

Multi-modal conditional diffusion model using signed distance functions for metal-organic frameworks generation

Junkil Park^{1#}, Youhan Lee^{2#} and Jihan Kim^{1,}*

¹ Department of Chemical and Biomolecular Engineering, Korea Advanced Institute of Science and Technology (KAIST), 291 Daehak-ro, Yuseong-gu, Daejeon 34141, Republic of Korea

² NVIDIA Corporation, Santa Clara, CA, USA

[#] Contributed Equally

^{*} Corresponding author: Jihan Kim (jihankim@kaist.ac.kr)

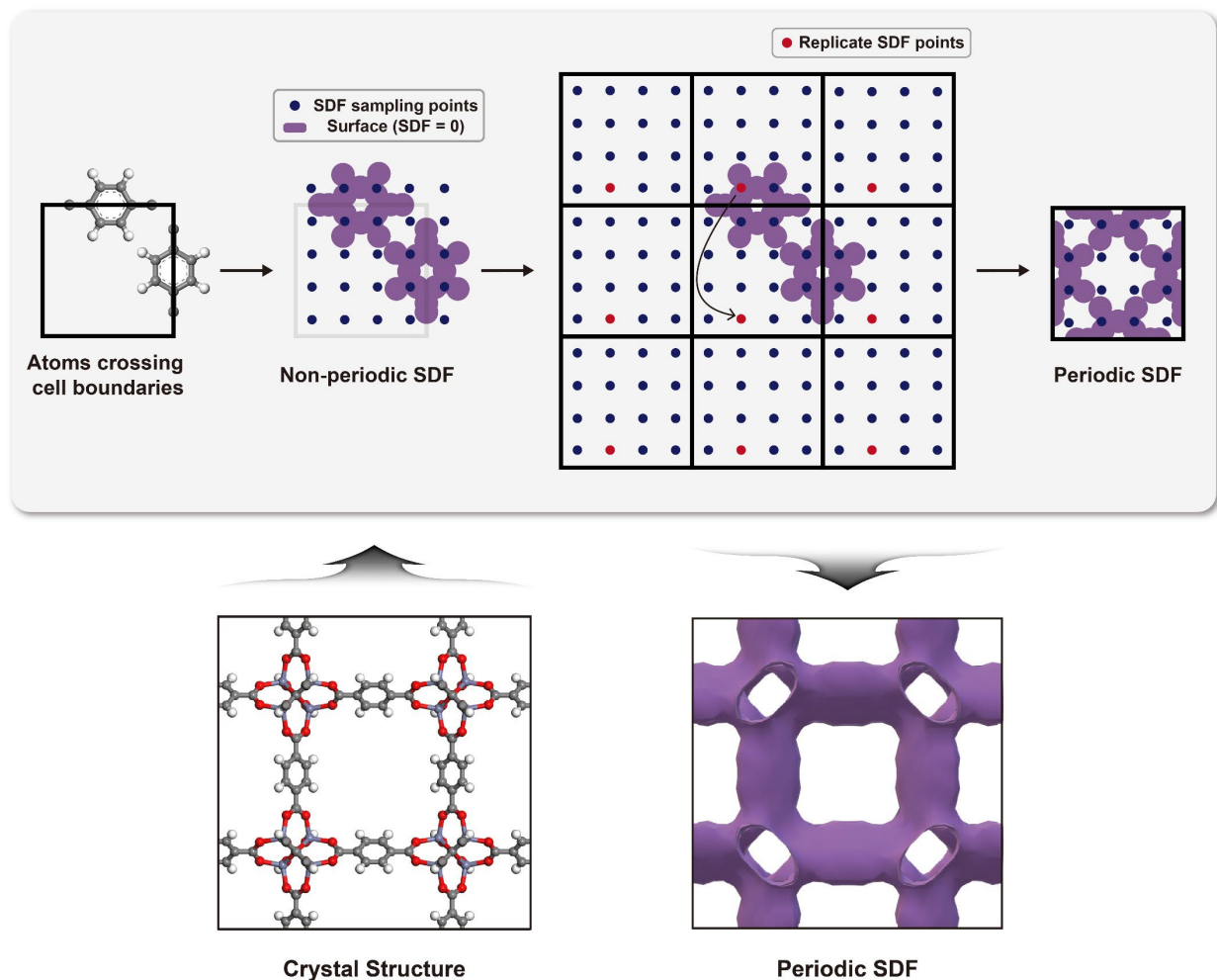
Electronic Supplementary Information

Supplementary Note #1. Details on Database Construction

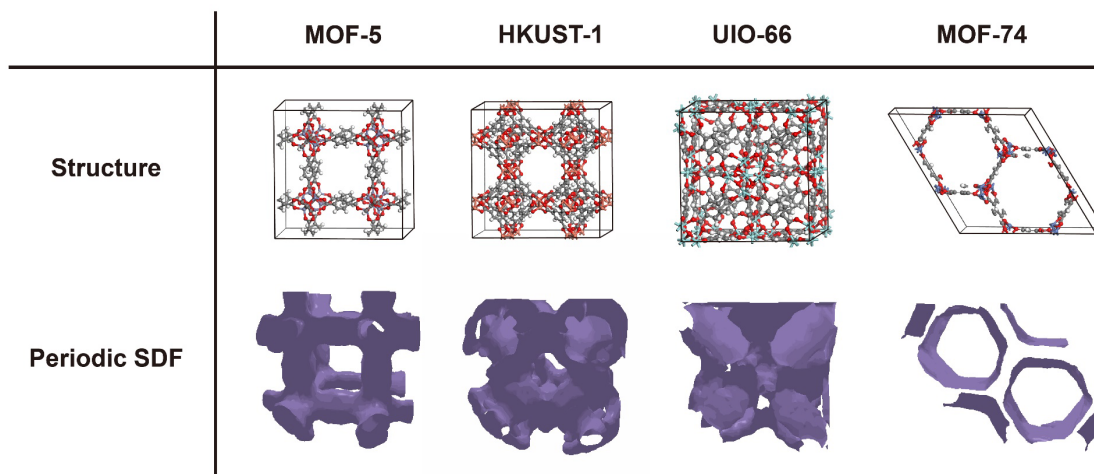
Supplementary Note #1-1. Preparation of Periodic SDF

Preparation procedures for periodic SDFs of MOFs are illustrated in Supplementary Fig. 1. The surface mesh of atoms comprising MOFs is obtained using the PyMOL¹ library, and the SDF can be sampled from the obtained mesh using the mesh-to-sdf² library. However, for the periodic SDF with periodic boundary conditions (PBCs), the sampling procedure of SDF is modified to be carried out in a 3x3x3 supercell where each sampling point has 27 replicates. To further apply PBCs, among the 27 replicate points, the minimum SDF value is retrieved and assigned to SDF of the original cell. A visualization of crystal structure and periodic SDF of MOF-5 is illustrated in Supplementary Fig. 1 as an example. As can be seen, the periodic SDF well describes the periodicity of the crystalline structure with considering connectivity between atoms crossing cell boundaries.

The resolution of the SDF is a tunable parameter, and there exist a trade-off between the accuracy and memory usage as the resolution increases. In this work, a resolution of 32^3 (which means each x,y, and z axis of the unit cell were divided into 32 partitions and used for the sampling) was chosen as it balances both representation accuracy and memory requirements. Supplementary Fig. 2 illustrates periodic SDFs of representative MOFs prepared with a resolution of 32^3 . It is visually confirmed that this resolution accurately represents the intrinsic pore shape of each structure.



Supplementary Fig. 1. Periodic SDFs for crystal structures. Procedure for applying periodic boundary conditions to SDFs. For Periodic SDF, the SDF is sampled in the expanded super cell, and the minimum value among the replicated sampling points is retrieved. The crystal structure of MOF-5 and its periodic SDFs are represented as an example. The periodic SDF accurately represents the pore surface with capturing the connectivity of the atoms across the cell boundaries.



Supplementary Fig. 2. Periodic SDFs of representative MOFs. Periodic SDFs of MOF-5, HKUST-1, UIO-66 and MOF-74 were prepared and visualized at a resolution of 32^3 . Their intrinsic and distinct pore structures are well represented.

Supplementary Note #1-2. MOF-SDF database Construction

A MOF database for the training of MOFFUISON is prepared using PORMAKE³, a software for the hypothetical MOF generation. PORMAKE allows for the easy generation of MOF structures using combinations of topologies and building blocks. In this work, to reduce computational costs, structures with more than 1500 atoms and/or with unit cell lengths larger than 60 Å were excluded from generation. Additionally, only MOFs consisting of up to two nodes and only one edge were considered to generate feasible structures. Structures without metal nodes were excluded to avoid generating COFs. All building blocks and topologies were uniformly and randomly selected during the generation process. To achieve a broad and uniform distribution of topologies within the database structures, the number of structures per topology was limited to a minimum of 100 and a maximum of 500. In total, 247,742 structures were generated, which are composed of 605 topologies, 432 metal nodes, 51 organic nodes and 220 edges. All generated structures were geometrically optimized using the LAMMPS⁴ package with the universal force field (UFF)⁵, following the same procedure as Kang et al.⁶, which is further described in Supplementary Note #5-3. Subsequently, the periodic SDFs of the optimized structures were obtained according to the procedures outlined in Supplementary Note #1-1.

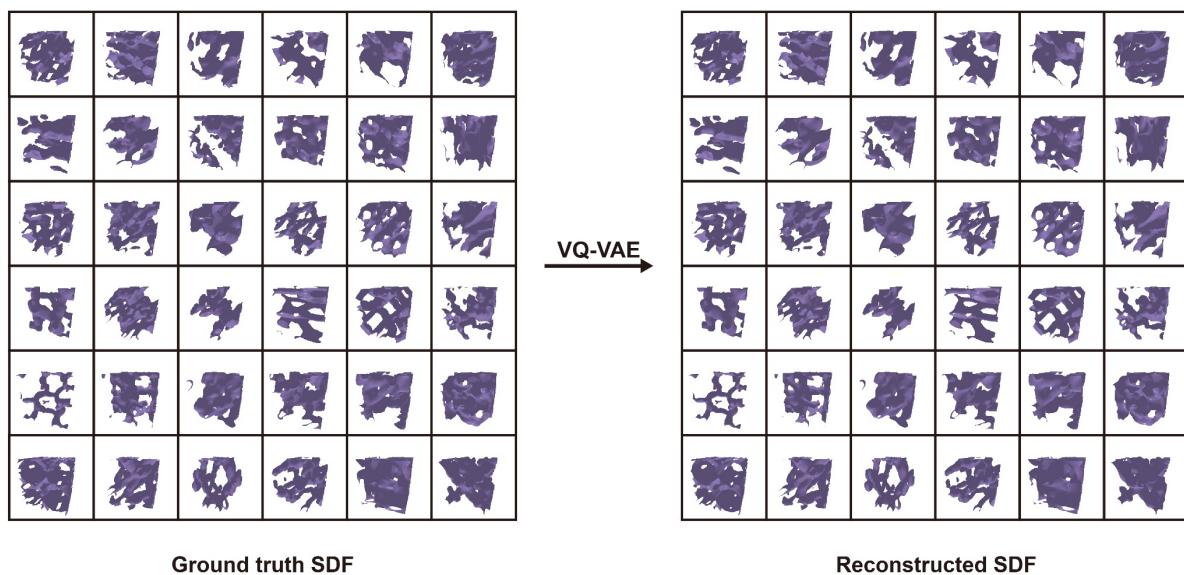
Supplementary Note #2. Details on Model Architecture

Supplementary Note #2-1. Vector Quantized-Variational AutoEncoder

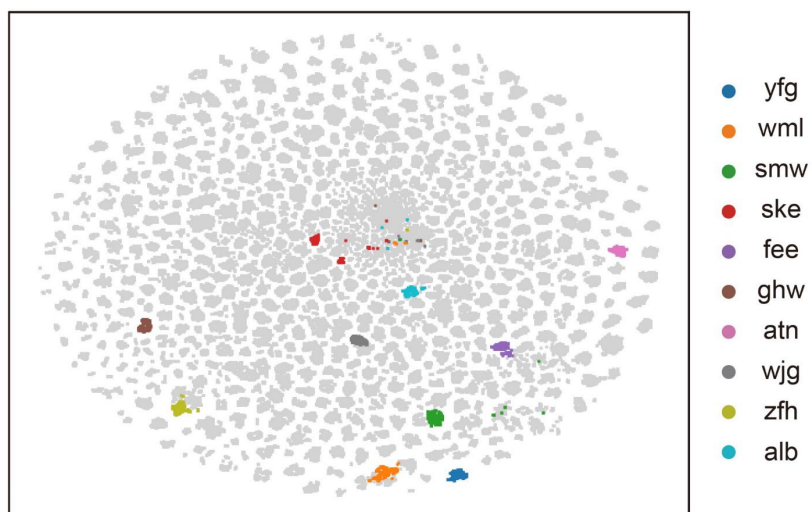
Within MOFFUSSION, the Vector Quantized-Variational AutoEncoder (VQ-VAE) maps SDFs of resolution 32^3 into a compressed latent space of resolution 8^3 . The architecture of VQ-VAE was adopted from Aaron et al.⁷, who first proposed the architecture. We trained the model by splitting train/test with a ratio of 8:2, and conducted further analysis to determine whether our trained VQ-VAE effectively captures the geometric characteristics of SDFs and maps them into appropriate latent vectors.

First, qualitative analysis on the trained VQ-VAE was conducted by comparing visual attributes of reconstructed (i.e., compression and restoration) SDFs through VQ-VAE with their original SDF. As shown in Supplementary Fig. 3, the reconstructed SDF showed almost identical visualizations, providing the most straightforward evidence of the performance of the trained VQ-VAE.

Additionally, the latent space of the trained VQ-VAE was explored using t-distributed stochastic neighbor embedding (t-SNE). The t-SNE plot of top 10 frequently appearing topologies were shown in Fig. 3b of the main text, and the MOFs with the same topology showed clustered distributions. The t-SNE plot of all MOFs is illustrated in Supplementary Fig. 4 with top 10 topologies colored as well. Clustering of each topology was observed even in the plot of the entire database, validating that the VQ-VAE was successfully trained with effectively capturing geometric features of SDFs.



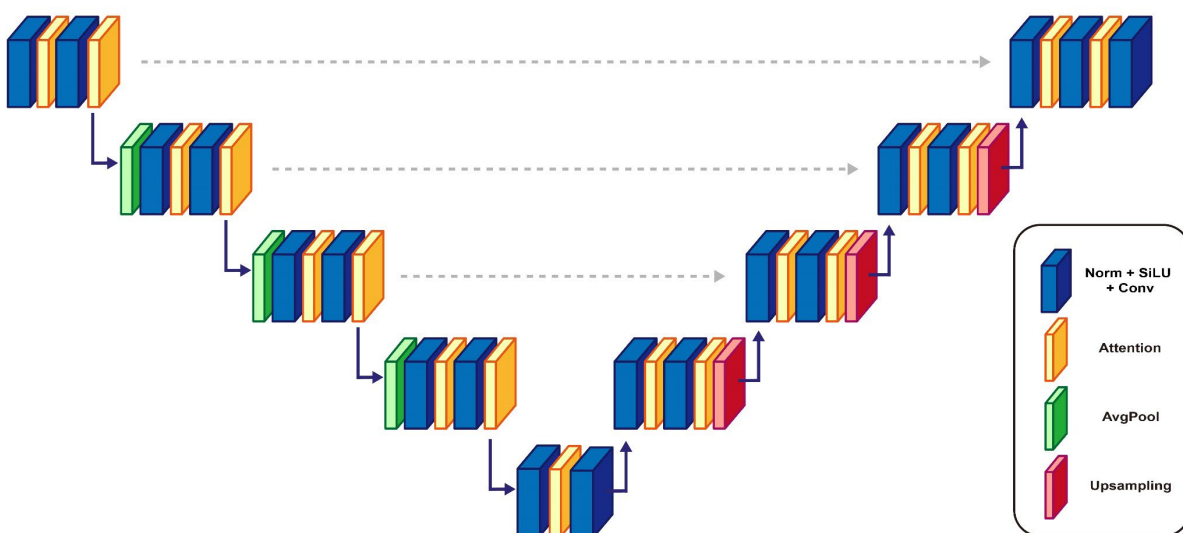
Supplementary Fig. 3. Reconstruction of SDFs through VQ-VAE. Qualitative comparison of ground truth SDFs of MOFs with their reconstructed SDFs through the trained VQ-VAE. The two set of SDFs display almost identical visual attributes.



Supplementary Fig. 4. A t-SNE plot of the trained VQ-VAE. A t-SNE plot of MOFs with the top 10 frequently appearing topologies, embedded using the trained VQ-VAE. Source data are provided as a Source Data file.

Supplementary Note #2-2. Diffusion Model

The detailed explanation on the diffusion model is provided in the Methods section of the main text. In Supplementary Fig. 5, we further introduced the 3D U-Net architecture used in MOFFUSION. We adopted the most common model shared by open AI⁸. As with the VQ-VAE architecture, circular padding option was used for convolutional layers to account for periodicity of crystalline materials during data processing. We randomly split the training and test dataset with a ratio of 8:2 and used for training.



Supplementary Fig. 5. 3D U-Net Architecture. The 3D U-Net architecture used within the diffusion model part of MOFFUSION.

Supplementary Note #2-3. MOF Constructor

As introduced in the main text, the MOF Constructor is a classification model that predicts the building components (i.e., topology, metal node, organic node, edge) based on the input SDF. MOF Constructor is the most unique component of MOFFUSION and plays a significant role in determining the performance of the workflow. Consequently, we conducted several experiments to achieve the highest possible performance for the MOF Constructor. We standardized its basic architecture to ResNet⁹, one of the most widely used architectures in image classification, and tested various approaches to identify the optimal setting. The dataset was split into train, test, and validation sets with a ratio of 8:1:1, and used for the training and accuracy comparison,

1) 1 channel SDF vs 4 channel SDF

When predicting the building blocks of a given SDF, using only the SDF of the entire structure might be insufficient. In this case, providing additional details on the shapes of individual building blocks (metal node, organic node, and edge) can be beneficial for making more accurate predictions. Therefore, we compared the model accuracy with using different numbers of SDF channels. For the 1 channel SDF, only the SDF of the entire structure was provided. For the 4 channel SDF, the SDFs of the entire structures, metal node, organic node, and edge were included (Supplementary Fig. 6a).

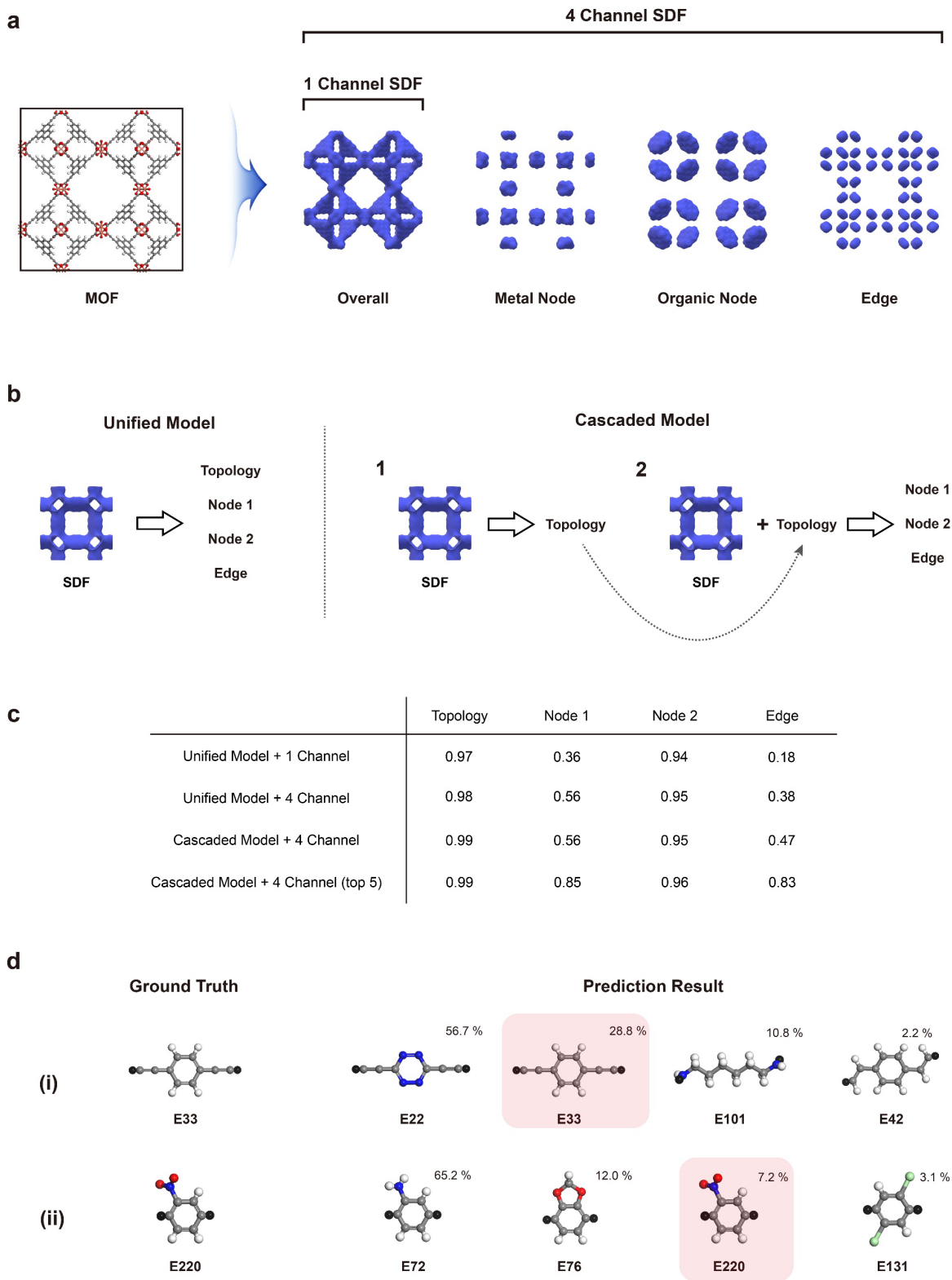
We tested the accuracy of MOF Constructor for both case and found that the accuracy was higher with the 4 channel SDF. (For this comparison, we used the unified model, which will be discussed in following section, Supplementary Fig. 6c) Therefore, even though it requires more memory, we decided to use the 4 channel SDF representation for the entire workflow.

2) Unified Model vs Cascaded Model

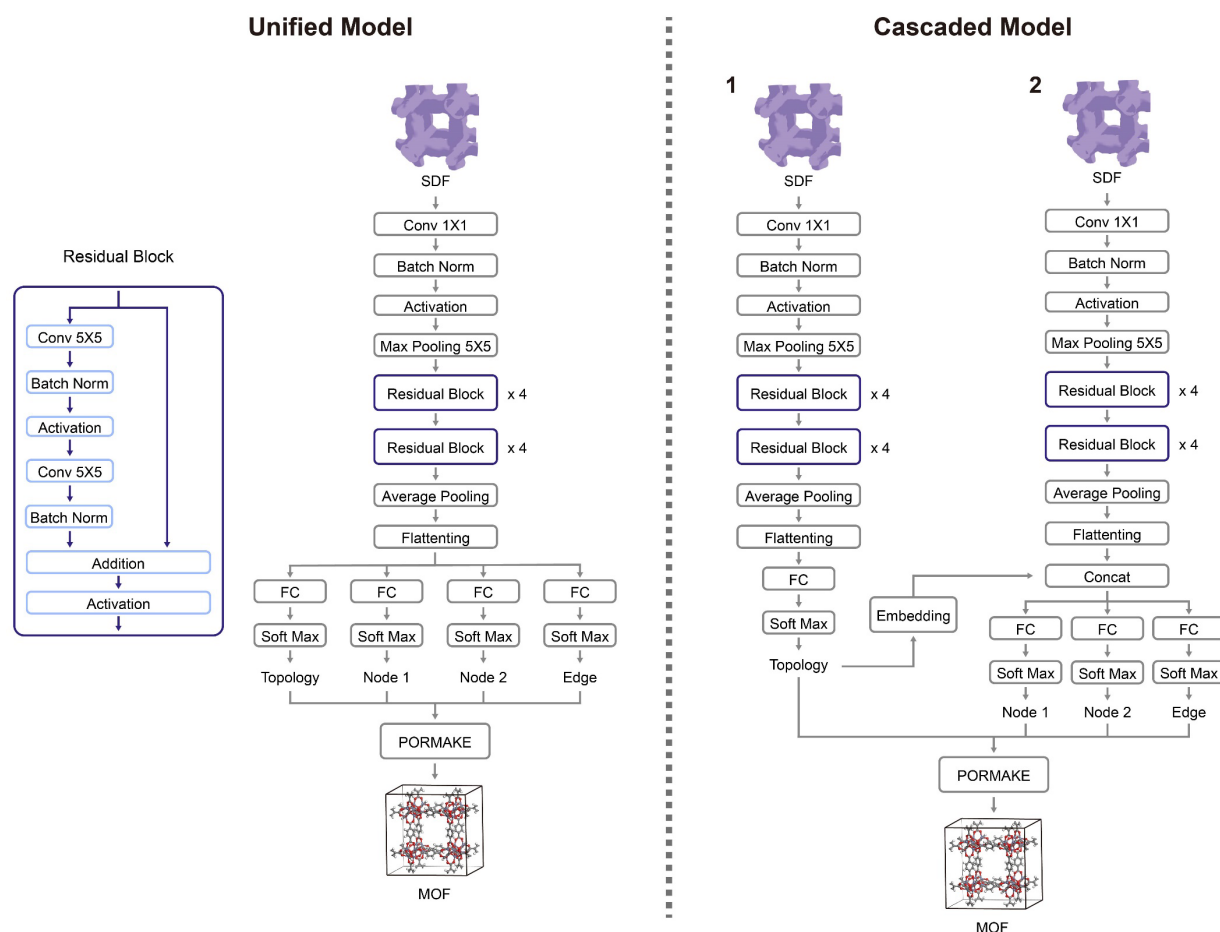
The four building components of MOFs to be predicted (i.e., topology, metal node, organic node, and edge) are not on the same hierarchical level. The topology of a MOF is at a higher hierarchical level and influences the determination of the other building blocks. Once topology of the structure is set, only certain nodes can be used. For example, if the topology is determined to be pcu, only metal nodes with 6 connect points can be placed. In that sense, predicting the topology first, and then predicting the rest of the components using pre-determined topology information can lead to more accurate and feasible structure predictions.

We compared two model architectures: the unified model and the cascaded model. The unified model predicts all components (topology and building blocks) simultaneously, while the cascaded model uses a hierarchical architecture that first predicts the topology and then predicts the building blocks using the information from the predicted topology (Supplementary Fig. 6b, Supplementary Fig. 7). We compared the accuracy of the two models and observed an increase in performance with the cascaded model as shown in Supplementary Fig. 6c. Therefore, we decided to use the MOF Constructor with the cascaded model architecture for MOFFUSION.

Edge prediction showed relatively low accuracy compared to other components (e.g., 0.47 for the cascaded with 4 channel case). We further investigated the causes of this decrease in accuracy. In Supplementary Fig. 6d, we visualized some cases of wrong edge predictions. As can be checked, MOF Constructor is confused between edges with similar chemical structures, rather than providing random incorrect output. In fact, for the top 5 predictions, the accuracy of the edge prediction increased to 0.83. Therefore, we concluded that current MOF Constructor is well trained, and even for incorrect predictions it provides structurally similar building blocks which would result in highly similar SDF.



Supplementary Fig. 6. Details on MOF Constructor. **a.** 1 channel and 4 channel SDF representation. Different numbers of channels were tested. The 1 channel SDF involves only the overall SDF, while the 4 channel SDF includes additional SDFs for the metal node, organic node and edge. **b.** Unified and cascaded architecture. The unified model predicts the topology, metal node, organic node, and edge simultaneously, whereas the cascaded model first predicts the topology and then predicts the building blocks using the predicted topology information. **c.** Accuracy of various methodologies. A comparison of the accuracies of different methodologies was conducted. The cascaded model with a 4 channel SDF showed the highest accuracy. The top 5 accuracy refers to when the correct answer is among the five categories that showed the highest confidence. **d.** Examples of wrong edge prediction. Two cases of wrong edge prediction cases are presented, along with the 4 candidates with the highest confidence score. In cases of incorrect predictions, confusion arises among edges that appear similar.

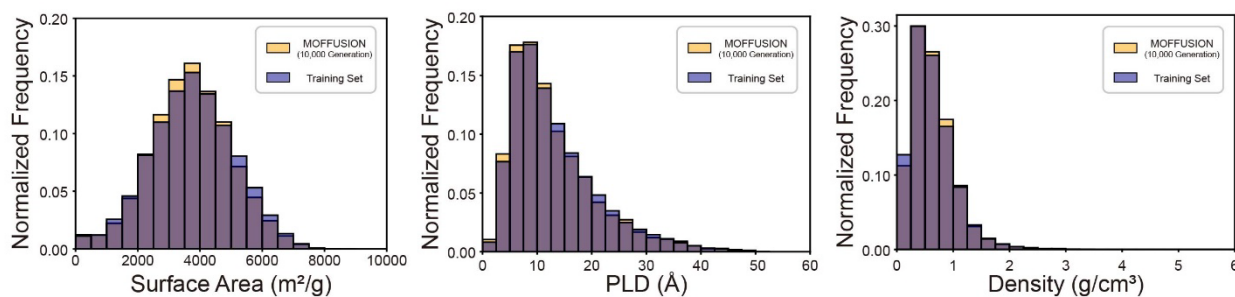


Supplementary Fig. 7. Architecture of unified and cascaded MOF Constructor. Unified and cased model of MOF Constructor is shown. Unified model makes prediction of topology, metal node, organic node, and edge simultaneously, while the cascaded model first predicts the topology, and then predicts remaining building blocks using the predicted topology information. After the prediction, MOF can be built using PORMAKE software. Both architectures are based on ResNet, one of the most widely used models for image classification.

Supplementary Note #3. Details on Unconditional Generation

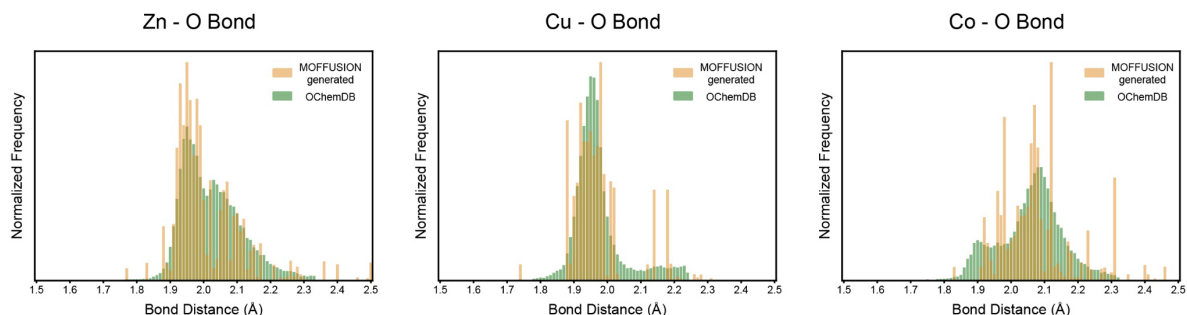
Supplementary Note #3-1. Unconditional Generation Result

Structural analysis was conducted on structures generated from 10,000 generation attempts by the trained MOFFUSION. Surface area, pore limiting diameter, and density were computed, and their distributions were shown alongside those of the structures comprising the training dataset (Supplementary Fig. 8). The distributions of MOFFUSION-generated structures showed highly similar distributions to those of the training set, validating that MOFFUSION has learned the intrinsic distribution of the training set and can generate structures resembling them. Qualitative analysis on the similarity between the structures generated by MOFFUSION and those in the training can be found in Table 1 of the main text and Supplementary Note #3-2.



Supplementary Fig. 8. Structural properties of generated structures by MOFFUSION. The distributions of surface area, pore limiting diameter, and density of structures obtained from 10,000 generation attempts by MOFFUSION are depicted. The distributions are visualized alongside the 247,742 structures comprising the training dataset. Two distributions showed an almost identical profile in all three descriptors. Source data are provided as a Source Data file.

In addition to the comparison with the structures in the training set, we further compared our generated structures to the experimental structures stored in the Open Chemistry Database (OChemDB). OChemDB offers various statistics on chemical structures (e.g., bond, angle, and torsion) of crystals. We compared the metal-oxygen bond distances of MOFFUSION generated structures and those comprising OChemDB. Specifically, we examined Zn, Cu, and Co atoms, which are among the most common metals in MOFs. As shown in Supplementary Fig. 9, the distributions of all three bond types in MOFFUSION generated structures exhibited similar profile to those in OChemDB, validating the reliability of the output structures and the entire workflow.



Supplementary Fig. 9. Distribution of metal-oxygen bond distances in Open Chemistry Database (OChemDB) and MOFFUSION generated structures. The distributions of metal-oxygen bond distances for Zn, Cu, and Co atoms in structures obtained from 10,000 generation attempts by MOFFUSION (orange) are compared with the structures from the OChemDB database (green). Source data are provided as a Source Data file.

Supplementary Note #3-2. Details on Database Diversity Comparison

We have compared the diversity of datasets used for training the generation models for MOFs in Table 1 of the main text. While we have detailed the procedure for dataset construction of MOFFUSION in Supplementary Note #1-2, further details on the dataset used for the training in precedent works would be more discussed in this section.

For training SMVAE¹⁰, the authors used their own hypothetical dataset, constructed by utilizing building blocks comprising CoRE MOF database¹¹. As noted in Table 1, they considered 153 topologies, 14 metal nodes, 47 organic nodes, and 372 edges in total. However, it should be noted that, to further augment data abundance, they expanded their edges sets through random functionalization. They functionalized their edges with 30 commonly used linker functional groups, expanding the number of generalizable edges to around 300k.

In the case of MOFDiff¹², a pre-built BW-DB dataset¹³ composed of 289k hypothetical MOF structures was used for training. The database was constructed specifically for investigating the CO₂ capture capabilities of MOFs. They utilized 12 topologies, 8 metal nodes, 65 organic nodes and 29 edges for construction. It is mentioned in the original BW-DB paper that a larger set of building blocks could have been used, but they felt that limiting the number of building blocks improved the synthetic viability of the structures.

For both cases, the ToBaCCo¹⁴ software was used for the construction of hypothetical MOF structures. ToBaCCo and PORMAKE (used in this work) operate in a similar manner, where both software generate hypothetical structures using a top-down method with predefined building blocks. It is worth mentioning that, as PORMAKE was reported later, PORMAKE's building block datasets contain every building block used in ToBaCCo.

Supplementary Note #3-3. Details on Performance Comparison

The generation performance of MOFFUSION was compared to previous generative models targeting MOFs, as shown in Table 1. As mentioned in the main text, a completely fair comparison is impossible because each model was trained on different datasets and used completely different workflows. However, a comparison was made as fairly as possible by focusing solely on the output structures generated by 10,000 generation trials of each model.

Structure validity is a metric widely used for assessing the quality of the generated structures through generative models. As described in the main text, the validity of the structures generated through MOFFUSION was assessed using the following metric. First, the structure must be able to be generated using PORMAKE. This includes: (i) the connection points of building blocks should be compatible with the topology, and (ii) building blocks need to be placed to a given site within a certain root mean square distance. (More details on the PORMAKE generation scheme can be found in Lee et al.³) Second, for the generated structures, geometry optimization is conducted to filter out those structures that are chemically non-feasible or suffer from excessive steric hindrance (which might lead to either failure in force field assignment or convergence). It is our opinion that our metric is more stringent compared to some precedent works, which considered the structure valid only if there was no atom overlap.¹⁵

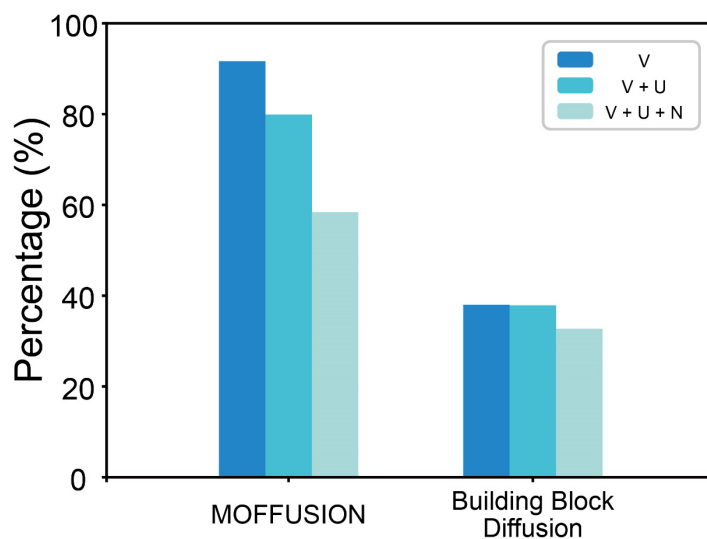
First, the validity of the structures was compared by examining the proportion of valid output structures over the generation trials. Fortunately, the original papers of SMVAE and MOFDIFF evaluated and reported the validity of their generated samples, so these values directly taken from their respective papers. (For SMVAE, this value refers the prior validity of the trained VAE model.) In the case of MOFFUSION, 10,000 generation attempts were made, and the proportion of valid structures was calculated. As also described in the main text, the validity of the generated structures

was assessed using the following metrics. In addition, uniqueness and novelty of the generated structures were further discussed in the main text.

The earth mover's distance (EMD) between the property distributions of generated materials and test materials was calculated and compared. This is a commonly used metric to evaluate generative models developed for materials generation. Property statistics for pore limiting diameter (PLD), accessible gravimetric surface area (SA), and density, which are key structural properties of porous materials, were compared. In the case of SMVAE, the structural properties of 45,000 randomly selected structures from their database were tabulated. Additionally, along with the trained final SMVAE model, a trained property predictor was provided, which predicts chemical properties based on the vector in the latent space. Therefore, we performed 10,000 generation runs and obtained the predicted PLD, SA, and density of 6,346 valid structures. We then calculated the EMD between these values and the 45,000 structures from the dataset. For MOFDIFF, we carried out 10,000 generation attempts and obtained the structural properties using Zeo++ software. Afterward, we calculated the EMD by comparing it with the BW-DB dataset, which was used for training MOFDIFF. For MOFFUSION, EMD between the generated valid structures from 10,000 generation attempts and the training set was calculated.

Supplementary Note #3-4. Ablation study of the SDFs

Within MOFFUSION, the SDFs is used to represent MOF structures and the generated SDFs through diffusion model are mapped back to MOF structures. To provide a quantitative evidence of the advantage of going through SDFs, we examined the performance of the workflow after excluding the SDF phase, allowing the diffusion model to generate building blocks directly. We utilized the D3PM¹⁶ architecture, one of the most widely used diffusion models specifically designed for discrete data. The model is designed to generate four categorical data (each corresponding to topology, metal node, organic node, and edge) simultaneously. The model is trained to learn the valid building block compositions from the training set, and to generate new compositions based on that. We referred this workflow (without SDFs) as “Building Block Diffusion”. The number of tunable parameters in Building Block Diffusion (790M) was set to comparable to that of MOFFUSION (690M). The exactly same training data was used with MOFFUSION, and the chemical space searchable for both workflows are identical. We assessed the validity, uniqueness, and novelty of 10,000 generated structures and confirmed that the original MOFFUSION workflow exhibits higher generation performance, as shown in Supplementary Fig. 10. We interpret this result to indicate that representing structures using SDFs as 32x32x32 continuous float data offers greater flexibility and a deeper understanding compared to representing them as discrete entities.



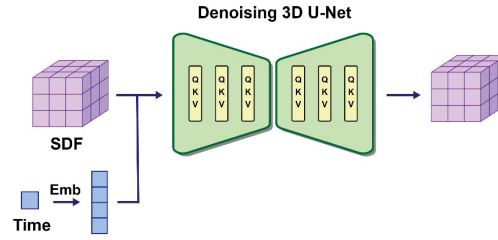
Supplementary Fig. 10. Comparison of 10,000 generated structures obtained from MOFFUSION and Building Block Diffusion, respectively. “V” refers to the ratio of valid structures, “V+U” indicates valid and unique structures, and “V+U+N” represents valid, unique and novel structures among the generated outputs. Further descriptions of the evaluation metrics can be found in the main text. Source data are provided as a Source Data file.

Supplementary Note #4. Details on Conditional Generation

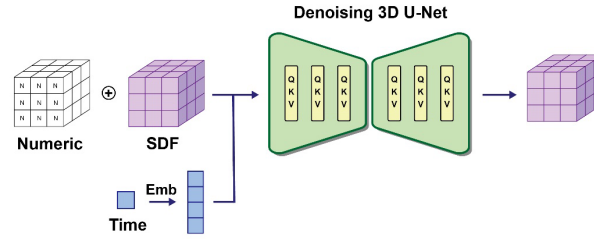
Supplementary Note #4-1. Conditioning of Diverse Modalities

MOFFUSION has been tested for conditional generation targeting various data modalities, including numeric, categorical, and text. The model architectures of the conditional MOFFUSION model for each data modality are depicted in Supplementary Fig. 11, along with comparisons to the unconditional case. In the unconditional model, SDF and diffusion time step embedding are used as inputs of 3D U-Net. For conditioning on numeric data, a tensor with the same size as the SDF but filled with the target numeric values is concatenated to the original SDF. In the case of categorical data, the target category is processed with an additional embedding layer, and then combined with the time step embedding. For text data, SciBERT¹⁷, a pre-trained BERT model for scientific literature, is used for natural language processing. The processed data is further refined with an additional projection layer for deeper and task-specific handling of text. The obtained text embedding is integrated into the diffusion process using a cross attention algorithm.

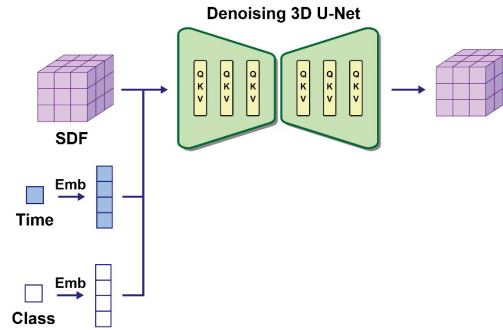
(i) Unconditional Model



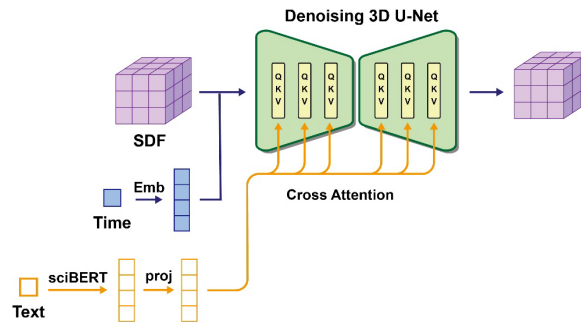
(ii) Conditional Model on Numeric Data



(iii) Conditional Model on Categorical Data



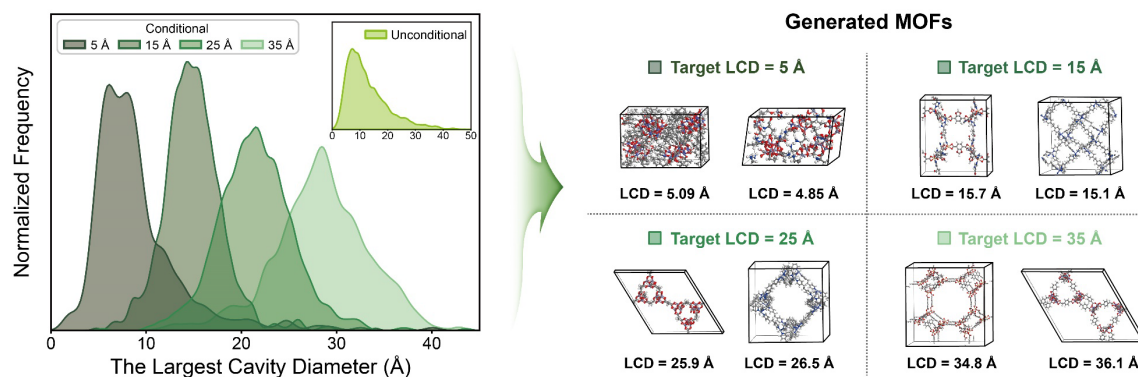
(iv) Conditional Model on Text Data



Supplementary Fig. 11. Conditioning mechanisms across diverse data modalities. Architectures of diffusion models for unconditional and conditional models of numeric, categorical, and text data.

Supplementary Note #4-2. Conditioning on the Largest Cavity Diameter

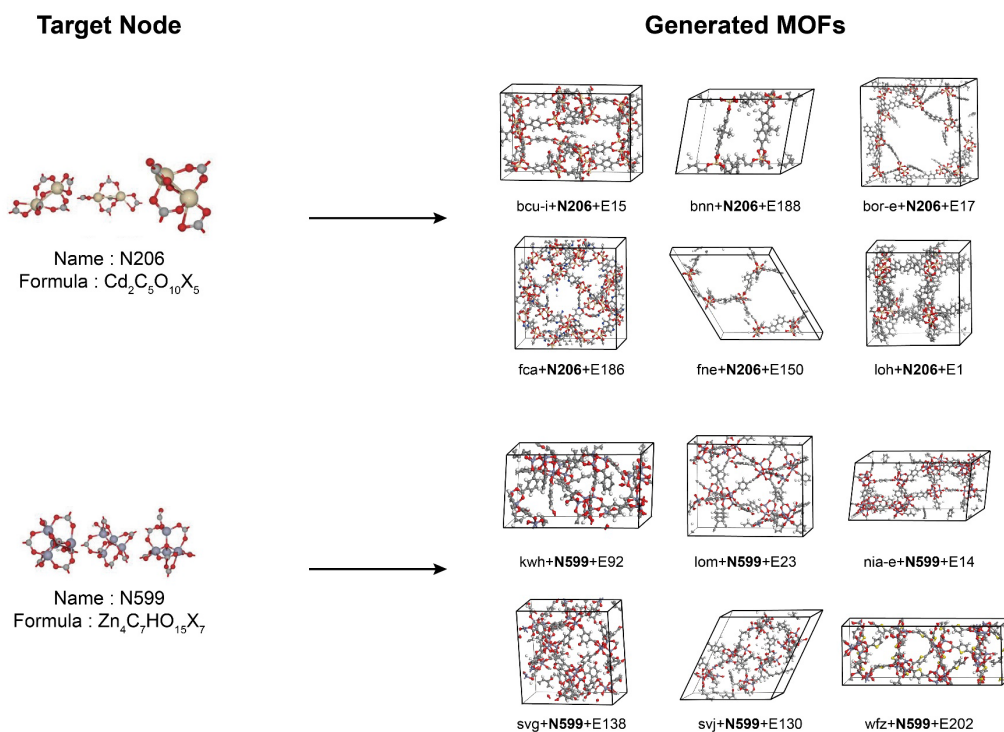
In addition to hydrogen working capacity (Fig. 3, main text), conditional generation on the largest cavity diameter (LCD) was tested as a representative of numeric data. The pore diameter a straightforward yet crucial parameter for porous materials. MOFFUSION was trained conditionally on the largest cavity diameter (LCD), and structures were generated with target LCD values of 5, 15, 25, and 35 Å for the validation. A total of 1,000 generation attempts were made for each target value, and distinctive distributions of LCD among generated structures with peaks located near the respective target values were observed (Supplementary Fig. 12). Representative samples were visualized as a sanity check, confirming that structures generated with larger target values exhibited large pore diameters, and vice versa.



Supplementary Fig. 12. Conditional generation on the largest cavity diameter (LCD). MOFFUSION was trained for the conditional generation targeted on LCD. Structures were conditionally generated with target LCD values of 5, 15, 25, and 35 Å. A total of 1,000 generation attempts were made for each target value, and the generated structures exhibited distinctive distributions with each peaks located near its target values. Among generated structures, representative samples were further visualized for a sanity check. Source data are provided as a Source Data file.

Supplementary Note #4-3. Conditioning on Metal Node Types

In addition to topology (Fig. 4a, main text), conditional generation on metal node types was tested as a representative of categorical data. Examples of conditional generation for N206 and N599, the two most frequently appearing nodes in the database, are illustrated in Supplementary Fig. 13. Once users designate their desired metal node types, corresponding structures with the designated metal nodes were successfully generated. In a test carried out with the top 10 most frequently occurring metal nodes and 1,000 generation trails each, our conditional model showed a 20 % accuracy in generating the desired metal nodes. Considering that there are a total of 432 metal nodes in the database, an accuracy of 20.4 % indicates that conditional model has a high tendency to generate structures with the desired metal nodes during the generation process. However, the lower accuracy achieved compared to conditioning on topology (96.0 %) might originate from the fact that it is easier to capture the visual attributes of topology compared to metal nodes, as many metal nodes have similar geometric structures (and corresponding SDFs). In addition, the lower prediction accuracy of MOF Constructor for metal nodes (Supplementary Fig. 6c) would also attributed to the relatively lower performance of conditioning on node types.



Supplementary Fig. 13. Conditional generation targeting metal node types. MOFFUSION was tested for conditional generation on metal node types, representing categorical data. The conditional generation results for N206 and N599, the two most frequently occurring metal nodes in the database, are illustrated as examples.

Supplementary Note #4-4. Details on Text Conditioning

For the text conditioning, we generated our own text database describing the characteristics of the structures. The text includes information about topology and the void fraction rounded up at 0.05 (Supplementary Fig. 14a). In Supplementary Fig. 14b, we illustrate how SciBERT handles given texts. We monitored how the SciBERT tokenizer converts the given text into tokens and then converts them back to a string. It was verified that sentences of interest are well restored, although there may be additional space after a float dot.

In this work, we used a fixed text format for training to check the capability of MOFFUSION for the text conditioning, but there is high flexibility in preparing text data. Utilizing language models such as ChatGPT¹⁸ would be a considerable option for generating natural and general text. Furthermore, it is noteworthy that it is known that language models such as BERT still have some difficulties in handling numeric data written in text, especially understanding their order relations. Although our model showed satisfactory performance in handling void fraction data in text format, the development of language models in handling numeric data would improve the text conditioning performance in MOFFUSION as well.

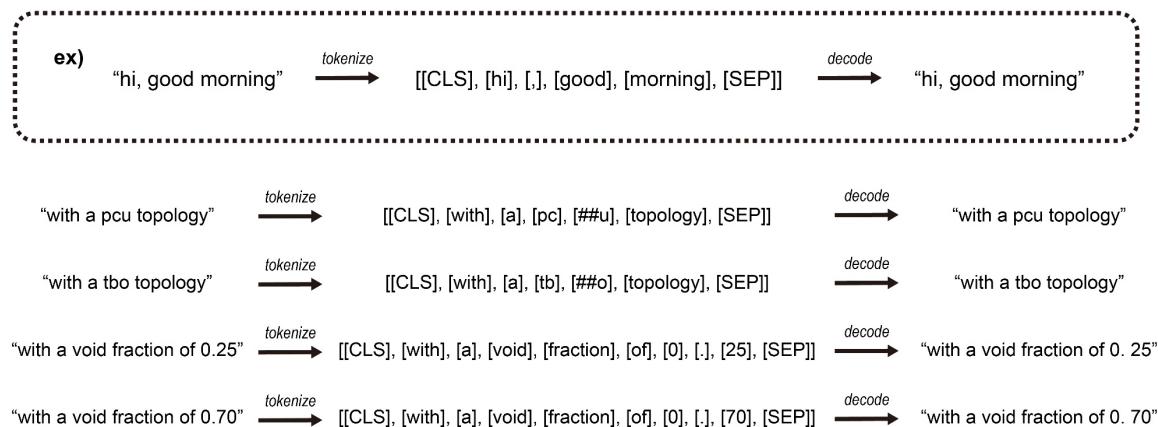
a

⋮

vmt+N664+E138 exhibits a vmt topology and a void fraction of 0.40
sow+N257+E56 exhibits a sow topology and a void fraction of 0.75
wjm+N482+E72 exhibits a wjm topology and a void fraction of 0.45
fob+N651+E175 exhibits a fob topology and a void fraction of 0.50
vma+N172+E154 exhibits a vma topology and a void fraction of 0.65
wys+N401+N485+E139 exhibits a wys topology and a void fraction of 0.35
sne+N684+E8 exhibits a sne topology and a void fraction of 0.60
loh+N651+E29 exhibits a loh topology and a void fraction of 0.30
vbk+N671+E186 exhibits a vbk topology and a void fraction of 0.55
vcy+N444+E204 exhibits a vcy topology and a void fraction of 0.20
wko+N667+E158 exhibits a wko topology and a void fraction of 0.40
zfa+N272+E71 exhibits a zfa topology and a void fraction of 0.45
fit+N510+N133+E154 exhibits a fit topology and a void fraction of 0.45
uof+N237+E92 exhibits a uof topology and a void fraction of 0.60
vcj+N284+E191 exhibits a vcj topology and a void fraction of 0.65
noz+N383+E91 exhibits a noz topology and a void fraction of 0.75

⋮

b



Supplementary Fig. 14. Details on text conditioning. **a.** Text dataset used for training the conditional model on text data. The text contains information on the topology and void fraction of the structures. In the case of void fraction, values were rounded to the nearest 0.05 units. **b.** Examples of SciBERT tokenizer handling text.

Supplementary Note #5. Details on Computational Methods

Supplementary Note #5-1. Details on Pore Structure Analysis

Geometric analysis on the structures, including pore radius, void fraction, surface area, and pore volume calculations, were carried out using Zeo++¹⁹ software. For those calculations that required a designated probe radius, a probe radius of 1.2 Å was used.

Supplementary Note #5-2. Details on Hydrogen Adsorption Simulation

Classical grand canonical Monte Carlo (GCMC) simulations were conducted for the calculation of hydrogen capacities of MOFs. Our in-house GPU code²⁰ was used, enabling rapid calculation of a large number of structures in a reasonable time. It has been demonstrated in previous works that our GPU code accurately simulates gas adsorption behavior, aligning well with experimental results or conventionally used GPU code. The pairwise interactions between atoms were modeled using the Lennard-Jones potential, with the universal force field applied to atoms comprising MOFs. Hydrogen molecules were modeled using the pseudo-Feynman–Hibbs-corrected (p-FH)^{21,22} united atom model. Lorentz–Berthelot mixing rule and a cutoff radius of 12.8 Å were imposed for the simulations and the MOF structures were deemed to be fixed.

Supplementary Note #5-3. Details on Geometry Optimization

The geometry optimization of the generated structures and those comprising training dataset was conducted using LAMMPS software as follow; energy minimization, relaxation of simulation box, and energy minimization again. Energy minimization was conducted with stopping tolerance for energy (E_{tol}) of 1e-6 and stopping tolerance for force (F_{tol}) of 1e-8 (both are unitless). During the

relaxation of simulation box, the dimensions of the simulation cell was gradually adjusted to achieve an isotropic pressure of 0.0 with a maximum change in the cell size of 0.001 per step.

Supplementary references

- 1 *pymol*, <<https://github.com/schrodinger/pymol-open-source>> (2006).
- 2 *mesh-to-sdf*, <https://github.com/marian42/mesh_to_sdf> (2020).
- 3 Lee, S. *et al.* Computational screening of trillions of metal–organic frameworks for high-performance methane storage. *ACS Applied Materials & Interfaces* **13**, 23647-23654 (2021).
- 4 Thompson, A. P. *et al.* LAMMPS-a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications* **271**, 108171 (2022).
- 5 Rappé, A. K., Casewit, C. J., Colwell, K., Goddard III, W. A. & Skiff, W. M. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American chemical society* **114**, 10024-10035 (1992).
- 6 Kang, Y., Park, H., Smit, B. & Kim, J. A multi-modal pre-training transformer for universal transfer learning in metal–organic frameworks. *Nature Machine Intelligence* **5**, 309-318 (2023).
- 7 Van Den Oord, A. & Vinyals, O. Neural discrete representation learning. *Advances in neural information processing systems* **30** (2017).
- 8 OpenAI. <<https://github.com/CompVis/latent-diffusion/blob/main/ldm/modules/diffusionmodules/openaimodel.py>> (2022).
- 9 He, K., Zhang, X., Ren, S. & Sun, J. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770-778 (2016).
- 10 Yao, Z. *et al.* Inverse design of nanoporous crystalline reticular materials with deep generative models. *Nature Machine Intelligence* **3**, 76-86 (2021).
- 11 Chung, Y. G. *et al.* Computation-ready, experimental metal–organic frameworks: A tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials* **26**, 6185-6192 (2014).
- 12 Fu, X., Xie, T., Rosen, A. S., Jaakkola, T. & Smith, J. MOFDiff: Coarse-grained Diffusion for Metal-Organic Framework Design. *arXiv preprint arXiv:2310.10732* (2023).
- 13 Boyd, P. G. *et al.* Data-driven design of metal–organic frameworks for wet flue gas CO₂ capture. *Nature* **576**, 253-256 (2019).
- 14 Colón, Y. J., Gómez-Gualdrón, D. A. & Snurr, R. Q. Topologically guided, automated construction of metal–organic frameworks and their evaluation for energy-related applications. *Crystal Growth & Design* **17**, 5801-5810 (2017).
- 15 Xie, T., Fu, X., Ganea, O.-E., Barzilay, R. & Jaakkola, T. Crystal diffusion variational autoencoder for periodic material generation. *arXiv preprint arXiv:2110.06197* (2021).
- 16 Austin, J., Johnson, D. D., Ho, J., Tarlow, D. & Van Den Berg, R. Structured denoising diffusion models in discrete state-spaces. *Advances in Neural Information Processing Systems* **34**, 17981-17993 (2021).
- 17 Beltagy, I., Lo, K. & Cohan, A. SciBERT: A pretrained language model for scientific text. *arXiv preprint arXiv:1903.10676* (2019).
- 18 Bubeck, S. *et al.* Sparks of artificial general intelligence: Early experiments with gpt-4. *arXiv preprint arXiv:2303.12712* (2023).
- 19 Willems, T. F., Rycroft, C. H., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Microporous and Mesoporous Materials* **149**, 134-141 (2012).

- 20 Kim, J. & Smit, B. Efficient Monte Carlo simulations of gas molecules inside porous materials. *Journal of Chemical Theory and Computation* **8**, 2336-2343 (2012).
- 21 Feynman, R., Hibbs, A. & Weiss, G. H. (American Institute of Physics, 1966).
- 22 Fischer, M., Hoffmann, F. & Fröba, M. Preferred hydrogen adsorption sites in various MOFs—a comparative computational study. *ChemPhysChem* **10**, 2647-2657 (2009).