

MOFormer: Self-Supervised Transformer Model for Metal–Organic Framework Property Prediction

Zhonglin Cao, Rishikesh Magar, Yuyang Wang, and Amir Barati Farimani*



Cite This: *J. Am. Chem. Soc.* 2023, 145, 2958–2967



Read Online

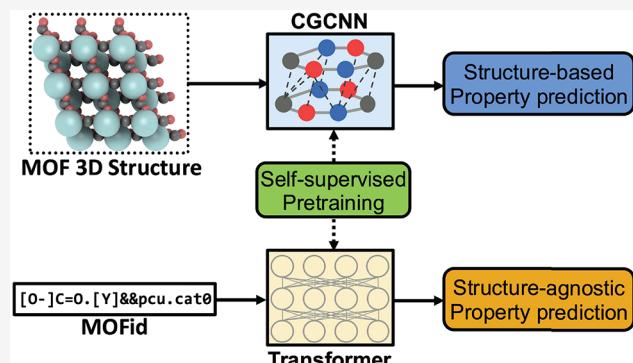
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: Metal–organic frameworks (MOFs) are materials with a high degree of porosity that can be used for many applications. However, the chemical space of MOFs is enormous due to the large variety of possible combinations of building blocks and topology. Discovering the optimal MOFs for specific applications requires an efficient and accurate search over countless potential candidates. Previous high-throughput screening methods using computational simulations like DFT can be time-consuming. Such methods also require the 3D atomic structures of MOFs, which adds one extra step when evaluating hypothetical MOFs. In this work, we propose a structure-agnostic deep learning method based on the Transformer model, named as MOFormer, for property predictions of MOFs. MOFormer takes a text string representation of MOF (MOFid) as input, thus circumventing the need of obtaining the 3D structure of a hypothetical MOF and accelerating the screening process. By comparing to other descriptors such as Stoichiometric-120 and revised autocorrelations, we demonstrate that MOFormer can achieve state-of-the-art structure-agnostic prediction accuracy on all benchmarks. Furthermore, we introduce a self-supervised learning framework that pretrains the MOFormer via maximizing the cross-correlation between its structure-agnostic representations and structure-based representations of the crystal graph convolutional neural network (CGCNN) on >400k publicly available MOF data. Benchmarks show that pretraining improves the prediction accuracy of both models on various downstream prediction tasks. Furthermore, we revealed that MOFormer can be more data-efficient on quantum-chemical property prediction than structure-based CGCNN when training data is limited. Overall, MOFormer provides a novel perspective on efficient MOF property prediction using deep learning.



INTRODUCTION

Metal–organic frameworks (MOFs) are a type of porous crystalline materials,^{1,2} which have been extensively researched during the past several decades. Research interests have been induced by the porous structure and versatile nature of MOFs on their potential applications such as gas adsorption,^{3–5} water harvesting and desalination,^{6–8} and energy storage.^{9–11} MOFs typically consist of several building blocks, including metal nodes and organic linkers.^{4,12,13} The assembly of those building blocks following certain topologies generates the two-dimensional or three-dimensional porous structures of MOFs. Because of the countless possible combinations of metal nodes, organic linkers, and topologies,^{13,14} there is a sheer number of MOFs with different physicochemical properties and surface chemistries. Given the enormous variety of possible MOF structures, rapidly and inexpensively selecting the potential top performers for each specific task can be challenging. High-throughput screening with computational tools such as molecular simulation^{5,15} or density functional theory (DFT)^{16,17} has been widely used to evaluate the properties of MOFs. Without the need to experimentally synthesize MOF structures, those computational tools

accelerate the screening process and allow researchers to screen hundreds of thousands of hypothetical MOF structures^{4,5} for their performance in different applications.

Recently, machine learning (ML) models have become increasingly popular in the field of MOF property prediction.^{18–25} The advantage of the ML models over the simulation methods is their instantaneous inference of the properties of MOFs. In contrast, the simulation methods require a computationally expensive rerun for every new MOF. In the past decade, multiple large-scale MOF data sets are released, including the CoRE MOF 2019,²⁶ hypothetical MOFs,⁵ and QMOF.^{27,28} These data sets contain the atomic structures of MOFs and their computed properties like CO₂ adsorption and band gap. These data sets are large enough to

Received: October 27, 2022

Published: January 27, 2023



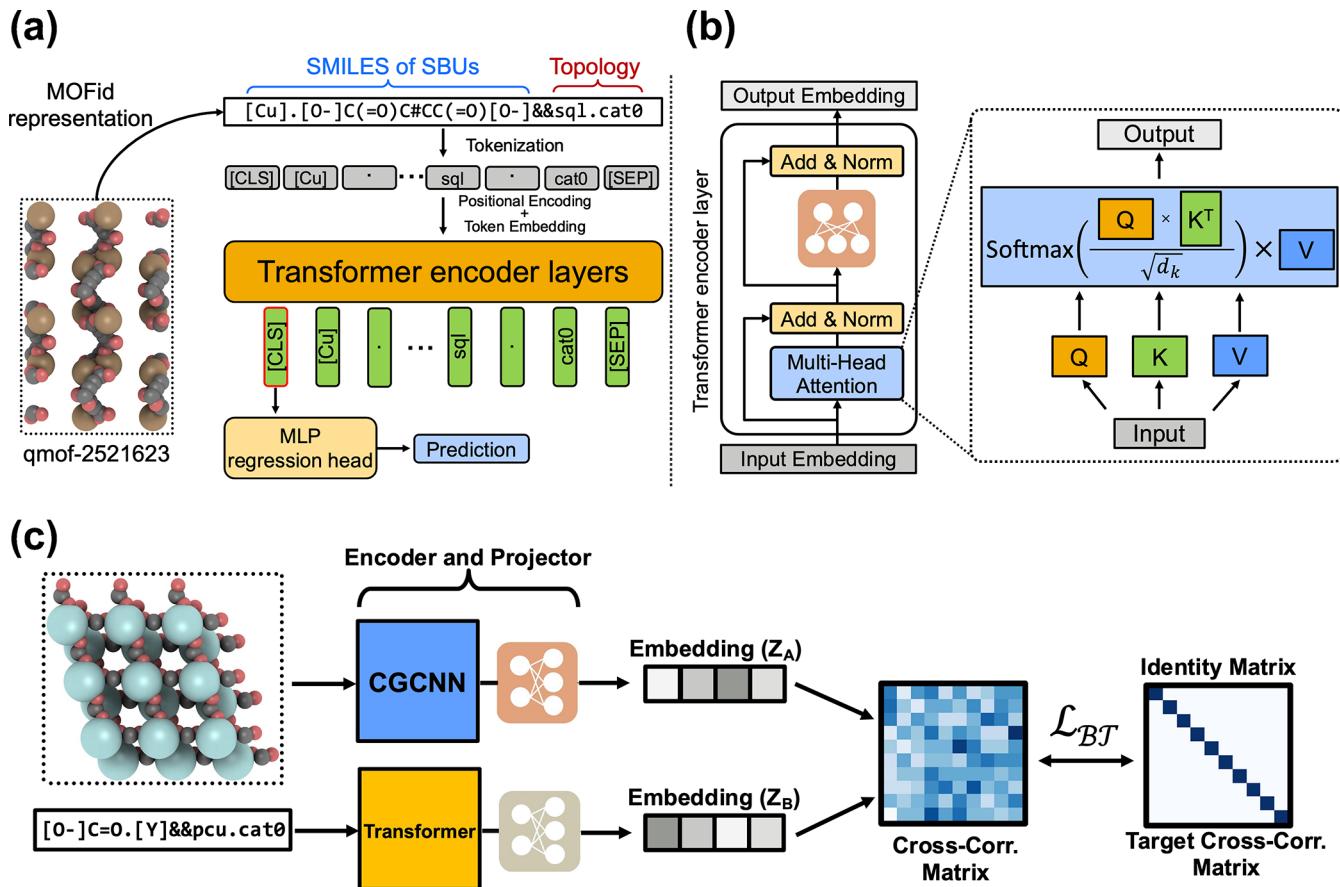


Figure 1. (a) The pipeline of the MOFormer model. A MOFid of a MOF (qmof-2521623 is used as an example) is the input to the model. The MOFid is converted into a tokenized sequence before being embedded and applied with positional encoding. The sequence is then fed into multiple Transformer encoder layers. The learned embedding of the first token will be used as input to an MLP regression head for downstream prediction tasks. (b) A schematic showing the details of each Transformer encoder layer. Embeddings of the sequence pass through the multihead scaled dot-product attention layer and then an MLP. Residue connection and layer normalization are adopted for both the attention and the MLP. (c) The self-supervised learning framework with CGCNN and MOFormer, respectively, for representation learning. The 3D structure and the MOFid of the same MOF are fed into the CGCNN and MOFormer, respectively, for representation learning. The MLP head following each models projects the representations into embeddings (Z_A and Z_B). A cross-correlation matrix is then constructed using the embeddings. Barlow Twins loss is applied to optimize the cross-correlation matrix to be as close as possible to an identity matrix.

train accurate data-driven ML models for the prediction of MOF properties. Handcrafted geometrical features such as large cavity diameter and pore limiting diameter have been used as input to a multilayer perceptron (MLP) to predict MOF properties.^{19,23} Although the training of MLP with a few layers can be fast, such a method suffers from underwhelming accuracy due to the simplicity of network architecture. Moreover, selecting features requires extensive domain knowledge from the researchers and optimized 3D structures of MOFs, thus making this method less generic. Given the aforementioned drawbacks, a novel method that can achieve high accuracy with a more generic input of MOF representations should be pursued. Wang et al.²⁹ utilize the crystal graph convolutional neural network (CGCNN)³⁰ to predict methane adsorption of MOFs. CGCNN is a prevalent model which has an architecture designed specifically for crystalline materials. It takes the element type and the 3D coordinates of atoms in the crystalline materials as input and constructs a crystal graph. CGCNN can extract features that encode rich chemical information through convolution operations on the crystal graph. However, obtaining the 3D structures of MOFs is a necessity when using the structure-based CGCNN model. In addition, some large MOF

structures consist of hundreds or even thousands of atoms, thus rendering crystal graphs of them memory-inefficient.

Enlightened by the fact that all MOFs are combinations of metal nodes, organic linkers, and topologies, Bucior et al.³¹ proposed a text string representation of MOFs called MOFid. The two core sections of a typical MOFid include the chemical information on building blocks and the topology and the catenation of the MOF structure. The building blocks are represented by an extensively used string representation of molecules called SMILES.³² The topology and catenation are each represented by a code adopted from the Reticular Chemistry Structure Resource (RCSR) database.³³ Therefore, MOFid is a concise text string representation of MOFs that preserves the chemical and the majority of the structural information through topology encoding. The MOFid text-based representation enables the application of language ML models that take text string as input for MOF property prediction.

In this work, we proposed and developed a Transformer-based language model for MOF property prediction. Transformer and its variants have become the top choice for natural language processing tasks since the publication in 2017 by Vaswani et al.³⁴ The multihead attention mechanism allows the

Transformer model to learn contextual information in a sequence without suffering from long-range dependency.^{35,36} With its success in processing long sequential data, Transformer and its other variants are also adopted for chemistry or bioinformatics applications such as molecular^{37–39} and protein⁴⁰ property prediction. The Transformer model in our work, named as MOFormer, takes a modified MOFid as input to make predictions of various MOF properties. The advantage of this method is that it does not require the 3D atomic structure of the MOF (structure-agnostic), thus enabling a much faster and more flexible exploration of the hypothetical MOF space. Specifically, MOFormer can be used to estimate properties of MOFs using only a hypothetically created MOFid. Predicting properties such as thermal conductivity is challenging for MOFormer because these properties are highly related to atom connections. However, we demonstrate that MOFormer can be the most accurate structure-agnostic model in predicting other properties such as band gap and gas adsorption. In practice, pretraining the Transformer model in a self-supervised manner^{40–44} can leverage a large quantity of unlabeled data to help the model learn a more robust representation of the sequence and further improve its performance in downstream tasks. To take advantage of pretraining, we also added a self-supervised learning framework, in which the MOFormer and the CGCNN models are jointly pretrained with >400k MOF structures. Benchmarks show that pretraining improves the prediction accuracy of both models. Dimensionality reduction tools are used to visualize the latent representation learned by both models to provide insight into their performance characteristics. Visualization of attention weights in MOFormer demonstrates that MOFormer learns MOF representations based on some key atoms and topology. Lastly, we compared the data efficiency of models to show which one is a better choice when training data is limited.

METHODS

MOFid Tokenization and Transformer. The MOFormer is built upon the encoder part of the Transformer model that takes a tokenized MOFid as input (Figure 1a). The MOFid tokenizer is a customized version of the SMILES tokenizer.⁴⁵ The SMILES strings of all secondary building units (SBUs) of the MOFid are tokenized by the SMILES tokenizer, while the topology and catenation section of the MOFid is separately tokenized based on the topology encoding adopted from RCSB.³³ The tokens from both sections are then connected by a separator token “&&”. The tokenization process follows the BERT⁴¹ to add a [CLS] token and a [SEP] token at the beginning and the end of the sequence to symbolize the start and the end, respectively. Since the tokenized sequences conform to a fixed length of 512, sequences longer than that are truncated, and sequences shorter than that are padded with special tokens [PAD]. None of the MOFids in the QMOF data set have a length over 512 tokens, and only 385 out of 102 858 MOFids (approximately 0.37%) in the hMOF data set have a length greater than 512 after tokenization.

A tokenized sequence is embedded and combined with a positional encoding (Figure 1a) to include information about the relative and absolute position of each token.⁴⁶ The position encoding is calculated by

$$\begin{aligned} \text{PE}_{(pos, 2i)} &= \sin\left(\frac{pos}{10000^{2i/d_{emb}}}\right) \\ \text{PE}_{(pos, 2i+1)} &= \cos\left(\frac{pos}{10000^{2i/d_{emb}}}\right) \end{aligned} \quad (1)$$

where pos is the position of the token in the sequence, i is the index of dimension, and $d_{emb} = 512$ is the embedding dimension. The Transformer model is a deep neural network model built upon the self-attention mechanism (detailed in the Supporting Information). Each of the Transformer encoder layers consists of a multihead attention layer followed by a simple feed-forward multilayer perceptron (MLP). Residue connection⁴⁷ and layer normalization⁴⁸ are adopted for both the attention and the MLP. In each head of the attention layer (Figure 1b), the input sequence embedding X is multiplied with three learnable weight vectors W_q , W_k , and W_v to be converted to the query, key, and value vector (Q , K , V). The scaled dot-product attention A is then calculated by the equation:

$$A = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (2)$$

where d_k is the dimension of Q and K (detailed in the Supporting Information). The randomly initialized W_q , W_k , and W_v vectors in each head allow the model to learn the contextual information between tokens in different representation subspaces.³⁴ Attentions from all heads are concatenated together and then fed into the MLP for the projected output embedding, which has the same size as the input embedding. Given that the self-attention mechanism can incorporate the information on the whole sequence into each one of the token embeddings, theoretically, any one of the embeddings can be used as a representation of the whole sequence. Therefore, we followed the common practice of related works^{37,38,49,50} to use the embedding of the first token, [CLS], for further supervised learning tasks. The MOFormer model in this work contains six encoder layers. A smaller model with three layers has been benchmarked on the QMOF data set to show it has lower accuracy than the six-layer model (Table S4), thus leading us to select the six-layer model.

Self-Supervised Pretraining with CGCNN. We introduce a self-supervised learning (SSL) paradigm for MOF representation learning. We designed the framework by taking into consideration two modalities of data including the text and graph information. One of the modalities is the text string representation (MOFid) that encapsulates building blocks' stoichiometry and bonds (SMILES) and the topology of the MOF. The text string information is processed by the MOFormer. One of the limitations of text string data is the lack of information about the geometry and the neighborhood of atoms creating an information bottleneck for the text-string-input-based models. The structure-agnostic nature of the text string input can prevent the MOFormer from achieving higher performance than the graph-based models. To mitigate such a limitation of the MOFormer framework, we introduce SSL pretraining with CGCNN.³⁰ The CGCNN model takes as input the 3D atomic structure of the MOF. The input to CGCNN contains the chemical information on all atoms in a MOF and the structure information in atomic resolution which is critical in property prediction tasks. To implement the SSL pipeline, we take inspiration from the Crystal Twins (CT) framework.⁵¹ The CT model makes use of the Barlow Twins loss function introduced by Zbontar et al.⁵² and SimSiam loss⁵³ functions. In this work, we use the Barlow Twins loss function on the embeddings generated from the MOFormer and CGCNN encoder. As shown in Figure 1C, we initially encode both the text string representation and graph representation with their respective encoders. The MOFormer will encode the text string representation, and the CGCNN will encode the graph representation. We generate an embedding of size 512 from both the encoders and use it to generate the cross-correlation matrix following eq 3

$$C_{ij} \stackrel{\Delta}{=} \frac{\sum_b Z_{b,i}^A Z_{b,j}^B}{\sqrt{(Z_{b,i}^A)^2} \sqrt{(Z_{b,j}^B)^2}} \quad (3)$$

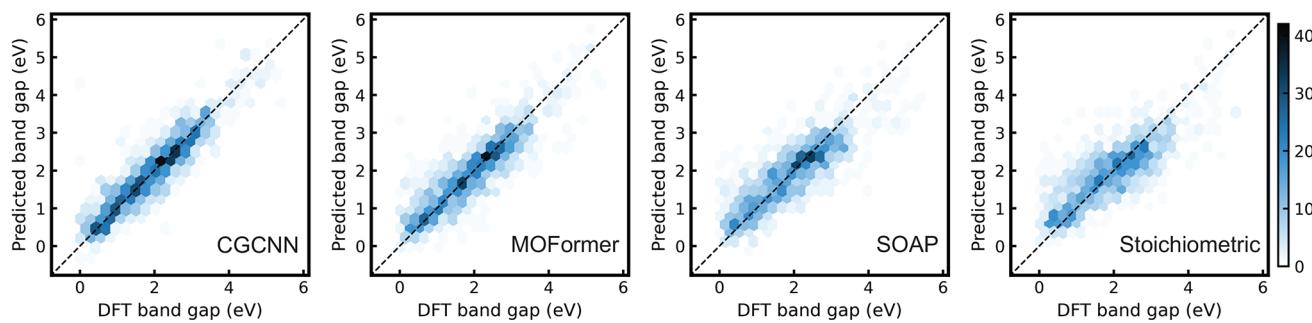
where b is the batch index and i, j index the 512-dimensional output from the projector (Z^A and Z^B), A is the graph representation, and B is text representation for the same MOF. Ideally, we want cross-correlation to be close to the identity matrix as both the

Table 1. Benchmark Performance of Different Models on the Band Gap Prediction of the QMOF Data Set^a

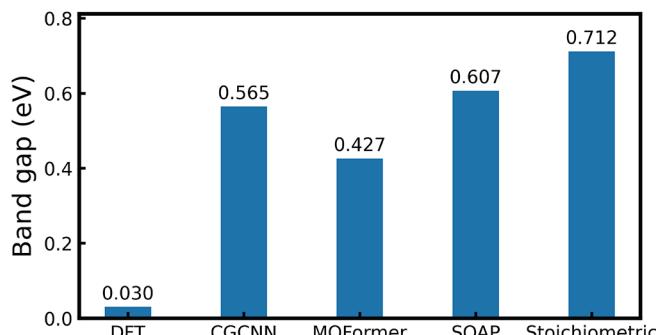
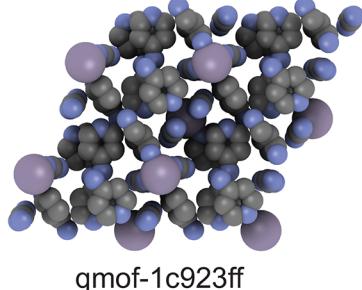
	CGCNN _{scratch}	CGCNN _{pretrain}	SOAP	MOFormer _{scratch}	MOFormer _{pretrain}	Stoichiometric-120	RACs
MAE (eV)	0.275 ± 0.015	0.256 ± 0.006	0.424 ± 0.007	0.387 ± 0.001	0.367 ± 0.005	0.466 ± 0.011	0.441 ± 0.008

^aMean absolute error (MAE, in the unit of eV) and standard deviation of three runs of different initial seeds of each model are reported. The left three models are structure-based, and the right four models are structure-agnostic. The best performance of each category is marked as bold.

(a)



(b)



(c)

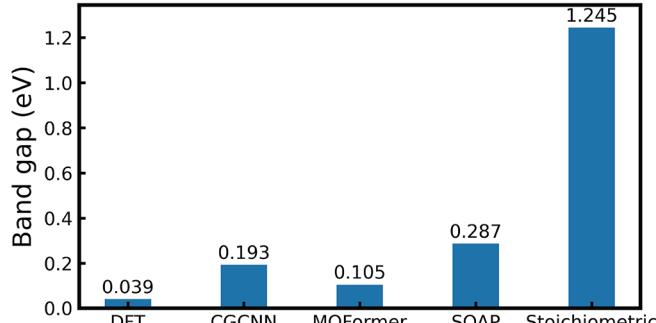
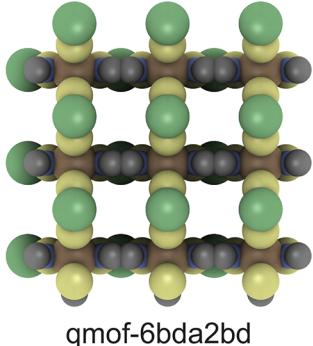


Figure 2. (a) The binned scatter plot shows the comparison between predicted and DFT-calculated band gap for MOFs in the QMOF data set. MOFs included in this figure are from the randomly split test set. The darker color of each hexagonal bin represents more data points in the bin. The dashed line represents perfect prediction. (b,c) Visualization⁶⁴ of the MOF structure with lowest (qmof-1c923ff) and the second lowest (qmof-6bda2bd) band gap in the test set. The bar plot shows the comparison between predictions made by different models.

representations generated from MOFormer and CGCNN are essentially capturing the same MOF. The Barlow Twins loss function, which we used for SSL pretraining (eq 4), tries to force the cross-correlation matrix to be the identity matrix.

$$L_{BT} \stackrel{\Delta}{=} \sum_i (1 - C_{ii})^2 + \lambda \sum_i \sum_{j \neq i} C_{ij}^2 \quad (4)$$

where C is the cross-correlation matrix of embeddings from the MOFormer and CGCNN; the cross-correlation matrix is given by eq 3. The λ used in this work is set to 0.0051.

Finally, after pretraining the models using SSL, the encoder weights are shared during the finetuning stage (Figure S1). The pretraining hyperparameter details are shown in Table S3. For finetuning, we initialize the model with pretrained weights and train the model for

200 epochs to generate the final prediction (Hyperparameters: Tables S1 and S2). The MOFormer and CGCNN models are finetuned separately. We observed that using an SSL pretraining framework improves the results of both CGCNN and MOFormer consistently for all the data sets.

Data Sets and Other Featurizations. Three public MOF data sets including the CORE MOF 2019,²⁶ the hypothetical MOFs⁵ (hMOF), and the Boyd&Woo⁴ are combined to create a large data set for the SSL pretraining. The pretraining data set only includes MOFs with both 3D structure and MOFid available. Since we consider the MOFid as a unique descriptor of each MOF, identical MOFids are defined as duplicated and are removed from the pretraining data set. After the removal of duplicated MOFs, the final pretraining data set has 413 535 unique MOFs. In the downstream prediction task, the

Table 2. Benchmark Performance of Different Models on Gas Adsorption Prediction of the hMOF Data Set^a

	CO ₂ 0.05 bar	CO ₂ 0.5 bar	CO ₂ 2.5 bar	CH ₄ 0.05 bar	CH ₄ 0.5 bar	CH ₄ 2.5 bar
CGCNN _{scratch}	0.126 ± 0.005	0.391 ± 0.017	0.818 ± 0.050	0.028 ± 0.001	0.121 ± 0.006	0.333 ± 0.017
CGCNN _{pretrain}	0.110 ± 0.001	0.330 ± 0.002	0.645 ± 0.003	0.025 ± 0.001	0.099 ± 0.001	0.258 ± 0.008
SOAP	0.115 ± 0.002	0.339 ± 0.004	0.666 ± 0.003	0.022 ± 0.001	0.106 ± 0.001	0.239 ± 0.002
MOFormer _{scratch}	0.178 ± 0.002	0.558 ± 0.001	1.000 ± 0.013	0.034 ± 0.000	0.174 ± 0.002	0.385 ± 0.003
MOFormer _{pretrain}	0.158 ± 0.001	0.545 ± 0.008	0.982 ± 0.011	0.033 ± 0.000	0.161 ± 0.011	0.384 ± 0.003
Stoichiometric-120	0.282 ± 0.002	0.983 ± 0.005	1.895 ± 0.003	0.050 ± 0.001	0.269 ± 0.001	0.631 ± 0.002
RACs	0.248 ± 0.002	0.842 ± 0.004	1.681 ± 0.004	0.044 ± 0.001	0.236 ± 0.002	0.570 ± 0.004

^aMean absolute error (mol kg⁻¹) and standard deviation of three runs of different initial seeds of each model are reported. The top three models are structure-based, and the bottom four models are structure-agnostic. The best performance of each category is marked as bold.

MOFormer and the CGCNN are trained on the quantum MOF^{27,28} (QMOF) and hMOF in a supervised manner. The QMOF data set contains 20 375 MOFs each with a label of a DFT-calculated band gap in eV. Only 7466 MOFs in the QMOF data set have a MOFid available. On the other hand, the hMOF has 137 652 MOFs, of which 102 858 have an available MOFid. The models are trained on hMOF with the labels of CO₂ and CH₄ adsorption in mol kg⁻¹ at 0.05, 0.5, and 2.5 bar of pressure. The benchmark data sets are split into training, test, and validation sets with a ratio of 0.7–0.15–0.15. During the training, the model with the best validation performance is recorded and then tested with the test set. According to the splitting rule, MOFormer has 5226–1119–1119 QMOF data and 72000–15428–15428 hMOF data, while other models have 14262–3056–3056 QMOF data and 96356–20647–20647 hMOF data in the training, validation, and test sets, respectively. Although the MOFs with an available MOFid form a subset of both benchmark data sets, the subset with MOFid shares the same distribution and has approximately the same mean and standard deviation compared with the original whole data set (Figures S2 and S3 in the Supporting Information). Therefore, it is fair to compare the performance of MOFormer and other models.

We also benchmarked the MOFormer and CGCNN against other non-DL-based featurization methods such as the Smooth Overlap of Atomic Positions^{54–56} (SOAP) and the Stoichiometric-120⁵⁷ features. SOAP is a structure-based featurization method, and the Stoichiometric-120 is a structure-agnostic featurization method based on the statistical properties of the MOF's stoichiometric formula. The parameters used for creating SOAP features are included in the Table S5. It is worth mentioning that the SOAP matrix of each MOF is converted into a single feature vector using the inner average. In addition to SOAP and Stoichiometric-120, we also benchmarked the performance of the revised autocorrelations (RACs^{13,58,59}) descriptor of MOFs. RACs are a descriptor based on the crystal graph and atom properties of MOFs. Since RACs do not require the 3D Cartesian coordinates of atoms as input, they can be considered as a structure-agnostic descriptor. RACs of MOFs are obtained using the mofdscribe⁶⁰ package. XGBoost⁶¹ model is used to make predictions using those handcrafted features.

RESULTS AND DISCUSSION

QMOF. The first data set we benchmark models on is the QMOF data set, in which the label for each MOF is the DFT-calculated band gap. A lower band gap value results in better conductivity of the MOF. Accurate prediction of the band gap can help to identify conductive MOFs which are useful in energy storage applications.^{62,63} The accuracy of models follows the rank of CGCNN > MOFormer > SOAP > RACs > Stoichiometric-120 (Table 1). MOFormer has a 21.2 and 16.9% lower MAE compared with the Stoichiometric-120 and RACs, respectively. It is worth noting that structure-agnostic MOFormer outperforms structure-based SOAP with a smaller size of the training set, indicating that MOFormer is capable of extracting critical features from the MOFid for energy-related property prediction. The pretraining helps to reduce the mean

absolute error (MAE) of CGCNN by 6.79% and MOFormer by 5.34%. The reduced error proves the improvement brought by the pretraining.

To better understand the superior performance of MOFormer and CGCNN in QMOF, we trained the four models with the same training set and then examined their performance on the same test set consisting of 1119 randomly selected MOFs. The binned scatter plot (Figure 2a) shows the comparison between the predicted and the DFT-calculated band gap. A darker color means more data fall in the bin. More predictions made by CGCNN and MOFormer are closer to the ground truth, especially for MOFs with a band gap ≤2 eV. The SOAP and Stoichiometric-120 are more likely to overpredict the lower band gap. This weakness of SOAP and Stoichiometric-120 can also be confirmed by the kernel density estimation of predicted values (Figure S4). The MOFs with the top-two lowest band gaps in this test set are the qmof-1c923ff (0.03 eV) and qmof-6bda2bd (0.039 eV). Band gaps predicted by MOFormer and CGCNN are much closer to the DFT-calculated value than predictions by SOAP and Stoichiometric-120 (Figure 2b,c), especially for qmof-6bda2bd. Accurately predicting the low band gap of MOFs can lead to the discovery of a conductive MOF, rendering pretrained MOFormer and CGCNN more valuable for prescreening MOFs.

hMOF. The models are also benchmarked on the hMOF data set with the labels of CO₂ and CH₄ adsorption under 0.05, 0.5, and 2.5 bar of pressure. Table 2 shows that pretrained MOFormer is constantly outperforming Stoichiometric-120 by achieving a 35–48% lower MAE. Pretrained MOFormer also achieves a 25–42% lower MAE than RACs. Pretrained CGCNN outperforms other models for the CO₂ adsorption prediction. The pretraining in average improves the accuracy of MOFormer by 4.3% and the CGCNN by 16.5% over all gas adsorption predictions. When obtaining the structure is relatively fast (e.g., using molecular mechanics optimization with UFF), CGCNN bears the promise for accurate gas adsorption prediction, which can be further improved by pretraining with MOFormer. It is worth mentioning that the prediction accuracy of MOFormer does not significantly drop with overlength MOFid (Figure S5a). SOAP has surprisingly low MAE for the gas adsorption prediction, outperforming pretrained CGCNN for two of the three CH₄ adsorption predictions and the CGCNN trained from scratch for all gas adsorption predictions. The outstanding performance of SOAP on hMOF can be attributed to the low variation of elements included in the hMOF data set. Only 11 different elements are present among all 137 652 hMOFs, which is very limited compared to 79 in the QMOF data set. A smaller number of elements results in a much

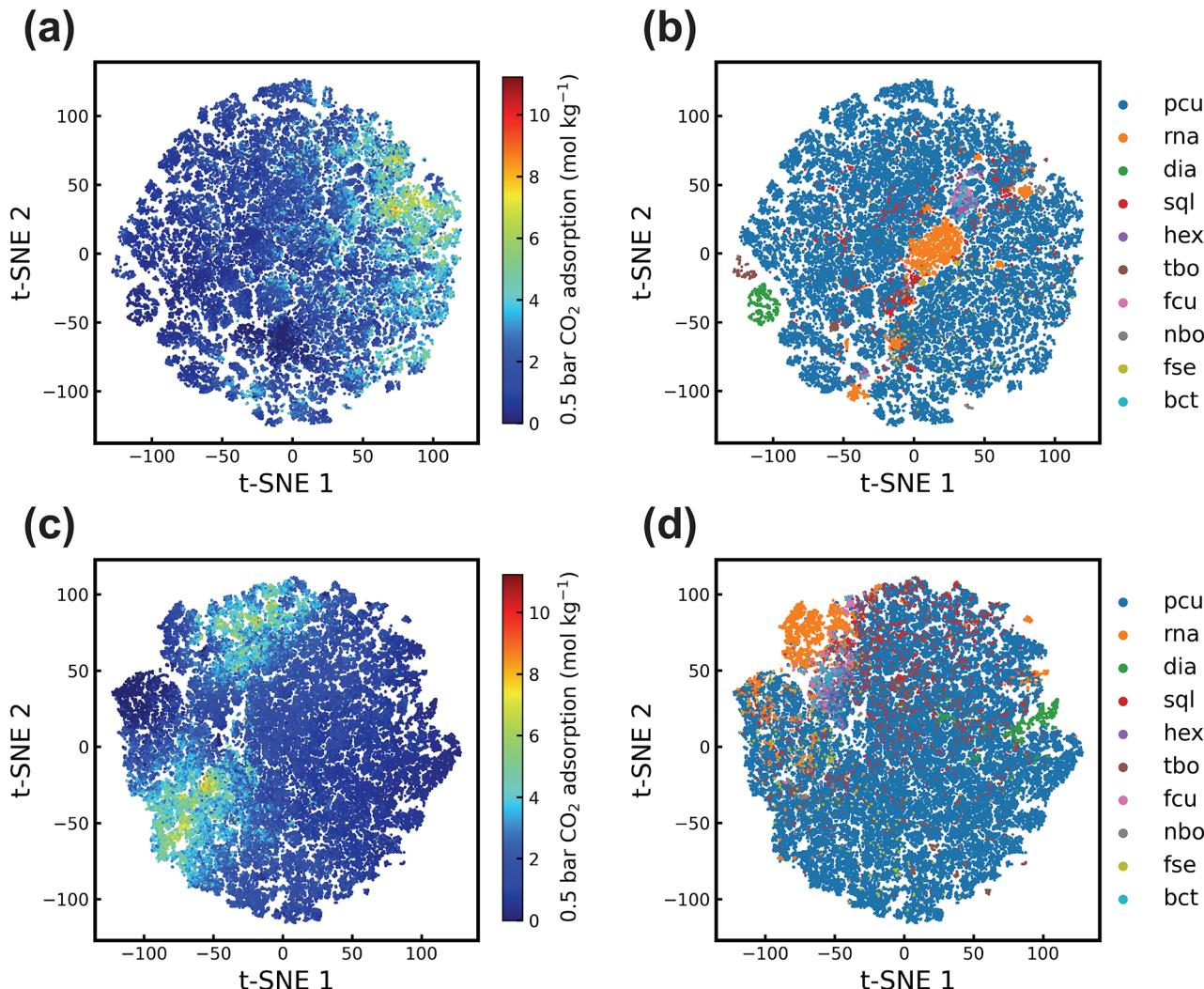


Figure 3. T-SNE⁶⁵ dimension-reduced visualization of MOF representations learned by (a,b) the MOFormer and (c,d) the CGCNN. Each data point in (a) and (c) is colored by its CO_2 adsorption at 0.5 bar of pressure, and each data point in (b) and (d) is colored by its topology. Only the MOF which has the top-10 most common topologies in the hMOF data set is shown.

smaller and less sparse SOAP feature vector (SOAP feature has a size of 2772 for hMOFs and 19 908 for QMOFs with the same parameters), thus leading to the high prediction accuracy of the following XGBoost regressor. However, the high accuracy of SOAP can hardly be sustainable when it is used in exploring more diverse hypothetical MOFs. When more elements are included in the data set, the SOAP feature vector size and sparsity increase drastically, rendering the data and model too large to be accommodated by the memory of local machines and a drop in prediction accuracy (Table S6). MOFormer and CGCNN will not suffer from such an issue, since their inputs remain invariant with increasing types of elements in the data set, making them better choices when exploring more diverse chemical space for MOFs.

The representations of MOFs learned by the MOFormer and CGCNN after finetuning are visualized to provide interpretability to the models (Figure 3). Each representation is projected to the 2D space using the dimension reduction tool t-SNE.⁶⁵ t-SNE clusters more similar data points together while placing less similar data points further away. Only MOFs which have the top-10 most common topologies in hMOF are included in Figure 3 because they take >99.7% of the whole data set. We can observe that CGCNN representations cluster

MOFs with high CO_2 adsorption more closely than MOFormer representations by comparing Figure 3a and 3c. This contributes to the higher prediction accuracy of CGCNN. On the other hand, MOFormer representation clusters MOFs with the same topology closer than CGCNN representation does. For example, the MOFs with dia (green) and tbo (brown) topologies form two clusters in the lower left corner of the MOFormer representation visualization (Figure 3b). Those MOFs are much more loosely clustered in the CGCNN representation visualization (Figure 3d). The MOFormer representation focusing on topology can be caused by the fact that gas adsorption is more dependent on the 3D structure of the MOF compared with its atom composition. The only structure-related information contained in the MOFid is the topology encoding. Therefore, more weights are on the topology after MOFormer is finetuned to predict the gas adsorption of MOFs. The input of CGCNN is the 3D structure of MOFs with atomic resolution; thus, CGCNN can rely less on the topology for gas adsorption prediction. The MOFormer model representations may fail to accurately predict the properties of MOFs with rare topologies (Figure S5b). Such a disadvantage can be alleviated by in the future increasing the topology diversity in the training data set.

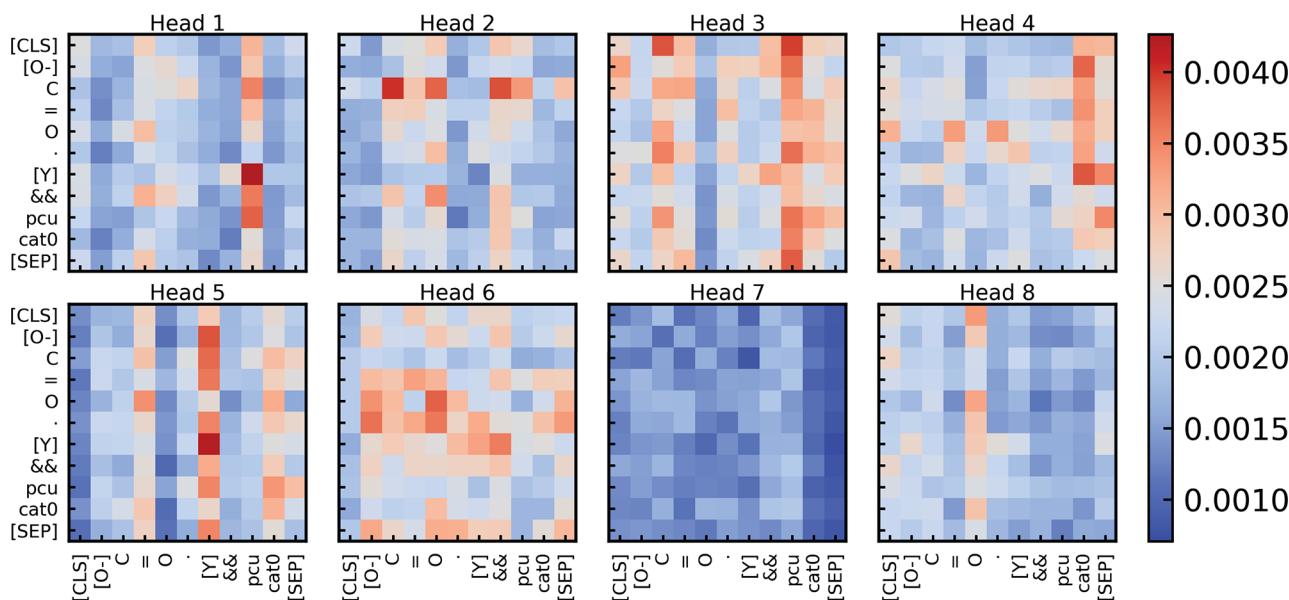


Figure 4. Heatmap of the attention between tokens (MOFid of qmof-ba40858) in different heads of the last MOFormer layer. We index each block in the heatmap as block_{n,ij}, where i, j, and n are the row, column, and head index, respectively. The value of block_{n,ij} represents the attention on the j-th token from the i-th token in n-th heatmap.

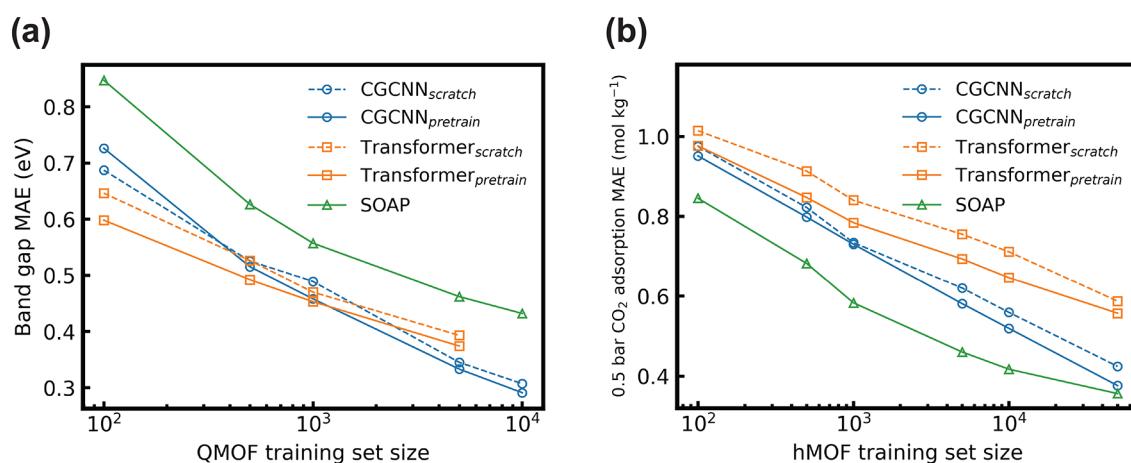


Figure 5. Data efficiency comparison between different models on the (a) QMOF and the (b) hMOF data set. The models are trained on a subset of the training set, while the validation and test set are kept the same. The subset sizes are 100, 500, 1000, 5000, 10 000, and 50 000 (hMOF only). Since only less than 7500 MOFs in QMOF have an available MOFid, the maximum training subset size for MOFormer on QMOF is 5000. Each data point is averaged over the results of three runs on randomly selected subsets drawn with different initial seeds.

Visualization of Attention Weights. Figure 4 demonstrates the attention maps between tokens of a MOFid (qmof-ba40858) in the last MOFormer layer after finetuning on band gap prediction. The attention map can serve as a visual interpretation of how the MOFormer learns MOF representations.⁶⁶ We observe a strong attention in head 5 from all tokens to the metal node ytterium, and to the topology encoding pcu in heads 1 and 3. The attention from the metal node to the topology encoding is especially high in head 1. Moreover, a large attention weight can be observed between tokens in the SMILES of the SBUs in head 6. Heads 1, 3, 5, and 8 also show large attention weights on the carbon and the oxygen atom and the double bond in the organic building block. The attention weight visualization shows that MOFormer learns a representation that emphasizes the contextual information between key components in the MOFid including

important atoms (e.g., Y, O, and C) and the topology, thus leading to more accurate prediction.

Data Efficiency Comparison. Obtaining high-quality MOF data using experimental or DFT methods can be time-consuming and expensive. A model with high data efficiency is ideal when the training data size is limited. We compared the data efficiency of different models on the QMOF and hMOF data sets (CO₂ adsorption at 0.5 bar pressure). For band gap prediction (Figure 5a), the pretrained MOFormer outperforms CGCNN when the training set size \leq 1000. This makes MOFormer more valuable in predicting quantum-chemical properties when the training data set is difficult to build (i.e., experimentally synthesized MOFs). Both MOFormer and CGCNN achieve higher accuracy than SOAP regardless of the training set size on the QMOF data set. For CO₂ adsorption prediction (Figure 5b), CGCNN constantly achieves higher accuracy than MOFormer regardless of the training set size,

indicating its higher data efficiency. CGCNN outperforms MOFormer on hMOF because the CO₂ adsorption correlates more with the MOF structure and the input to CGCNN provides more structural information than MOFid. SOAP achieves higher data efficiency than CGCNN and MOFormer on hMOF but is eventually caught up by CGCNN after the training set size exceeds 50 000. Figure 5a,b shows that pretraining consistently improves the data efficiency of MOFormer and CGCNN. Moreover, SOAP is shown to have diminishing improvement with increasing training set size, but CGCNN and MOFormer do not suffer from such an issue.

CONCLUSION

In summary, we propose a Transformer-based model, named as MOFormer, for structure-agnostic MOF property prediction. Taking only MOFid as input, the MOFormer model is expected to expedite the exploration of hypothetical MOFs. We also introduce a self-supervised learning framework to jointly pretrain the MOFormer and CGCNN model on a large unlabeled MOF data set to enhance their prediction accuracy in downstream tasks. Compared with other structure-agnostic methods Stoichiometric-120 and RACs, MOFormer achieves 21.4% and 16.9% higher accuracy on band gap prediction as well as 35–48% and 25–42% higher accuracy on various gas adsorption prediction tasks, respectively. MOFormer even outperforms the structure-based SOAP method in band gap prediction with less training data. The pretraining is further shown to improve the accuracy of MOFormer by 5.34% and 4.3% on average and CGCNN by 6.79% and 16.5% on average, for band gap and gas adsorption prediction, respectively. MOFormer and CGCNN are shown to be less likely to overpredict the band gap of MOFs compared with SOAP and Stoichiometric-120, making them better choices for prescreening conductive MOFs for energy applications. When used for gas adsorption prediction of MOFs, MOFormer relies more on the topology information compared with CGCNN because of the strong correlation between the label and the structure of MOF. Visualization of the attention weights in the last MOFormer layer reveals that the attention layers in MOFormer focus more on several important atoms and the topology to learn the representation of a MOF. Lastly, MOFormer is shown to be more data-efficient than CGCNN for band gap prediction when the training set size ≤ 1000 . As a structure-agnostic model, MOFormer can make rapid and accurate inferences on the property of MOFs (especially for quantum-chemical properties) using an arbitrarily constructed MOFid as input. Therefore, MOFormer can serve as a tool for exploring the vast chemical space of hypothetical MOFs.

ASSOCIATED CONTENT

Data Availability Statement

The Python code as well as data used in this work can be found on GitHub: <https://github.com/zcao0420/MOFormer>.

Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.2c11420>.

Transformer and self-attention mechanism. Details of CGCNN and MOFormer model. Details of the self-supervised pretraining. Effect of model size on prediction accuracy. Distribution of the benchmark data sets. Parameters for SOAP feature vector creation

and the effect of SOAP vector length on model accuracy. Kernel density estimation of band gap prediction from different models. Prediction accuracy with extra long MOFid or rare topology (PDF)

AUTHOR INFORMATION

Corresponding Author

Amir Barati Farimani – Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States; Department of Chemical Engineering and Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;  orcid.org/0000-0002-2952-8576; Email: barati@cmu.edu

Authors

Zhonglin Cao – Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;  orcid.org/0000-0003-2096-1178

Rishikesh Magar – Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;  orcid.org/0000-0001-6216-0518

Yuyang Wang – Department of Mechanical Engineering, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, United States;  orcid.org/0000-0003-0723-6246

Complete contact information is available at:

<https://pubs.acs.org/10.1021/jacs.2c11420>

Author Contributions

Z.C. and R.M. hold joint first authorship.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by the start-up fund from Mechanical Engineering Department at CMU.

REFERENCES

- (1) James, S. L. Metal-organic frameworks. *Chem. Soc. Rev.* **2003**, *32*, 276–288.
- (2) Zhou, H.-C.; Long, J. R.; Yaghi, O. M. Introduction to metal-organic frameworks. *Chem. Rev.* **2012**, *112*, 673–674.
- (3) Ahmed, A.; Seth, S.; Purewal, J.; Wong-Foy, A. G.; Veenstra, M.; Matzger, A. J.; Siegel, D. J. Exceptional hydrogen storage achieved by screening nearly half a million metal-organic frameworks. *Nat. Commun.* **2019**, *10*, 1568.
- (4) Boyd, P. G.; Chidambaram, A.; García-Díez, E.; Ireland, C. P.; Daff, T. D.; Bounds, R.; G ladysiak, A.; Schouwink, P.; Moosavi, S. M.; Maroto-Valer, M. M.; et al. Data-driven design of metal-organic frameworks for wet flue gas CO₂ capture. *Nature* **2019**, *576*, 253–256.
- (5) Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-scale screening of hypothetical metal-organic frameworks. *Nature Chem.* **2012**, *4*, 83–89.
- (6) Almassad, H. A.; Abaza, R. I.; Siwwan, L.; Al-Maythalony, B.; Cordova, K. E. Environmentally adaptive MOF-based device enables continuous self-optimizing atmospheric water harvesting. *Nat. Commun.* **2022**, *13*, 4873.
- (7) Hanikel, N.; Prévet, M. S.; Yaghi, O. M. MOF water harvesters. *Nature Nanotechnol.* **2020**, *15*, 348–355.
- (8) Cao, Z.; Liu, V.; Barati Farimani, A. Water desalination with two-dimensional metal-organic framework membranes. *Nano Lett.* **2019**, *19*, 8638–8643.

- (9) Baumann, A. E.; Burns, D. A.; Liu, B.; Thoi, V. S. Metal-organic framework functionalization and design strategies for advanced electrochemical energy storage devices. *Communications Chemistry* **2019**, *2*, 86.
- (10) Zhao, Y.; Song, Z.; Li, X.; Sun, Q.; Cheng, N.; Lawes, S.; Sun, X. Metal organic frameworks for energy storage and conversion. *Energy storage materials* **2016**, *2*, 35–62.
- (11) Xu, G.; Nie, P.; Dou, H.; Ding, B.; Li, L.; Zhang, X. Exploring metal organic frameworks for energy storage in batteries and supercapacitors. *Mater. Today* **2017**, *20*, 191–209.
- (12) Sharp, C. H.; Bukowski, B. C.; Li, H.; Johnson, E. M.; Ilic, S.; Morris, A. J.; Gersappe, D.; Snurr, R. Q.; Morris, J. R. Nanoconfinement and mass transport in metal–organic frameworks. *Chem. Soc. Rev.* **2021**, *50*, 11530–11558.
- (13) Moosavi, S. M.; Nandy, A.; Jablonka, K. M.; Ongari, D.; Janet, J. P.; Boyd, P. G.; Lee, Y.; Smit, B.; Kulik, H. J. Understanding the diversity of the metal-organic framework ecosystem. *Nat. Commun.* **2020**, *11*, 4068.
- (14) Falcaro, P.; Hill, A. J.; Nairn, K. M.; Jasieniak, J.; Mardel, J. I.; Bastow, T. J.; Mayo, S. C.; Gimona, M.; Gomez, D.; Whitfield, H. J.; et al. A new method to position and functionalize metal-organic framework crystals. *Nat. Commun.* **2011**, *2*, 237.
- (15) Ren, E.; Guilbaud, P.; Coudert, F.-X. High-throughput computational screening of nanoporous materials in targeted applications. *Digital Discovery* **2022**, *1*, 355–374.
- (16) Canepa, P.; Arter, C. A.; Conwill, E. M.; Johnson, D. H.; Shoemaker, B. A.; Soliman, K. Z.; Thonhauser, T. High-throughput screening of small-molecule adsorption in MOF. *Journal of Materials Chemistry A* **2013**, *1*, 13597–13604.
- (17) Rosen, A. S.; Notestein, J. M.; Snurr, R. Q. Identifying promising metal–organic frameworks for heterogeneous catalysis via high-throughput periodic density functional theory. *Journal of computational chemistry* **2019**, *40*, 1305–1318.
- (18) Fung, V.; Zhang, J.; Juarez, E.; Sumpter, B. G. Benchmarking graph neural networks for materials chemistry. *npj Computational Materials* **2021**, *7*, 84.
- (19) Burner, J.; Schwiedrzik, L.; Krykunov, M.; Luo, J.; Boyd, P. G.; Woo, T. K. High-Performing Deep Learning Regression Models for Predicting Low-Pressure CO₂ Adsorption Properties of Metal–Organic Frameworks. *J. Phys. Chem. C* **2020**, *124*, 27996–28005.
- (20) Altintas, C.; Altundal, O. F.; Keskin, S.; Yildirim, R. Machine learning meets with metal organic frameworks for gas storage and separation. *J. Chem. Inf. Model.* **2021**, *61*, 2131–2146.
- (21) Lee, S.; Kim, B.; Cho, H.; Lee, H.; Lee, S. Y.; Cho, E. S.; Kim, J. Computational screening of trillions of metal–organic frameworks for high-performance methane storage. *ACS Appl. Mater. Interfaces* **2021**, *13*, 23647–23654.
- (22) Choudhary, K.; Yildirim, T.; Siderius, D. W.; Kusne, A. G.; McDannald, A.; Ortiz-Montalvo, D. L. Graph neural network predictions of metal organic framework CO₂ adsorption properties. *Comput. Mater. Sci.* **2022**, *210*, 111388.
- (23) Moghadam, P. Z.; Rogge, S. M.; Li, A.; Chow, C.-M.; Wieme, J.; Moharrami, N.; Aragones-Anglada, M.; Conduit, G.; Gomez-Gualdrón, D. A.; Van Speybroeck, V.; et al. Structure-mechanical stability relations of metal-organic frameworks via machine learning. *Matter* **2019**, *1*, 219–234.
- (24) Nandy, A.; Duan, C.; Kulik, H. J. Using Machine Learning and Data Mining to Leverage Community Knowledge for the Engineering of Stable Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2021**, *143*, 17535–17547.
- (25) Moosavi, S. M.; Novotny, B. Á.; Ongari, D.; Moubarak, E.; Asgari, M.; Kadioglu, Ö.; Charalambous, C.; Ortega-Guerrero, A.; Farmahini, A. H.; Sarkisov, L.; Garcia, S.; Noé, F.; Smit, B. A data-science approach to predict the heat capacity of nanoporous materials. *Nat. Mater.* **2022**, *21*, 1419–1425.
- (26) Chung, Y. G.; Haldoupis, E.; Bucior, B. J.; Haranczyk, M.; Lee, S.; Zhang, H.; Vogiatzis, K. D.; Milisavljevic, M.; Ling, S.; Camp, J. S.; et al. Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019. *Journal of Chemical & Engineering Data* **2019**, *64*, 5985–5998.
- (27) Rosen, A. S.; Iyer, S. M.; Ray, D.; Yao, Z.; Aspuru-Guzik, A.; Gagliardi, L.; Notestein, J. M.; Snurr, R. Q. Machine learning the quantum-chemical properties of metal–organic frameworks for accelerated materials discovery. *Matter* **2021**, *4*, 1578–1597.
- (28) Rosen, A. S.; Fung, V.; Huck, P.; O'Donnell, C. T.; Horton, M. K.; Truhlar, D. G.; Persson, K. A.; Notestein, J. M.; Snurr, R. Q. High-throughput predictions of metal–organic framework electronic properties: theoretical challenges, graph neural networks, and data exploration. *npj Computational Materials* **2022**, *8*, 112.
- (29) Wang, R.; Zhong, Y.; Bi, L.; Yang, M.; Xu, D. Accelerating Discovery of Metal–Organic Frameworks for Methane Adsorption with Hierarchical Screening and Deep Learning. *ACS Appl. Mater. Interfaces* **2020**, *12*, 52797–52807.
- (30) Xie, T.; Grossman, J. C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters* **2018**, *120*, 145301.
- (31) Bucior, B. J.; Rosen, A. S.; Haranczyk, M.; Yao, Z.; Ziebel, M. E.; Farha, O. K.; Hupp, J. T.; Siepmann, J. I.; Aspuru-Guzik, A.; Snurr, R. Q. Identification schemes for metal–organic frameworks to enable rapid search and cheminformatics analysis. *Cryst. Growth Des.* **2019**, *19*, 6682–6697.
- (32) Weininger, D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of chemical information and computer sciences* **1988**, *28*, 31–36.
- (33) O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The reticular chemistry structure resource (RCSR) database of, and symbols for, crystal nets. *Accounts of chemical research* **2008**, *41*, 1782–1789.
- (34) Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, 5998–6008.
- (35) Bahdanau, D.; Cho, K.; Bengio, Y. Neural machine translation by jointly learning to align and translate. *arXiv*, 2014; arXiv:1409.0473 [cs.CL].
- (36) Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
- (37) Schwaller, P.; Laino, T.; Gaudin, T.; Bolgar, P.; Hunter, C. A.; Bekas, C.; Lee, A. A. Molecular transformer: a model for uncertainty-calibrated chemical reaction prediction. *ACS central science* **2019**, *5*, 1572–1583.
- (38) Schwaller, P.; Probst, D.; Vaucher, A. C.; Nair, V. H.; Kreutter, D.; Laino, T.; Reymond, J.-L. Mapping the space of chemical reactions using attention-based neural networks. *Nature Machine Intelligence* **2021**, *3*, 144–152.
- (39) Xu, C.; Wang, Y.; Farimani, A. B. TransPolymer: a Transformer-based Language Model for Polymer Property Predictions. *arXiv*, 2022; arXiv:2209.01307 [cs.LG].
- (40) Elnaggar, A.; Heinzinger, M.; Dallago, C.; Rehawi, G.; Wang, Y.; Jones, L.; Gibbs, T.; Feher, T.; Angerer, C.; Steinegger, M.; Bhowmik, D.; Rost, B. ProtTrans: Toward Understanding the Language of Life Through Self-Supervised Learning. *IEEE Trans Pattern Anal Mach Intell.* **2022**, *44*, 7112–7127.
- (41) Devlin, J.; Chang, M.-W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv*, 2018; arXiv:1810.04805 [cs.CL].
- (42) Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv*, 2019; arXiv:1907.11692 [cs.CL].
- (43) Wang, Y.; Wang, J.; Cao, Z.; Barati Farimani, A. Molecular contrastive learning of representations via graph neural networks. *Nature Machine Intelligence* **2022**, *4*, 279–287.
- (44) Wang, Y.; Magar, R.; Liang, C.; Barati Farimani, A. Improving Molecular Contrastive Learning via Faulty Negative Mitigation and Decomposed Fragment Contrast. *J. Chem. Inf. Model.* **2022**, *62*, 2713–2725.

- (45) Schwaller, P.; Gaudin, T.; Lanyi, D.; Bekas, C.; Laino, T. Found in Translation": predicting outcomes of complex organic chemistry reactions using neural sequence-to-sequence models. *Chemical science* **2018**, *9*, 6091–6098.
- (46) Haviv, A.; Ram, O.; Press, O.; Izsak, P.; Levy, O. Transformer Language Models without Positional Encodings Still Learn Positional Information. *arXiv*, 2022; arXiv:2203.16634 [cs.CL].
- (47) He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *Proceedings of the IEEE conference on computer vision and pattern recognition* **2016**, 770–778.
- (48) Ba, J. L.; Kiros, J. R.; Hinton, G. E. Layer normalization. *arXiv*, 2016; arXiv:1607.06450 [stat.ML].
- (49) Schwaller, P.; Hoover, B.; Reymond, J.-L.; Strobel, H.; Laino, T. Extraction of organic chemistry grammar from unsupervised learning of chemical reactions. *Science Advances* **2021**, *7*, No. eabe4166.
- (50) Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S., et al. An image is worth 16×16 words: Transformers for image recognition at scale. *arXiv*, 2020; arXiv:2010.11929 [cs.CV].
- (51) Magar, R.; Wang, Y.; Barati Farimani, A. Crystal twins: self-supervised learning for crystalline material property prediction. *npj Computational Materials* **2022**, *8*, 231.
- (52) Zbontar, J.; Jing, L.; Misra, I.; LeCun, Y.; Deny, S. Barlow twins: Self-supervised learning via redundancy reduction. International Conference on Machine Learning. *Proceedings of the 38th International Conference on Machine Learning*, 2021; pp 12310–12320.
- (53) Chen, X.; He, K. Exploring simple siamese representation learning. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021; pp 15750–15758.
- (54) Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments. *Phys. Rev. B* **2013**, *87*, 184115.
- (55) Bartók, A. P.; De, S.; Poelking, C.; Bernstein, N.; Kermode, J. R.; Csányi, G.; Ceriotti, M. Machine learning unifies the modeling of materials and molecules. *Science advances* **2017**, *3*, No. e1701816.
- (56) Himanen, L.; Jäger, M. O. J.; Morooka, E. V.; Federici Canova, F.; Ranawat, Y. S.; Gao, D. Z.; Rinke, P.; Foster, A. S. DScribe: Library of descriptors for machine learning in materials science. *Comput. Phys. Commun.* **2020**, *247*, 106949.
- (57) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **2014**, *89*, 094104.
- (58) Janet, J. P.; Kulik, H. J. Resolving transition metal chemical space: Feature selection for machine learning and structure–property relationships. *J. Phys. Chem. A* **2017**, *121*, 8939–8954.
- (59) Nandy, A.; Duan, C.; Janet, J. P.; Gugler, S.; Kulik, H. J. Strategies and Software for Machine Learning Accelerated Discovery in Transition Metal Chemistry. *Ind. Eng. Chem. Res.* **2018**, *57*, 13973–13986.
- (60) Jablonka, K. M.; Rosen, A. S.; Krishnapriyan, A. S.; Smit, B. An ecosystem for digital reticular chemistry. *ChemRxiv*, 2022.
- (61) Chen, T.; Guestrin, C. Xgboost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 2016; pp 785–794.
- (62) Xie, L. S.; Skorupskii, G.; Dincă, M. Electrically conductive metal–organic frameworks. *Chem. Rev.* **2020**, *120*, 8536–8580.
- (63) Sheberla, D.; Bachman, J. C.; Elias, J. S.; Sun, C.-J.; Shao-Horn, Y.; Dincă, M. Conductive MOF electrodes for stable supercapacitors with high areal capacitance. *Nature materials* **2017**, *16*, 220–224.
- (64) Dubbeldam, D.; Calero, S.; Vlugt, T. J. iRASPA: GPU-accelerated visualization software for materials scientists. *Mol. Simul.* **2018**, *44*, 653–676.
- (65) van der Maaten, L.; Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **2008**, *9*, 2579–2605.
- (66) Vig, J. A multiscale visualization of attention in the transformer model. *arXiv*, 2019; arXiv:1906.05714 [cs.HC].

Quantum Informed Machine-Learning Potentials for Molecular Dynamics Simulations of CO₂'s Chemisorption and Diffusion in Mg-MOF-74
Bowen Zheng, Binquan Luan, et al.
MARCH 08, 2023
ACS NANO
Get More Suggestions >

READ ↗

Computational Design of Metal–Organic Frameworks with Unprecedented High Hydrogen Working Capacity and High Synthesizability
Junkil Park, Jihan Kim, et al.
DECEMBER 23, 2022
CHEMISTRY OF MATERIALS

READ ↗

Effect of Spatial Heterogeneity on the Unusual Uptake Behavior of Multivariate-Metal–Organic Frameworks
Soyeon Ko, Kyung Min Choi, et al.
JANUARY 29, 2023
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

READ ↗

Quantum Informed Machine-Learning Potentials for Molecular Dynamics Simulations of CO₂'s Chemisorption and Diffusion in Mg-MOF-74
Bowen Zheng, Binquan Luan, et al.
MARCH 08, 2023
ACS NANO

READ ↗