (approximate 8:1:1; see Table 1). The optimal hyper-parameters were selected based on minimizing the error compared to the validation set. After optimization, the model was used to predict the band gap values for the testing set. Fig. 2 shows the performance of the CGCNN-FS model with the training set, validation set, and testing set. Fig. 2**a** shows the correlation plot between the established PBE band gap values (taken from the Materials Project database and based on DFT) and the predicted band gaps from the CGCNN-FS for the training set. The plot indicates that the data points are evenly distributed along the diagonal with a coefficient of determination (R^2) of 0.96. Fig. 2**b** shows the statistical histogram of the gap differences, where the deviations are almost all concentrated at approximately 0 eV. Fig. 2**c-d** show the correlation plots of the validation and testing sets, where the predicted band gaps are also consistent with the values obtained from the Materials Project, achieving an R^2 of 0.84 for the validation set and 0.89 for the testing set. Fig. 2 shows that the band gap values predicted by the CGCNN-FS are in good agreement with the DFT calculations, without systematic over- or underestimation. To reflect the actual prediction error between the DFT calculations and the CGCNN-FS predictions, we calculated the mean absolute error (MAE) and mean squared error (MSE), as shown in Table 1. The MAE of the training set, validation set, and testing set are 0.20 eV, 0.37 eV, and 0.37 eV, respectively, and the MSE of the training set, validation set, and testing set were 0.09 eV, 0.40 eV, and 0.35 eV, respectively. In addition, very few data points that are off the base line are inevitable but acceptable. Such low error values indicate that the relationship between the 1,503 crystal structures and their band gaps was well learned by the CGCNN-FS, and the complex DFT calculation can effectively be avoided at the prediction stage. The distribution of PBE band gaps of 1,503 semiconductor materials is shown in Fig.S1 in Supplementary Materials (SM).

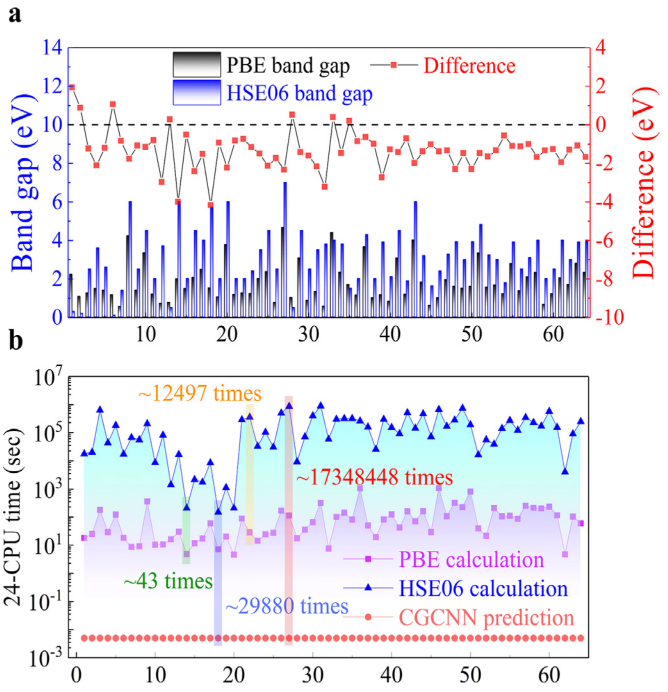


Fig. 3. Comparison of accuracy and computation speed. **a** PBE and HSE06 band gap values of the selected 64 structures and the difference between them (PBE - HSE06). **b** Comparison of calculation time for the PBE functional, HSE06 functional, and CGCNN (FS and TL) predictions.

Through the above data analysis, the PBE band gap prediction model, CGCNN-FS, was determined to be well trained. The hyper-parameters from the neural network were used as the initial points of the HSE06 band gap prediction model, CGCNN-TL. In other words, the hyper-parameters for the CGCNN-TL model were not initialized randomly. Next, these parameters were fine-tuned based on the HSE06 band gap data set through the training, validation, and testing stages to obtain a well-trained transfer learning model.

2.3. HSE06 band gap prediction based on CGCNN-TL

To obtain a HSE06 band gap prediction model, we randomly selected 64 crystal structures (the size of a single batch for the CGCNN-FS model), accounting for approximately 4.26% of the 1,503 total structures, and then calculated their HSE06 band gaps using high-throughput calculations. Fig. 3**a** shows the PBE and HSE06 band gaps of the 64 structures. The HSE06 band gap values are generally larger than the PBE values, with most of the differences (PBE minus HSE06, the red squares) concentrating between 0 eV and −2.0 eV. These results are consistent with the conclusion that the PBE functional significantly underestimates the band gap values. The HSE06 method has the advantage of high precision in calculating the band gap values, but its high computational cost is a barrier to wide adoption. Fig. 3**b** shows a comparison of the calculation

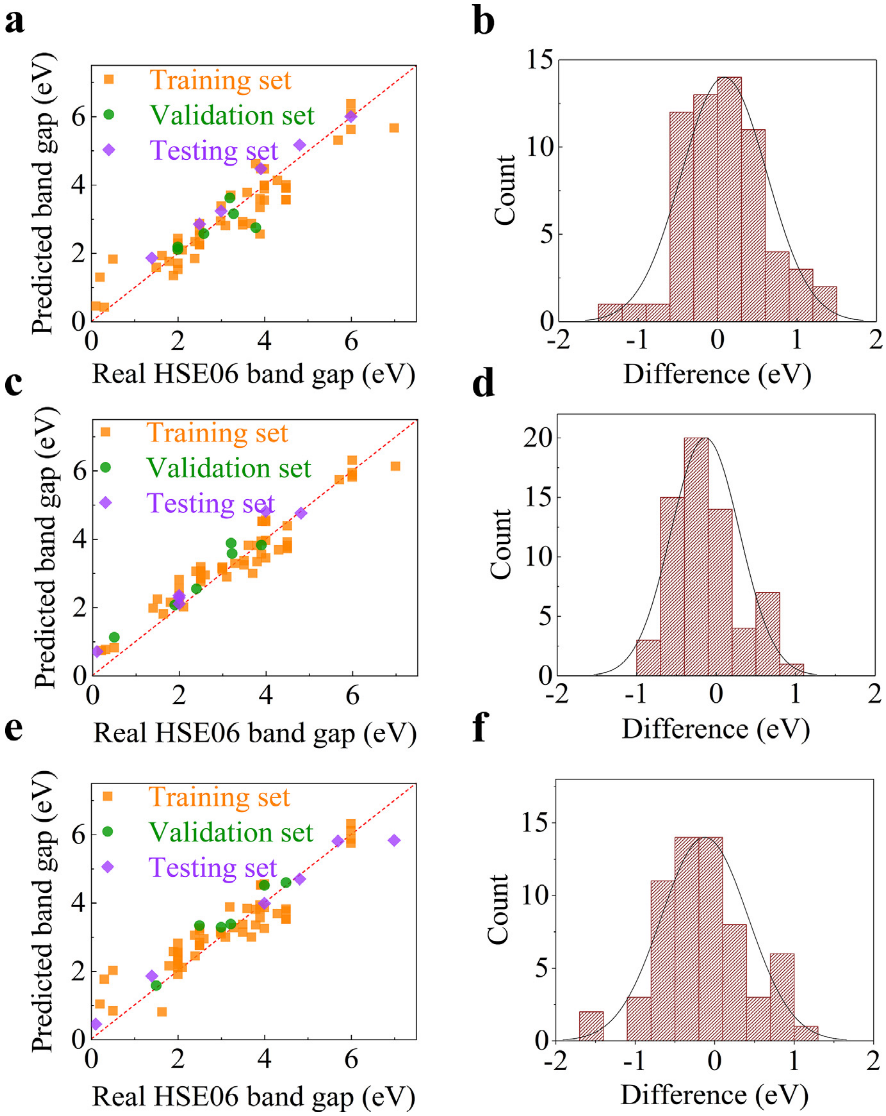


Fig. 4. Performance of the CGCNN-TL model in three independent tests, including the training set, validation set, and testing set. **a, c, e.** The correlation plots between the HSE06 band gap values (calculated by DFT) and the predicted values from the CGCNN-TL model using the training set, validation set, and testing set, respectively. **b, d, f** Statistical histograms of the difference between the band gap values.

time for the 64 selected structures, based on the PBE functional, HSE06 functional, and CGCNN (FS and TL) predictions, represented by purple squares, blue triangles, and red dots, respectively. The green shadow in Fig. 3**b** indicates that the calculation speed for HSE06 values is at least 43 times slower than that of PBE, while the yellow shadow indicates that the calculation speed for HSE06 values is at most 12,497 times slower than that of PBE.

The 64 structures with their corresponding HSE06 band gap values were used as the dataset, and the well-trained CGCNN-FS model was used to fine-tune the parameters of the CGCNN-FS model with TL technology to obtain the CGCNN-TL model. In the CGCNN-TL model, the network architecture was the same as in the CGCNN-FS model, but the way that neurons updated in the network was different. Only a few hyperparameters in the network that needed to be adjusted since a number of parameters were fixed when using only 64 structures (see more details in Methods). Because the 64 structures were selected from the total data set of 1,503 structures, the overall feature vectors for the crystal structures calculated by the convolutional layers and the pooling layers were exactly the same. Thus, the parameters of the CGCNN-FS model were taken as the initial point of the CGCNN-TL model, and the parameters inside the convolutional layers, pooling layers, and full connection layers were fixed (meaning they do not change with iteration). Only the parameters between the pooling layer and the first full connection layer, the second full connection layer, and the output layer were fine-tuned during the training process (as shown in Fig. 1**c**). The distributions of PBE and HSE06 band gaps of selected 64 semiconductor materials are shown in Fig.S2 and Fig.S3 in SM, respectively.

Fig. 4 shows the performance of the CGCNN-TL model. With the relatively small data set to prevent overfitting, we performed three independent and strict validations and tests. We divided the data set three times according to the ratio of 8:1:1, and conducted three rounds of training, validation, and testing. Fig. 4**a** shows the performance of the model during the first independent test. The predicted values in the training set, validation set, and testing set are evenly distributed along the diagonal, with R^2 values of 0.94, 0.72, and 0.99, respectively (see Table 1 for more details). The statistical histograms of the difference in the band gap values are presented in Fig. 4**b,** showing that the deviations are almost all concentrated around 0 eV. Similarly, Fig. 4**c-f** also show this phenomenon in the results of the second and third independent tests. Furthermore, as shown in Table 1, the MAE for the testing sets of the three independent tests are 0.32, 0.37, and 0.35 eV, with an average MAE of 0.35 eV; and the MSE for the testing sets of the three independent tests are 0.13, 0.21, and 0.28 eV, with an average MSE of 0.21 eV; the R^2 values for the results of the testing sets are 0.99, 0.98, and 0.98, with an average R^2 of 0.98. From the correlation plots in Fig. 4, these 64 HSE06 band gaps are concentrated between 2–5 eV. Thus, the

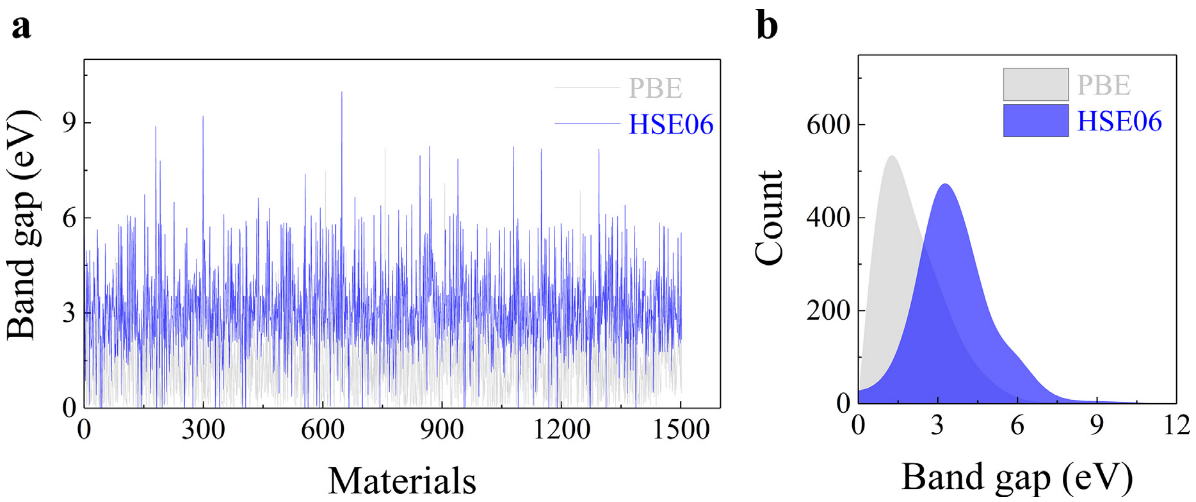


Fig. 5. Predicted HSE06 band gap values for all 1,503 crystal materials. **a** Comparison of the HSE06 PBE band gap values. **b** Statistical distribution plots of the HSE06 and PBE band gaps.

phenomenon that the datasets with the band gap smaller than 1.0 eV systematically deviates from the correlation line may be caused by the uneven band gap distribution. However, in terms of MAE and MSE, as well as the statistical histogram in Fig. 4, these abnormal data have little predictive effect on the overall results. Considering the importance of band gap values of these data points for energy materials, we did not exclude them. The above analysis indicates that the HSE06 band gap values predicted by the CGCNN-TL model are consistent with the DFT calculations using the HSE06 functional. Importantly, our three independent validations and tests (totally 36 data points) demonstrate the reliability of the transfer learning model even when data sets are scarce.

In addition, we compared the computation time for the CGCNN-TL model and the HSE06 calculations. In Fig. 3, the blue shadow indicates that the HSE06 calculation speed is at least 29,880 times ($\sim10^4$) slower than the CGCNN-TL prediction time, while the red shadow indicates that the HSE06 calculation time is at most 17,348,448 times ($\sim10^8$) slower than the CGCNN-TL prediction time. These results show that the well-trained CGCNN-TL model can be used for the ultra-fast prediction of HSE06 band gap values, ensuring high prediction accuracy with a greatly shortened computation time.

2.4. Candidate materials in energy systems

By employing the well-trained CGCNN-TL model, the HSE06 band gap values of the remaining 1,439 crystal materials, including all the binary metallic oxides, nitrides, and sulfides, can be predicted in milliseconds. Fig. 5**a** shows the predicted HSE06 band gap values from the CGCNN-TL model as well as the PBE band gaps. The majority of the HSE06 band gap values are larger than the PBE band gaps. The statistical distributions of the HSE06 and PBE band gap values are plotted in Fig. 5**b**, showing significant distribution shifts. This work is effectively a precise update to the band gap values for the 1,503 crystalline materials since the HSE06 calculations are generally considered the closest to experimental results[38,39,56–59]. In semiconductor structures, different material band gap values correspond to different research applications. For example, potential carrier transport layer materials for PSCs require band gaps in the visible wavelength range of 0.9–1.6 eV, electrode materials require narrow band gaps (less than 0.9 eV), well-performed sensitive materials have band gap values of 2.0–4.0 eV, and some high-power conversion devices require a very wide band gap greater than 4.0 eV.

Among the 1,503 materials, 75 are identified to have band gaps in the visible wavelength range of 0.9–1.6 eV, which could be potential transport layer materials for PSCs; 33 have band gaps less than 0.9 eV, which are promising electrode and electrocatalytic materials; 299 have band gaps greater than 4.0 eV, providing a new opportunity for using bipolar oxide materials as switches to manage large amounts of electrical energy; and 114 have band gaps of approximately 3.0 eV, which may be further explored as superior sensing materials. The candidate materials are listed in Table 2. The corresponding IDs, chemical formulas, PBE band gap values, the predicted HSE06 band gap values for the 75 candidate materials (with band gaps between 0.9 and 1.6 eV), the 33 candidate materials (with band gaps less than 0.9 eV), the 299 candidate materials (with band gaps more than 4.0 eV), and the 114 candidate materials (with the band gaps around 3.0 eV) are listed in Table 3, Table S1-S3 of SM, respectively. From these four tables, there are large differences between PBE and HSE06 band gap values for some materials, for example, see N-243, O-585, S-168, N-57, and O-861 in Table 3. This suggests that the traditional PBE band gap calculation can produce forecasting errors, which may hinder the development of and application for some semiconductor materials. The crystallographic information files (CIFs) and the corresponding PBE (taken from the MP database) and HSE06 (predicted in this work) band gaps of all 1,503 crystal materials are provided in the SM.

Table 2
Screened new binary semiconductor materials and their potential applications, where 75 structures with band gaps of 0.9–1.6 eV are available for carrier transport materials, 114 structures with band gaps of about 3.0 eV are available for sensing materials, and 299 structures with band gaps larger than 4.0 eV has potential applications in power switch materials.

Band gap (HSE06)	No. of screened materials	Potential application
0.9-1.6 eV	75	Carrier transport material, photocatalyst, etc.
< 0.9 eV	33	Electrode materials, electro-catalysts, etc.
> 4.0 eV	299	Power switch materials, electrolyte materials, etc.
~ 3.0 eV	114	Sensing material, etc.

Table 3
The IDs, chemical formulas, PBE band gaps, and HSE06 band gaps for the 75 candidate materials. The predicted HSE06 band gaps of these materials are in the visible wavelength range of 0.9–1.6 eV, which indicates that they may be potential carrier transport layer materials in PSCs.

ID	Chemical formula	PBE E_g (eV)	HSE06 E_g (eV)	ID	Chemical formula	PBE E_g (eV)	HSE06 E_g (eV)
N-243	Cu_3N	0.19	1.36	O-658	FeO_2	0.87	1.02
N-442	AgN_3	1.67	1.43	O-660	Fe_2O_3	1.42	1.11
N-488	HfN_2	1.39	1.07	O-662	Fe_2O_3	1.00	0.96
N-529	WN_2	0.57	1.04	O-665	Fe_2O_3	0.26	1.13
N-534	WN_2	0.15	0.93	O-672	FeO	0.45	1.35
N-546	WN_2	0.17	1.05	O-680	Fe_2O_3	1.08	1.29
N-57	$Ti_{19}N_{25}$	0.01	1.30	O-696	Fe_2O_3	1.39	0.98
N-68	TiN	0.01	1.13	O-701	Fe_2O_3	1.12	1.11
N-70	Ti_3N_4	0.10	1.17	O-704	FeO_2	0.95	1.10
O-1336	W_3O_8	1.70	1.58	O-707	Fe_7O_9	1.14	1.17
O-1339	WO_2	1.25	1.10	O-716	Fe_2O_3	0.18	1.04
O-1373	W_3O_8	0.73	1.53	O-718	Fe_2O_3	0.27	1.26
O-1381	W_2O_3	0.02	1.25	O-733	$Fe_{43}O_{64}$	0.77	0.95
O-1383	WO_2	1.73	1.54	O-739	Fe_2O_3	0.12	1.23
O-1396	WO_2	2.23	1.41	O-743	Fe_2O_3	1.41	1.18
O-1397	WO_2	1.40	1.22	O-744	Fe_2O_3	2.02	0.95
O-1408	WO_2	1.40	1.41	O-758	Fe_2O_3	1.49	0.93
O-1409	WO_2	1.61	1.56	O-772	CoO	0.22	1.52
O-1412	WO_2	2.47	1.46	O-776	CoO	0.35	1.57
O-1416	W_3O_8	1.82	1.55	O-784	$Co_{29}O_{40}$	0.26	1.40
O-1464	HgO	1.19	1.34	O-786	$Co_{23}O_{32}$	0.30	1.59
O-1468	HgO	1.20	1.31	O-796	CoO	0.27	1.47
O-1469	HgO	1.30	1.35	O-808	CoO	0.26	1.41
O-1470	HgO	1.24	1.29	O-816	Co_2O_3	0.15	0.98
O-407	VO	0.81	1.45	O-861	CuO	0.15	1.39
O-531	Cr_3O_4	0.97	1.20	O-882	CuO	0.01	1.39
O-539	Cr_2O_3	2.43	1.52	O-939	Rb_2O_2	0.01	1.41
O-543	CrO	0.17	1.45	S-158	Fe_7S_{12}	0.17	1.59
O-545	Cr_2O_3	2.13	1.51	S-168	FeS_2	0.46	1.58
O-552	Cr_2O_3	1.77	1.57	S-172	Fe_3S_4	0.00	1.53
O-561	CrO	0.50	1.22	S-182	Fe_7S_8	0.02	1.11
O-567	CrO	0.34	1.09	S-225	Cu_2S	0.12	1.11
O-570	CrO_2	0.34	1.60	S-505	RuS_2	0.67	1.53
O-584	Cr_9O_{13}	0.61	1.17	S-507	Rh_2S_3	0.42	1.03
O-585	Cr_5O_7	0.78	1.17	S-508	Rh_2S_3	0.19	1.17
O-590	Cr_2O_3	0.59	1.19	S-635	HgS	1.70	1.37
O-652	Fe_2O_3	1.54	0.99	S-639	HgS	1.68	1.37
O-657	Fe_2O_3	0.14	1.40				

3. Conclusions

As an important feature of semiconductor, band gap is difficult to be accurately measured due to the long experimental period and high computational cost. In this work, we developed a deep learning model that can perform ultra-fast predictions of the band gaps at the HSE06 level in milliseconds for all the binary metallic oxides, nitrides, and sulfides. The data set we used was taken from the Materials Project database and included the structures for 1,503 semiconductors and their band gap values at the PBE functional level. The well-trained CGCNN-FS model predicted the PBE band gap at the PBE level with a MAE of 0.37 eV, a MSE of 0.35 eV, and a R^2 of 0.89. We then randomly selected 64 structures (< 5% of the data set) from the 1,503 semiconductors learned by the CGCNN-FS model and calculated their HSE06 band gap values. Using the TL strategy, we quickly trained the CGCNN-TL model with these 64 structures, achieving a MAE of 0.35 eV, a MSE of 0.21 eV, and a R^2 of 0.98. Finally, the well-trained CGCNN-TL model was used to predict the band gap values of the 1,503 semiconductor materials at the high-precision HSE06 level with a computation speed at least 10^4 times faster than the traditional calculation method. Importantly, based on the predicted HSE06 band gaps, we identified several semiconductor materials with potential applications to various fields that might have been ignored by the traditional PBE band gap predictions. For example, we found 75 materials with band gaps in the visible wavelength range of 0.9–1.6 eV, which might have potential as transport layer materials for PSCs. 33 materials had band gap less than 0.9 eV that were promising electrode and electrocatalytic materials. We also identified 299 semiconductor materials with band gaps greater than 4.0 eV, which might provide a new opportunity for using bipolar oxide materials as switches to manage a large amount of electrical energy and high-power electrons, and for solar-blind ultraviolet photodetectors. Moreover, at a band gap of approximately 3.0 eV, we found 114 structures that might be superior sensing materials. Table S4 in SM lists the IDs, chemical formulas, PBE band gaps and the predicted HSE06 band gaps in this work of all 1503 materials. The deep learning method proposed in this paper can not only provide high-precision electronic characteristic parameters of corresponding materials, but also provide a prediction method that can predict high-precision HSE06 band gap values of more materials within milliseconds. Although the deep learning method proposed in this paper still has inevitable errors, the errors can be further reduced if more effective data and new algorithms are incorporated in the future.

In summary, this work is the first to demonstrate the feasibility of transfer learning for band gap prediction, from the low-precision PBE functional to the high-precision HSE06 functional, and does not require large amounts of data. The CGCNN-TL model developed in this study has overcome the issues of low precision or overfitting caused by small datasets, and could play a powerful role in remedying the lack of data caused by inaccessible data sets. The entire materials database contains thousands of materials, including many organic and inorganic crystals. However, these structures are underutilized, and might be overlooked by researchers in relevant fields, thus affecting the development of new materials.

In addition, the proposed model can not only quickly and accurately identify binary semiconductor materials with potential applications in energy systems, but also expand the material database to include more organic/inorganic crystal structures (rather than just binary oxides, nitrides, and sulfides). Researchers can use all available material databases as datasets, including their own crystal structures and PBE band gaps, and then extract a small fraction of these structures (likely 10%–20%) to calculate their HSE06 band gaps. The transfer from PBE to HSE band gaps can be achieved by using the method presented in this paper. In addition, the method can also predict many properties of materials, such as stability, mechanical properties, optical properties, etc. Taking the screening of semiconductor materials as an example, the model proposed in this study can provide strong guidance for experiments and has far-reaching significance for the identification of other new materials.

4. Methods

4.1. Data sets

The data set used in this work was taken from the Materials Project[55] and contained all the binary metallic oxides, nitrides, and sulfides, covering a variety of crystal structures (Cmmm, $P6_3/mmc$, I4/m, …, $P6_3mc$), with 2,877 crystal structures in total. In this work, we focused on high-precision band gap prediction; thus 1,503 semiconductor materials ($E_g > 0$) were selected, and their band gaps, ranging from approximately 0 to 8 eV (see Fig. S1 for more information), were calculated by using the PBE method. Among these semiconductor materials, there are 979 oxides, 356 sulfides, and 168 nitrides and these materials contain many metal elements (Al, Zr, K, Ti, …, Y). All of the data can be downloaded using the Materials API Pymatgen[60]. In the transfer learning process, we randomly selected 64 structures (the size of a single batch), including 44 oxides, 12 sulfides, and 8 nitrides, and calculated their band gaps using the HSE06 method. The distribution of the PBE and HSE06 band gap values is shown in Fig. S2-3. In order to make the data easier to query, and to uniquely encode each crystal material during training, the 1,503 semiconductor materials were tagged by type and numbered sequentially, for example, oxides O-1, …, O-979. The crystallographic information files (CIFs) are provided in the SM.

4.2. CGCNN model

Crystal graph convolutional neural network (CGCNN) is a popular method of machine learning and has been increasingly applied to the tasks of structure-property prediction in periodic crystal systems[50]. In the crystal graph, each node i was represented by a feature vector $\boldsymbol{v}_i$, encoding the property of the atoms corresponding to node i. Similarly, each edge $(i, j)_k$, represented by $\boldsymbol{u}(i, j)_k$, denoted the kth bond connecting atom i and atom j.

The convolutional layers iteratively updated the atomic feature vector $\boldsymbol{v}_i$ by incorporating neighboring atoms and bonds using the nonlinear graph convolution function:

$$\boldsymbol{v}_i^{(t+1)} = Conv\left(\boldsymbol{v}_i^{(t)},\ \boldsymbol{v}_j^{(t)}, \boldsymbol{u}_{(i,\ j)_k}\right) \tag{1}$$

To obtain the feature vector $v_i^{(R)}$ for each atom. The pooling layer was used to produce an overall feature vector $\boldsymbol{v}_c$ for the crystal (or descriptors), which was represented by the pooling function:

$$\boldsymbol{v}_c = Pool\left(\boldsymbol{v}_0^{(0)}, \boldsymbol{v}_1^{(0)}, \ldots, \boldsymbol{v}_N^{(0)}, \ldots, \boldsymbol{v}_N^{(R)}\right). \tag{2}$$

Then, the overall feature vectors were served as inputs of full connection neural network layers. More details for constructing crystal graphs were provided in Note. S1. To choose suitable hyper-parameters to obtain the best model and prevent overfitting, we applied a train-validation scheme to optimize the prediction of the crystal formation energy. The data set was divided randomly into a training set (80% of the data), validation set (10%), and test set (10%). Next, models with different hyperparameters were trained with the training set via optimizers, with a learning rate of 0.01, the number of hidden atomic features in convolutional layer at 64, the number of hidden features after pooling at 64, the number of convolution layers at 2, and the number of hidden layers after pooling at 2. For the CGCNN-FS model, the batch size was set to 64, and at 10 for the CGCNN-TL. The resulting neural network weights were used to predict the properties of crystals in the validation set. Finally, the best model was used to predict the values for the testing set.

Furthermore, since the 64 structures were selected from the total data set of 1,503 structures, the overall feature vectors for the crystal structures calculated by the convolutional layers and the pooling layers were exactly the same. Thus, the parameters of the CGCNN-FS model were taken as the initial point of the CGCNN-TL model, and the parameters inside the convolutional layers and pooling layers were fixed (meaning they do not change with iteration). At the same time, to avoid overfitting among thousands of parameters in the training full connection layers, we also fixed the parameters of full connection layer, thus only the parameters between the pooling layer, the first full connection layer, the second full connection layer, and the output layer are fine-tuned during the training process (as shown in Fig. 1**c**).

4.3. DFT details

All DFT calculations were conducted using the Vienna Ab initio Simulation Package (VASP)[61]. For the optimization of the geometric structure, the projected augmented wave (PAW) method[62,63] was applied to describe ion–electron interactions along with the PBE exchange-correlation function within the generalized gradient approximation (GGA)[64]. For the electronic structure calculations, the nonlocal screened Coulomb potential hybrid density HSE06 functional was performed on the smaller dataset[38,39]. In all calculations, the DFT-D3 method was employed for the van der Waals correction[65,66]. The cutoff energy for the plane-wave basis was set as 500 eV[67,68]. The structure optimization process ran until an energy convergence threshold of 10^{-5} eV and atomic force less than 0.05 eV/Å was reached. As for the materials that can form a variety of crystal structures, the Brillouin zone integration was performed using a Monkhorst-Pack mesh that referred to the Materials Project calculations.

Data availability

The datasets generated during and/or analyzed during the current study are available from the Supplementary Materials and Materials Project (https://www.materialsproject.org).

Author contributions

Conceptualization: Zhilong Wang, Jinjin Li; Methodology: Zhilong Wang, Qingxun Wang, Yanqiang Han; Writing - original draft preparation: Zhilong Wang, Qingxun Wang, Yanqiang Han, Yan Ma, Hua Zhao, Andrzej Nowak; Writing - review and editing: Jinjin Li; Funding acquisition: Jinjin Li; Supervision: Jinjin Li.

Declaration of Competing Interest

J.J.L., Z.L.W., Y.Q.H., and Q.X.W. have filed a patent related to this work: Chinese

Application No. 202010986874.4, dated 18 September 2020.

Acknowledgements

The authors are grateful for the financial support provided by the National Natural Science Foundation of China (No.21901157), and the SJTU Global Strategic Partnership Fund (2020 SJTU-HUJI).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ensm.2021.04.006.

References

[1] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559 (2018) 547–555.
[2] J. Schmidt, M.R.G. Marques, S. Botti, M.A.L Marques, Recent advances and applications of machine learning in solid-state materials science, npj Comput Mater 5 (2019) 83.
[3] C. Chen, et al., A critical review of machine learning of energy materials, Adv. Energy Mater 10 (2020) 1903242.
[4] S. Cestellos-Blanco, H. Zhang, J.M. Kim, Y. Shen, P. Yang, Photosynthetic semiconductor biohybrids for solar-driven biocatalysis, Nat Catal 3 (2020) 245–255.
[5] A. Chaves, et al., Bandgap engineering of two-dimensional semiconductor materials, npj 2D Mater Appl 4 (2020) 29.
[6] O. Voznyy, B.R. Sutherland, A.H. Ip, D. Zhitomirsky, E.H. Sargent, Engineering charge transport by heterostructuring solution-processed semiconductors, Nat Rev Mater 2 (2017) 17026.
[7] F.P. García de Arquer, A. Armin, P. Meredith, E.H. Sargent, Solution-processed semiconductors for next-generation photodetectors, Nat Rev Mater 2 (2017) 16100.
[8] O.A. von Lilienfeld, K.-R. Müller, A. Tkatchenko, Exploring chemical compound space with quantum-based machine learning, Nat Rev Chem 4 (2020) 347–358.
[9] H. Bronstein, C.B. Nielsen, B.C. Schroeder, I. McCulloch, The role of chemical design in the performance of organic semiconductors, Nat Rev Chem 4 (2020) 66–77.
[10] Y. Wang, J. Schwartz, J. Gim, R. Hovden, Z. Mi, Stable unassisted solar water splitting on semiconductor photocathodes protected by multifunctional GaN nanostructures, ACS Energy Lett 4 (2019) 1541–1548.
[11] Y. Zhang, et al., Synthesis and characterization of spinel cobaltite (Co_3O_4) thin films for function as hole transport materials in organometallic halide perovskite solar cells, ACS Appl. Energy Mater. 3 (2020) 3755–3769.
[12] L. Johnson, et al., The role of LiO_2 solubility in O_2 reduction in aprotic solvents and its consequences for Li–O_2 batteries, Nat Chem 6 (2014) 1091–1099.
[13] C. Liu, et al., Growth of NaO_2 in highly efficient Na–O_2 batteries revealed by synchrotron In operando X-ray diffraction, ACS Energy Lett 2 (2017) 2440–2444.
[14] D. Dang, et al., A bi-functional WO_3-based anode enables both energy storage and conversion in an intermediate-temperature fuel cell, Energy Storage Mater 12 (2018) 79–84.
[15] M. Jiang, et al., Defect-engineered MnO_2 enhancing oxygen reduction reaction for high performance Al-air batteries, Energy Storage Mater 18 (2019) 34–42.
[16] Y. Jiao, et al., Recent progress and prospects of Li-CO_2 batteries: mechanisms, catalysts and electrolytes, Energy Storage Mater 34 (2021) 148–170.
[17] H. Chen, L. Sun, G.-D. Li, X. Zou, Well-tuned surface oxygen chemistry of cation off-Stoichiometric spinel oxides for highly selective and sensitive formaldehyde detection, Chem. Mater. 30 (2018) 2018–2027.
[18] D. Kozawa, et al., Evidence for fast interlayer energy transfer in $MoSe_2/WS_2$ heterostructures, Nano Lett 16 (2016) 4087–4093.
[19] E. Avigad, L. Etgar, Studying the effect of MoO_3 in hole-conductor-free perovskite solar cells, ACS Energy Lett 3 (2018) 2240–2245.
[20] T. Pham, G. Li, E. Bekyarova, M.E. Itkis, A. Mulchandani, MoS_2-based optoelectronic gas sensor with sub-parts-per-billion limit of NO_2 gas detection, ACS Nano 13 (2019) 3196–3205.
[21] L. Tian, et al., Understanding the role of surface states on mesoporous NiO films, J. Am. Chem. Soc. 142 (2020) 18668–18678.
[22] M.L. Divya, S. Natarajan, Y.-S. Lee, V. Aravindan, Highly perforated V_2O_5 cathode with restricted lithiation toward building “rocking-chair” type cell with graphite anode recovered from spent Li-ion batteries, Small 16 (2020) 2002624.
[23] M. Xiao, K.P. Musselman, W.W. Duley, Y.N. Zhou, Reliable and low-power multilevel resistive switching in TiO_2 nanorod arrays structured with a TiO_x seed layer, ACS Appl. Mater. Interfaces 9 (2017) 4808–4817.
[24] J. Zhang, et al., Fabrication and interfacial electronic structure of wide bandgap NiO and Ga_2O_3 p–n heterojunction, ACS Appl. Electron. Mater. 2 (2020) 456–463.
[25] R. Woods-Robinson, et al., Wide band gap chalcogenide semiconductors, Chem. Rev. 120 (2020) 4007–4055.
[26] A. Staerz, et al., WO_3-based gas sensors: identifying inherent qualities and understanding the sensing mechanism, ACS Sens 5 (2020) 1624–1633.
[27] V. Galstyan, et al., A novel approach for green synthesis of WO_3 nanomaterials and their highly selective chemical sensing properties, J. Mater. Chem. A 8 (2020) 20373–20385.
[28] K. Kim, P.gyu Choi, T. Itoh, Y. Masuda, Catalyst-free highly sensitive SnO_2 nanosheet gas sensors for parts per billion-level detection of acetone, ACS Appl. Mater. Interfaces 12 (2020) 51637–51644.
[29] K. Suematsu, et al., Selective detection of toluene using pulse-driven SnO_2 micro gas sensors, ACS Appl. Electron. Mater. 2 (2020) 2913–2920.
[30] N.J. O'Brien, et al., In situ activation of an indium(III) triazenide precursor for epitaxial growth of indium nitride by atomic layer deposition, Chem. Mater. 32 (2020) 4481–4489.
[31] S.G. Bishop, et al., Room-temperature quantum emitter in aluminum nitride, ACS Photonics 7 (2020) 1636–1641.
[32] J.E.N Swallow, et al., Influence of polymorphism on the electronic structure of Ga_2O_3, Chem. Mater. 32 (2020) 8460–8470.
[33] F. Rahman, et al., Dual selective gas sensing characteristics of 2D α-MoO_{3-x} via a facile transfer process, ACS Appl. Mater. Interfaces 11 (2019) 40189–40195.
[34] P. Borlido, et al., Large-scale benchmark of exchange–correlation functionals for the determination of electronic band gaps of solids, J. Chem. Theory Comput. 15 (2019) 5069–5079.
[35] E.I. Marchenko, et al., Database of two-dimensional hybrid perovskite materials: open-access collection of crystal structures, band gaps, and atomic partial charges predicted by machine learning, Chem. Mater. 32 (2020) 7383–7388.
[36] R. Dong, A. Jacob, S. Bourdais, S. Sanvito, High-throughput bandstructure simulations of van der Waals hetero-bilayers formed by 1T and 2H monolayers, npj 2D Mater Appl 5 (2021) 26.
[37] Y. Zhuo, A. Mansouri Tehrani, J. Brgoch, Predicting the band gaps of inorganic solids by machine learning, J. Phys. Chem. Lett. 9 (2018) 1668–1673.
[38] J. Heyd, G.E. Scuseria, Efficient hybrid density functional calculations in solids: Assessment of the Heyd–Scuseria–Ernzerhof screened Coulomb hybrid functional, J. Chem. Phys. 121 (2004) 1187–1192.
[39] A.J. Garza, G.E. Scuseria, Predicting band gaps with hybrid density functionals, J. Phys. Chem. Lett. 7 (2016) 4165–4170.
[40] S. Kim, et al., A band-gap database for semiconducting inorganic materials calculated with hybrid functional, Sci Data 7 (2020) 387.
[41] M. Gerosa, et al., Electronic structure and phase stability of oxide semiconductors: performance of dielectric-dependent hybrid functional DFT, benchmarked against GW band structure calculations and experiments, Phys. Rev. B 91 (2015) 155201.
[42] B. Liu, et al., Rationalizing the interphase stability of Li|doped-$Li_7La_3Zr_2O_{12}$ via automated reaction screening and machine learning, J. Mater. Chem A 7 (2019) 19961–19969.
[43] Y. Liu, J.-M. Wu, M. Avdeev, S.-Q. Shi, Multi-layer feature selection incorporating weighted score-based expert knowledge toward modeling materials with targeted properties, Adv. Theory Simul. 3 (2020) 1900215.
[44] K.M. Jablonka, D. Ongari, S.M. Moosavi, B. Smit, Big-data science in porous materials: materials genomics and machine learning, Chem. Rev. 120 (2020) 8066–8129.
[45] V. Gladkikh, et al., Machine learning for predicting the band gaps of ABX_3 perovskites from elemental properties, J. Phys. Chem. C 124 (2020) 8905–8918.
[46] Y. Han, Z. Wang, J. Li, Neural networks accelerate the ab initio prediction of solid—solid phase transitions at high pressures, J. Phys. Chem. Lett. 12 (2021) 132–137.
[47] Z. Wang, H. Zhang, J. Li, Accelerated discovery of stable spinels in energy systems via machine learning, Nano Energy 81 (2021) 105665.
[48] H. Zhang, Z. Wang, J. Ren, J. Liu, J. Li, Ultra-fast and accurate binding energy prediction of shuttle effect-suppressive sulfur hosts for lithium-sulfur batteries using machine learning, Energy Storage Mater 35 (2021) 88–98.
[49] A. Wang, et al., Identifying chemical factors affecting reaction kinetics in Li-air battery via ab initio calculations and machine learning, Energy Storage Mater 35 (2021) 595–601.
[50] T. Xie, J.C. Grossman, Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties, Phys. Rev. Lett. 120 (2018) 145301.
[51] S. Shi, et al., Multi-scale computation methods: their applications in lithium-ion battery research and development, Chinese Phys. B 25 (2016) 018212.
[52] Y. Liu, B. Guo, X. Zou, Y. Li, S. Shi, Machine learning assisted materials design and discovery for rechargeable batteries, Energy Storage Mater 31 (2020) 434–450.
[53] D. Jha, et al., Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning, Nat. Commun. 10 (2019) 5316.
[54] J.S. Smith, et al., Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning, Nat. Commun. 10 (2019) 2903.
[55] A. Jain, et al., The Materials Project: a materials genome approach to accelerating materials innovation, APL Mater 1 (2013) 011002.
[56] J.P. Perdew, et al., Restoring the density-gradient expansion for exchange in solids and surfaces, Phys. Rev. Lett. 100 (2008) 136406.
[57] Y. Hinuma, et al., Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis, Nat Commun 7 (2016) 11962.
[58] M. Sluydts, M. Pieters, J. Vanhellemont, V. Van Speybroeck, S. Cottenier, High--throughput screening of extrinsic point defect properties in Si and Ge: database and applications, Chem. Mater. 29 (2017) 975–984.
[59] P. Borlido, et al., Exchange-correlation functionals for band gaps of solids: benchmark, reparametrization and machine learning, npj Comput Mater 6 (2020) 96.
[60] S.P. Ong, et al., Python materials genomics (pymatgen): a robust, open-source python library for materials analysis, Comput Mater Sci 68 (2013) 314–319.
[61] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Phys. Rev. B 54 (1996) 11169–11186.
[62] P.E. Blöchl, Projector augmented-wave method, Phys. Rev. B 50 (1994) 17953–17979.
[63] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Phys. Rev. B 59 (1999) 1758–1775.
[64] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized gradient approximation made simple, Phys. Rev. Lett. 77 (1996) 3865–3868.
[65] K. Lee, É.D. Murray, L. Kong, B.I. Lundqvist, D.C. Langreth, Higher-accuracy van der Waals density functional, Phys. Rev. B 82 (2010) 081101.
[66] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu, J. Chem. Phys. 132 (2010) 154104.
[67] E.B. Isaacs, C. Wolverton, Inverse band structure design via materials database screening: application to square planar thermoelectrics, Chem. Mater. 30 (2018) 1540–1546.
[68] B.D. Lee, et al., Discovery of lead-free hybrid organic/inorganic perovskites using metaheuristic-driven DFT calculations, Chem. Mater. 33 (2021) 782–798.