Rational Design of Deep Learning Networks Based on Fusion Strategy for Improved Materials Property Predictions

Hongwei Du, Jian Hui[*], Lanting Zhang[*], Hong Wang[*]

1 School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China.

2 Zhangjiang Institute for Advanced Study, Shanghai Jiao Tong University, Shanghai 201203, China.

3 Materials Genome Initiative Center, Shanghai Jiao Tong University, Shanghai 200240, China.


Corresponding authors

Correspondence to: Hong Wang.

hongwang2@sjtu.edu.cn (Hong Wang);

lantingzh@sjtu.edu.cn (Lanting Zhang); hj20151107@sjtu.edu.cn (Jian Hui).

ABSTRACT

The success of machine learning in predicting material properties depends largely on the design of the model. However, the current designs of deep learning models in materials science have the following prominent problems. Firstly, the model design lacks a rational guidance strategy and heavily relies on a large amount of trial and error. Secondly, numerous deep learning models are utilized across various fields, each with its own advantages and disadvantages. Therefore, it is important to incorporate a fusion strategy to fully leverage them and further expand the design strategies of the models. To address these problems, we analyze that the main reason is the lack of new feedback method rich in physical insights. In this study, we developed a feedback method called Chemical Environment Clustering Vector (CECV) of compounds at different thresholds, which is rich in physical insights. Based on CECV, we rationally designed the Long Short-Term Memory and Gated Recurrent Unit fused with Deep Convolutional Neural Network(L-G-DCNN) to explore the field of structure-agnostic materials property predictions. L-G-DCNN accurately captures the interactions between elements in compounds, enabling more accurate and efficient predictions of material properties. Our results demonstrate that the performance of the L-G-DCNN surpasses the current state-of-the-art structure-agnostic models across 28 benchmark datasets, exhibiting superior sample efficiency and faster convergence speed. By employing different visualization methods, we demonstrate that the fusion strategy based on CECV significantly enhances the comprehension of the L-G-DCNN model design and provides a fresh perspective for researchers in the field of materials informatics.
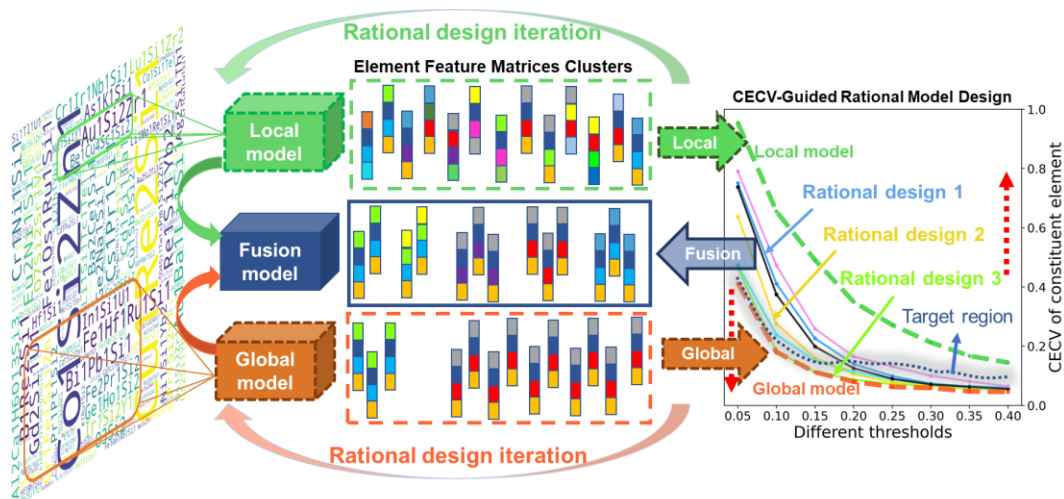
INTRODUCTION

Traditional materials discovery relies on the expensive and time-consuming experimental process, which becomes infeasible due to the vastness of the materials space. To address this challenge, computational methods such as density functional theory (DFT)[1] have provided a more economical approach to calculating material properties. However, DFT simulations require the atomic coordinates of synthesized and characterized materials as input, limiting their effectiveness in exploring the unconstrained chemical space. Machine learning (ML) can predict the properties of unknown materials based on existing data and guide search efforts, enabling the generation of more data to further refine ML models, offering assistance in various materials science challenges[2].

Traditionally, building effective ML models requires developing a representation that is suitable for input data. There are various methods exist to encode information about material composition or crystal structure[3, 4]. However, the predictive accuracy of these methods is limited by the ability to design material representations that encompass all necessary domain knowledge for correct predictions. Deep learning (DL) provides one avenue by reducing the reliance on feature engineering with physically relevant features, allowing for more accurate prediction models. However, defining appropriate input descriptors and model designs remains a key obstacle to widespread adoption. In the study of descriptors, material science primarily includes descriptors based on crystal structure and stoichiometric information. The use of structure-based descriptors[5, 6] implies that the resulting models are limited by the same structural bottlenecks as DFT methods when searching for new compounds. The use of stoichiometric-based descriptors circumvents the structural bottleneck and allows for the learning of interactions between elements in compounds to predict material properties[4, 7]. In terms of model design, the current optimization methods primarily rely on statistical feedback methods. However, these methods lack the incorporation of rich physical insights, such as spectra or imaging feedback similar to those obtained through material characterization techniques like x-ray diffraction (XRD), x-ray photoelectron spectroscopy (XPS), scanning electron microscopy (SEM), transmission electron microscopy (TEM), etc. Moreover, there are many deep learning models applied in various fields, each with its own advantages and disadvantages. It is important to incorporate a fusion strategy to fully utilize them and further expand the design strategies of the models.

In this paper, we rationally designed the long short-term memory[8] and gated recurrent unit[9] fused with deep convolutional neural network[10] (L-G-DCNN) based on the chemical environment clustering vector (CECV) feedback method. The network consists of embedding layer, global feature extraction layer, local feature extraction layer, fusion layer, and output layer, each of which can be independently designed and then combined. Based on the CECV method, our main goal is to design a flexible and scalable model fusion strategy that can adapt to the complexity of the materials field, including varying data sizes, generalization requirements, and prediction accuracy.

We compared the performance of L-G-DCNN with representation learning from stoichiometry

(Roost)[11], compositionally restricted attention-based network (Crabnet)[12], formula graph self-attention network for materials discovery (Finder)[6] and random forest (RF) in predicting material properties. Our model has lower errors and higher sample efficiency. Furthermore, through ablation studies, we have demonstrated the primary function of each layer in the L-G-DCNN. Finally, By systematically deconstructing the L-G-DCNN model and employing visualization techniques, we have enhanced the understanding of the model fusion process, identified critical elemental interaction pairs, and provided intuitive explanations for the property prediction.
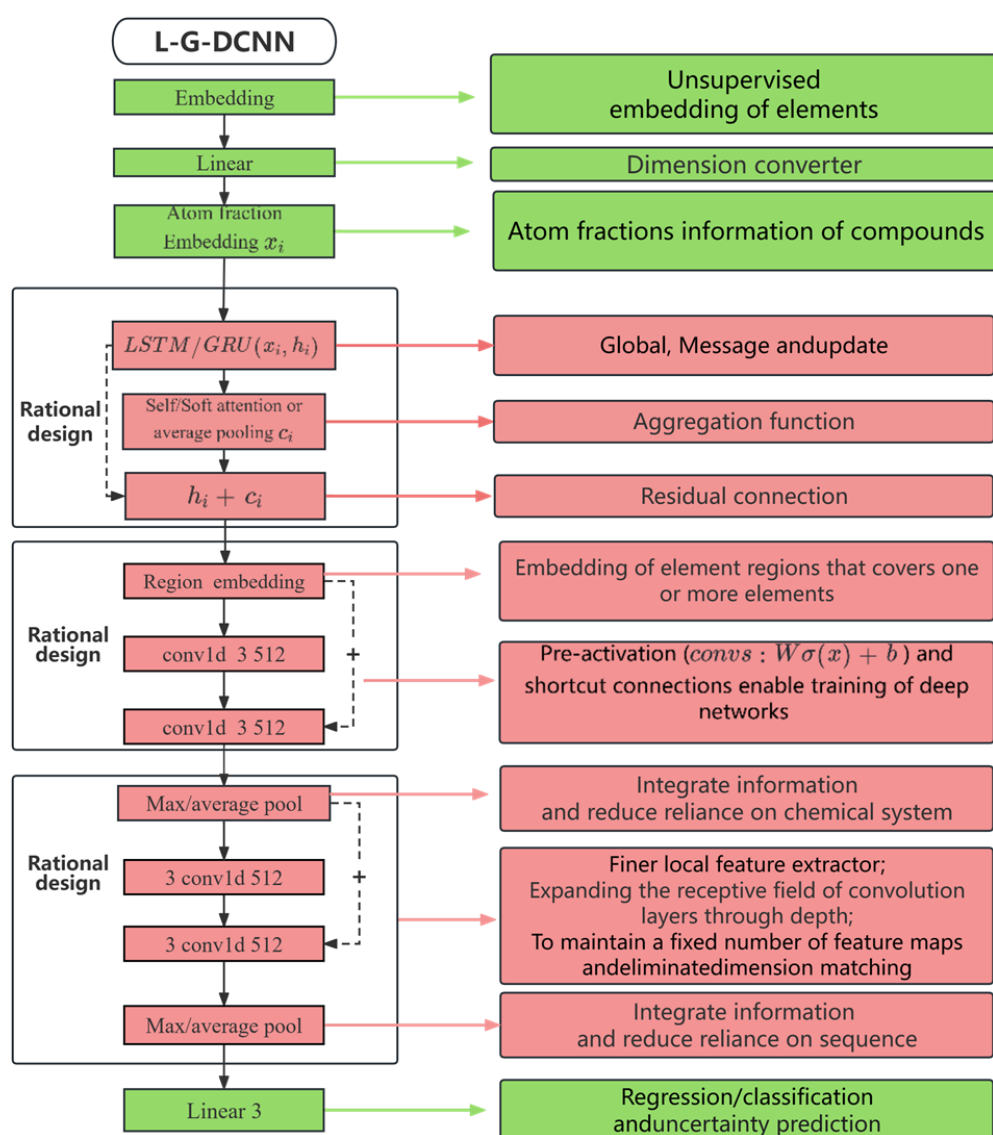
METHODS

Figure 1 show the architecture introduction of the L-G-DCNN model and the functions of each layer.

The CrabNet architecture involves several components, including the input element-derived matrix (EDM), self-attention layers, updated and final element representations (EDM′ and EDM″), residual network, and the final model output. The calculation steps for element contributions and

prediction of targets and uncertainties are also outlined. After the Transformer encoder updates the element representations, each EDM'' passes through a fully connected residual network, transforming the EDMs into the shape (nelements, nelements, 3). The final three vectors, namely element-proto-contributions, element-uncertainties, and element-logits, are defined. The element scaling factor is obtained by applying the sigmoid function to the element-logits. The element-contributions are then derived by multiplying the element-proto-contributions by their respective scaling factor, resulting in element contributions. The mean of the element contributions is then calculated and output as the predicted property value for each compound.

The architecture of Finder involves the processing of a formula graph through multiple messages passing layers followed by a post-processing neural network. Each message passing layer is coupled with a global attention pooling layer to enable residual connections to subsequent layers. The core operations of the architecture are executed during the message phase, which involves predicting directional edge attributes, allowing information to cascade from neighboring nodes to the edges. These edge features, along with end-node attributes, contribute to a message vector. Each message vector is weighted by a self-attention mechanism that quantifies the importance of other nodes for the current message vector. The aggregate step summarizes all messages around a given node via a local pooling function. Finally, at the update step, the aggregated message vector is added to the initial node attribute, completing one cycle of information flow.

Common graph-based models, such as Roost and Finder, represent chemical formulas as weighted graphs, where nodes correspond to different elements in the formula and each node is weighted by the corresponding element stoichiometry. Then, message-passing operations are used to propagate contextual information about different elements in the formula between nodes in the graph to update node representations. In this paper, the L-G-DCNN is obtained by fusing the global and local models (LSTM||GRU and DCNN, where LSTM||GRU represents the stack of LSTM and GRU.). The element embedding learned by L-G-DCNN from the data will perceive the types of composing elements and the element context. Among them, LSTM and DCNN are reference models before fusion. The introduction of reference models is beneficial for us to design feedback rich in physical information, enabling rational model design.

As shown in Figure 1, chemical compositions are inputted by their constituent element's atomic number and fraction. Atomic numbers are used to retrieve element representations, which can be achieved through mat2vec or one-hot encoding. We compared one-hot and mat2vec element embeddings and found that mat2vec performed better on the extended test set. Therefore, we use mat2vec as the default source of chemical information for each element. Mat2vec is based on using unsupervised word embedding methods to encode knowledge in materials science literature into information-dense word embeddings. These embeddings can capture latent knowledge in materials science literature, including the basic structure of elements in the periodic table and the relationship between material properties, and can be used to better understand and predict material properties. To achieve dimension matching, we pass the obtained element embeddings through a fully connected network, with the default dimension of 512. Chemical stoichiometric information is used to obtain atomic fraction embeddings. The chemical stoichiometry of each element is represented by two fraction embeddings, with the first part representing stoichiometry and the second part using logarithmic scaling to map stoichiometry. This logarithmic transformation preserves small fraction amounts and better adapts to doped systems. The L-G-DCNN takes a three-dimensional tensor as input, generated by adding atomic fraction embeddings and element embedding matrices element-

wise. Each row corresponds to an element and the columns contain element embeddings. The chemical compositions are batched on the third dimension. This section is inspired by Crabnet. For more details about the methods, please refer to Supplementary Table S8.

RESULTS AND DISCUSSION

Rational Model Design

Chemical Environment Classification Vector (CECV)

Model \ Thresholds	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4
L-G-DCNN-V1	426	149	81	51	40	35	23	18
L-G-DCNN	598	241	127	73	47	34	29	21
L-ATT-RES-DCNN	709	374	192	107	67	42	32	27
LDCNN	696	338	165	93	58	39	31	24
DCNN	748	435	224	132	99	68	49	33
DCNN-K1	910	615	453	314	237	179	134	112
LSTM	376	147	80	48	31	27	16	14

Figure 2 illustrates the process of obtaining EFM and how, based on this, the CECVs of all models before and after fusion under different thresholds are obtained, thereby achieving rational model design guided by CECV.

DL models can effectively capture the relationship (f) between the composition and structure of compounds and their properties. However, these models are often considered as black boxes, making it challenging to rationally design DL models. If we view the process of fine-tuning the model based on evaluation feedback as an approximation of the function (f), traditional model evaluation metrics such as mean absolute error (MAE) and root mean square Error (RMSE) only provide statistical analyses of the model's performance on the test set. They lack the ability to offer more physical insights and limited insights into how the model learns from materials data and understands physical laws. In order to address these limitations, we need to innovate the method of model evaluation and obtain feedback that contains rich physical insights. We can consider the relationship between the CECV of compounds at different thresholds and their properties as a function (f'). Discovering the function (f') is equivalent to finding the target function (f). Thus, if our fine-tuned model is closer to the target function (f), the CECV of compounds at different thresholds will be more accurate.

Next, we will explain the acquisition of Element Feature Matrices (EFM) and CECV, along

with the process guiding the iterative design of the model, as depicted in Figure 2:

1. Selection and initialization of model components: As shown in Figure 2, step1, elect model components from the Common Model Library, including global models (such as LSTM, GRU, RNN) and local models (such as deep convolutional neural network (DCNN), deep convolutional neural network with a 1x1 kernel size (DCNN-K1), CNN, TextCNN[13], ResNet[14]), as well as attention and pooling components[15]. Initialize the composition model, such as the fusion of LSTM and DCNN (LDCNN), and set linear and activation layers for post-processing.

2. Model training and testing: As shown in Figure 2, step2, based on the OQMD enthalpy dataset, use the selected model components (such as LSTM, DCNN-K1, DCNN, LDCNN) for training and validation. Divide the dataset into training, validation, and test sets in a 70/15/15 ratio, and train all models under unified parameter settings.

3. Extraction of compound EFM: As shown in Figure 2, step3, for 2374 compounds containing silicon elements in the test dataset of OQMD enthalpy (covering binary to heptenary compounds), use the trained models (LSTM, DCNN-K1, DCNN, LDCNN) to extract the EFM of the compounds. Each compound corresponds to a 512-dimensional EFM. It is worth noting that we only use compounds containing silicon elements as an example here, but other element compounds can also be set, or compounds with different elements can be randomly selected.

4. CECV generation: As shown in Figure 2, step4, based on the extracted EFM, cluster all Si-containing compounds, set different clustering threshold ranges as [0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4], and then generate a set of CECV, recording the number of clusters corresponding to each threshold under different models. For example, LSTM obtains the clustering vector within the set threshold: [376, 147, 80, 48, 31, 27, 16, 14].

5. Visualization and definition of target region: As shown in Figure 2, step5, plot the graph, with the clustering threshold on the x axis and the number of CECVs obtained by each model on the y axis, showing the performance of different models under different clustering thresholds. Mark the LSTM and DCNN-K1 as global and local reference model, defining the target region between the LSTM and DCNN-K1. This step helps to intuitively understand the tendencies and trade-offs of the model in global and local feature extraction.

6. Iterative optimization of model components: As shown in Figure 2, step6, identify the deviation between the current model's CECV and the target region, and decide whether to strengthen the weight of the global feature extractor or the local feature extractor weight, guiding targeted adjustments of model components from the model library to make the CECV of the new model closer to the target region, achieving rational optimization of model design.
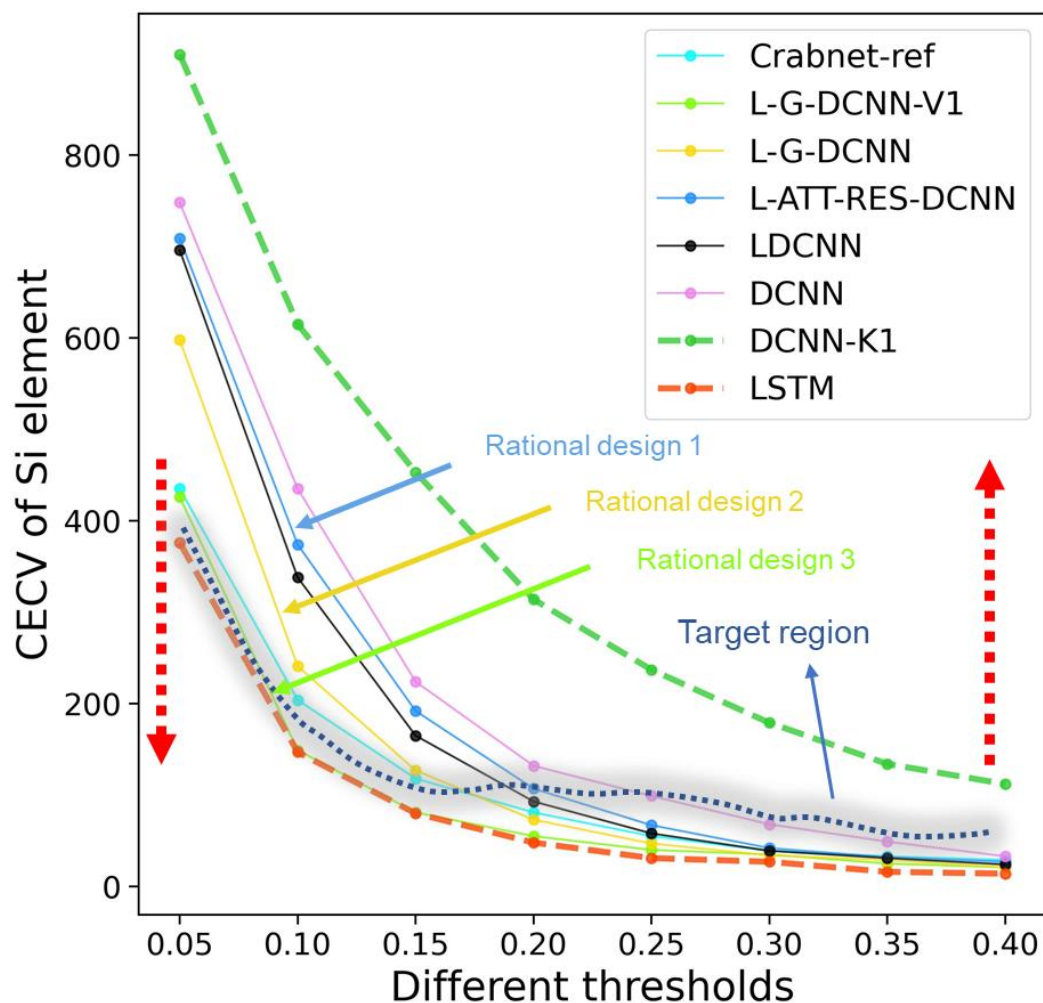
Figure 3 shows the CECV of silicon-containing compounds the OQMD Formation enthalpy test datasets based on the fusion strategy, using LSTM and DCNN-K1 as global and local reference models. CECV were used as the feedback method to achieve rational model design. The shaded region in the Figure marks the CECV of the optimal fusion model. The blue, yellow, and green solid arrows mark the rational design from LDCNN to L-G-DCNN-V1. Two red dashed arrows represent the rational design direction of the model at the LSTM and DCNN-K1 ends. Crabnet-ref is a reference model, which performs best on the OQMD Formation enthalpy.

We selected 10 datasets with a size greater than 20,000 from 18 extended datasets (Supplementary Figure S1) from different sources, including AFLOW[16], Material Projects[17] (MP), OQMD[18], and experimental datasets[19], as the benchmark dataset to ensure the robustness of the test results. The compounds in the dataset have different lengths, and the goal of DL models is to accurately capture the interactions between elements within compounds, including global and local interactions, based on the embedded representations of element and stoichiometry information. For the global interactions between elements in compounds, we provide three options: LSTM[8], GRU[20], and Recurrent Neural Network (RNN)[21]. They can all capture long-range interactions between elements. LSTM and GRU introduce gating mechanisms to better solve the problem of gradient vanishing or explosion in RNN, and can not only sensitively respond to short-range information correlations, but also capture long-range correlation information. Next, we tested three candidate models on 10 benchmark datasets, as shown in Supplementary Figure S2(a). Obviously, LSTM and GRU performed better than RNN in terms of overall MAE results, especially on datasets at the 10^5

level. Therefore, this paper initially chooses LSTM as the candidate global feature extractor. Additionally, we extract internal vector representations of all 51,242 compounds in the OQMD Bandgap test datasets from the last layer of LSTM and RNN respectively, use Uniform Manifold Approximation and Projection (UMAP)[22] for dimensionality reduction, and visualize compounds to validate the above conclusion, that RNN is not good at capturing long-range interactions and LSTM performs better. See supplementary Figure S3(a) and (b) for details.

In order to achieve more finer local interaction between elements, we designed several convolutional neural networks as shown in supplementary Figure S2 (b). They are respectively DCNN[23], where the red box represents the window, the kernel size and padding parameters are 3 and 1, and the pooling layer does not perform downsampling[24], so DCNN covers the embedding regions of multiple elements to achieve local information extraction; DCNN-K1, where the kernel size and padding parameters are 1 and 0, and the pooling layer does not perform downsampling, Therefore, DCNN-K1 is a special case of DCNN that minimizes the local coverage ability of DCNN; Deep Pyramid Convolutional Neural Networks (DPCNN)[10], where the kernel size and padding parameters are 3 and 1, and the alternating use of pooling layers with a stride of 2 for downsampling[25] results in a pyramid-shaped reduction in internal data size for each layer. Therefore, DCNN can effectively cover more element regions to represent longer element correlations in compounds. However, downsampling will lose some of the original input information; Convolutional Neural Networks[26](CNN), where the kernel size and padding parameters are 3 and 1, respectively, and no downsampling was performed in the pooling layer. The CNN only transformed the original compound embedding into a localized embedding representation covering multiple elements, but the main difference was that it lacked depth.

We tested four candidate models on 10 benchmark datasets, as shown in supplementary table S1. Obviously, DCNN, DPCNN and CNN performed better than DCNN-K1 in terms of overall MAE results. Therefore, we discarded DCNN-K1 and selected DCNN, DPCNN, and CNN as candidate local feature extractors. However, through visual analysis comparing DNN and DCNN-K1, it was found that DCNN-K1 was entirely incapable of capturing any interactions between elements within compounds, focusing solely on individual elements. This can be observed by comparing its output EFM with the EFM at the input layer. Their UMAP two-dimensional projections are nearly identical. Thus, we chose DCNN-K1 as the reference local model. For more detailed visual analysis of DCNN and DCNN-K1, see Supplementary Figure S2(c) and Figure s S3(c) and (d). Next, we heuristically fused them with LSTM to obtain LCNN, LDPCNN, and LDCNN. We evaluated their MAE results on the benchmark test set, as shown in Supplementary Figure S2 (d). The results clearly indicate that LDCNN has a smaller MAE value, thus determining it as the initial fusion architecture. For the optimization of convolution kernel hyperparameters for LDCNN, please refer to Supplementary Table S2.

The focus of this paper is on how to improve LDCNN rationally. Generally, we introduce commonly used model components, such as residual networks[14], pooling, and attention networks[27], heuristically, and observe their performance on the test dataset as the evaluation metric to achieve network design. This conventional method is feasible, but it is very inefficient, lacks explanation and rational guidance. Based on the fusion strategy, this paper designs discrete models with global and local components LSTM and DCNN-K1 as references, and converts the model evaluation method into a cluster problem of the chemical environment of the constituent elements in compounds at different thresholds, achieving rational model design. Here, we extracted the EFM of

the learned compounds from the model and computed the similarity between EFM at different thresholds. The resulting Pearson correlation matrix was used to generate CECV for compounds at different thresholds[28]. Thresholds is defined as the tolerance for distinguishing the chemical environment of compounds, with default values ranging from 0.05 to 0.4 in increments of 0.05.

As shown in Figure 3, two discrete reference models, represented by red and green dashed lines, respectively, show the CECV of silicon-containing compounds in the OQMD Formation enthalpy dataset at different thresholds for LSTM and DCNN-K1. Based on the characteristics of LSTM and DCNN-K1, they have different requirements for CECV in different thresholds. For the global feature extractor LSTM, its performance approaches the optimal solution within the thresholds<0.15 range, because the smaller the thresholds, the lower the tolerance for distinguishing element chemical environments, and the higher the requirement. If the model focuses too much on local details, it is easy to overestimate CECV, so the weight of the global extractor needs to be increased to adjust the model and achieve more accurate identification of element chemical environments, as shown by the red downward dashed arrow on the left side of Figure 3. Similarly, for the local feature extractor DCNN-K1, its performance approaches the optimal solution within the thresholds>0.15 range, because the larger the thresholds, the higher the tolerance for distinguishing element chemical environments, and the lower the requirement. If the model focuses too much on global information, it is easy to underestimate CECV. Therefore, the weight of the local extractor needs to be increased to adjust the model, as shown by the red upward dashed arrow on the right side of Figure 3. Based on the above strategies, we can draw the fusion target region, which is marked by the shaded region in the Figure 3.

Figure 3 shows that the black line in LDCNN achieves better fusion compared to the two discrete reference models. LDCNN has significantly reduced CECV compared to DCNN-K1 within the thresholds range <0.15, very close to the shadow region. However, within the thresholds range >0.15, its improvement in CECV is limited, with much deviations from the shadow region, resulting in the average performance of LDCNN. Based on the guidance of Figure 3, we identified that the key of the problem lies in adding a model component that can increase the weight of the LSTM part of LDCNN. Therefore, a rational design 1 was proposed, introducing soft attention and residual connections to optimize the LSTM part of LDCNN, resulting in L-ATT-RES-DCNN as shown by the cyan arrows in Figure 3. However, L-ATT-RES-DCNN achieved an increase in CECV throughout the entire thresholds range, indicating that the introduction of soft attention and residual connections in the LSTM part did not effectively increase the weight of the LSTM part. Therefore, a rational design 2 was proposed, introducing stacked residual GRU and softmax attention[29] to obtain L-G-DCNN. This design significantly reduced CECV in the thresholds range <0.15 and greatly improved the performance of L-G-DCNN, as shown by the yellow arrows in Figure 3. We can see that L-G-DCNN still has room for improvement within the range of thresholds <0.15. Therefore, a rational design 3, L-G-DCNN-V1, was introduced in L-G-DCNN by incorporating residual connections and max pooling, as indicated by the green arrow in Figure 3. L-G-DCNN-V1 continues to reduce CECV within the range of thresholds <0.15, entering the target shadow region and achieving optimization. However, there is also a decline in CECV within the range of thresholds >0.15. Overall, the target region has not yet been reached. From LDCNN to L-G-DCNN-V1, based on the CECV, we have improved the performance of the model through three-step design, achieving a rational and efficient design.

Finally, we tested LDCNN, L-G-DCNN, and L-G-DCNN-V1 on the test set of OQMD

Formation enthalpy, and obtained the MAE results as shown in Supplementary Figure S9 (d), (e) and (f), which are 0.0349, 0.0332 and 0.0317, respectively. This is consistent with the evaluation results of CECV, confirming the correctness of the CECV method. In the Supplementary Table S3, we obtained the CECV of silicon-containing compounds on five datasets, which is consistent with the results of MAE, further demonstrating the stability of the CECV method. Throughout the process of rational design, the evolutionary details of the fusion model architecture can be found in Supplementary Figure S4.

Aside from learning about EFM in materials science, DL methods can also associate EFM with specific material properties. A useful aspect for chemistry and materials science applications is its interpretability and reproducibility of known chemical intuition. We extracted the EFM of 89 elements from the last layer of the model based on the OQMD formation enthalpy dataset. Pearson correlations between the EFM and the elements are shown in Supplementary Figure S9. Combining the results from Figure 3 and Figure S9, we observe an improvement in the model's MAE with the fusion of the model under CECV guidance. At the same time, the Pearson correlation between EFM correctly recovers the periodicity of elements in the periodic table.

Properties	L-G-DCNN	L-G-DCNN-V1	Finder-ref	Roost-ref	Crabnet-ref	RF-ref
AFLOW Bulk modulus	**8.221**	8.317	8.598	8.820	8.692	11.907
AFLOW Debye temperature	**32.210**	33.030	35.259	37.167	33.646	36.484
AFLOW Shear modulus	**8.946**	9.110	9.528	9.983	9.082	10.094
AFLOW Thermal conductivity	**2.197**	2.259	2.376	2.703	2.318	2.658
AFLOW Thermal expansion	**3.61e-06**	3.728e-06	3.679e-06	3.69e-06	3.85e-06	5.44e-06
AFLOW Band gap	0.310	**0.291**	0.322	0.337	0.301	0.384
AFLOW Energy per atom	0.090	0.089	**0.086**	0.086	0.093	0.224
Bartel Decomposition enthalpy	0.062	**0.059**	0.062	0.067	0.063	0.076
Bartel Formation enthalpy	0.057	0.058	**0.054**	0.055	0.059	0.100
MP Energy above the convex hull	**0.083**	0.085	0.092	0.094	0.089	0.126
MP Magnetization of the unit cell	**2.002**	2.104	2.197	2.507	2.105	2.732
MP Bulk modulus	10.612	10.850	**10.514**	11.395	11.209	14.358
MP Shear modulus	**11.701**	11.770	13.320	12.797	12.787	12.777
MP Elastic anisotropy	**8.018**	8.260	8.653	8.082	8.263	11.691
OQMD Band gap	**0.041**	0.043	0.055	0.088	0.049	0.060
OQMD Energy per atom	0.035	**0.032**	0.036	0.032	0.033	0.141
OQMD Formation enthalpy	0.033	0.032	0.033	0.032	**0.031**	0.083
OQMD Volume per atom	0.299	0.279	0.031	0.296	**0.277**	0.544

Figure 4 compares the MAE values of the L-G-DCNN series models with the reference models Roost, Crabnet, and Finder on the Extended dataset. Cells are colored based on the relative MAE within each row (blue indicating better performance, red indicating worse performance). The best MAE values are highlighted in bold.

We found that currently, models that only utilize compositional information without manual feature engineering, in chronological order, mainly include Roost[11] (2020), Crabnet[12] (2021), and Finder[6] (2022). Consequently, we have selected Crabnet and Finder, which are based on GNN and transformer architectures respectively, to serve as our baseline models for comparative analysis. We train them using the default model parameters provided by their respective repositories. As representatives of traditional machine learning, we use the Magpie-featurized[30] RF model as the baseline model. In this section, we evaluate the structure-agnostic L-G-DCNN series models (L-G-DCNN and L-G-DCNN-V1), and compare them with Roost, Crabnet, Finder and the RF.

Figure 4 displays the MAE scores of all models on the held-out test dataset, indicating that L-G-DCNN outperforms most benchmark tasks. Furthermore, the testing results of L-G-DCNN-V1 show improvement compared to L-G-DCNN in larger datasets and perform similarly to L-G-DCNN in smaller datasets. This once again reaffirms the significance of our rational guidance strategy. It is

worth noting that, based on the guidance of CECV strategy, to further reduce the CECV on the LSTM end while increasing the CECV on the DCNN-K1 end, we modified the attention component in L-G-DCNN and introduced self-attention mechanism to obtain L-G-DCNN-V2, which further improved the test results, especially surpassing Roost and Finder, reaching the level of Crabnet in large datasets, as shown in Supplementary Figure S10. Supplementary Figure S11 presents the CECV for the silicon-containing compounds across different thresholds of L-G-DCNN-V2 and Crabnet, based on the OQMD formation enthalpy dataset. CECV explain the the performance improvement of L-G-DCNN-V2. Detailed data for Figure S11 can be found in Supplementary Table S4. Considering the different architectures and modeling philosophies of Roost, Crabnet, Finder and L-G-DCNN, it is encouraging to see that all methods converge to the same performance. This further confirms that rational guidance for model design based on CECV maximizes complementarity and improves fusion between discrete models, thereby enhancing our understanding of the model design. The training progress of L-G-DCNN and Crabnet is compared in Supplementary Figure S12.

Structure-agnostic Model Evaluation
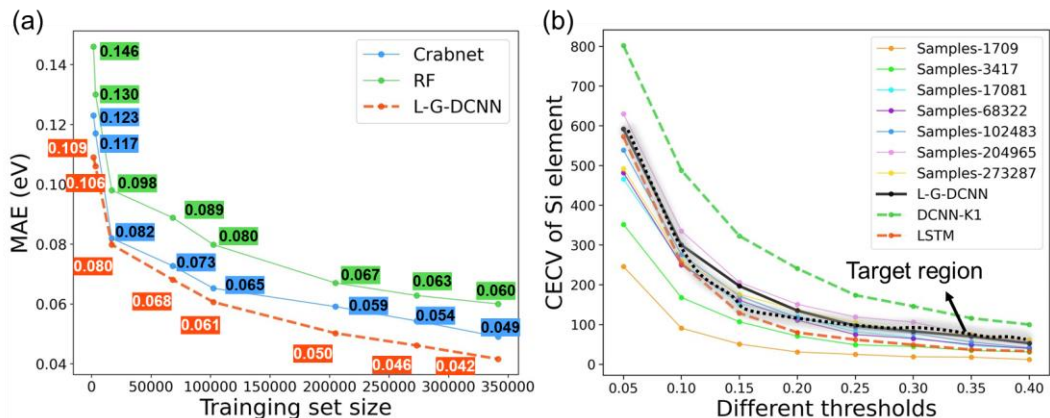
Figure 5 shows the sample efficiency of the model-agnostic models evaluated on the OQMD Band gap dataset. (a), L-G-DCNN outperforms Crabnet and RF and exhibits a steeper gradient descent. (b) shows the performance of the CECV of silicon-containing compounds across different thresholds for models trained on different training set sizes on the OQMD Band gap dataset. The best performing model, L-G-DCNN, is highlighted with a bold black line.

Material data, particularly experimental measurements, are often limited in size. This raises concerns about the ability of DL models in general to learn from significantly undersampled datasets and still provide reasonably accurate out-of-database predictions. The sample efficiency of L-G-DCNN was evaluated by examining its performance at different training set sizes. Figure 5(a) shows the prediction MAE curve of the structure-agnostic model on the OQMD bandgap test set. L-G-DCNN achieved the lowest error scores across all training set levels (from 10^3 to $\approx 3 \times 10^5$). Traditional ML models that use interpretable features, such as RF, typically perform well on small datasets. Although inheriting the DL paradigm, our model still outperforms RF when the training set size reaches $10^2 \sim 10^3$. Furthermore, the MAE curve of L-G-DCNN exhibits a steeper gradient descent compared to Crabnet.

Generally, the main factors that impact model performance are the model architecture and data size. Our CECV method explains the influence of data size on model performance. Figure 5(b)

illustrates that, with a fixed model architecture, as the training set size increases, the CECV of the model converges towards the target region, which matches the traditional MAE statistical evaluation results shown in Figure 5(a).
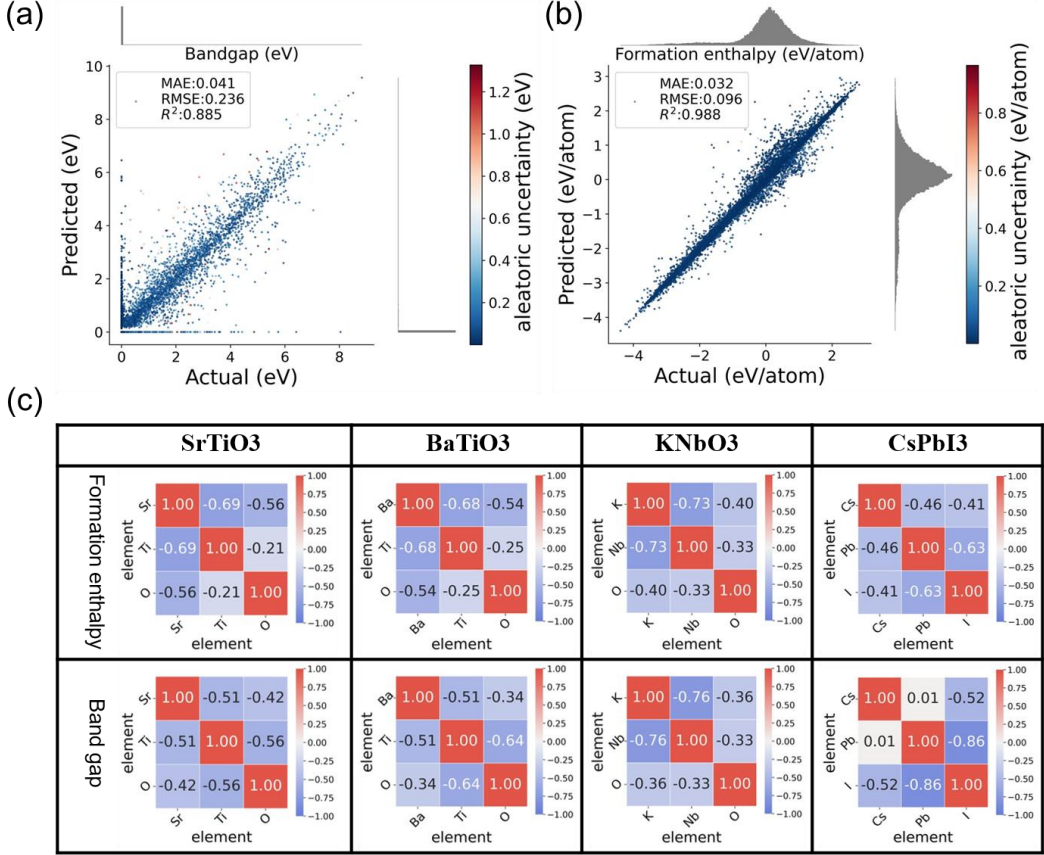

Figure 6. Parity plots for the (a) OQMD band gap and (b) OQMD formation enthalpy of the test set for the structure-agnostic L-G-DCNN. The L-G-DCNN exhibits good predictive error and uncertainty. The test data and predicted marginal distributions are shown on the minor axis. RMSE-root-mean-square error, R^2 score-determination coefficient. (c) EFM of perovskite materials $SrTiO_3$, $BaTiO_3$, $KNbO_3$, and $CsPbI_3$ visualized by probing OQMD formation enthalpy and OQMD band gap L-G-DCNN models.

In the parity plots of OQMD band gap and OQMD Formation enthalpy shown in Figure 6(a) and (b), L-G-DCNN achieved good prediction results. For band gap, most of the values in the dataset are distributed around 0. The RMSE is 0.236, much larger than the 0.096 for Formation enthalpy, indicating a higher deviation from the true values in the prediction of band gap, especially in the range close to 0. For Formation enthalpy, the prediction results are better when the target value is negative. This is because the computational materials database tends to report more stable materials, which usually have negative values. As expected, the aleatoric uncertainty of relatively inaccurate predictions is higher than the uncertainty of samples close to the perfect prediction.

We investigate whether the EFM indeed captures some chemical insights that are not explicitly input into the model. Figure 6(c) presents the EFM of four well-known perovskites, obtained from the final layer of the trained OQMD Formation enthalpy and OQMD band gap models, respectively. From the OQMD Formation enthalpy model, we find that compositionally and structurally similar materials such as $SrTiO_3$ and $BaTiO_3$ have similar EFM. Interestingly, the EFM of $KNbO_3$, which has a different composition but a similar structure, is comparable to the EFMs of the aforementioned materials. The EFM of halide perovskite $CsPbI_3$ is considerably different from those of its oxide

counterparts. A consistent trend is observed from the OQMD band gap model, but different EFM are obtained for the same material. Obviously, EFM entries are determined by the constituent element types and the context. Incorporating EFM that reflect interactions between elements adds another dimension for materials similarity analysis. Parity plots for other properties and comparison of L-G-DCNN and Crabnet on the OQMD band gap test datasets are shown in Supplementary Figure S13.

Evaluation On Matbench Datasets

Property	L-G-DCNN	L-G-DCNN-V1	Crabnet-ref	Finder-ref	Dataset size
Dielectric	**0.310**	0.313	0.323	0.3204	4,764
Exp band gap	**0.323**	0.329	0.346	*	4,604
Jdft2d	**44.280**	44.437	45.610	47.961	636
Log gvrh	**0.089**	0.093	0.101	0.099	10,987
Log kvrh	**0.069**	0.071	0.076	0.076	10,987
MP formation enthapy	0.081	**0.078**	0.086	0.084	132,752
MP band gap	0.239	0.248	0.265	**0.230**	106,113
Perovskites	0.401	**0.390**	0.407	0.645	18,928
Phonons	47.615	49.409	55.111	**46.575**	1,265
Yield strength	**90.577**	92.661	107.316	*	312

Table 1. Performance comparison on the Matbench suite. The best performing models in each domain are indicated in bold. Crabnet and Finder's performance metrics are reported in Matbench leaderboard at https://matbench.materialsproject.org/ . It should be noted that the datasets are used as-is (e.g. preprocessing such as removing duplicate compositions and outliers have not been applied) for consistent comparison[31]. *jdft2d* - exfoliation energy (meV/atom);[32] *phonons* - phonon DOS peak frequency;[33] *dielectric* - refractive index (unitless) ;[34] *log gvrh* - base 10 logarithm of the DFT Voigt-Reuss-Hill average shear moduli in GPa;[35] *log kvrh* - base 10 logarithm of the DFT Voigt-Reuss-Hill average bulk moduli in GPa;[35] *perovskites* - perovskite formation energy, in eV per atom;[36] *MP band gap* - The band gap as calculated by PBE DFT from the Materials Project, in eV;[37] *MP formation enthapy* -Formation energy in eV as calculated by the Materials Project, in eV per atom;[37] *Exp band gap* - experimentally measured gap, in eV;[38] Yield strength -experimentally measured steel yield strengths, in MPa[17]

Matbench serves as a common test dataset for ML models and includes 13 materials property prediction tasks. Here, we select 10 regression tasks. Two recent structure-agnostic models Finder and Crabnet are selected for comparison. We follow the fivefold cross validation strategy with the same random seed variable recommended in the original study to evaluate our algorithm[31]. Fivefold cross validation conducts an outer test loop with 20% test data and 80% training + validation data, and an inner validation process for model weight parameter tuning for each outer fold. L-G-DCNN divides the overall fold into 70% training data, 10% validation data, and 20% test data.

Table 1 shows the performance comparison on the Matbench suite. L-G-DCNN and L-G-DCNN-V1 achieve the best MAE scores in eight out of 10 structure-agnostic tasks. Our model outperforms Crabnet and Finder in small datasets ($<10^4$) and large datasets ($>10^5$).

Ablation Study

Benchmark properties	DCNN layers in L-G-DCNN.			Model1	Model2	Model3	Model4	Model5	Model6
	2	4	6						
AFLOW Energy per atom	**0.0937**	0.0942	0.0939	0.0941	0.0948	0.183	0.0988	0.0940	0.0952
Bartel Formation enthalpy	**0.0582**	0.0583	0.0592	0.0601	0.0594	0.0794	0.0594	0.0588	0.0603
MP Energy above hull	**0.0851**	0.0878	0.0861	0.0918	0.0898	0.0949	0.0913	0.0860	0.0861
OQMD Formation enthalpy	**0.0332**	0.323	0.0329	0.0331	0.0339	0.0385	0.0353	0.0310	0.035

Table 2. Four intricately designed datasets MAEs of different model architectures considered in ablation study. The reference model (L-G-DCNN-2B) is indicated in bold. Other models differ from the default architecture of L-G-DCNN as follows. Model 1 - uses one-hot node embeddings; Model 2 - post-processing network added; Model 3 - all residual connections removed; Model 4 - only the residual connections coming from region embedding layer removed; Model 5 - soft-attention component after the LSTM and GRU layers replaced with a self-attention mechanism. Model 6 - soft-attention and GRU components removed from the networks, resulting in the LDCNN model.

To comprehensively and accurately validate the performance of different components of the model in ablation experiments on different data sources and volumes, we carefully selected four datasets: AFLOW Energy per atom, MP Energy above convex hull, OQMD Formation enthalpy, and Bartel Formation enthalpy. These datasets are from AFLOW, MP, OQMD calculations[39], and experimental data sources, with dataset sizes ranging from 10^4 to 10^5. Table 2 shows the results of ablation study on the four datasets. Although the depth of convolutional neural networks does not significantly affect the overall performance of structure-agnostic models, a 4-layer DCNN compared to the default 2-layer DCNN can achieve the MAE decrease on large datasets (level 10^5). We investigate one-hot element embeddings and observe that MAE remains unchanged on the 10^5 level dataset but significantly increases on the other 10^4 level datasets (see Model 1 in Table 2). This indicates that element embedding capturing prior knowledge still help navigate to a lower minimum in the error surface, although this accuracy gap is expected to narrow as the database size grows. Generally, post-processing neural networks are considered as an important component of the model, such as Crabnet and Finder, which both include designed post-processing networks. However, L-G-DCNN does not include a specifically designed post-processing network, but simply incorporates a fully connected network into its output to reshape it as (batch size, sequence length, 3). In Model 2, we added a post-processing neural network, which includes four linear layers, residual connections, LeakyReLU activation function[40], and an output linear layer. We observed that adding a post-processing network impairs performance.

In Model 3, we remove all residual connections and observe a substantial error increase in structure-agnostic models. We then remove only the residual connections from region embedding layers and keep the rest of the residual connections to obtain Model 4. Compared to removing all residual connections, the error of the Model 4 is reduced. This indicates that residual components can achieve lower errors in the fusion architecture. In Model 5, we replace the soft-attention component after the LSTM‖GRU with a self-attention component. Compared to the reference model, Model 5 shows a significant improvement in MAE for the OQMD Formation enthalpy dataset (0.0310), while performance decreases in other datasets. Model 6 is obtained by removing soft-attention and GRU components. The MAE increases significantly, highlighting the importance of designing

LSTM‖GRU combinations with soft attention. In conclusion, the residual components, global feature extractor with attention, and number of DCNN layers are key factors determining model performance.
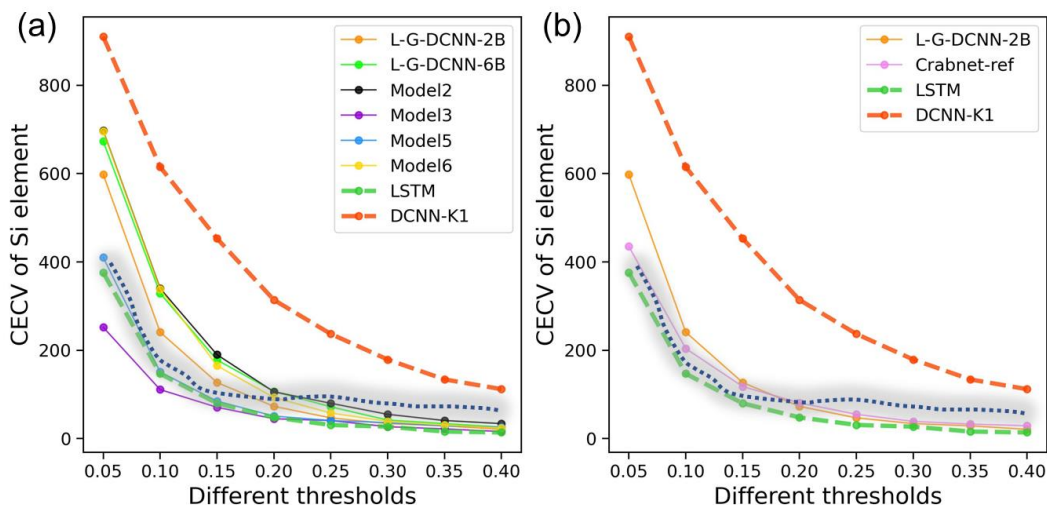

Figure 7 (a) CECV distributions of silicon-containing compounds in different thresholds for Model2, Model3, Model5, Model6, L-G-DCNN2B, and L-G-DCNN6B on the OQMD Formation enthalpy test datasets, where LSTM and DCNN-K1 are fusion reference models. (b) CECV distributions of silicon-containing compounds in different thresholds for the default model L-G-DCNN-2B and Crabnet.

To further understand how different components of the model in the ablation experiment affect the changes in CECV, we plotted the CECV of the silicon-containing compounds at different thresholds on the OQMD Formation enthalpy dataset for the above part of the model, as shown in Figure 7. The raw data for Figure 7 is provided in Supplementary Table S5. First, based on Figure 7(a), we analyzed the effect of increasing the number of DCNN layers in the reference model. By observing the changes in CECV of L-G-DCNN-2B and L-G-DCNN-6B in the entire thresholds range, it can be seen that the CECV of L-G-DCNN-6B is significantly increase compared to L-G-DCNN-2B. This may be because the increase in the number of DCNN layers caused the local ability of the model to be excessively corrected. The situation of Model2 is similar to that of L-G-DCNN-6B, both of which have excessively enhanced local ability on the DCNN blocks. Model6 is different, as it adjusts the GRU components on the LSTM side, directly reducing the global ability of the model, which in turn leads to a decrease in CECV in the smaller thresholds range. Note that the local ability of Model 6 has not been significantly enhanced, so its CECV is relatively smaller than L-G-DCNN-6B and Model2 in the larger thresholds range. Model3 removes all residual components, greatly reducing the fusion ability of the model, so its CECV directly exceeds the reference fusion model. Model5 enters the target shadow region in the smaller thresholds range, so its performance is greatly improved, and the MAE value reaches 0.0310, which is very close to Crabnet. This indicates that self-attention has a significant effect on OQMD Formation enthalpy. In Figure 7(b), we compared L-G-DCNN with the Crabnet. Overall, Crabnet's CECV is closer to the target shadow region, indicating better MAE performance. Based on the above analysis, we can see that the CECV evaluation method is stable and provides more feedback information than traditional statistical metrics, which helps in the rational design and understanding of the model.

Visualizing The Understanding Of Model Fusion Process

The Contribution of Elements in Property Prediction as a Function of Composition Before and After Fusion
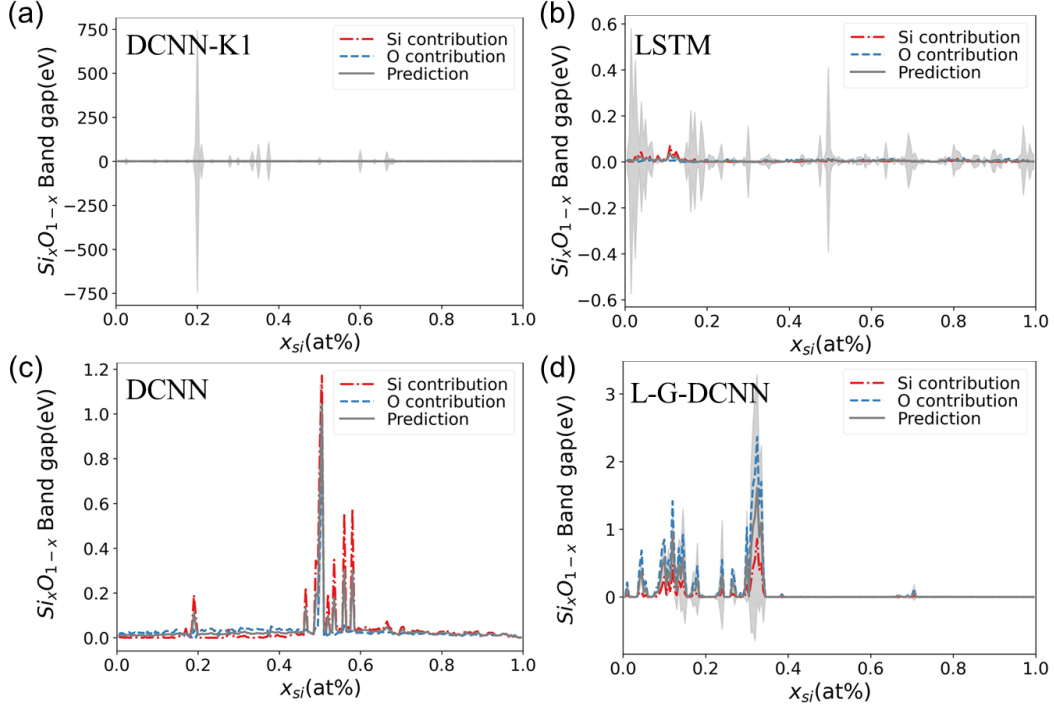

Figure 8 illustrates the prediction results for the Si_xO_{1-x} system from both the discrete models (LSTM, DCNN-K1, and DCNN) and the fusion model (L-G-DCNN), highlighting the contribution of elements as constituents to property predictions before and after model fusion.

For the band gap prediction of the Si_xO_{1-x} system across the entire stoichiometry range, the results of the discrete models (LSTM, DCNN-K1, and DCNN) and the fusion model (L-G-DCNN) are presented in Figure 8. Several observations can be made: Firstly, with the fusion of models, the uncertainty gradually decreases. Specifically, the discrete reference models LSTM and DCNN-K1 deviate significantly from the DFT computational results, with DCNN predicting a band gap peak within 0.4 to 0.6, while the DFT results fall within 0.2 to 0.4, overestimating the band gap values. Secondly, with the fusion of models, the overall prediction results gradually approach the DFT calculations. The DFT results can be found in Supplementary Figure S14. Lastly, Figure 8(a) and (b) effectively demonstrate the global and local characteristics of LSTM and DCNN-K1. In particular, the predictions of the reference model LSTM, acting as a global feature extractor, are notably lower than the actual values, while the predictions of the reference model DCNN-K1, serving as a local feature extractor, are significantly higher than the actual values. With the progress of fusion and appropriate adjustments to the global and local feature extractors, it is evident that the predictions of DCNN and L-G-DCNN show significant improvements, as shown in Figure 8(c) and (d).

To evaluate the prediction ability of L-G-DCNN for different material properties in diverse

chemical stoichiometry ranges of multicomponent compounds, we obtained 18 binary chemical systems AxO1-x and 19 ternary chemical systems $Li_xAl_{1-x-y}O_y$ from MP. The A elements were sampled according to different groups of the periodic table, such as metalloid, alkaline, alkali, post-transition metal, and transition metal groups. In the supplementary Figure S15 show the MAE distributions of L-G-DCNN and the Crabnet, respectively, for six different material properties (including Formation enthalpy, Band gap, Energy above convex hull, Magnetic moment, Shear modulus, and Bulk modulus) across the diverse stoichiometry ranges of the binary and ternary chemical systems. It is observed that L-G-DCNN outperforms Crabnet for all properties except bulk modulus. For raw data, please refer to Supplementary Table S6.

Comparison of the Distribution of Overall Element Contributions Differences in Property Prediction Before and After Fusion
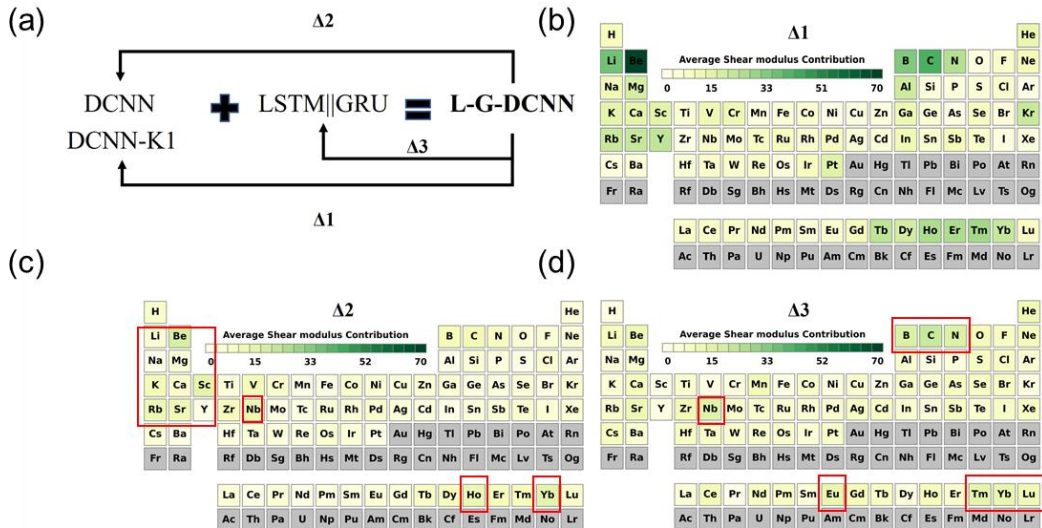
Figure 9 (a) shows a schematic of the fused discrete DCNN and LSTM model, denoted as L-G-DCNN, where DCNN-K1 serves as the reference model for local feature extraction. Δ1, Δ2, and Δ3 represent the average contribution differences of all elements to the shear modulus predictions relative to L-G-DCNN, computed from the Materials Project (MP) shear modulus dataset for DCNN-K1, DCNN, and LSTM, respectively. As shown in (b), (c) and (d), on a periodic table, the darker colored elements contribute more towards the shear modulus value of a compound relative to L-G-DCNN.

Similar to Crabnet, L-G-DCNN uses the vector representation of each element to directly predict the contribution of element pairs to property prediction. Figure 9 (a) shows the fusion process of the model, where Δ1, Δ2, and Δ3 represent (L-G-DCNN - DCNN-K1), (L-G-DCNN - DCNN), and (L-G-DCNN - LSTM), respectively. Figure 9 (b), (c), and (d) show the difference in the average contribution of each element of the discrete models trained on MP shear modulus relative to the L-G-DCNN model. The larger the difference, the darker the color. Δ1 has much darker color than Δ2 and Δ3, indicating that the element contribution estimated by DCNN-K1 is inaccurate. Δ2 and Δ3 are similar in color, with different distribution of element contributions, as highlighted by the red box in the Figure, with larger range of Δ2 and smaller range of Δ1, complementing each other. However, the complementary effect of some elements is not good, such as Nb and Yb. Please refer

to Supplementary Figure S16 for the average element contribution maps predicted by DCNN-K1, DCNN, LSTM, and L-G-DCNN. The visualization results intuitively indicate which elements are expected to have the greatest impact on the shear modulus properties of the compound. This exploratory data analysis based on fusion strategy helps us deeply understand the L-G-DCNN model and elemental behavior.

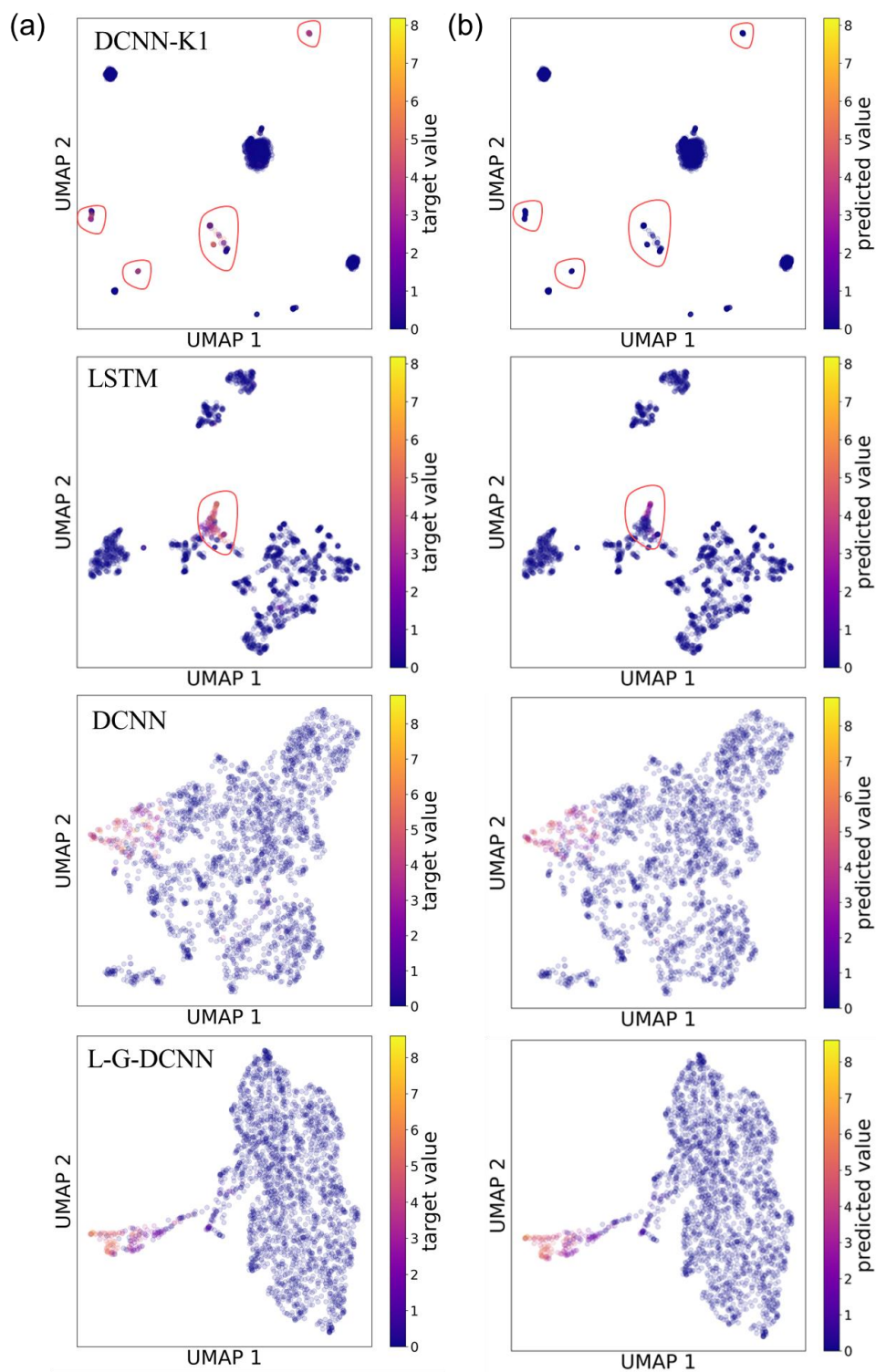**Visualization of the Evolution Process of Compound EFM Before and After Fusion**

Figure 10 (a) and (b) displays the two-dimensional projections of the EFM for silicon-containing compounds in 2374 different chemical environments, based on UMAP, using the OQMD band gap test dataset. Each point in the Figure is colored based on the predicted and computed band gap values. The points with significant differences between predicted and computed values in DCNN-K1 and LSTM are marked with red circles.

We extracted the EFM of compounds containing Si from the L-G-DCNN trained on the OQMD band gap dataset, which already encapsulate the learned feature representations of the L-G-DCNN. If the L-G-DCNN has learned the clusters of chemical environments containing Si compounds during training, these EFMs will reflect these clusters. By applying UMAP to these EFMs for dimensionality reduction, a two-dimensional visualization can be obtained, showcasing the relative positions of data points in the original high-dimensional space and indirectly revealing the clustering results. As shown in Figure 10, displays the two-dimensional projections of the EFM for silicon-containing compounds in 2374 different chemical environments, based on UMAP, using the OQMD band gap test dataset. Each point in the Figure is colored based on the predicted and DFT computed band gap values. As the model fusion, the predicted results on the OQMD band gap dataset become increasingly closer to the DFT values, with the red circles marking the inaccurate predictions of Si-containing compounds by the LSTM before fusion. Figure 10 demonstrates that with model fusion, the L-G-DCNN not only achieves accurate band gap predictions but also accurately captures the chemical environments of Si-containing compounds, effectively distinguishing between chemical environments with significant band gap differences.
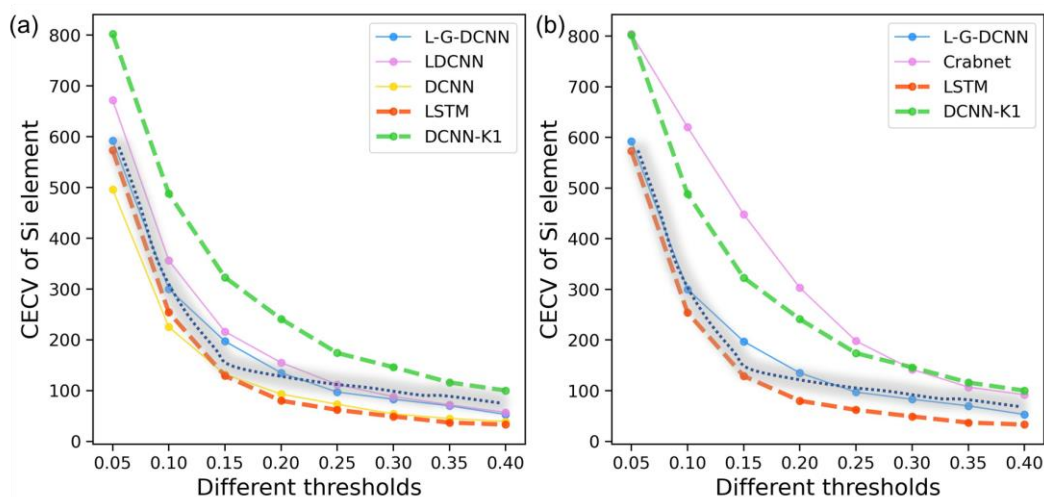
On the OQMD bandgap dataset, Figure 11(a) shows the CECV of silicon-containing compounds at different thresholds obtained by DCNN-K1, LSTM, DCNN, LDCNN, and L-G-DCNN. (b) CECV of silicon-containing compounds at different thresholds obtained by L-G-DCNN and Crabnet. The shaded region with dashed lines is the target reference region.

Figure 11 displays the CECV obtained based on the EFM of silicon-containing compounds at different thresholds for models. If the model can accurately identify the chemical environments of all silicon-containing compounds, the CECV of the model should be closest to the target region represented by the dashed shadow in Figure 11(a). We found that the L-G-DCNN represented by the blue line in the Figure performs better than the LDCNN. In addition, the MAE obtained by DCNN-K1 to L-G-DCNN on the OQMD Bandgap test set are 0.148, 0.115, 0.0531, 0.0428, and 0.0410, respectively, with L-G-DCNN producing the best results, which is consistent with the above conclusion. Finally, we compared the CECV at different thresholds of L-G-DCNN and Crabnet, as shown in Figure 11(b), and found that L-G-DCNN is closer to the target region, and performs better. For raw data, please refer to Supplementary Table S7.

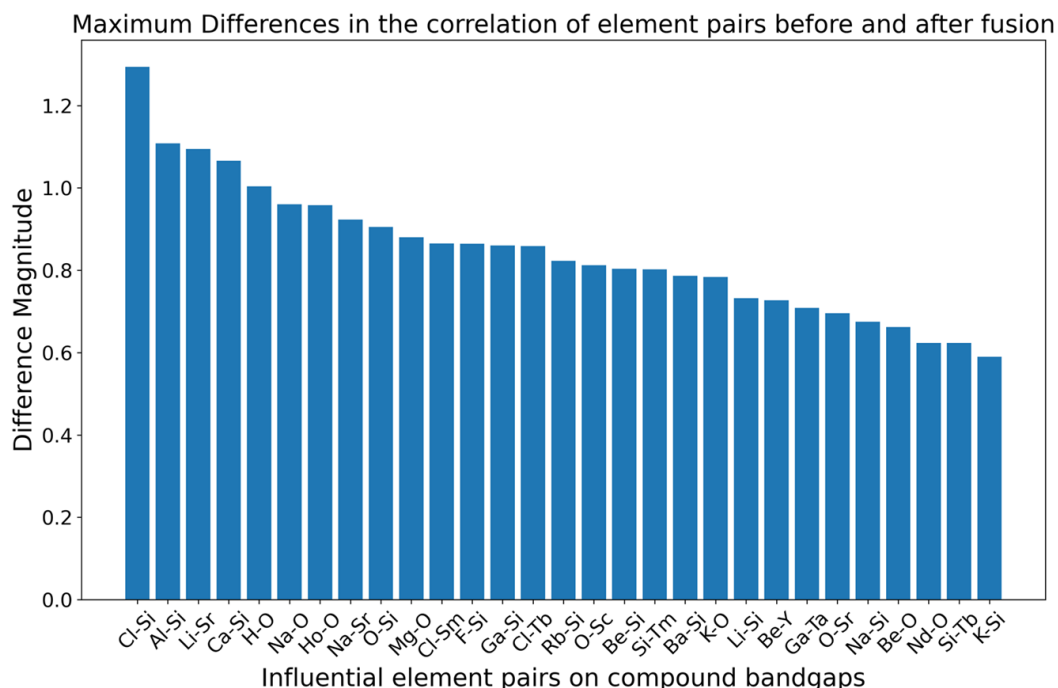**Identifying Critical Elemental Interactions Pairs for Band gap Prediction**

Figure 12 summarizes the most influential element interactions pairs on the band gap among the 56 compounds before and after model fusion. From the Figure, it can be observed that most element interactions pairs involve elements with high electronegativity such as O, Cl, and F.

In this paper, we have explored a range of visualization techniques to explain the relationship between model representations, chemical environments, and periodic trends. Our goal is to provide an intuitive way to understand the crucial factors that influence the material properties. Although these visualization techniques can only capture macroscopic changes in periodic trends and clustering, interpretability remains a pervasive challenge for deep learning models. Here, we provide some tentative explanations. We found that the L-G-DCNN based on the fusion strategy can be decomposed into LSTM and DCNN through a process similar to dismantling building blocks, thereby providing more information about the process of model performance improvement from "poor" to "good". For example, based on the OQMD band gap test dataset, we first filtered out 56 compounds containing Si elements that performed very poorly in the LSTM compared to the L-G-DCNN. Then, we compared the EFM plots of these 56 compounds based on LSTM and L-G-DCNN to obtain (LSTM) - (L-G-DCNN) ΔEFM. Through comparison, we can determine which elements interactions are most important for the band gap property. As depicted in Figure 12, we summarize the elements interaction that have the greatest impact on the band gap among these 56 compounds. It can be seen that most of the elements interaction pairs contain elements with high electronegativity such as oxygen (O), chlorine (Cl), and fluorine (F), which are very conducive to forming strong covalent bonds, and this has a significant impact on the band gap, which is consistent with our physical intuition, indicating that L-G-DCNN is a high effective deep learning model for predicting material properties from chemical composition data.

CONCLUSION

Designing deep learning models for materials is challenging due to the limitations of traditional statistical feedback methods. These methods cannot provide the rich physical insights like the material characterization techniques[41]. In this paper, based on the fusion strategy, we design a feedback method CECV enriched with physical insights to address the limitations of rational model design. The L-G-DCNN achieves a deep fusion of global (LSTM||GRU) and local (DCNN) feature extractors guided by CECV feedback. This enables us to accurately capture the important interactions between elements in compound systems. To evaluate the performance of L-G-DCNN, we compared it with models such as Roost, Crabnet, Finder and RF on a total of 28 datasets. Our results indicate that L-G-DCNN consistently achieves higher predictive accuracy. Furthermore, our model outperforms other models in terms of sample efficiency and learning efficiency. It particularly excels in small data learning tasks that are common in materials science. Additionally, through examining the contribution of elements in property prediction as a function of composition before and after fusion, comparing the distribution of overall element contributions differences in property prediction before and after fusion, and visualizing the evolution process of compound EFM before and after fusion, we have deepened our understanding of the model fusion process. Finally, by identifying critical elemental interaction pairs for band gap prediction, we have enhanced the comprehension of how interactions between elements significantly influence the properties of materials. This multi-faceted approach not only clarifies the understanding of model fusion but also illuminates the pivotal role of elemental interactions in material property predictions. The following points provide a specific description of how L-G-DCNN can further enhance this field:

1. Innovative framework for expanding deep learning model design: L-G-DCNN introduces a novel fusion architecture that combines the strengths of LSTM for global feature extraction and DCNN for local feature capture. This design differs from previous models like Roost, Finder, and CrabNet by offering a more holistic analysis of material properties. This shift from singular to comprehensive in model design promotes the deep application of deep learning models in the field of materials.

2. Transparent demonstration of model performance evolution: The modular design of L-G-DCNN enables a clear visual representation of the model's improvement, transitioning from individual LSTM and DCNN components to an fusion model. This transparent evolution not only demonstrates improved performance but also a more nuanced capture of elemental interactions. By contrasting LSTM and L-G-DCNN predictions, we can pinpoint critical elemental interactions that influence material properties, thereby deepening our comprehension of the relationship between element chemical environments and material properties, and facilitating the strategic design of novel materials.

3. Efficient and rational optimization strategy based on CECV: CECV introduces the chemical environment, offering a new method for optimizing surrogate models in small datasets. It enriches the physical basis of model design and guides rational model design through feedback mechanisms. For instance, In alloy component design, traditional methods like Gaussian Process (GP) surrogate models struggle with large search spaces and limited data, often leading to inaccurate RL reward calculations and poor generalizability. L-G-DCNN, after training on the same small dataset, can calculate the CECV of LSTM, DCNN, and L-G-DCNN, evaluating differences and knowing

specifically how to adjust, leading to a more rational and efficient fusion strategy. In an active learning framework, CECV steers model design and experiment selection, optimizing the use of limited resources and scarce data, accelerating the transformation from data to knowledge.

AUTHOR INFORMATION

Corresponding Authors

Hong Wang - *School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China;* Email: <u>hongwang2@sjtu.edu.cn</u>

Lanting Zhang - *School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*; Email: lantingzh@sjtu.edu.cn

Jian Hui - *School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*; Email: hj20151107@sjtu.edu.cn

Authors

Hongwei Du - *School of Materials Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240, China*

Author Contributions

Hongwei Du, Jian Hui and Hong Wang devised the idea for the paper. Hongwei Du implemented the idea and conducted the code design and visualizations. Hongwei Du, Jian Hui and Hong Wang interpreted the results and prepared the manuscript. All authors discussed the results.

Notes

The authors declare no competing financial interest.

ASSOCIATED CONTENT

The source code and data of this paper is available at: <u>https://github.com/dhw059/DeepModelFusion</u>.

Supporting Information

Distribution of training, validation, and test data in extended datasets and Matbench; explanation of the rational model design process; L-G-DCNN model parity plots and comparative analysis with Crabnet on OQMD band gap dataset; CECV analysis of silicon-containing compounds across thresholds in ablation study models on OQMD formation enthalpy dataset; comparative predictions and DFT validation of Si_xO_{1-x} band gaps by L-G-DCNN and Crabnet on OQMD data with elemental contribution and uncertainty analysis; elemental contributions to shear modulus predictions across the periodic table: insights from MP dataset analysis; silicon compound CECV assessment across DCNN-K1, LSTM, DCNN, LDCNN, L-G-DCNN, and Crabnet models on OQMD bandgap data; comparative parameter analysis of L-GDCNN, CrabNet, and Finder models from fusion, transformer, and GNN architectures.

ACKNOWLEDGMENTS

REFERENCES

(1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Physical Review* **1964**, *136* (3B), B864-B871. DOI: 10.1103/PhysRev.136.B864. Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, *140* (4A), A1133-A1138. DOI: 10.1103/PhysRev.140.A1133.

(2) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwaraknath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; et al. Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials* **2018**, *3* (5), 5-20. DOI: 10.1038/s41578-018-0005-z.

(3) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Physical Review Letters* **2015**, *114* (10), 105503. DOI: 10.1103/PhysRevLett.114.105503. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M. SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2018**, *2* (8), 083802. DOI: 10.1103/PhysRevMaterials.2.083802. Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; et al. Matminer: An open source toolkit for materials data mining. *Computational Materials Science* **2018**, *152*, 60-69. DOI: https://doi.org/10.1016/j.commatsci.2018.05.018.

(4) Kauwe, S. K.; Graser, J.; Vazquez, A.; Sparks, T. D. Machine Learning Prediction of Heat Capacity for Solid Inorganics. *Integrating Materials and Manufacturing Innovation* **2018**, *7* (2), 43-51. DOI: 10.1007/s40192-018-0108-9.

(5) Schütt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Müller, K. R. SchNet – A deep learning architecture for molecules and materials. *The Journal of Chemical Physics* **2018**, *148* (24), 241722. DOI: 10.1063/1.5019779 (acccessed 12/27/2023). Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters* **2018**, *120* (14), 145301. DOI: 10.1103/PhysRevLett.120.145301. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chemistry of Materials* **2019**, *31* (9), 3564-3572. DOI: 10.1021/acs.chemmater.9b01294. Willatt, M. J.; Musil, F.; Ceriotti, M. Atom-density representations for machine learning. *The Journal of Chemical Physics* **2019**, *150* (15), 154110. DOI: 10.1063/1.5090481 (acccessed 1/2/2024). Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Physical Review B* **2013**, *87* (18), 184115. DOI: 10.1103/PhysRevB.87.184115.

(6) Ihalage, A.; Hao, Y. Formula Graph Self-Attention Network for Representation-Domain Independent Materials Discovery. *Advanced Science* **2022**, *9* (18), 2200164. DOI: https://doi.org/10.1002/advs.202200164 (acccessed 2023/12/26).

(7) Graser, J.; Kauwe, S. K.; Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chemistry of Materials* **2018**, *30* (11), 3601-3612. DOI: 10.1021/acs.chemmater.7b05304. Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Brgoch, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *Journal of the American Chemical Society* **2018**, *140* (31), 9844-9853. DOI: 10.1021/jacs.8b02717. Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chemistry of Materials* **2010**, *22* (12), 3762-3767. DOI: 10.1021/cm100795d. Choudhary, K.; DeCost, B.; Tavazza, F. Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape. *Physical Review Materials* **2018**, *2* (8), 083801. DOI: 10.1103/PhysRevMaterials.2.083801. Gaultois, M. W.; Oliynyk, A. O.; Mar, A.; Sparks, T. D.; Mulholland, G. J.; Meredig, B. Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Materials* **2016**, *4* (5). DOI: 10.1063/1.4952607 (acccessed 1/7/2024). de Jong, M.; Chen, W.; Notestine, R.; Persson, K.; Ceder, G.; Jain, A.; Asta, M.; Gamst, A. A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-nary Inorganic Polycrystalline Compounds. *Scientific Reports* **2016**, *6* (1), 34256. DOI: 10.1038/srep34256. Oliynyk, A. O.; Antono, E.; Sparks, T. D.; Ghadbeigi, L.; Gaultois, M. W.; Meredig, B.; Mar, A. High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chemistry of Materials* **2016**, *28* (20), 7324-7331. DOI: 10.1021/acs.chemmater.6b02724.

(8) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9* (8), 1735-1780. DOI: 10.1162/neco.1997.9.8.1735.

(9) Cho, K.; Merrienboer, B. v.; Gülçehre, Ç.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In *Conference on Empirical Methods in Natural Language Processing*, 2014.

(10) Johnson, R.; Zhang, T. Deep Pyramid Convolutional Neural Networks for Text Categorization. In Proceedings of the 55th Annual Meeting of the Association for
  Computational Linguistics (Volume 1: Long Papers), 2017.

(11) Goodall, R. E. A.; Lee, A. A. Predicting materials properties without crystal structure: deep representation learning from stoichiometry. *Nature Communications* **2020**, *11* (1), 6280. DOI: 10.1038/s41467-020-19964-7.

(12) Wang, A. Y.-T.; Kauwe, S. K.; Murdock, R. J.; Sparks, T. D. Compositionally restricted attention-based network for materials property predictions. *npj Computational Materials* **2021**, *7* (1), 77. DOI: 10.1038/s41524-021-00545-1.

(13) Chen, Y. Convolutional Neural Network for Sentence Classification. 2015.

(14) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* **2015**, 770-778.

(15) Vaswani, A.; Shazeer, N. M.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is All you Need. In *Neural Information Processing Systems*, 2017.

(16) Clement, C. L.; Kauwe, S. K.; Sparks, T. D. Benchmark AFLOW Data Sets for Machine Learning. *Integrating Materials and Manufacturing Innovation* **2020**, *9* (2), 153-156. DOI:

10.1007/s40192-020-00174-4.

(17) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. *npj Computational Materials* **2020**, *6* (1), 138. DOI: 10.1038/s41524-020-00406-3.

(18) Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C. The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, *1* (1), 15010. DOI: 10.1038/npjcompumats.2015.10.

(19) Jha, D.; Ward, L.; Yang, Z.; Wolverton, C.; Foster, I.; Liao, W.-k.; Choudhary, A.; Agrawal, A. IRNet. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019.

(20) Ballakur, A. A.; Arya, A. Empirical Evaluation of Gated Recurrent Neural Network Architectures in Aviation Delay Prediction. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)*, 14-16 Oct. 2020, 2020; pp 1-7. DOI: 10.1109/ICCCS49678.2020.9276855.

(21) Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *ArXiv* **2014**, *abs/1409.2329*.

(22) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software* **2018**, *3* (29). DOI: 10.21105/joss.00861.

(23) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **2017**, *60* (6), 84-90. DOI: 10.1145/3065386.

(24) Stergiou, A.; Poppe, R.; Kalliatakis, G. *Refining activation downsampling with SoftPool*; 2021.

(25) Dumitrescu, D.; Boiangiu, C.-A. A Study of Image Upsampling and Downsampling Filters. *Comput.* **2019**, *8*, 30.

(26) Deng, J.; Dong, W.; Socher, R.; Li, L. J.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, 20-25 June 2009, 2009; pp 248-255. DOI: 10.1109/CVPR.2009.5206848.

(27) Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. R. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *ArXiv* **2018**, *abs/1807.02758*.

(28) Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **2010**, *2* (11), 559-572. DOI: 10.1080/14786440109462720.

(29) Zhang, Y.; Li, K.; Li, K.; Zhong, B.; Fu, Y. R. Residual Non-local Attention Networks for Image Restoration. *ArXiv* **2019**, *abs/1903.10082*.

(30) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Computational Materials* **2016**, *2* (1), 16028. DOI: 10.1038/npjcompumats.2016.28.

(31) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking materials property prediction methods: the Matbench test set and Automatminer reference algorithm. *NPJ COMPUTATIONAL MATERIALS* **2020**, *6* (1). DOI: 10.1038/s41524-020-00406-3.

(32) Choudhary, K.; Garrity, K. F.; Reid, A. C. E.; DeCost, B.; Biacchi, A. J.; Walker, A. H. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.; Centrone, A.; et al. The joint automated repository for various integrated simulations (JARVIS) for data-driven materials design. *NPJ COMPUTATIONAL MATERIALS* **2020**, *6* (1). DOI: 10.1038/s41524-020-00440-1.

(33) Petretto, G.; Dwaraknath, S.; Miranda, H. P. C.; Winston, D.; Giantomassi, M.; van Setten, M.

J.; Gonze, X.; Persson, K. A.; Hautier, G.; Rignanese, G. M. High-throughput density-functional perturbation theory phonons for inorganic materials. *SCIENTIFIC DATA* **2018**, *5*. DOI: 10.1038/sdata.2018.65.

(34) Petousis, I.; Mrdjenovich, D.; Ballouz, E.; Liu, M.; Winston, D.; Chen, W.; Graf, T.; Schladt, T. D.; Persson, K. A.; Prinz, F. B. High-throughput screening of inorganic compounds for the discovery of novel dielectric and optical materials. *SCIENTIFIC DATA* **2017**, *4*. DOI: 10.1038/sdata.2016.134.

(35) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.; Gamst, A.; Sluiter, M.; Ande, C. K.; van der Zwaag, S.; Plata, J. J.; et al. Charting the complete elastic properties of inorganic crystalline compounds. *SCIENTIFIC DATA* **2015**, *2*. DOI: 10.1038/sdata.2015.9.

(36) Castelli, I. E.; Landis, D. D.; Thygesen, K. S.; Dahl, S.; Chorkendorff, I.; Jaramillo, T. F.; Jacobsen, K. W. New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *ENERGY & ENVIRONMENTAL SCIENCE* **2012**, *5* (10), 9034-9043. DOI: 10.1039/c2ee22341d.

(37) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A. The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *COMPUTATIONAL MATERIALS SCIENCE* **2015**, *97*, 209-215. DOI: 10.1016/j.commatsci.2014.10.037.

(38) Zhuo, Y.; Mansouri Tehrani, A.; Brgoch, J. Predicting the Band Gaps of Inorganic Solids by Machine Learning. *The Journal of Physical Chemistry Letters* **2018**, *9* (7), 1668-1673. DOI: 10.1021/acs.jpclett.8b00124.

(39) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S. T.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, *1*, 011002-011002. Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G. A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **2011**, *50* (8), 2295-2310. DOI: https://doi.org/10.1016/j.commatsci.2011.02.023. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65* (11), 1501-1509. DOI: 10.1007/s11837-013-0755-4.

(40) Maas, A. L. Rectifier Nonlinearities Improve Neural Network Acoustic Models. 2013.

(41) Oviedo, F.; Ren, Z.; Sun, S.; Settens, C.; Liu, Z.; Hartono, N. T. P.; Ramasamy, S.; DeCost, B. L.; Tian, S. I. P.; Romano, G.; et al. Fast and interpretable classification of small X-ray diffraction datasets using data augmentation and deep neural networks. *npj Computational Materials* **2019**, *5* (1), 60. DOI: 10.1038/s41524-019-0196-x. Ghosh, K.; Stuke, A.; Todorović, M.; Jørgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke, P. Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Adv Sci (Weinh)* **2019**, *6* (9), 1801367. DOI: 10.1002/advs.201801367   From NLM. Dinic, F.; Singh, K.; Dong, T.; Rezazadeh, M.; Wang, Z.; Khosrozadeh, A.; Yuan, T.; Voznyy, O. Applied Machine Learning for Developing Next-Generation Functional Materials. *Advanced Functional Materials* **2021**, *31*. Ghosh, K.; Stuke, A.; Todorović, M.; Jørgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke, P. Deep Learning Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Advanced Science* **2019**, *6*.