# Chapter 4

# Results & Discussion

THE PROPOSED FRAMEWORK is tested on the University of Ottawa (UO) database, for predicting $CO_2$ uptake in metal-organic frameworks (MOFs), the gas that mainly "triggered" the development of energy-based descriptors. In order to evaluate the transferability of the approach, a different host-guest system is also examined. We apply the suggested approach in the database created by Mercado et al. (2018), for predicting $CH_4$ uptake in covalent organic frameworks (COFs). In both cases, the resulting machine learning (ML) models are compared with conventional ones, built upon geometric descriptors. In the rest of this chapter, results from these comparisons are presented, followed by discussion for improvements of the proposed framework. Before delving into the results, we first take a look at RetNet, the 3D convolutional neural network (CNN) under the hood, that takes as input a voxelized potential energy surface (PES) and outputs a prediction for a gas adsorption property, hereon gas uptake.

## 4.1 Visualizing RetNet

Figure 4.1 illustrates the processing a voxelized PES undergoes, as it is passing through RetNet. For the purpose of this visualization, we use the model trained on the MOFs dataset with the largest training set size (see Section 3.1). Moreover, for the ease of visualization, only some feature maps of RetNet are visualized. Please note, that each feature map of a given layer, combines all the feature maps of the precedent layer. The only exception are the pooling layers, which just downsample the feature maps from the previous layers.

For example, each feature map of the Conv2 layer takes into account all the twelve feature maps of Conv1 layer. In contrast, the feature maps of the MaxPool1 layer, are just downsampled versions of the corresponding feature maps in Conv2 layer. Although feature maps of CNNs are not meant to be interpreted by humans—especially the ones found deeper in the network—it is worth noticing that early Conv layers (i.e. Conv1 and Conv2) emphasize the texture of the structure. For instance, the third feature map of Conv1 layer delineates the skeleton of the framework.

Moving towards the output layer, the alternation of MaxPool and Conv layers continues until the Flatten layer, which just flattens out and concatenates[1] all feature maps from Conv2 layer into a single vector of size 3240. This vector is then processed by a fully connected neural network (FCNN)—i.e. the stack of Dense and Output layers—to give the final prediction. Since the Output layer is really nothing more than a linear layer, all that RetNet does is the following:

[1] Given $m$ feature maps of size $n \times n \times n$, a Flatten layer converts them into a vector of size $mn^3$.

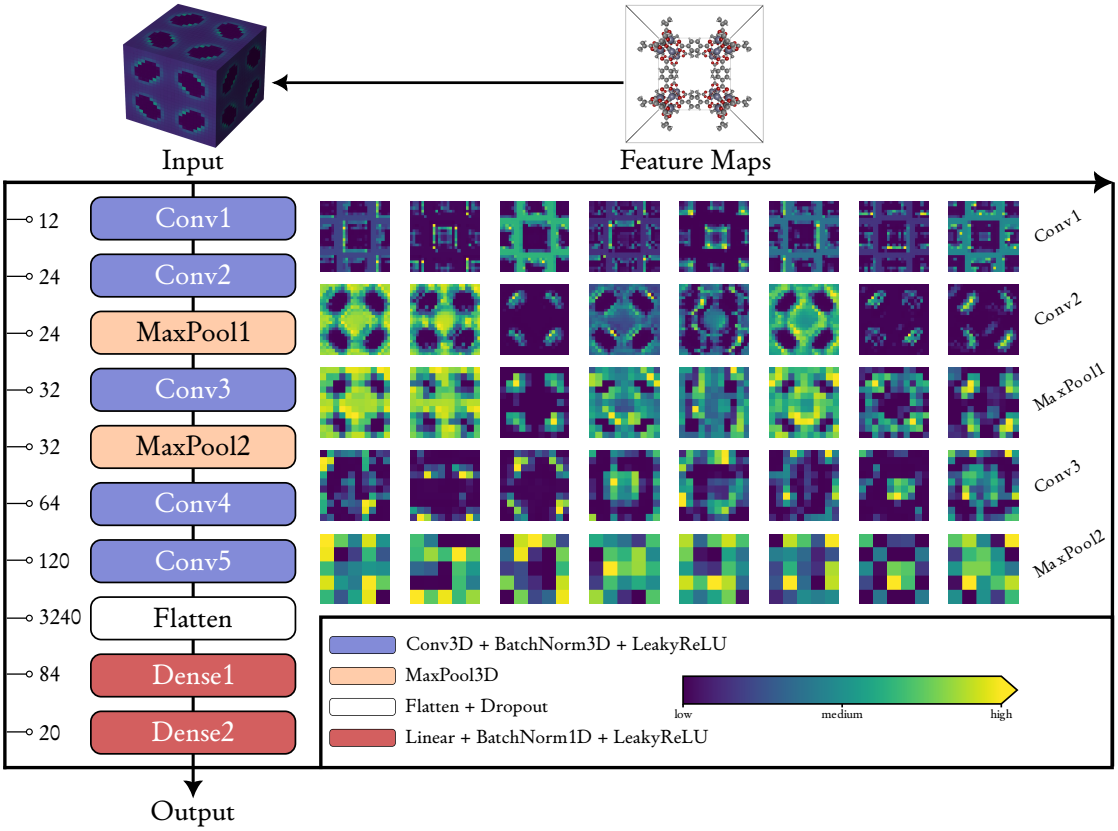
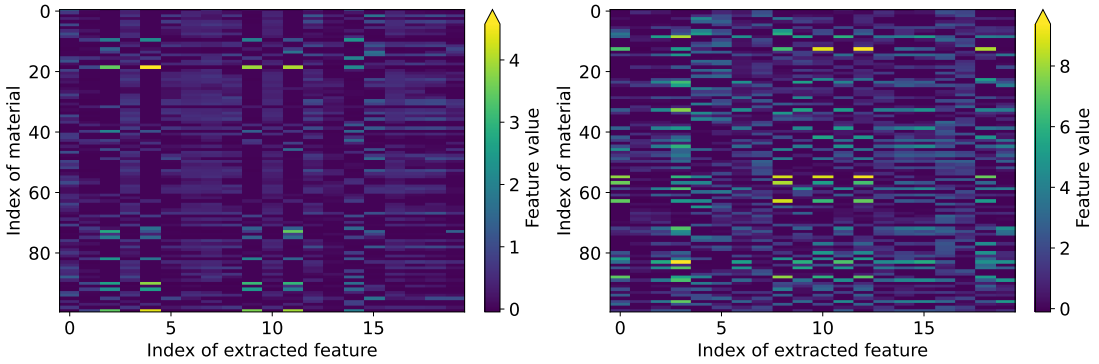



**Figure 4.1:** Forward pass of IRMOF-1 through RetNet. For the sake of visualization, only slices (feature maps are 3D matrices) of eight feature maps from the first five layers are visualized. For Conv1 layer, the fifth slice is presented, while for the remaining layers, the first slice is presented. The IRMOF-1 structure was visualized with the iRASPA software (Dubbeldam et al. 2018).

$$\underbrace{\boldsymbol{x}}_{\substack{\text{PES}\\ \text{input}}} \longrightarrow \underbrace{\phi(\boldsymbol{x};\boldsymbol{\theta})}_{\substack{\text{fingerprint}\\ \text{feature extraction}}} \longrightarrow \overbrace{\boldsymbol{\beta}^\top \phi(\boldsymbol{x};\boldsymbol{\theta}) + \beta_0}^{\substack{\text{gas uptake}\\ \text{output}}} \tag{4.1}$$

Equation 4.1 says that RetNet, *starting from the PES, extracts a fingerprint—that is, a high level representation of the PES—and then predicts the gas uptake by using a linear model on top of this fingerprint*. All intermediate layers between Input and Output layer participate in this feature extraction step, with the Dense2 layer determining the size of the fingerprint, which is a vector of size 20, i.e. $\phi(\boldsymbol{x}) \in \mathbb{R}^{20}$ (see Figure 4.2). The fact that *this fingerprint extraction step is learnable*—the parameters $\boldsymbol{\theta}$ of $\phi$ are learned during the training phase—*is what fundamentally distinguishes the proposed approach from methods that use hand-crafted fingerprints* (see Section 1.3). *In these methods the fingerprint or extraction step is fixed*, and based on some heuristic, such as energy histograms (Bucior et al. 2019) or average interactions (Fanourgakis, Gkagkas, Tylianakis, and Froudakis 2020). *Hereon, feature extraction from*




**(A)** Fingerprints extracted from the MOFs dataset. **(B)** Fingerprints extracted from the COFs dataset.

**FIGURE 4.2:** Output of the last LeakyReLU layer of RetNet trained on MOFs (left) and COFs (right) datasets, with the corresponding maximum training set size. The fingerprints of the first 100 materials in the training set are depicted.

*the PES is no longer fixed, but is an essential part of the training phase.*

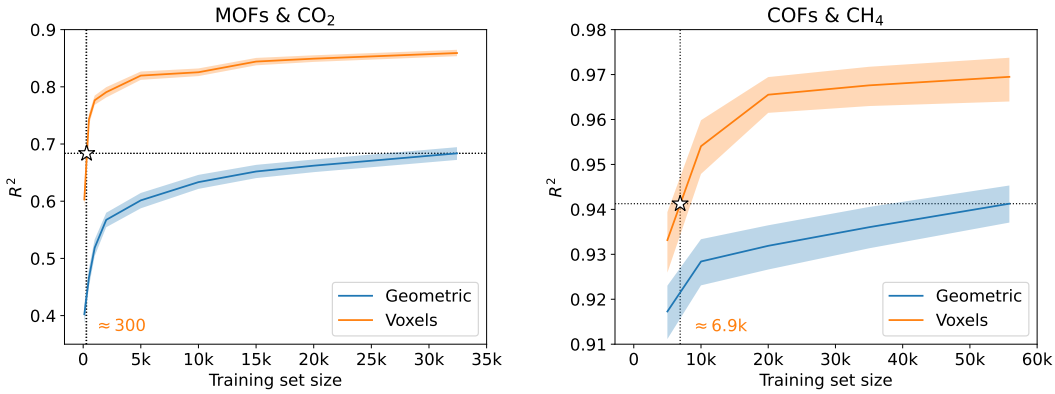
## 4.2 Learning Curves

The learning curves of the conventional models—built upon geometric descriptors—and the proposed ones–built upon energy voxels—are shown in Figure 4.3. As it can be seen from Figure 4.3a, in the MOF-$CO_2$ case, the CNN model achieves an $R^2$ score of 0.859, outperforming the conventional model, which shows an $R^2$ score of 0.690. This amounts to around 25 % increase in accuracy, even with such a coarse approximation of the PES[2]. Moreover, from the same figure, one can notice that the proposed model reaches the peak performance of the conventional one—that is, the performance when trained with the maximum training set size—by requiring two orders of magnitude less training data, around 300.

Analogous results are observed when examining the COF-$CH_4$ case. Again the CNN model performs better, showing an $R^2$ score 0.969 compared to 0.941 for the conventional one. Similar to the previous case, a substantially smaller amount of training data are required—one order of magnitude less training, around 6900—for the CNN model to match the performance of the conventional model.

The fact that in both cases, the learning curves of the proposed models lie above the corresponding ones of the conventional models, should be credited to the following factors: i). The increased informativeness of the voxelized the voxelized PES—in comparison to geometric descriptors. ii). The ability of CNNs to handle images and image-like data, such as the voxelized PES, which is essentially a single channel 3D image. iii). The data augmentation technique, which was applied during the CNN training (see Section 3.3.3).

[2]In this work, all host-guest interactions were modeled with the Lennard-Jones (LJ) potential (see Section 3.2), which neglects electrostatic interactions.




**(A)** Learning curves for MOFs & $CO_2$.

**(B)** Learning curves for COFs & $CH_4$.

**FIGURE 4.3:** Performance ($R^2$ score) on test set as function of the training set size for conventional and CNN models. Shaded areas correspond to the 95 % confidence interval (CI). The $x$-coordinate of the white star denotes the trainig set size where the CNN model reaches the performance of the conventional one, the $y$-coordinate. "Geometric" stands for geometric descriptors, while "Voxels" stands for energy voxels.

## 4.3 Discussion

It is worth mentioning the increase in performance, approximately 13 %, of the CNN model in the COFs-$CH_4$ case ($R^2 = 0.969$) compared to the MOFs-$CO_2$ case ($R^2 = 0.859$). In contrast to $CO_2$, which exhibits strong electrostatic interactions with the framework atoms, $CH_4$ lacks dipole or quadrupole moment. Given that the same resolution—i.e. the same grid size—was used in both cases and that the LJ potential doesn't account for electrostatic interactions, this performance gap should be attributed to the absence of the latter in the voxelized PES. *In other words, the extra "contrast" that such strong interactions add to the energy image of the masterial, is missing from the voxelized PES. As such, a straightforward approach to improve the performance of the proposed approach, especially for adsorbates like $CO_2$, $H_2$ and $H_2S$, is to include this type of interactions into the voxelized PES*. Of course, there is no free lunch, since these refinements require the assignment of partial charges to each framework atom, which is a computationally expensive task. Luckily, ML-based approaches have already been developed (**Bleiziffer2018**; **Raza2020**; **Kancharlapalli2021**), which can assign partial charges rapidly and with high fidelity, enabling the efficient construction of a more accurate voxelized PES.

Improving the input, and as such, the performance of the suggested pipeline is a major concern, but not the only one. *What about the data efficiency of the pipeline?* Imagine that we are asked to predict $CH_4$ uptake at various thermodynamic conditions. A naive approach would be to collect training data and retraining from scratch the CNN for every thermodynamic condition, which is of course a laborious task. *Can we do something smarter?* Well, the fact that the proposed framework uses a deep learning (DL) algorithm under the hood, opens the door for applying ***transfer learning*** *techniques*. In a nutshell, transfer learning(**Zhuang2019**; **Ma2020**; **Kang2023**) is based on the following idea: *a violist can learn to play piano faster than others, since both the piano and the violin are musical instruments, and may share some common knowledge.* Translating this to neural networks (NNs), a pre-trained NN

on an original task—known as the *source task*—may require less training data to perform well on a new task—known as the *target task*—if there is some *similarity between the tasks*. Coming back to our "imaginary" scenario, all we have to do is to train the CNN once in a specific thermodynamic condition[3] and then fine-tune this pre-trained model on the other conditions.

**ADD citations**

Throughout this work we focused on gas adsorption, but of course this doesn't mean we are not interested in predicting other properties of reticular materials. *What if we are asked to predict properties such as band gap or bulk modulus?* In that case, quantities such as *electron density* are more informative over host-guest interactions with regards to the aforementioned properties. This entails that the *voxelized electron density* should substitute the voxelized PES, as input to the 3D CNN. Nevertheless, ***wouldn't be great if all properties could be predicted from one and only one input?*** If our aim is to predict *different properties for the same structure, shouldn't the structure itself be used as input?* Currently, the approaches to tackle this challenge are based on *text representations* and *crystal graphs*. The main drawback of these approaches, is their inability to represent exactly the structure, that is the *exact arrangement of the atoms in the 3D space*. ***Point clouds*** are a natural way to solve this problem, since they are just *a set of coordinates and associated features*. In our context, the coordinates are the coordinates of the atoms, and the associated features are the types of the atoms. It should be emphasized, that *a point cloud is not another mathematical representation of a material—in the sense of a descriptor—it is the material itself*[4]. Therefore, an answer to the original question is to *couple point clouds with neural networks that can handle such kind of input*.

[3] Preferably, the one where we have more labeled training data.

[4] Same ideas apply for molecules and in general, for any chemical system.