An ML-based Generative Workflow for Metal-Organic Framework Synthesis

BINGHAMTON UNIVERSITY
THOMAS J. WATSON COLLEGE OF ENGINEERING AND APPLIED SCIENCE

Shehtab Zaman*, Christopher Owen⁺, Jacob Barkovitch*, Kevin Phillips*, Musen Zhou‡, Jianzhong Wu‡, Kenneth Chiu*, Michael Lawler⁺

* Department of Computer Science, Watson School of Engineering, Binghamton University
⁺ Department of Physics, Harpur College, Binghamton University
‡ Department of Chemical and Environmental Engineering, University of California, Riverside

Introduction

In recent years Metal-organic frameworks (MOFs) have emerged as a substantial class of crystalline structures with extremely high porosity, inner surface area, and variability of the organic and inorganic components. MOFs have applications in gas separation, gas purification, and electrolytic catalysis, among other fields. The creation of better MOFs for these purposes represents a multibillion-dollar engineering challenge.

We develop a ML based workflow to generate and characterize MOFs. We use a generative adversarial network, a deep generative model, to synthesize periodic energy grids.

We also focus on the characterization of MOFs to speed up workflow. We use a dimensionality reduction algorithms to take 3D coordinates down to 2D image representations to utilize existing deep learning vision algorithms. We use transfer learning to predict geometric properties such as largest cavity diameter, pore limiting diameter, and accessible surface area.

We characterize the electronic properties of MOFs as well. We use a Graph CNN on the 3D coordinates to predict the geometric properties and potential energies of MOFs. We use physics guided graph kernel that takes advantage of the local interactions of quantum systems to predict bond level energy functions and predict potential energies an order of magnitude faster than density functional theory.

Datasets

We use the computation ready experimental metal organic framework [1] (CoRE MOF) dataset for our generative and predictive tasks. The energy grids are generated by processing the atomic coordinates obtained from the dataset.

The dataset for the Graph model is constructed using Quantum Espresso [2], an ab initio software for electronic structure and energy calculations. We used the FIGXAU from the CoRE MOF database. We found the ground state configuration using the Kjpaw [3] pseudopotentials and the Perdew-Burke-Ernzerhof(PBE) [4] exchange-correlation functional. From this ground state configuration, random fluctuations were introduced by allowing each atom to randomly move any rational number between (Å) either on its x,y or z axis. 47,617 new atomic configurations were generated and a Self-Consistent Field Calculation (SCF) was done for each one.

Generation: Energy Grids

As we know, generating novel MOFs is an important challenge to solve. We use a Wasserstein Generative Adversarial Network with 3D convolutions to generate new MOF energy grids. We treat the energy grids as a blueprint informing us of potential

The energy grids are obtained using the Feynman-Hibbs 4th order corrected Lennard-Jones potential energy. The energy grids directly correlate to the hydrogen adsorption capabilities of MOF.

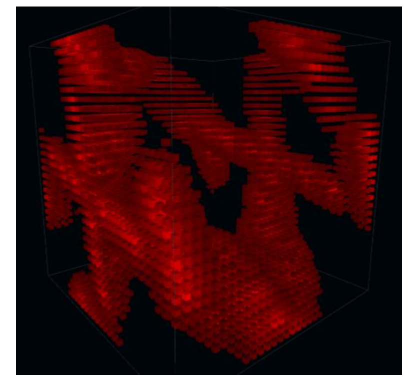


Fig. Energy grid representation of the unit cell of a MOF. Each voxel represents the energy value for that region.

A GAN is used to generate novel energy grids. The GAN is compromised of two separate neural networks. The first is a generator which tries to create realistic energy grids. The second is a discriminator that tries to distinguish between real energy grids (from the dataset) and fakes (created by the generator).

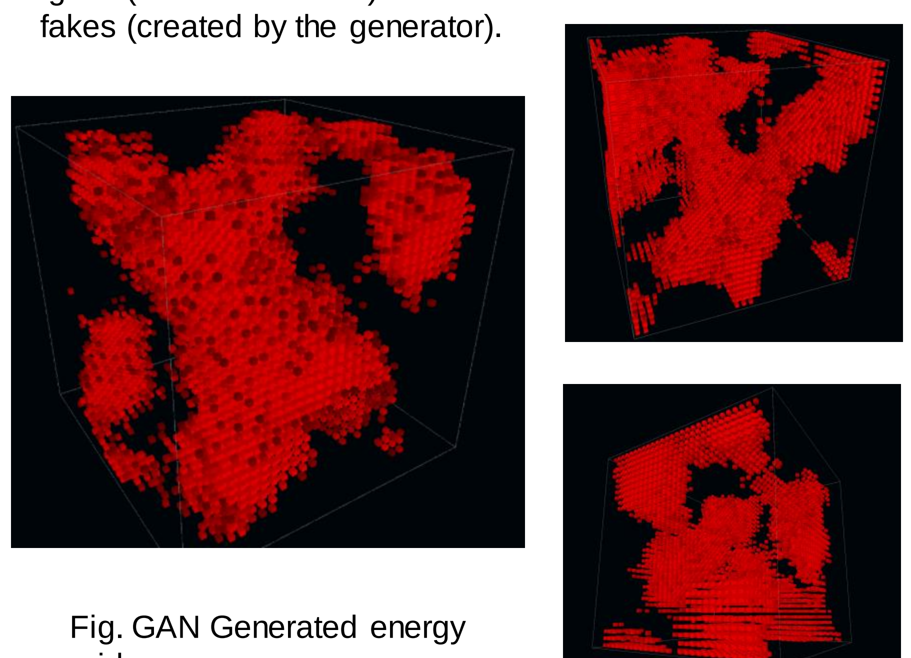
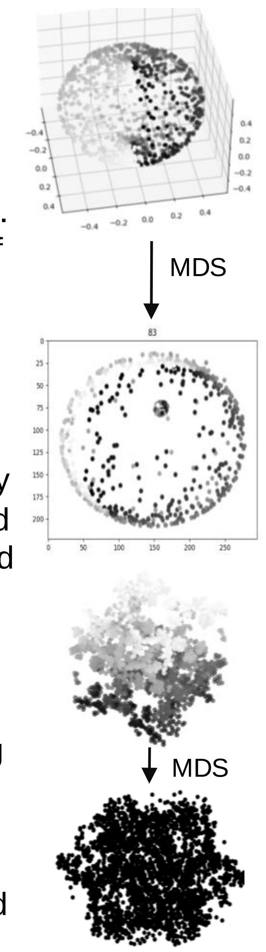


Fig. GAN Generated energy grids

Characterization: Geometric Properties

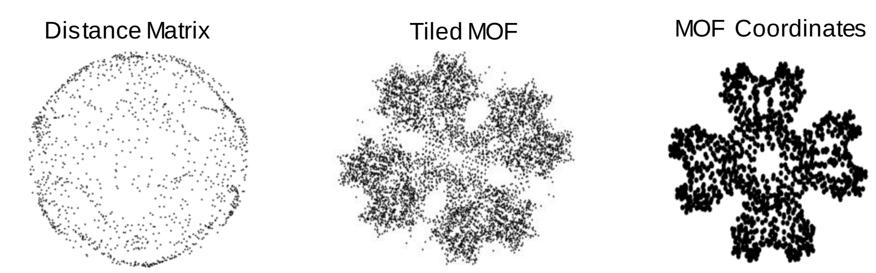
There are many pretrained 2D CNN models. As a result, we use multi-dimensional scaling (MDS) to translate MOF 3D coordinates to 2D. MDS can preserve the geometric properties of the 3D data and illustrate it in the 2D images. Some of these geometric properties include, largest cavity diameter (LCD), Pore Limiting Diameter (PLD) and Henry's Constant.



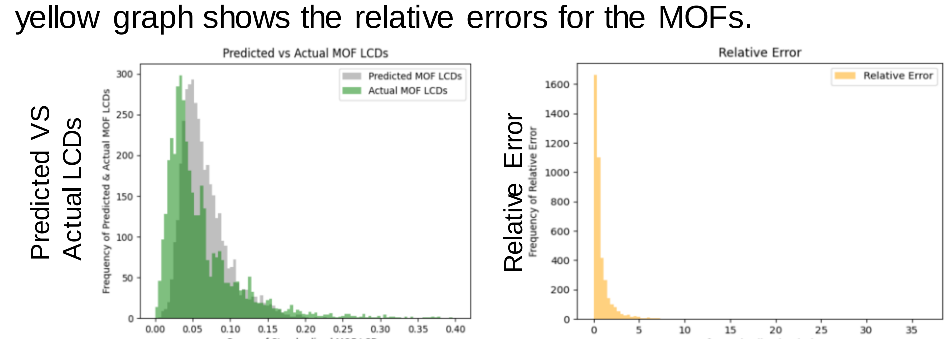
We initially tested on spheres which were easy to generate and calculate LCDs. We then used MOFs from the CoRE MOF Dataset and tested several different methods.

We first ran MDS on each MOF's coordinates. Our second method was tiling the MOFs to capture hidden cavities. The third was passing the distance matrix of the MOF into MDS.

We decided to use the distance matrix MOFs for testing because of their fast generation and similarity to the synthesized spheres.

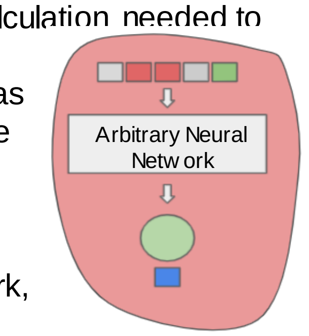


Our results were good on the MOFs generated with the distance matrix as seen in the graphs below. We did test on the tiled MOFs and single MOFs, but they were both less accurate. The average percent error was 85% with a median error of 45%. The yellow graph shows the relative errors for the MOFs.

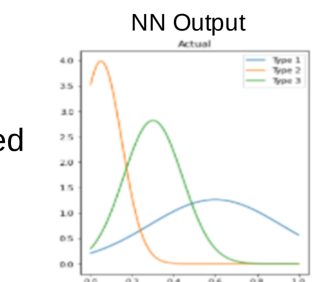


Characterization: Electronic Properties

The potential energy is a fundamental calculation needed to design MOFs for many applications. It is currently computed via techniques such as density functional theory (DFT), which are prohibitively expensive and not suitable for high-throughput screening.

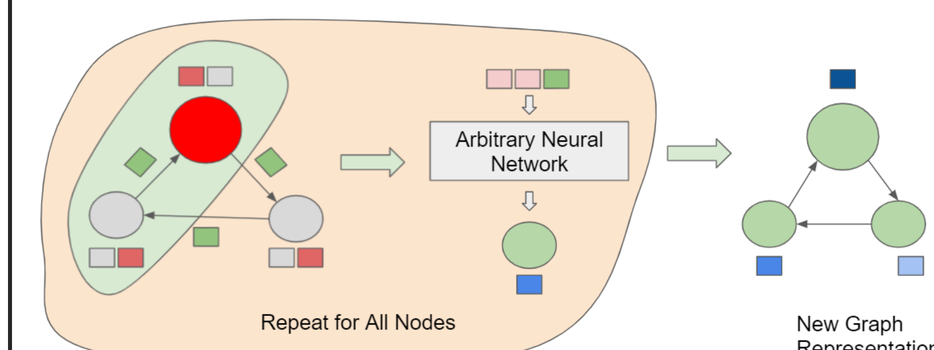


We propose a graph convolutional network, MOFGCN, to predict potential energies of MOFs.



We represent the crystal structures using graphs, such that each atom is represented by a node, and the edge represents the nearest image distance between the two atoms.

We can predict the potential energy and generate bond level energies.



Works Cited

[1] Chung, Yongchul G., et al. "Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019." Journal of Chemical & Engineering Data 64.12 (2019): 5985-5998.
[2] Giannozzi, Paolo, et al. "QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials." Journal of physics: Condensed matter 21.39 (2009): 395502.
[3] Dal Corso, Andrea. "A pseudopotential plane waves program (pwscf) and some case studies." Quantum-mechanical Ab-initio calculation of the properties of crystalline materials. Springer, Berlin, Heidelberg, 1996. 155-178.
[4] Perdew, John P., Kieron Burke, and Matthias Ernzerhof. "Generalized gradient approximation made simple." Physical review letters 77.18 (1996): 3865.

This material is based upon work supported by the National Science Foundation under Grant No. DMR-1940243. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Contact

For any questions feel free to contact Shehtab Zaman
szaman5@binghamton.edu