

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

Published as part of *Journal of Chemical Theory and Computation virtual special issue "Machine Learning and Statistical Mechanics: Shared Synergies for Next Generation of Chemical Theory and Computation"*.

Hongwei Du, Jian Hui,* Lanting Zhang,* and Hong Wang*



Cite This: <https://doi.org/10.1021/acs.jctc.4c00187>



Read Online

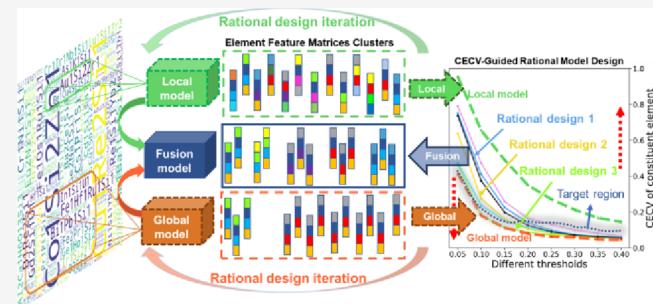
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: The success of machine learning in predicting material properties is largely dependent on the design of the model. However, the current designs of deep learning models in materials science have the following prominent problems. First, the model design lacks a rational guidance strategy and heavily relies on a large amount of trial and error. Second, 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 a new feedback method rich in physical insights. In this study, we developed a feedback method called the 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 material property predictions. L-G-DCNN accurately captures the interactions between elements in compounds, enabling more accurate and efficient predictions of the 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 data sets, 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 an 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 an 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 the development of a representation that is suitable for input data. 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 of the 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, materials 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

Received: February 14, 2024

Revised: July 4, 2024

Accepted: July 8, 2024

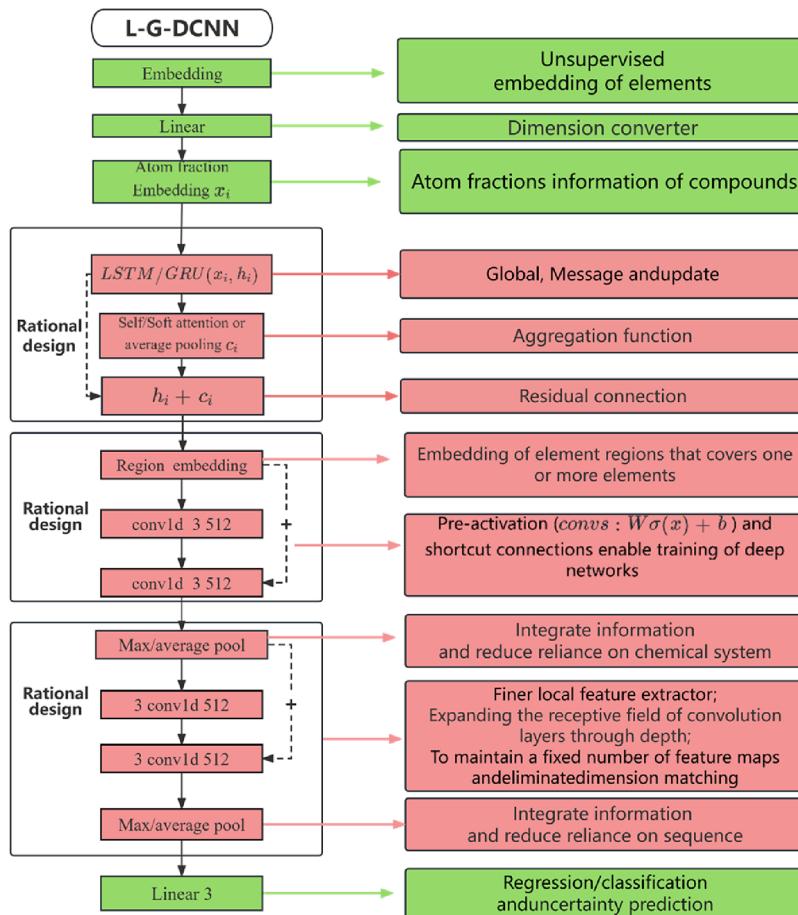


Figure 1. Architecture introduction of the L-G-DCNN model and the functions of each layer.

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 such as 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 memory⁸ and gated recurrent unit⁹ fused with deep convolutional neural network¹⁰ (L-G-DCNN) based on the chemical environment clustering vector (CECV) feedback method. The network consists of an 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 a 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

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 final model output. The calculation steps for element contributions and the 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 elementproto-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 postprocessing neural network. Each message

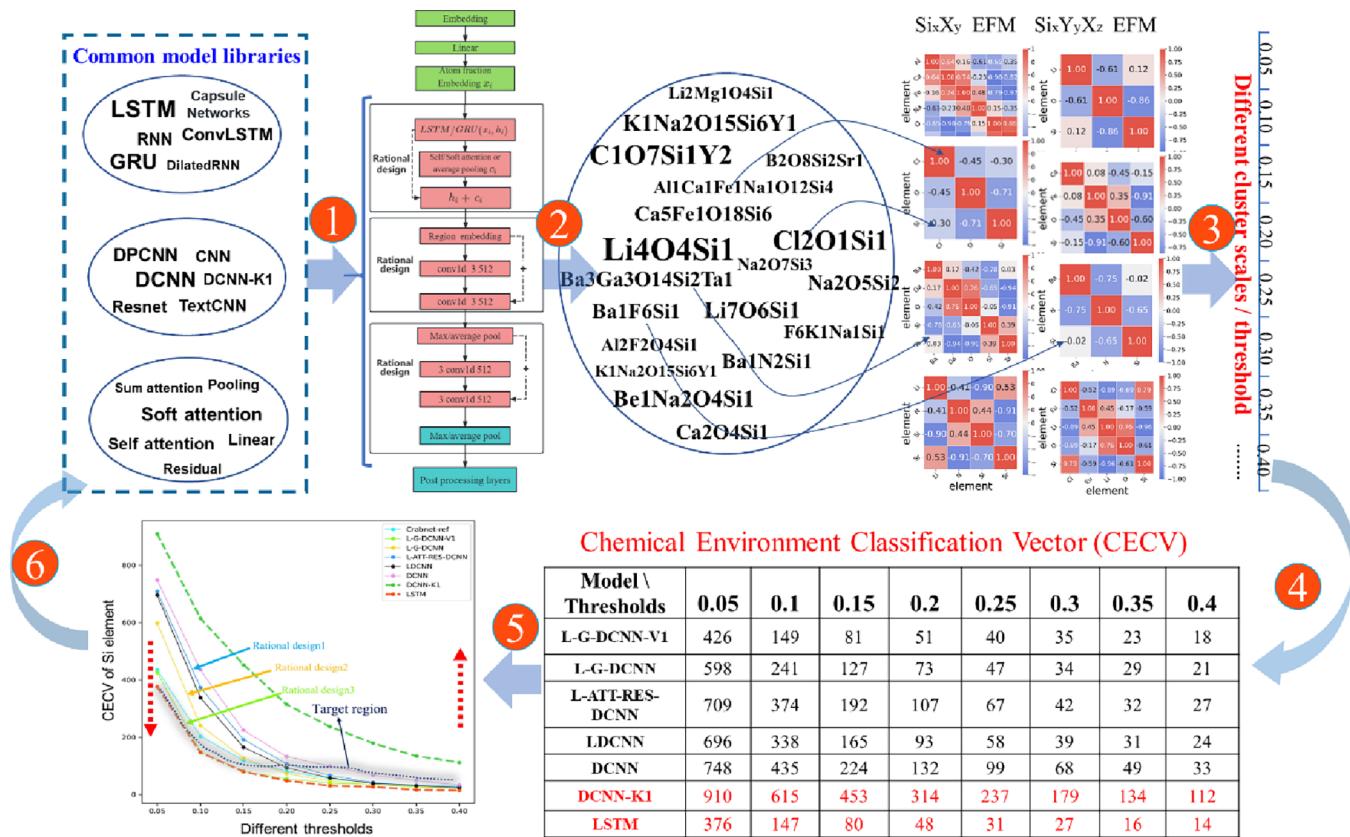


Figure 2. Process of obtaining EFM and how, based on this, the CECVs of all models before and after fusion under different thresholds are obtained, thereby achieving rational model design guided by CECV.

passing layer is coupled to 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 and 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 quantified 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, in 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 the 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 allow for the perception of 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 the materials science literature into information-dense word embeddings. These embeddings can capture latent knowledge in the 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 the 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 the 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 mapped 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. 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 design rational 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 values of compounds at different thresholds will be more accurate.

Next, we 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:

- Selection and initialization of model components: As shown in Figure 2, step 1, 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 1×1 kernel size (DCNN-K1), CNN, TextCNN,¹³ ResNet¹⁴), 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 postprocessing.
- Model training and testing: As shown in Figure 2, step 2, based on the OQMD enthalpy data set, use the selected model components (such as LSTM, DCNN-K1, DCNN, LDCNN) for training and validation. Divide the data set into training, validation, and test sets in a 70/15/15 ratio, and train all models under unified parameter settings.
- Extraction of compound EFM: As shown in Figure 2, step 3, for 2374 compounds containing silicon elements in the test data set 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.
- CECV generation: As shown in Figure 2, step 4, 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].
- Visualization and definition of target region: As shown in Figure 2, step 5, 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, step 6, 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.

We selected 10 data sets with a size greater than 20,000 from 18 extended data sets (Supplementary Figure S1) from different sources, including AFLOW,¹⁶ Material Projects¹⁷ (MP), OQMD,¹⁸ and experimental data sets,¹⁹ as the benchmark data set to ensure the robustness of the test results. The compounds in the data set 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: LSTM⁸, 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 data sets, as shown in Supplementary Figure S2(a). Obviously, LSTM and GRU performed better than RNN in terms of overall MAE results, especially on data sets at the 10^5 level. Therefore, this paper initially chose LSTM as the candidate global feature extractor. Additionally, we extract internal vector representations of all 51,242 compounds in the OQMD bandgap test data sets 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 Figure S3(a,b) for details.

In order to achieve a much finer local interaction between elements, we designed several convolutional neural networks as shown in 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 downsampling²⁵ results in a pyramid-shaped reduction in internal data size for each layer. Therefore, DCNN can effectively cover more element regions to

300 represent longer element correlations in compounds. However,
 301 downsampling will lose some of the original input information;
 302 Convolutional Neural Networks²⁶ (CNN), where the kernel
 303 size and padding parameters are 3 and 1, respectively, and no
 304 downsampling was performed in the pooling layer. The CNN
 305 only transformed the original compound embedding into a
 306 localized embedding representation covering multiple elements,
 307 but the main difference was that it lacked depth.

308 We tested four candidate models on 10 benchmark data sets,
 309 as shown in [Supplementary Table S1](#). Obviously, DCNN,
 310 DPCNN, and CNN performed better than DCNN-K1 in terms
 311 of overall MAE results. Therefore, we discarded DCNN-K1 and
 312 selected DCNN, DPCNN, and CNN as candidate local feature
 313 extractors. However, through visual analysis comparing DNN
 314 and DCNN-K1, it was found that DCNN-K1 was entirely
 315 incapable of capturing any interactions between elements within
 316 the compounds, focusing solely on individual elements. This can
 317 be observed by comparing its output EFM to the EFM at the
 318 input layer. Their UMAP two-dimensional projections are
 319 nearly identical. Thus, we chose DCNN-K1 as the reference
 320 local model. For more detailed visual analysis of DCNN and
 321 DCNN-K1, see [Supplementary Figures S2\(c\)](#) and [S3\(c,d\)](#).
 322 Next, we heuristically fused them with LSTM to obtain LDCNN,
 323 LDPCNN, and LDCNN. We evaluated their MAE results on the
 324 benchmark test set, as shown in [Figure S2\(d\)](#). The results clearly
 325 indicate that LDCNN has a smaller MAE value, thus
 326 determining it as the initial fusion architecture. For the
 327 optimization of convolution kernel hyperparameters for
 328 LDCNN, please refer to [Table S2](#).

329 The focus of this work is on how to improve LDCNN
 330 rationally. Generally, we introduce commonly used model
 331 components, such as residual networks,¹⁴ pooling, and attention
 332 networks,²⁷ heuristically, and observe their performance on the
 333 test data set as the evaluation metric to achieve network design.
 334 This conventional method is feasible, but it is very inefficient and
 335 lacks explanation and rational guidance. Based on the fusion
 336 strategy, this paper designs discrete models with global and local
 337 components LSTM and DCNN-K1 as references and converts
 338 the model evaluation method into a cluster problem of the
 339 chemical environment of the constituent elements in com-
 340 pounds at different thresholds, achieving rational model design.
 341 Here, we extracted the EFM of the learned compounds from the
 342 model and computed the similarity between EFM values at
 343 different thresholds. The resulting Pearson correlation matrix
 344 was used to generate CECV for compounds at different
 345 thresholds.²⁸ Thresholds are defined as the tolerance for
 346 distinguishing the chemical environment of compounds, with
 347 default values ranging from 0.05 to 0.4 in increments of 0.05.

348 As shown in [Figure 3](#), two discrete reference models,
 349 represented by red and green dashed lines, respectively, show
 350 the CECV of silicon-containing compounds in the OQMD
 351 formation enthalpy data set at different thresholds for LSTM
 352 and DCNN-K1. Based on the characteristics of LSTM and
 353 DCNN-K1, they have different requirements for CECV in
 354 different thresholds. For the global feature extractor LSTM, its
 355 performance approaches the optimal solution within the
 356 thresholds <0.15 range, because the smaller the thresholds, the
 357 lower the tolerance for distinguishing element chemical
 358 environments, and the higher the requirement. If the model
 359 focuses too much on local details, it is easy to overestimate
 360 CECV, so the weight of the global extractor needs to be
 361 increased to adjust the model and achieve a more accurate
 362 identification of element chemical environments, as shown by

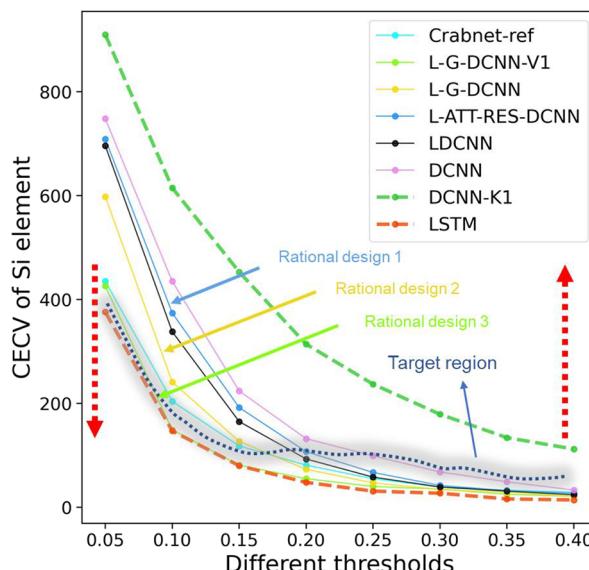


Figure 3. CECV of silicon-containing compounds in the OQMD formation enthalpy test data sets based on the fusion strategy, using LSTM and DCNN-K1 as global and local reference models. CECV was used as the feedback method to achieve rational model design. The shaded region in the figure marks the CECV 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.

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 [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 threshold range <0.15, very close to the shadow region.
 However, within the threshold 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, 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 threshold 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,
 rational design 2 was proposed, introducing stacked residual
 GRU and softmax attention²⁹ to obtain L-G-DCNN. This
 design significantly reduced CECV in the threshold range <0.15
 and greatly improved the performance of L-G-DCNN, as shown

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. MAE values of the L-G-DCNN series models with the reference models Roost, Crabnet, and Finder on the extended data set. 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.

396 by the yellow arrows in Figure 3. We can see that L-G-DCNN
 397 still has room for improvement within the range of thresholds
 398 <0.15. Therefore, rational design 3, L-G-DCNN-V1, was
 399 introduced in L-G-DCNN by incorporating residual con-
 400 nections and maximum pooling, as indicated by the green
 401 arrow in Figure 3. L-G-DCNN-V1 continues to reduce CECV
 402 within the range of thresholds <0.15, entering the target shadow
 403 region and achieving optimization. However, there is also a
 404 decline in CECV within the range of thresholds >0.15. Overall,
 405 the target region has not yet been reached. From LDCNN to L-
 406 G-DCNN-V1, based on the CECV, we have improved the
 407 performance of the model through three-step design, achieving a
 408 rational and efficient design.

409 Finally, we tested LDCNN, L-G-DCNN, and L-G-DCNN-V1
 410 on the test set of the OQMD formation enthalpy and obtained
 411 the MAE results as shown in Figure S9 (d–f), which are 0.0349,
 412 0.0332, and 0.0317, respectively. This is consistent with the
 413 evaluation results of CECV, confirming the correctness of the
 414 CECV method. In Supporting Information Table S3, we
 415 obtained the CECV of silicon-containing compounds on five
 416 data sets, which is consistent with the results of MAE, further
 417 demonstrating the stability of the CECV method. Throughout
 418 the process of rational design, the evolutionary details of the
 419 fusion model architecture can be found in Figure S4.

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

433 We found that currently, models that only utilize composi-
 434 tional information without manual feature engineering, in

chronological order, mainly include Roost¹¹ (2020), Crabnet¹² 435
 436 (2021), and Finder⁶ (2022). Consequently, we have selected 437
 438 Crabnet and Finder, which are based on GNN and transformer 439
 440 architectures, respectively, to serve as our baseline models for 441
 442 comparative analysis. We train them using the default model 443
 444 parameters provided by their respective repositories. As 445
 446 representatives of traditional machine learning, we use the 447
 448 Magpie-featurized³⁰ RF model as the baseline model. In this 449
 450 section, we evaluate the structure-agnostic L-G-DCNN series 451
 452 models (L-G-DCNN and L-G-DCNN-V1) and compare them 453
 454 with Roost, Crabnet, Finder and the RF. 455

456 Figure 4 displays the MAE scores of all models on the held-out 446 f4
 447 test data set, indicating that L-G-DCNN outperforms most 448
 449 benchmark tasks. Furthermore, the testing results of L-G- 450
 451 DCNN-V1 show improvement compared to L-G-DCNN in 452
 453 larger data sets and perform similarly to L-G-DCNN in smaller 454
 455 data sets. This once again reaffirms the significance of our 456
 457 rational guidance strategy. It is worth noting that, based on the 458
 459 guidance of CECV strategy, to further reduce the CECV on the 460
 461 LSTM end while increasing the CECV on the DCNN-K1 end, 462
 463 we modified the attention component in L-G-DCNN and 464
 464 introduced self-attention mechanism to obtain L-G-DCNN-V2, 465
 466 which further improved the test results, especially surpassing 467
 467 Roost and Finder, reaching the level of Crabnet in large data sets, 468
 469 as shown in Supplementary Figure S10. Supplementary Figure 459
 460 S11 presents the CECV for the silicon-containing compounds 461
 462 across different thresholds of L-G-DCNN-V2 and Crabnet, 463
 464 based on the OQMD formation enthalpy data set. CECV 465
 466 explains the performance improvement of L-G-DCNN-V2. 467
 467 Detailed data for Figure S11 can be found in Supplementary 468
 469 Table S4. Considering the different architectures and modeling 470
 471 philosophies of Roost, Crabnet, Finder and L-G-DCNN, it is 472
 472 encouraging to see that all methods converge to the same 473
 473 performance. This further confirms that rational guidance for 474
 474 model design based on CECV maximizes complementarity and 475
 475 improves fusion between discrete models, thereby enhancing 476
 476 our understanding of the model design. The training progress of 477
 477 L-G-DCNN and Crabnet is compared in Figure S12. 478

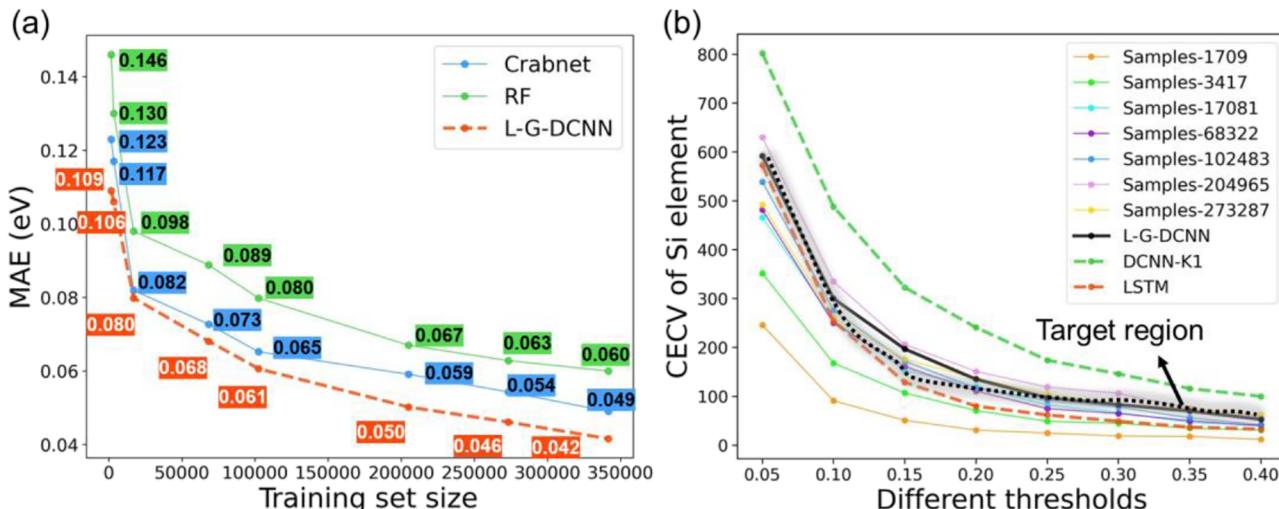


Figure 5. Sample efficiency of the model-agnostic models evaluated on the OQMD Band gap data set. (a) L-G-DCNN outperforms Crabnet and RF and exhibits a steeper gradient descent. (b) Performance of the CECV of silicon-containing compounds across different thresholds for models trained on different training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

473 **Structure-Agnostic Model Evaluation.** Material data,
 474 particularly experimental measurements, are often limited in
 475 size. This raises concerns about the ability of DL models in
 476 general to learn from significantly undersampled data sets and
 477 still provide reasonably accurate out-of-database predictions.
 478 The sample efficiency of L-G-DCNN was evaluated by
 479 examining its performance at different training set sizes. Figure
 480 5(a) shows the prediction MAE curve of the structure-agnostic
 481 model on the OQMD bandgap test set. L-G-DCNN achieved
 482 the lowest error scores across all training set levels (from 10^3 to
 483 $\approx 3 \times 10^5$). Traditional ML models that use interpretable
 484 features, such as RF, typically perform well on small data sets.
 485 Although inheriting the DL paradigm, our model still outper-
 486 forms RF when the training set size reaches 10^2 – 10^3 .
 487 Furthermore, the MAE curve of L-G-DCNN exhibits a steeper
 488 gradient descent compared to Crabnet.

489 Generally, the main factors that impact the model perform-
 490 ance are the model architecture and data size. Our CECV
 491 method explains the influence of data size on model perform-
 492 ance. Figure 5(b) illustrates that, with a fixed model architecture,
 493 as the training set size increases, the CECV of the model
 494 converges toward the target region, which matches the
 495 traditional MAE statistical evaluation results shown in Figure
 496 5(a).

497 In the parity plots of the OQMD band gap and the enthalpy of
 498 formation of the OQMD shown in Figure 6(a,b), L-G-DCNN
 499 achieved good prediction results. For the band gap, most of the
 500 values in the data set are distributed around 0. The RMSE is
 501 0.236, much larger than the 0.096 for formation enthalpy,
 502 indicating a higher deviation from the true values in the
 503 prediction of band gap, especially in the range close to 0. For
 504 formation enthalpy, the prediction results are better when the
 505 target value is negative. This is because the computational
 506 materials database tends to report more stable materials, which
 507 usually have negative values. As expected, the aleatoric
 508 uncertainty of relatively inaccurate predictions is higher than
 509 the uncertainty of samples close to the perfect prediction.
 510 We investigate whether the EFM indeed captures some
 511 chemical insights that are not explicitly input into the model.
 512 Figure 6(c) presents the EFM of four well-known perovskites
 513 obtained from the final layer of the trained OQMD formation

514 enthalpy and OQMD band gap models, respectively. From the 515 OQMD formation enthalpy model, we find that compositionally 516 and structurally similar materials such as SrTiO_3 and BaTiO_3 517 have a similar EFM. Interestingly, the EFM of KNbO_3 , which 518 has a different composition but a similar structure, is comparable 519 to the EFMs of the aforementioned materials. The EFM of 519 halide perovskite CsPbI_3 is considerably different from that of its 520 oxide counterparts. A consistent trend is observed from the 521 OQMD band gap model, but different EFMs are obtained for 522 the same material. Obviously, EFM entries are determined by 523 the constituent element types and the context. Incorporating 524 EFM that reflect interactions between elements adds another 525 dimension for materials similarity analysis. Parity plots for other 526 properties and comparison of L-G-DCNN and Crabnet on the 527 OQMD band gap test data sets are shown in Supporting 528 Information Figure S13.

529 **Evaluation on Matbench Data Sets.** Matbench serves as a 530 common test data set for ML models and includes¹³ material 531 property prediction tasks. Here, we select 10 regression tasks. 532 Two recent structure-agnostic models, Finder and Crabnet, are 533 selected for comparison. We follow the 5-fold cross validation 534 strategy with the same random seed variable recommended in 535 the original study to evaluate our algorithm.³¹ 5-fold cross 536 validation conducts an outer test loop with 20% test data and 537 80% training + validation data, and an inner validation process 538 for model weight parameter tuning for each outer fold. L-G- 539 DCNN divides the overall fold into 70% training data, 10% 540 validation data, and 20% test data.

541 Table 1 shows a performance comparison on the Matbench 542 t1 suite. L-G-DCNN and L-G-DCNN-V1 achieve the best MAE 543 scores in 8 out of 10 structure-agnostic tasks. Our model 544 outperforms Crabnet and Finder in small data sets ($< 10^4$) and 545 large data sets ($> 10^5$).

546 **Ablation Study.** To comprehensively and accurately 547 validate the performance of different components of the 548 model in ablation experiments on different data sources and 549 volumes, we carefully selected four data sets: AFLOW energy 550 per atom, MP energy above the convex hull, the OQMD 551 formation enthalpy, and the Bartel formation enthalpy. These 552 data sets are from AFLOW, MP, OQMD calculations,³⁹ and 553 experimental data sources, with data set sizes ranging from 10^4 to 554 t2

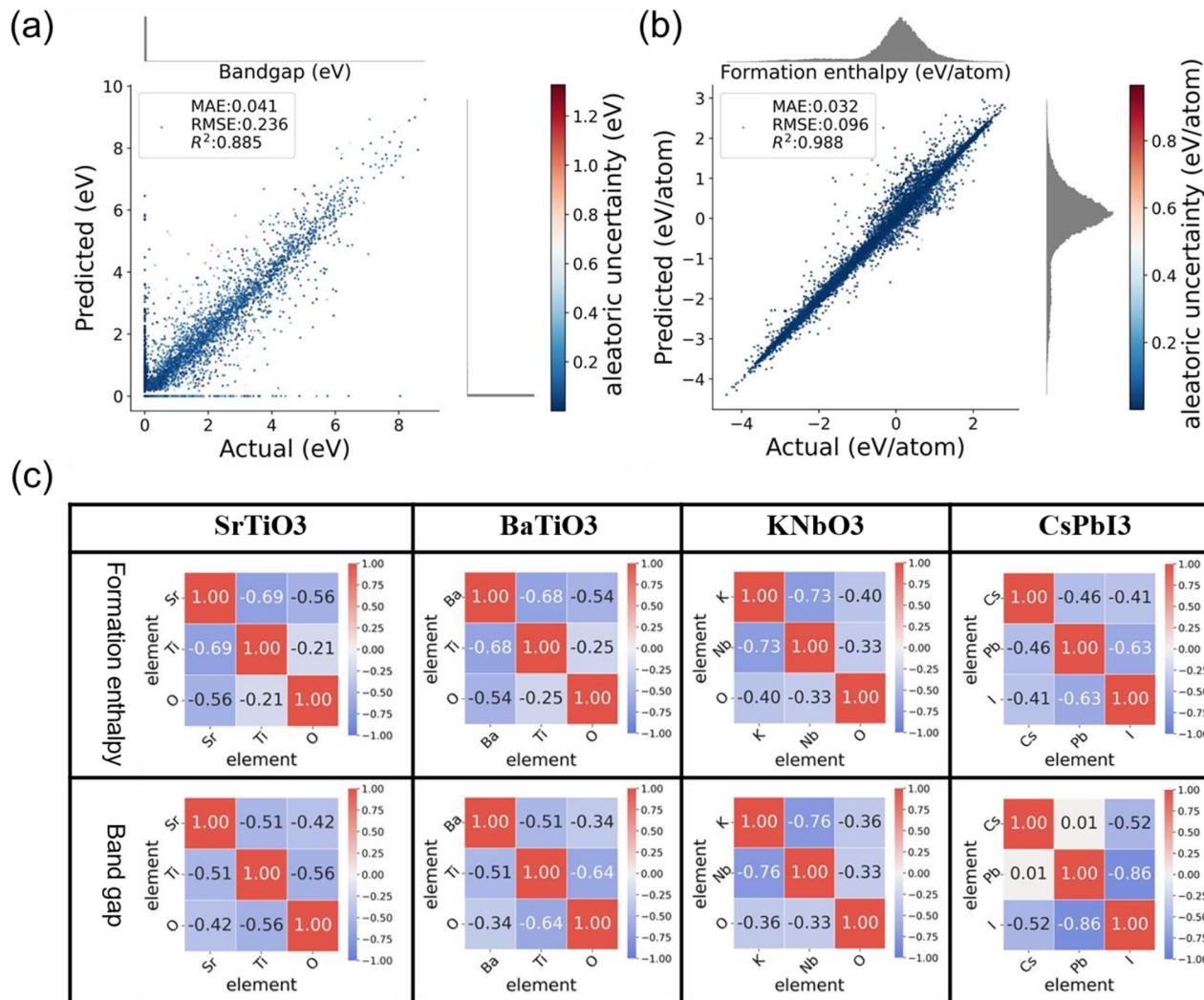


Figure 6. Parity plots for the (a) OQMD band gap and (b) OQMD formation enthalpy of the test set for the structure-agnostic L-G-DCNN. The L-G-DCNN exhibits a good predictive error and uncertainty. The test data and predicted marginal distributions are shown on the minor axis. RMSE-root-mean-square error, R^2 score-determination coefficient. (c) EFM of perovskite materials SrTiO₃, BaTiO₃, KNbO₃, and CsPbI₃ visualized by probing OQMD formation enthalpy and OQMD band gap L-G-DCNN models.

555 10^5 . **Table 2** shows the results of the ablation study on the four
 556 data sets. Although the depth of convolutional neural networks
 557 does not significantly affect the overall performance of structure-
 558 agnostic models, a 4-layer DCNN compared to the default 2-
 559 layer DCNN can achieve the MAE decrease on large data sets
 560 (level 10^5). We investigate one-hot element embeddings and
 561 observe that MAE remains unchanged on the 10^5 level data set
 562 but significantly increases on the other 10^4 level data sets (see
 563 Model 1 in **Table 2**). This indicates that element embedding
 564 capturing prior knowledge still helps navigate to a lower
 565 minimum in the error surface, although this accuracy gap is
 566 expected to narrow as the database size grows. Generally,
 567 postprocessing neural networks are considered as an important
 568 component of the model, such as Crabnet and Finder, which
 569 both include designed postprocessing networks. However, L-G-
 570 DCNN does not include a specifically designed postprocessing
 571 network, but simply incorporates a fully connected network into
 572 its output to reshape it (batch size, sequence length, 3). In
 573 Model 2, we added a postprocessing neural network, which
 574 includes four linear layers, residual connections, LeakyReLU

activation function,⁴⁰ and an output linear layer. We observed
 575 that adding a postprocessing network impairs performance.
 576

In Model 3, we remove all residual connections and observe a
 577 substantial error increase in structure-agnostic models. We then
 578 remove only the residual connections from the region
 579 embedding layers and keep the rest of the residual connections
 580 to obtain Model 4. Compared to removing all residual
 581 connections, the error of Model 4 is reduced. This indicates
 582 that residual components can achieve lower errors in the fusion
 583 architecture. In Model 5, we replace the soft-attention
 584 component after the LSTM||GRU with a self-attention
 585 component. Compared to the reference model, Model 5
 586 shows a significant improvement in MAE for the OQMD
 587 formation enthalpy data set (0.0310), while performance
 588 decreases in other data sets. Model 6 is obtained by removing
 589 soft-attention and GRU components. The MAE increases
 590 significantly, highlighting the importance of designing LSTM||
 591 GRU combinations with soft attention. In conclusion, the
 592 residual components, global feature extractor with attention, and
 593 number of DCNN layers are key factors determining the model
 594 performance.

Table 1. Performance Comparison of the Matbench Suite^a

property	L-G-DCNN	DCNN-V1	Crabnet-ref	Finder-ref	data set 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 enthalpy	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

^aThe 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 data sets 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: band gap as calculated by PBE DFT from the Materials Project, in eV;³⁷ MP formation enthalpy: 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

leads to a decrease in CECV in the smaller thresholds range.⁶¹⁴ Note that the local ability of Model 6 has not been significantly⁶¹⁵ enhanced, so its CECV is relatively smaller than L-G-DCNN-6B⁶¹⁶ and Model 2 in the larger thresholds range. Model 3 removes all⁶¹⁷ residual components, greatly reducing the fusion ability of the⁶¹⁸ model, so its CECV directly exceeds the reference fusion model.⁶¹⁹ Model 5 enters the target shadow region in the smaller threshold⁶²⁰ range; therefore, its performance is greatly improved, and the⁶²¹ MAE value reaches 0.0310, which is very close to that of⁶²² Crabnet. This indicates that self-attention has a significant effect⁶²³ on the OQMD formation enthalpy. In Figure 7(b), we⁶²⁴ compared L-G-DCNN with the Crabnet. Overall, Crabnet's⁶²⁵ CECV is closer to the target shadow region, indicating better⁶²⁶ MAE performance. Based on the above analysis, we can see that⁶²⁷ the CECV evaluation method is stable and provides more⁶²⁸ feedback information than traditional statistical metrics, which⁶²⁹ helps in the rational design and understanding of the model.⁶³⁰

Visualizing The Understanding of Model Fusion⁶³¹

Process. *Contribution of Elements in Property Prediction⁶³² as a Function of Composition before and after Fusion.* For⁶³³ the band gap prediction of the $\text{Si}_x\text{O}_{1-x}$ system across the entire⁶³⁴ stoichiometry range, the results of the discrete models (LSTM,⁶³⁵ DCNN-K1, and DCNN) and the fusion model (L-G-DCNN)⁶³⁶ are presented in Figure 8. Several observations can be made:⁶³⁷ First, with the fusion of models, the uncertainty gradually⁶³⁸ decreases. Specifically, the discrete reference models LSTM and⁶³⁹ DCNN-K1 deviate significantly from the DFT computational⁶⁴⁰ results, with DCNN predicting a band gap peak within 0.4 to 0.6,⁶⁴¹ while the DFT results fall within 0.2 to 0.4, overestimating the⁶⁴² band gap values. Second, with the fusion of models, the overall⁶⁴³ prediction results gradually approach the DFT calculations. The⁶⁴⁴ DFT results can be found in Supplementary Figure S14. Lastly,⁶⁴⁵ Figure 8(a,b) effectively demonstrates the global and local⁶⁴⁶ characteristics of LSTM and DCNN-K1. In particular, the⁶⁴⁷ predictions of the reference model LSTM, acting as a global⁶⁴⁸ feature extractor, are notably lower than the actual values, while⁶⁴⁹ the predictions of the reference model DCNN-K1, serving as a⁶⁵⁰ local feature extractor, are significantly higher than the actual⁶⁵¹ values. With the progress of fusion and appropriate adjustments⁶⁵² to the global and local feature extractors, it is evident that the⁶⁵³ predictions of DCNN and L-G-DCNN show significant⁶⁵⁴ improvements, as shown in Figure 8(c,d).⁶⁵⁵

To evaluate the prediction ability of L-G-DCNN for different⁶⁵⁶ material properties in diverse chemical stoichiometry ranges of⁶⁵⁷ multicomponent compounds, we obtained 18 binary chemical⁶⁵⁸ systems A_xO_{1-x} and 19 ternary chemical systems $\text{Li}_x\text{Al}_{1-x-y}\text{O}_y$ ⁶⁵⁹ from MP. The A elements were sampled according to different⁶⁶⁰ groups of the periodic table, such as metalloid, alkaline, alkali,⁶⁶¹ post-transition metal, and transition metal groups. In the⁶⁶²

596 To further understand how different components of the
597 model in the ablation experiment affect the changes in CECV,
598 we plotted the CECV of the silicon-containing compounds at
599 different thresholds on the OQMD Formation enthalpy data set
600 for the above part of the model, as shown in Figure 7. The raw
601 data for Figure 7 are provided in Supporting Information Table
602 S5. First, based on Figure 7(a), we analyzed the effect of
603 increasing the number of DCNN layers in the reference model.
604 By observing the changes in CECV of L-G-DCNN-2B and L-G-
605 DCNN-6B in the entire threshold range, it can be seen that the
606 CECV of L-G-DCNN-6B is significantly increased compared to
607 L-G-DCNN-2B. This may be because the increase in the
608 number of DCNN layers caused the local ability of the model to
609 be excessively corrected. The situation of Model 2 is similar to
610 that of L-G-DCNN-6B, both of which have an excessively
611 enhanced local ability on the DCNN blocks. Model 6 is
612 different, as it adjusts the GRU components on the LSTM side,
613 directly reducing the global ability of the model, which in turn

Table 2. Four Intricately Designed Data Set MAEs of Different Model Architectures Considered in the Ablation Study^a

benchmark properties	DCNN layers in L-G-DCNN			Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	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

^aThe 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: postprocessing 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.

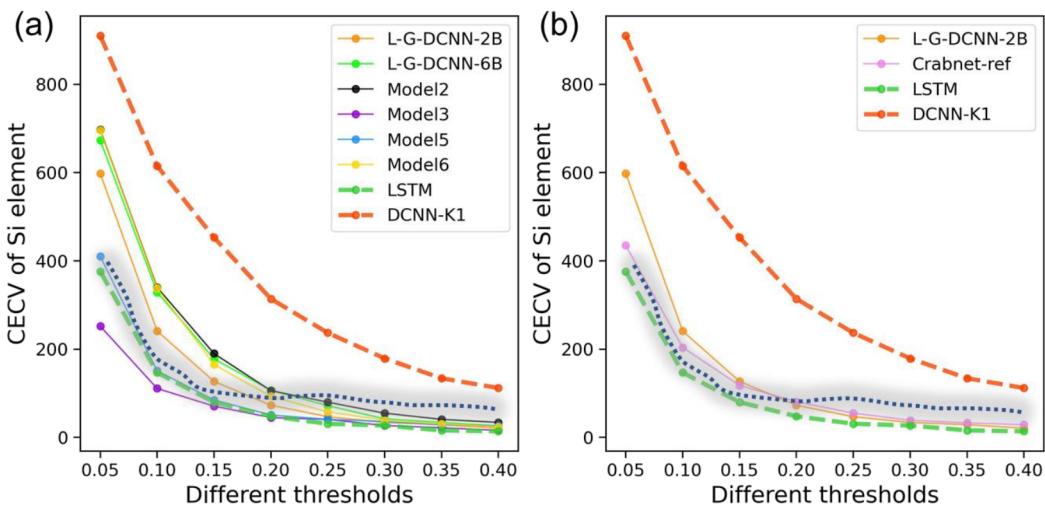


Figure 7. (a) CECV distributions of silicon-containing compounds in different thresholds for Model 2, Model 3, Model 5, Model 6, L-G-DCNN2B, and L-G-DCNN6B on the OQMD formation enthalpy test data sets, where LSTM and DCNN-K1 are fusion reference models. (b) CECV distributions of silicon-containing compounds in different thresholds for the default model L-G-DCNN-2B and Crabnet.

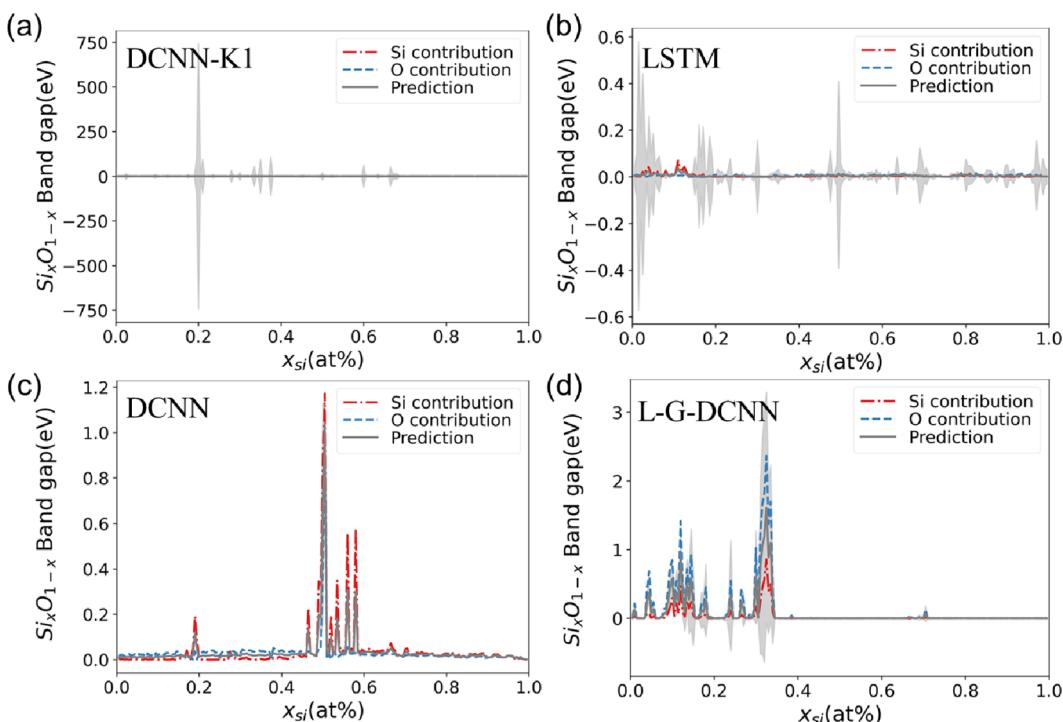


Figure 8. Prediction results for the Si_xO_{1-x} system from both the discrete models (LSTM, DCNN-K1, and DCNN) and the fusion model (L-G-DCNN), highlighting the contribution of elements as constituents to property predictions before and after model fusion.

Supplementary Figure S15 shows 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 Table S6.

Comparison of the Distribution of Overall Element Contributions Differences in Property Prediction before and after Fusion. Similar to Crabnet, L-G-DCNN uses the vector representation of each element to directly predict the contribution of element pairs to the property prediction. Figure

9(a) shows the fusion process of the model, where $\Delta 1$, $\Delta 2$, and $\Delta 3$ represent (L-G-DCNN - DCNN-K1), (L-G-DCNN - DCNN), and (L-G-DCNN - LSTM), respectively. Figure 9(b-d) shows 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. $\Delta 1$ has a much darker color than $\Delta 2$ and $\Delta 3$, indicating that the element contribution estimated by DCNN-K1 is inaccurate. $\Delta 2$ and $\Delta 3$ are similar in color, with different distribution of element contributions, as highlighted by the red box in the figure, with larger range of $\Delta 2$ and smaller range of $\Delta 1$, complementing each other. However, the complementary effect of some elements is not good, such as Nb and Yb. Please

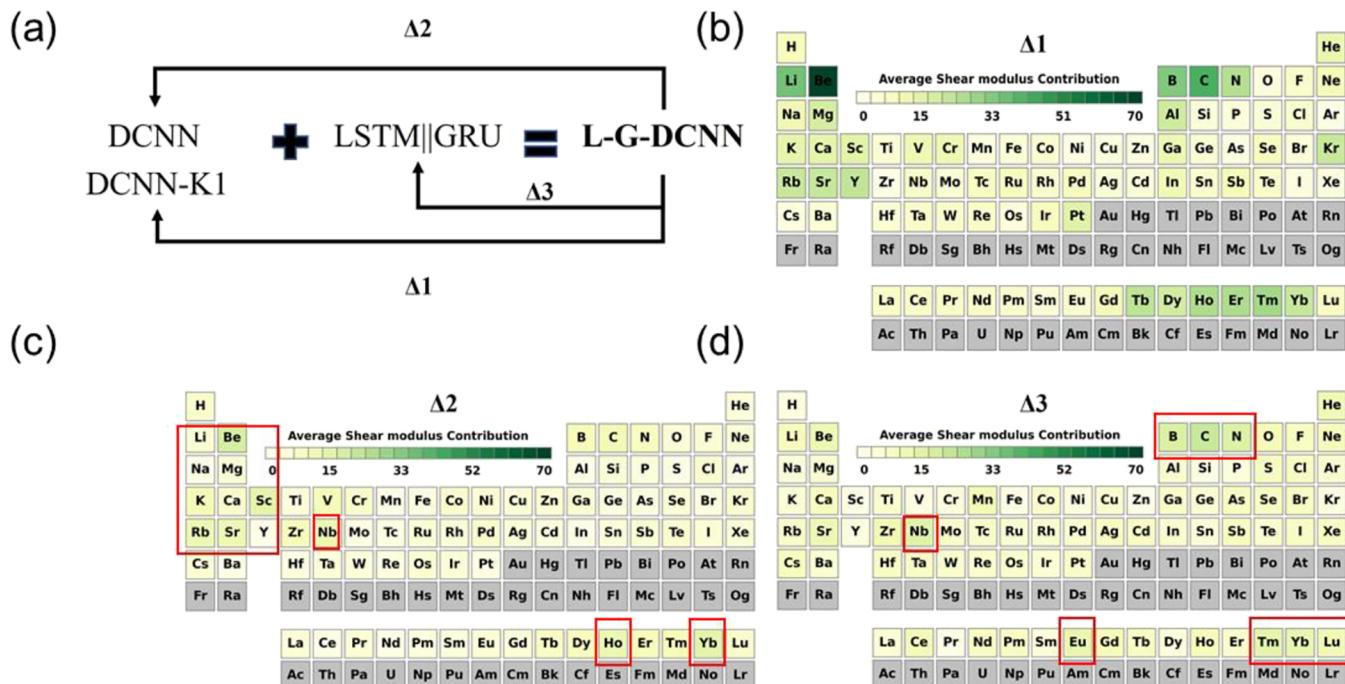


Figure 9. (a) Schematic of the fused discrete DCNN and LSTM model, denoted as L-G-DCNN, where DCNN-K1 serves as the reference model for local feature extraction. $\Delta 1$, $\Delta 2$, and $\Delta 3$ represent the average contribution differences of all elements to the shear modulus predictions relative to L-G-DCNN, computed from the Materials Project (MP) shear modulus data set for DCNN-K1, DCNN, and LSTM, respectively. As shown in (b)–(d), on a periodic table, the darker colored elements contribute more toward the shear modulus value of a compound relative to L-G-DCNN.

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 a 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. We extracted the EFM of compounds containing Si from the L-G-DCNN trained on the OQMD band gap data set, 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. **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 data set. 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 data set 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.

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, then 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 **Table S7**.

Identifying Critical Elemental Interactions Pairs for Band Gap Prediction. In this article, we explored a range of visualization techniques to explain the relationship between model representations, chemical environments, and periodic trends. The goal of this work is to provide an intuitive way to understand the crucial factors that influence the material properties. Although these visualization techniques can capture only 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 data set, 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) Δ EFM. Through compar-

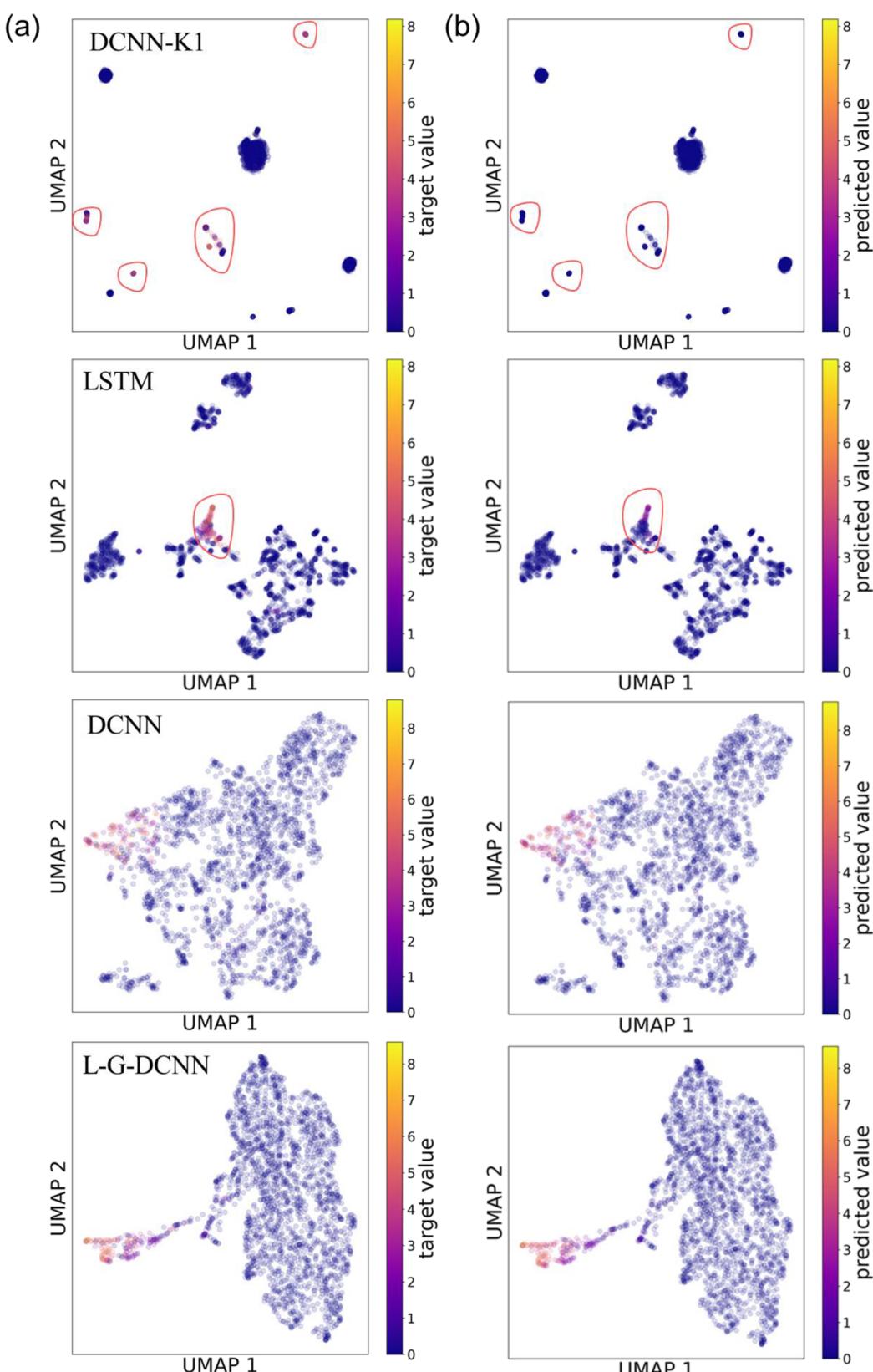


Figure 10. (a,b) Two-dimensional projections of the EFM for silicon-containing compounds in 2374 different chemical environments, based on UMAP, using the OQMD band gap data set. Each point in the figure is colored based on the predicted and computed band gap values. The points with significant differences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

ison, we can determine which element interactions are most

we summarize the elements interaction that have the greatest

757 impact on the band gap among these 56 compounds. It can be 759

760

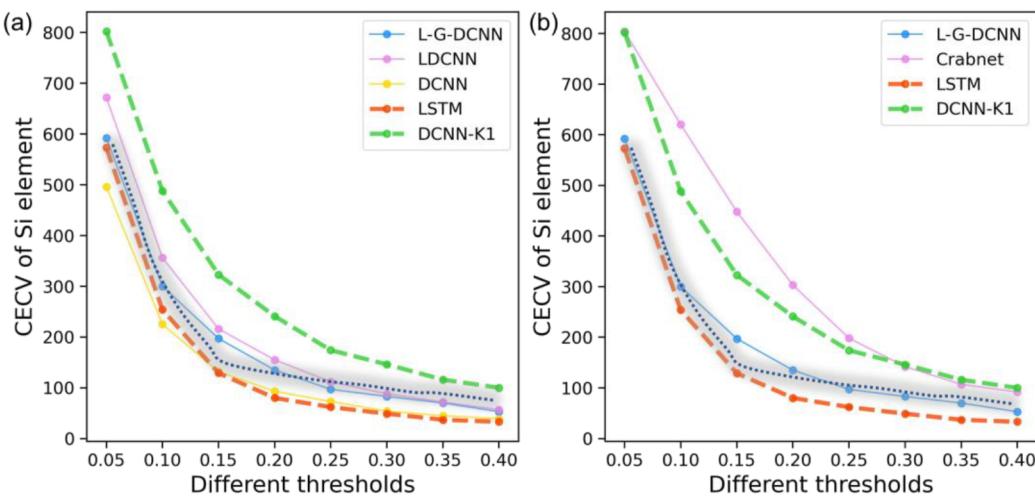


Figure 11. On the OQMD bandgap data set, (a) CECV of silicon-containing compounds at different thresholds obtained by DCNN-K1, LSTM, DCNN, LDCNN, and L-G-DCNN. (b) CECV of silicon-containing compounds at different thresholds obtained by L-G-DCNN and Crabnet. The shaded region with dashed lines is the target reference region.

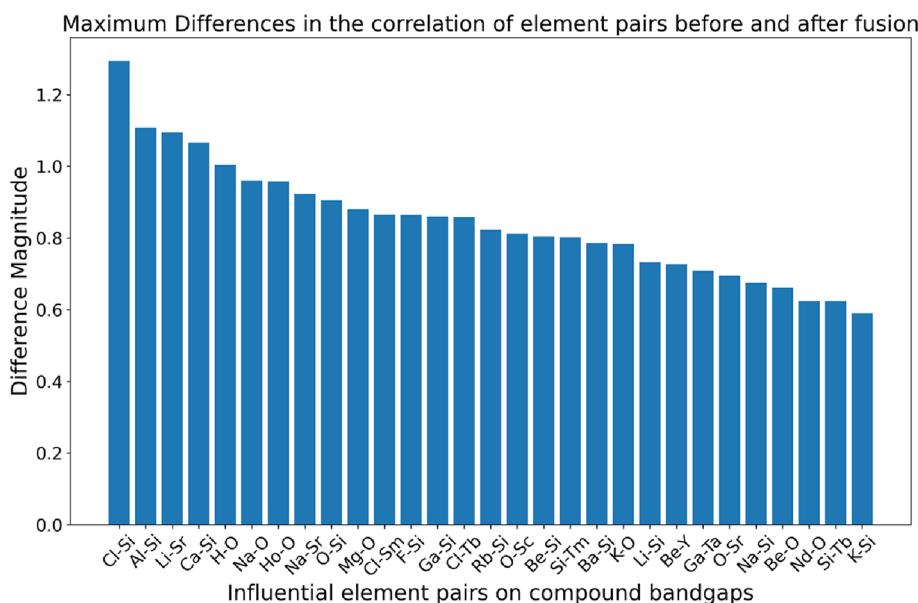


Figure 12. Most influential element interaction pairs on the band gap among the S6 compounds before and after model fusion. From the figure, it can be observed that most element interactions pairs involve elements with high electronegativity such as O, Cl, and F.

seen that most of the element 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 highly effective deep learning model for predicting material properties from chemical composition data.

CONCLUSIONS

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 among 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 data sets. 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 contribution 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 the materials.

796 This multifaceted approach not only clarifies the understanding
797 of model fusion but also illuminates the pivotal role of elemental
798 interactions in material property predictions. The following
799 points provide a specific description of how L-G-DCNN can
800 further enhance this field:

- 801 1. Innovative framework for expanding deep learning model
802 design: L-G-DCNN introduces a novel fusion architecture
803 that combines the strengths of LSTM for global
804 feature extraction and DCNN for local feature capture.
805 This design differs from previous models like Roost,
806 Finder, and Crabnet by offering a more holistic analysis of
807 material properties. This shift from singular to comprehensive
808 in model design promotes the deep application of
809 deep learning models in the field of materials.
- 810 2. Transparent demonstration of model performance
811 evolution: The modular design of L-G-DCNN enables a
812 clear visual representation of the model's improvement,
813 transitioning from individual LSTM and DCNN
814 components to a fusion model. This transparent evolution
815 not only demonstrates improved performance but also a
816 more nuanced capture of elemental interactions. By
817 contrasting LSTM and L-G-DCNN predictions, we can
818 pinpoint critical elemental interactions that influence
819 material properties, thereby deepening our comprehension
820 of the relationship between element chemical
821 environments and material properties and facilitating
822 the strategic design of novel materials.
- 823 3. Efficient and rational optimization strategy based on
824 CECV: CECV introduces the chemical environment,
825 offering a new method for optimizing surrogate models in
826 small data sets. It enriches the physical basis of model
827 design and guides rational model design through feedback
828 mechanisms. For instance, in alloy component design,
829 traditional methods like Gaussian Process (GP) surrogate
830 models struggle with large search spaces and limited data,
831 often leading to inaccurate RL reward calculations and
832 poor generalizability. L-G-DCNN, after training on the
833 same small data set, can calculate the CECV of LSTM,
834 DCNN, and L-G-DCNN, evaluating differences and
835 knowing specifically how to adjust, leading to a more
836 rational and efficient fusion strategy. In an active learning
837 framework, CECV steers model design and experiment
838 selection, optimizing the use of limited resources and
839 scarce data, accelerating the transformation from data to
840 knowledge.

841 ■ ASSOCIATED CONTENT

842 ◀ Supporting Information

843 The Supporting Information is available free of charge at
844 <https://pubs.acs.org/doi/10.1021/acs.jctc.4c00187>.

845 Distribution of training, validation, and test data in
846 extended data sets and Matbench; explanation of the
847 rational model design process; L-G-DCNN model parity
848 plots and comparative analysis with Crabnet on OQMD
849 band gap data set; CECV analysis of silicon-containing
850 compounds across thresholds in ablation study models on
851 OQMD formation enthalpy data set; comparative
852 predictions and DFT validation of $\text{Si}_x\text{O}_{1-x}$ band gaps by
853 L-G-DCNN and Crabnet on OQMD data with elemental
854 contribution and uncertainty analysis; elemental
855 contributions to shear modulus predictions across the
856 periodic table: insights from MP data set analysis; silicon

857 compound CECV assessment across DCNN-K1, LSTM, 857
858 DCNN, LDCNN, L-G-DCNN, and Crabnet models on 858
859 OQMD bandgap data; comparative parameter analysis of 859
860 L-GDCNN, Crabnet, and Finder models from fusion, 860
861 transformer, and GNN architectures (PDF) 861

862 Two-dimensional T-SNE plots of elemental embeddings; 862
863 EFM of the XMn_2O_4 compound; Pearson correlations 863
864 between elemental embedding vectors; MAE of reference 864
865 Roost, Crabnet, and random forest on the held-out test 865
866 data sets; CECV of silicon-containing compounds at 866
867 different thresholds; training progress curves of L-G- 867
868 DCNN and Crabnet; parity plots of the L-G-DCNN; 868
869 predictions of L-G-DCNN and Crabnet trained on 869
870 OQMD band gap data for the $\text{Si}_x\text{O}_{1-x}$ system; average 870
871 contribution of all elements to the prediction of shear 871
872 modulus (PDF) 872

873 ■ AUTHOR INFORMATION

874 Corresponding Authors

Jian Hui – School of Materials Science and Engineering and 875
Materials Genome Initiative Center, Shanghai Jiao Tong 876
University, Shanghai 200240, China; Zhangjiang Institute for 877
Advanced Study, Shanghai Jiao Tong University, Shanghai 878
201203, China; Email: hj20151107@sjtu.edu.cn 879

Lanting Zhang – School of Materials Science and Engineering 880
and Materials Genome Initiative Center, Shanghai Jiao Tong 881
University, Shanghai 200240, China; Zhangjiang Institute for 882
Advanced Study, Shanghai Jiao Tong University, Shanghai 883
201203, China;  orcid.org/0000-0003-2547-1952; 884
Email: rantingzh@sjtu.edu.cn 885

Hong Wang – School of Materials Science and Engineering and 886
Materials Genome Initiative Center, Shanghai Jiao Tong 887
University, Shanghai 200240, China; Zhangjiang Institute for 888
Advanced Study, Shanghai Jiao Tong University, Shanghai 889
201203, China;  orcid.org/0000-0003-1986-512X; 890
Email: hongwang2@sjtu.edu.cn 891

892 Author

Hongwei Du – School of Materials Science and Engineering and 893
Materials Genome Initiative Center, Shanghai Jiao Tong 894
University, Shanghai 200240, China; Zhangjiang Institute for 895
Advanced Study, Shanghai Jiao Tong University, Shanghai 896
201203, China 897

Complete contact information is available at: 898
<https://pubs.acs.org/10.1021/acs.jctc.4c00187> 899

900 Author Contributions

H.D. and H.W. devised the idea for the paper. H.D. 901
implemented the idea and conducted the code design and 902
visualizations. H.D. and H.W. interpreted the results and 903
prepared the manuscript. All authors discussed the results. 904

905 Funding

The source code and data of this paper is available at: <https://github.com/dhw059/DeepModelFusion>. 906 907

908 Notes

The authors declare no competing financial interest. 909

910 ■ ACKNOWLEDGMENTS

We are grateful for the financial support from the National Key 911
Research and Development Program of China (Grant 912
Nos.2021YFB3702102). The computations in this paper were 913

run on the π 2.0 cluster supported by the Center for High Performance Computing at Shanghai Jiao Tong University.

REFERENCES

- (1) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138.
- (2) Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwarakanath, 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.
- (3) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114* (10), No. 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), No. 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. *Comput. Mater. Sci.* **2018**, *152*, 60–69.
- (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.
- (5) Schütt, K. T.; Sauceda, H. E.; Kindermans, P. J.; Tkatchenko, A.; Müller, K. R. SchNet – A deep learning architecture for molecules and materials. *J. Chem. Phys.* **2018**, *148* (24), 241722. Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), No. 145301. Chen, C.; Ye, W.; Zuo, Y.; Zheng, C.; Ong, S. P. Graph Networks as a Universal Machine Learning Framework for Molecules and Crystals. *Chem. Mater.* **2019**, *31* (9), 3564–3572. Willatt, M. J.; Musil, F.; Ceriotti, M. Atom-density representations for machine learning. *J. Chem. Phys.* **2019**, *150* (15), 154110. Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87* (18), No. 184115.
- (6) Ihalage, A.; Hao, Y. Formula Graph Self-Attention Network for Representation-Domain Independent Materials Discovery. *Advanced Science* **2022**, *9* (18), 2200164.
- (7) Graser, J.; Kauwe, S. K.; Sparks, T. D. Machine Learning and Energy Minimization Approaches for Crystal Structure Predictions: A Review and New Horizons. *Chem. Mater.* **2018**, *30* (11), 3601–3612. Mansouri Tehrani, A.; Oliynyk, A. O.; Parry, M.; Rizvi, Z.; Couper, S.; Lin, F.; Miyagi, L.; Sparks, T. D.; Bröggel, J. Machine Learning Directed Search for Ultraincompressible, Superhard Materials. *J. Am. Chem. Soc.* **2018**, *140* (31), 9844–9853. 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. *Chem. Mater.* **2010**, *22* (12), 3762–3767. 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), No. 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), 053213 (accessed 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. *Sci. Rep.* **2016**, *6* (1), 34256. 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. *Chem. Mater.* **2016**, *28* (20), 7324–7331.
- (8) Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **1997**, *9* (8), 1735–1780.
- (9) Cho, K.; Merriënboer, B. v.; Gülcühre, 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. *Nat. Commun.* **2020**, *11* (1), 6280.
- (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.
- (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. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015*, 770778.
- (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.
- (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.
- (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.
- (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: .
- (21) Zaremba, W.; Sutskever, I.; Vinyals, O. Recurrent Neural Network Regularization. *ArXiv* **2014**, DOI: .
- (22) McInnes, L.; Healy, J.; Saul, N.; Großberger, L. UMAP: Uniform Manifold Approximation and Projection. *arXiv* **2018**, *3* (29). DOI: .
- (23) Krizhevsky, A.; Sutskever, I.; Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM* **2017**, *60* (6), 84–90.
- (24) Stergiou, A.; Poppe, R.; Kalliatakis, G. Refining activation downsampling with SoftPool; 2021.
- (25) Dumitrescu, D.; Boianiu, 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: .
- (27) Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y.; Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y.; Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y.; Ferrari, V.; Hebert, M.; Sminchisescu, C.; Weiss, Y. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. *ArXiv* **2018**, DOI: .
- (28) Pearson, K. On lines and planes of closest fit to systems of points in space. *London, Edinburgh Dublin Philos. Mag. J. Sci.* **2010**, *2* (11), 559–572.
- (29) Zhang, Y.; Li, K.; Li, K.; Zhong, B.; Fu, Y. R. Residual Non-local Attention Networks for Image Restoration. *ArXiv* **2019**, DOI: .
- (30) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2016**, *2* (1), 16028.
- (31) Dunn, A.; Wang, Q.; Ganose, A.; Dopp, D.; Jain, A. Benchmarking materials property prediction methods: the Matbench

- 1050 test set and Automatminer reference algorithm. *npj Comput. Mater.*
1051 **2020**, *6* (1), 138.
1052 (32) Choudhary, K.; Garrity, K. F.; Reid, A. C. E.; DeCost, B.; Biacchi,
1053 A. J.; Walker, A. H. R.; Trautt, Z.; Hattrick-Simpers, J.; Kusne, A. G.;
1054 Centrone, A.; et al. The joint automated repository for various
1055 integrated simulations (JARVIS) for data-driven materials design. *npj*
1056 *Comput. Mater.* **2020**, *6* (1), 173.
1057 (33) Petretto, G.; Dwarakanath, S.; P.C. Miranda, H.; Winston, D.;
1058 Giantomassi, M.; van Setten, M. J.; Gonze, X.; Persson, K. A.; Hautier,
1059 G.; Rignanese, G. M. High-throughput density-functional perturbation
1060 theory phonons for inorganic materials. *Sci. Data* **2018**, *5*, 180065.
1061 (34) Petousis, I.; Mrdjenovich, D.; Ballouz, E.; Liu, M.; Winston, D.;
1062 Chen, W.; Graf, T.; Schladt, T. D.; Persson, K. A.; Prinz, F. B. High-
1063 throughput screening of inorganic compounds for the discovery of
1064 novel dielectric and optical materials. *Sci. Data* **2017**, *4*, 160134.
1065 (35) de Jong, M.; Chen, W.; Angsten, T.; Jain, A.; Notestine, R.;
1066 Gamst, A.; Sluiter, M.; Krishna Ande, C.; van der Zwaag, S.; Plata, J. J.;
1067 et al. Charting the complete elastic properties of inorganic crystalline
1068 compounds. *Sci Data* **2015**, *2*, 150009.
1069 (36) Castelli, I. E.; Landis, D. D.; Thygesen, K. S.; Dahl, S.;
1070 Chorkendorff, I.; Jaramillo, T. F.; Jacobsen, K. W. New cubic
1071 perovskites for one- and two-photon water splitting using the
1072 computational materials repository. *ENERGY & ENVIRONMENTAL*
1073 *SCIENCE* **2012**, *5* (10), 9034–9043.
1074 (37) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder,
1075 G.; Persson, K. A. The Materials Application Programming Interface
1076 (API): A simple, flexible and efficient API for materials data based on
1077 REpresentational State Transfer (REST) principles. *COMPUTA-
1078 TIONAL MATERIALS SCIENCE* **2015**, *97*, 209–215.
1079 (38) Zhuo, Y.; Mansouri Tehrani, A.; Brugge, J. Predicting the Band
1080 Gaps of Inorganic Solids by Machine Learning. *J. Phys. Chem. Lett.*
1081 **2018**, *9* (7), 1668–1673.
1082 (39) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.;
1083 Dacek, S. T.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al.
1084 Commentary: The Materials Project: A materials genome approach to
1085 accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002–
1086 011002. Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.;
1087 Mueller, T.; Persson, K. A.; Ceder, G. A high-throughput infrastructure
1088 for density functional theory calculations. *Comput. Mater. Sci.* **2011**, *50*
1089 (8), 2295–2310. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.;
1090 Wolverton, C. Materials Design and Discovery with High-Throughput
1091 Density Functional Theory: The Open Quantum Materials Database
1092 (OQMD). *JOM* **2013**, *65* (11), 1501–1509.
1093 (40) Maas, A. L. Rectifier Nonlinearities Improve Neural Network
1094 Acoustic Models. *2013*.
1095 (41) Oviedo, F.; Ren, Z.; Sun, S.; Settens, C.; Liu, Z.; Hartono, N. T.
1096 P.; Ramasamy, S.; DeCost, B. L.; Tian, S. I. P.; Romano, G.; et al. Fast
1097 and interpretable classification of small X-ray diffraction datasets using
1098 data augmentation and deep neural networks. *npj Computational*
1099 *Materials* **2019**, *5* (1), 60. Ghosh, K.; Stuke, A.; Todorović, M.;
1100 Jørgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke, P. Deep Learning
1101 Spectroscopy: Neural Networks for Molecular Excitation Spectra. *Adv.*
1102 *Sci. (Weinh)* **2019**, *6* (9), 1801367. Dinic, F.; Singh, K.; Dong, T.;
1103 Rezazadeh, M.; Wang, Z.; Khosrozadeh, A.; Yuan, T.; Voznyy,
1104 O. Applied Machine Learning for Developing Next-Generation Func-
1105 tional Materials. *Adv. Funct. Mater.* **2021**, *31*2104195 Ghosh, K.; Stuke,
1106 A.; Todorović, M.; Jørgensen, P. B.; Schmidt, M. N.; Vehtari, A.; Rinke,
1107 P. Deep Learning Spectroscopy: Neural Networks for Molecular
1108 Excitation Spectra. *Adv. Sci* **2019**, *6*1801367