**Showcasing research from Profs. Peng Bai and Subhransu Maji's laboratories, Colleges of Engineering and of Information and Computer Sciences, University of Massachusetts Amherst, Massachusetts, United States.**

ZeoNet: 3D convolutional neural networks for predicting adsorption in nanoporous zeolites

A deep learning framework, ZeoNet, was developed that can rapidly and accurately predict the separation performance of zeolite materials for large, flexible molecules. ZeoNet borrows ideas from the computer vision community by viewing materials as 3-dimensional images. On a standard GPU, ZeoNet can process more than eight structures per second, compared to hours per structure using state-of-art computer simulations, with an accuracy of 9.3 kJ mol^{-1} in adsorption free energy for n-octadecane.
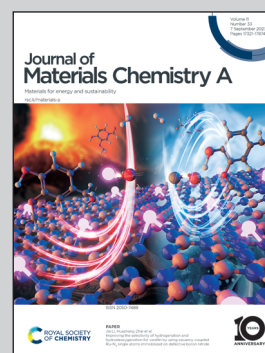
ROYAL SOCIETY OF CHEMISTRY

rsc.li/materials-a

Check for updates

ZeoNet: 3D convolutional neural networks for predicting adsorption in nanoporous zeolites†

Yachan Liu,‡[a] Gustavo Perez, [ID] ‡[b] Zezhou Cheng,[b] Aaron Sun,[b] Samuel C. Hoover,[a] Wei Fan, [ID] [a] Subhransu Maji*[b] and Peng Bai [ID] *[a]

Zeolites are one of the most widely used materials in the chemical industry due to their nanometer-sized pores that can adsorb and react upon molecules selectively. With hundreds of known framework topologies and hundreds of thousands of computationally predicted structures, the ability to rapidly predict zeolite performance allows researchers to prioritize their efforts on the most promising structures for a given application. Although the accuracy of forcefield-based atomistic simulations has advanced significantly in the past two decades, these simulations can be computationally expensive, especially for long-chain, complex molecules. We present ZeoNet, a representation learning framework using convolutional neural networks (ConvNets) and 3D volumetric representations for predicting adsorption in zeolites. ZeoNet was trained on the task of predicting Henry's constants for adsorption, k_H, of *n*-octadecane in more than 330 000 known and predicted zeolite materials. Employing a 3D grid based on the distances to solvent-accessible surfaces, a volumetric representation that can be generated efficiently, the best-performing ZeoNet achieved a correlation coefficient $r^2 = 0.977$ and a mean-squared error MSE = 3.8 in $\ln k_H$, which corresponds to an error of 9.3 kJ mol^{-1} in adsorption free energy. In comparison, a model based on hand-designed geometric features has values of $r^2 = 0.783$ and MSE = 35.7. ZeoNet is also relatively efficient and can process ≈ 8 structures per second on an Nvidia RTX 2080TI GPU, orders of magnitude faster than forcefield-based simulations. A systematic analysis was conducted to investigate how the choice of ConvNet architectures, the linear dimension (L) and spatial resolution (Δd) of the distance grids, batch size, optimizer, and learning rate impact the model performance. We found that ConvNets based on the ResNet architecture offer the best tradeoff between expressiveness and efficiency. The performance for all models reaches a plateau at $L = 30-45$ Å and depends less sensitively on grid resolution, with a small benefit around $\Delta d = 0.30-0.45$ Å. Finally, saliency maps were visualized to identify which regions of the materials contributed the most to model predictions. It was found, interestingly, that the predictions are driven primarily by the accessible pore volume rather than the region occupied by the framework atoms.

[a]*Department of Chemical Engineering, University of Massachusetts Amherst, Amherst, Massachusetts 01003, USA. E-mail: pengbai@umass.edu*

[b]*College of Information and Computer Sciences, University of Massachusetts Amherst, Amherst, Massachusetts 01002, USA. E-mail: smaji@cs.umass.edu*

Peng Bai is an Assistant Professor of Chemical Engineering at the University of Massachusetts Amherst. He obtained his BSc degree from Tsinghua University in China and received both his PhD degree and post-doctoral training from the University of Minnesota. His group's research focuses on the development and application of molecular simulation, first-principles, and data science methods. Current interests include the catalytic upcycling of polymers, separation with nanoporous materials, and ion conduction in solid-state batteries. He is the recipient of the NSF CAREER Award and an ACS PRF Doctoral New Investigator.

17570 | *J. Mater. Chem. A*, 2023, **11**, 17570–17580

This journal is © The Royal Society of Chemistry 2023

1 Introduction

Nanoporous materials such as zeolites and metal organic frameworks (MOFs) are important adsorbents and catalysts in the chemical industry due to their numerous applications such as gas storage, separation, and shape-selective catalysis.[1,2] However, finding the best zeolite for a given application is challenging since the relationship between performance and structure is often unknown, and the space of potential structures is large. To date there are over 250 known zeolite framework topologies[3] and hundreds of thousands of computationally predicted structures.[4] Although the development of accurate, transferable intermolecular potentials[5–7] have enabled the computational predictions of adsorption performance in zeolites for a diverse range of applications,[8–12] physics-based simulations still require significant computational resources, especially when large materials databases or complex mixtures are involved.[8,13]

Machine learning (ML) is increasingly being used to predict structure–property relationships in a data-driven manner. Such efforts have roots in quantitative structure–activity relationships (QSAR) for drug design[14] and other molecular property predictions. These cheminformatics and ML approaches have often used features of atoms and their connectivity such as electronegativity, bond order, molecular weights, and surface area as descriptors. Along a similar line but adapting for extended crystalline materials, Gaillac et al.[15] selected 22 local descriptors, 19 global descriptors, and seven porosity descriptors including bond lengths, densities, pore volume, and accessible surface area to predict the mechanical properties of zeolites. Anderson et al.[16] built a multi-layer perceptron (MLP) model using six textural properties (e.g., helium void fraction, gravimetric surface area, largest cavity diameter, pore limiting diameter, inverse framework density, and the pore size standard deviation) together with the number density of 17 distinct MOF chemical moieties to predict the adsorption isotherms in MOFs. While conceptually intuitive, these features are high-level coarse-grained properties that may not be able to accurately capture phenomena dominated by structural details of a material. Adsorption by all-silica zeolites is one such example: the materials are chemically identical, which all consist of corner-sharing SiO_4 tetrahedra, and their dramatic molecular shape selectivity is completely controlled by how framework atoms are arranged in space.[1,17,18]

Given the materials structures and an accurate intermolecular potential, the quantitative prediction of adsorption in porous materials is, to a large extent, a solved problem through the use of molecular simulations.[1] For many adsorption systems, the assumption of a rigid framework structure allows one to pre-tabulate the energies felt by a probe molecule on a regular grid, a practice that improves the simulation efficiency by allowing the framework–sorbate interactions to be interpolated rather than computed. In other words, the energy grid contains complete information about a solid material that can be considered as rigid. Based on this insight, energy grids have been used as the input for ML models by Snurr et al.[19,20] The interaction energy of a hydrogen probe at each grid point within the MOF unit cell was calculated and then summarized as an energy histogram. Bins of the energy histogram were used as the input to train a regression model to predict hydrogen and methane uptake with an accuracy within 3 g L^{-1}. Then, they extended this method to gas mixtures such as binary mixtures of Ke and Xr, and short linear alkanes up to propane. The selectivity for Xe over Kr in Xe/Kr mixtures and single-component adsorption of ethane and propane can be predicted in good agreement with grand-canonical Monte Carlo simulations. However, energy grids are computationally relatively expensive to calculate and condensing them into histograms may also lose 3D structural information.
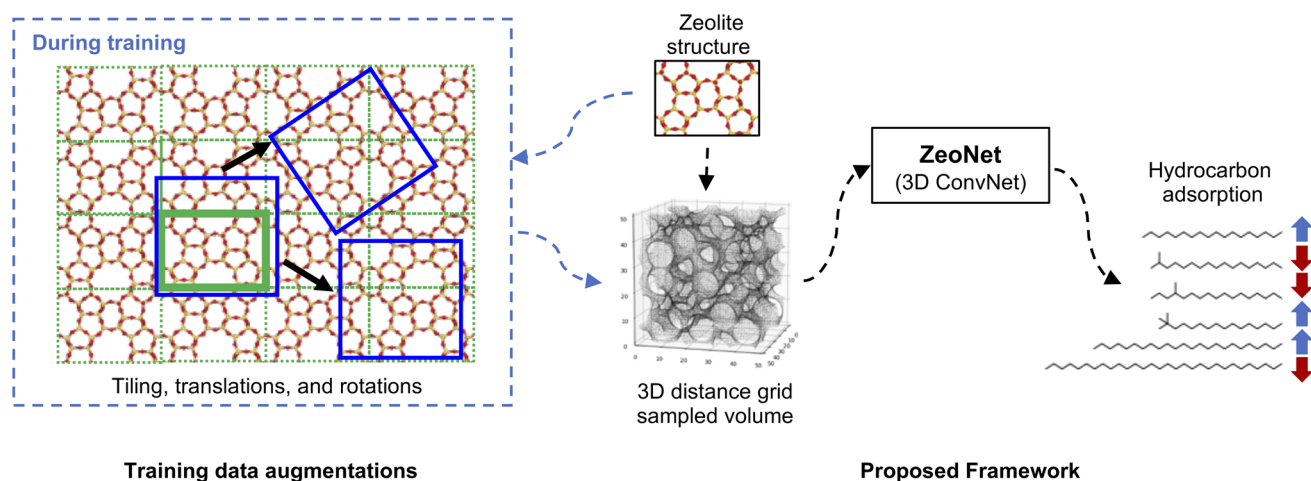

Fig. 1 The ZeoNet pipeline for predicting adsorption in zeolites. The unit cell of a zeolite is replicated to obtain an extended material structure. A fixed-size volumetric chunk with random origins and orientations is converted to a distance grid representation, which is fed to train 3D ConvNets using data collected from physical simulations in a supervised learning setting. In this work, Henry's constants for *n*-octadecane (C18) adsorption in more than 330 000 known and computationally predicted zeolites were used as the training data. For inference, ZeoNet is applied in a feed-forward manner on the distance grids without the translation and rotation augmentations.

This journal is © The Royal Society of Chemistry 2023

J. Mater. Chem. A, 2023, **11**, 17570–17580 | **17571**

Table 1 Model performance for predicting C18 adsorption in zeolites, comparing the accuracy and efficiency of a multi-layer perceptron (MLP) and extreme gradient boosting (XGBoost) trained on geometric features with various ZeoNet architectures operating on distance grids. The optimal input representations for each model are given in parentheses, (Δd, L) in Å (see the main text for details)

Model	r^2	MSE	Time [s per sample]
MLP	0.783	35.7	0.0049
XGBoost	0.841	26.2	0.001
3D AlexNet (1, 100)	0.944	9.2	0.14
3D VGG (0.45, 45)	0.961	6.4	0.14
3D ResNet (0.45, 45)	0.973	4.4	0.13
3D DenseNet (0.45, 45)	0.977	3.8	0.14
MC simulations			≈ 1 hour

To represent 3D structures directly, Lin *et al.*[21] pioneered the use of 3D ConvNets with a binary occupancy grid, in which each grid location was marked as either zero or one depending on its distance to the nearest framework atom. They used a LeNet/AlexNet based network to predict methane adsorption isotherms and were able to achieve an MSE of 0.015 mol kg^{-1} in loadings. This approach was recently extended to CO_2 adsorption in MOFs.[22] Using both structural and energy grids, Kim *et al.* developed a generative adversarial network to produce plausible zeolite structures with user-specified heats of adsorption for methane.[23,24] While these studies demonstrate the utility of modern ConvNets in representing materials structures, they have focused on the adsorption of small, relatively rigid molecules. It remains unclear how well 3D ConvNets perform for large, flexible molecules whose properties are expected to be influenced not only by the local pore dimensions, but also their larger structural features.

In this work, we propose a 3D structural representation learning method, ZeoNet, for the task of predicting the adsorption of *n*-octadecane, a long-chain hydrocarbon molecule, in all-silica zeolites (see Fig. 1). We carried out a systematic evaluation of 3D ConvNets and benchmarked them against MLP and XGBoost regressors trained on high-level descriptors. Four 3D ConvNets were tested, which were 3D variants of the popular AlexNet,[25] VGG Net,[26] ResNet,[27] and DenseNet.[28] Two volumetric representations, one based on binary occupancy grids and the other based on distance grids, were compared. The effect of grid resolution, input size, and other hyperparameters such as batch size, learning rate, and optimizer were examined. As summarized in Table 1, ZeoNet vastly outperformed the MLP and XGBoost regressors and among the various 3D ConvNets, modern deep networks provided significant improvement in model accuracy compared to older AlexNet without sacrificing inference speeds.

2 The ZeoNet framework

2.1 Adsorption dataset

To study the ability of 3D ConvNets in capturing spatial correlations of materials structures, the dataset of long-chain hydrocarbon adsorption was selected. This dataset was produced from a computational screening study that used Monte Carlo (MC) simulations to predict the adsorption of three normal alkanes from C18 to C30 and mono- and di-branched C18 isomers.[8] The adsorption at both the infinite-dilution regime (as characterized by Henry's constants, k_H, and heats of adsorption) and a high-pressure, liquid regime (as characterized by the loadings at $p = 3$ MPa for an equimolar, six-component mixture) was calculated. The intermolecular potentials used in this study were developed for a diverse range of molecules and zeolite structures and their accuracy has been validated extensively against experiments.[7,29] In total, the study included 402 experimentally synthesized structures catalogued by the Structure Commission of the International Zeolite Association (IZASC)[3] and 331 172 computationally predicted structures from the Predicted Crystallography Open Database (PCOD).[4] Here, we focus on *n*-octadecane (C18), a linear hydrocarbon molecule that has a length of $\sim$2.2 nm when fully extended, and predicting ln k_H, as k_H scales exponentially with the adsorption free energy. Therefore, zeolites for which $k_H = 0$ were removed, leaving 100 520 structures (269 IZA zeolites and 100 251 PCOD zeolites). It is also worth noting that due to the stochastic nature of the simulations, the adsorption estimates have statistical uncertainties, not unlike experimental measurements, and zeolites with higher adsorption strengths tend to have smaller uncertainties; fortunately these are precisely the structures more important for the application.

The dataset was initially split randomly into 60% (60 312) for training, 20% (20 104) for validation, and 20% (20 104) for testing. The test set was then sub-divided in order to determine the minimum size needed to reach the level of precision desired for model evaluation. The first subdivision included ten sets, each containing 2000 samples, the second included four sets of 5000 samples each, and the third included two sets of 10 000 samples each. Based on the results discussed later, 10 104 testing samples were moved to the training set, resulting in a training/validation/testing split of roughly 7 : 2 : 1.

2.2 Volumetric grids and high-level feature descriptors

Zeo++, version 0.3,[30] was used to calculate distance grids with a probe radius of 1.2 Å and a grid resolution of 0.15 Å, while distance grids with lower resolutions were obtained *via* downsampling using the trilinear interpolation. In a distance grid, each grid location is assigned its shortest distance to the solvent-accessible surface formed by zeolite framework atoms. In this calculation, Si and O atoms have radii of 2.1 and 1.52 Å, respectively. The distance can be positive or negative, depending on whether the grid locations lie outside or inside the solvent-accessible surface. To construct the binary occupancy grid, we simply assign a value of one to all grid locations where distances are non-positive and zero where they are positive.

Zeo++ was also used to calculate the pore-limiting diameter (PLD, unit Å), the largest-cavity diameter (LCD, unit Å), surface area (unit m^2 g^{-1}), and pore volume (unit cm^3 g^{-1}) for each zeolite using a spherical probe with a radius of 1.2 Å, as well as the number density of framework Si atoms (ρ_{Si}, unit number per nm^3). These high-level aggregate feature descriptors were used

17572 | *J. Mater. Chem. A*, 2023, **11**, 17570–17580

This journal is © The Royal Society of Chemistry 2023

to construct a MLP regressor and an XGBoost regressor as the performance baseline.

2.3 ConvNet architectures

3D variants of four ConvNet architectures, which have been used extensively for image recognition, were evaluated. These architectures are designed to work with primarily RGB images and employ 2D convolutions. To operate on 3D data, we replace the 2D convolutions and pooling operations in these networks with their 3D variants similar to prior work that has extended these architectures to handle spatio-temporal data (*e.g.*, for video understanding[31]). We briefly describe these architectures below.

2.3.1 AlexNet. AlexNet was the first large-scale model trained for image classification and won the 2012 ImageNet Challenge.[25] Our implementation consists of seven 3D convolutional (Conv) layers and two fully-connected (FC) layers. Each Conv layer is followed by batch normalization and ReLU activation. Two max pooling layers are inserted after the second and fourth Conv layers. All conv filters have 16 channels, a kernel size of 3, a stride of 1, and a padding of 1. The max pooling layers have a kernel size of 2 and a stride of 2.

2.3.2 VGG16. The VGG architectures[26] were introduced as deeper variants of AlexNet with several design changes and outperformed AlexNet on the ImageNet challenge. This VGG16 architecture consists of five blocks and three FC layers. A dropout of 0.5 is added after each of the first two FC layers. The first two blocks each contain two Conv layers and the latter three contain three Conv layers. All Conv layers have a kernel size of 3, a stride of 1, and a padding of 1, which is followed by batch normalization, and ReLU activation. Each block is terminated by a max-pooling layer with a kernel size of 2 and a stride of 2. The first block has 64 output channels and each subsequent block doubles the number of output channels, until it reaches 512.

2.3.3 ResNet18. He *et al.* introduced residual blocks with skip connections for training substantially deeper networks.[27] The ResNet18 architecture used in this work consists of a Conv layer with a kernel size of 7, a stride of 2, and a padding of 3, followed by a max pooling layer with a kernel size of 3, a stride of 2, a padding of 1, and a dilation of 1. This is followed by four modules that each contain two residual blocks, and finally, an average pooling layer and a FC layer. Each residual block contains two Conv layers with a kernel size of 3, a stride of 1 or 2, and a padding of 1. The output channels of the first residual module is 64, and is doubled in each subsequent residual module by including a 1×1 Conv layer in the first skip connection while the height, width, and depth are halved in the last Conv layer. Batch normalization and ReLU activation are used after all Conv layers.

2.3.4 DenseNet121. Extending the idea of residual connections, Huang *et al.*[28] proposed densely-connected networks in which each layer's output is concatenated in all subsequent layers in a feed-forward fashion. The DenseNet121 architecture used here consists of a Conv block, six dense blocks, a transition block, 12 dense blocks, a transition block, 24 dense blocks, a transition block, 16 dense blocks, and a FC layer. The first Conv layer has a kernel size of 7, a stride of 2, and a padding of 3. All dense blocks are identical, containing two Conv blocks, each using the modified ResNet structure[32] of batch normalization, ReLU activation, and convolution. The Conv layer in the first block has a kernel size of 1 and a stride of 1 and that in the second block has a kernel size of 3, a stride of 1, and a padding of 1. The transition block contains batch normalization, ReLU activation, a Conv layer with a kernel size of 1 and a stride of 1, and an average pooling layer with a kernel size of 2 and a stride of 2, hence reducing the number of the output channels. The number of output channels of the three transition blocks are 128, 256, and 512, respectively.

2.4 Training

All models were trained to predict ln k_H using the mean squared error (MSE) as the loss function. The baseline MLP and XGBoost models used high-level aggregate features including PLD, LCD, density of framework Si atoms, surface area, and pore volume as input, while the four 3D ConvNets used distance grids as input. During training of 3D ConvNets, random translations up to full unit cell lattice lengths and rotations covering all possible spherical angles were applied as data augmentation techniques. The resulting 3D grid was then tiled and cropped to create the desired input size (see Fig. 1). The grids at this stage have the same lattice system as the materials themselves but were re-sampled into a cubic lattice. Trilinear interpolation was used for the translation, rotation, and re-sampling operations. For all modeling work, PyTorch v1.11.0 was used with an Nvidia RTX 2080TI or A100 GPU as the accelerator. All 3D ConvNets were trained for a total of 30 epochs with a batch size up to what is allowed by the GPU memory. Apart from the section on hyper-parameter optimization, the Adam optimizer was used with a learning rate of 0.001 and a batch size of 16 for AlexNet and ResNet18, 4 for VGG16, and 8 for DenseNet121.

3 Results and discussion

3.1 How large does the test set size need to be?

To maximize the number of training samples while also ensuring that the test set is large enough to allow for precise estimates of model performance, test sets of different sizes were used to evaluate an AlexNet model pre-trained using 60 312 training and 20 104 validation samples. As shown in Table 2, increasing the test set size from 2000 to 5000 leads to roughly seven times more precise estimate of the model performance. With 5000 or 10 000 testing samples, r^2 is accurate to the third

Table 2 Mean, standard deviation, and spread of r^2 and MSE for a trained AlexNet model as evaluated using test sets of different sizes

Test set	10×2000	4×5000	$2 \times 10\,000$	$1 \times 20\,104$
r^2	0.665 ± 0.012	0.665 ± 0.002	0.665 ± 0.002	0.665
Spread	[0.639, 0.682]	[0.662, 0.667]	[0.663, 0.667]	
MSE	56 ± 3	55.7 ± 0.4	55.7 ± 0.3	55.8
Spread	[51, 60]	[55.3, 56.4]	[55.4, 56.0]	

This journal is © The Royal Society of Chemistry 2023

J. Mater. Chem. A, 2023, **11**, 17570–17580 | **17573**

Table 3 Statistics of the geometric features for all zeolites from the IZA and PCOD databases for which $k_H > 0$ and channels are externally accessible to a probe of radius 1.2 Å

	Min	Max	Mean	Median	SD
All zeolites					
PLD (Å)	2.8	28.8	5.3	4.8	1.8
LCD (Å)	3.6	29.2	6.7	6.3	1.7
ρ_{Si} (number per nm^3)	8.3	26.3	17.9	18.0	1.7
Surface area (m^2 g^{-1})	60	2038	700	676	200
Pore volume (cm^3 g^{-1})	0.0014	0.6684	0.06	0.0471	0.04
IZA zeolites					
PLD (Å)	2.9	12.4	5.4	5.2	1.6
LCD (Å)	4.4	16.9	7.6	7.1	2.0
ρ_{Si} (number per nm^3)	10.8	26.3	16.6	16.9	2.0
Surface area (m^2 g^{-1})	350	2038	1000	927	300
Pore volume (cm^3 g^{-1})	0.0211	0.3483	0.09	0.0774	0.05
PCOD zeolites					
PLD (Å)	2.8	28.8	5.3	4.8	1.8
LCD (Å)	3.6	29.2	6.7	6.3	1.7
ρ_{Si} (number per nm^3)	8.3	26.3	17.9	18.0	1.7
Surface area (m^2 g^{-1})	60	1826	700	676	200
Pore volume (cm^3 g^{-1})	0.0014	0.6684	0.06	0.0470	0.04

decimal digit and MSE is accurate to the first decimal digit, which we consider adequate for comparing subsequent benchmarks.

3.2 Materials characteristics and MLP/XGBoost performance

Porous materials are conventionally characterized using geometric concepts. Viewing framework atoms as spheres of different radii, one can define the pore volume and surface area to be the unoccupied space (a 3D property) and exposed surface (a 2D property) per unit mass of the material. If a spherical probe is placed in the free space, the radius of the largest sphere that can fit at a given location is defined as the local pore diameter (a 1D property), and since the interior of zeolites is not uniform, one can further distinguish between the pore-limiting diameter (PLD) and the largest-cavity diameter (LCD), which are the smallest and largest local pore diameters across an entire zeolite, respectively. Table 3 gives a summary of the descriptive statistics of these high-level geometric features for all materials in the dataset and Fig. 2 compares the distributions of all the known zeolites and the computationally predicted ones. As shown in Fig. 2 and also noted by Pophale *et al.*,[4] the computationally generated PCOD database contains a larger amount of smaller-pore zeolites, coincident with higher Si atom density, lower surface area, and smaller pore volume. A fraction of these zeolites contain channel systems inaccessible externally by a probe with a radius of 1.2 Å, which is given a value of zero in Fig. 2.

The above geometric features are often used in scatter plots to construct structure–property relationships, although the
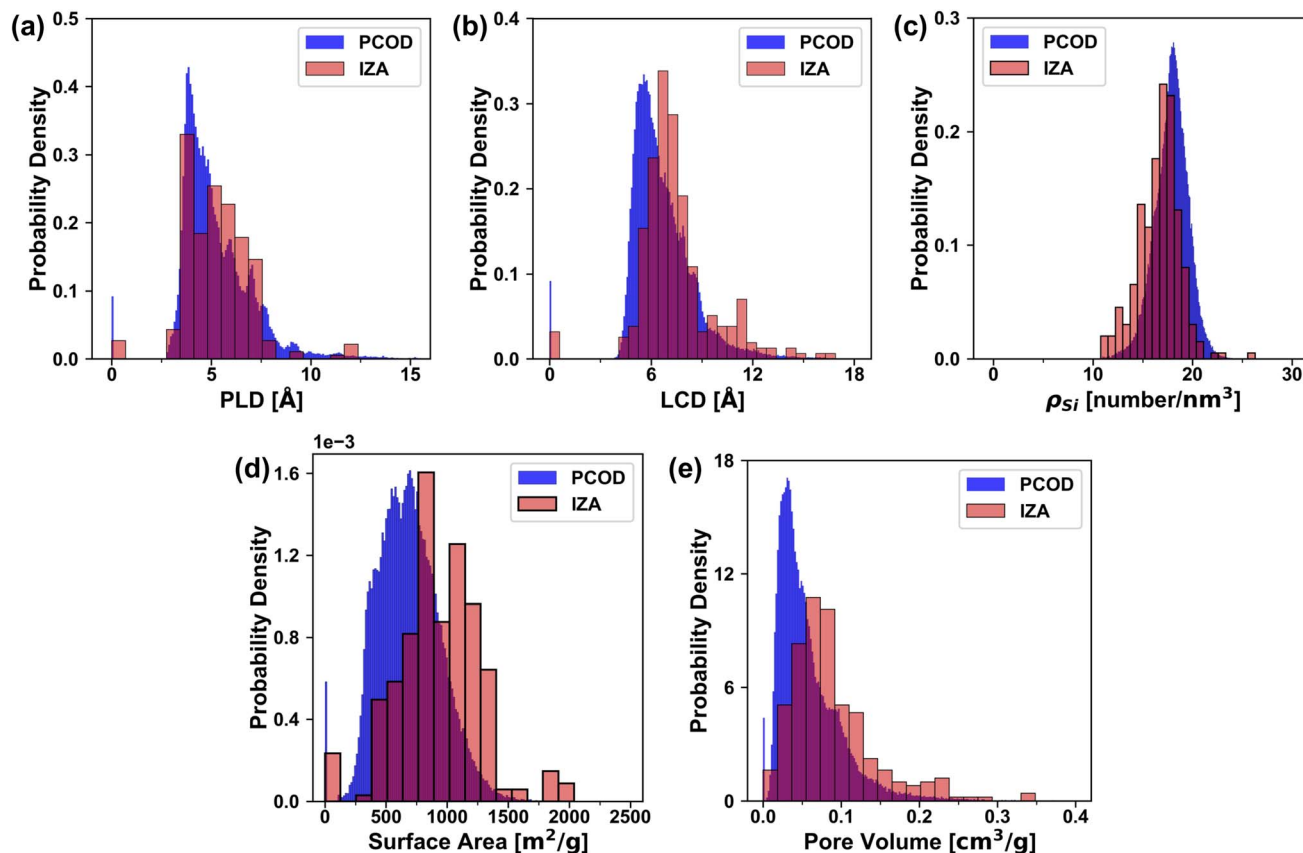

Fig. 2 Probability density for the distributions of five geometric features: (a) PLD, (b) LCD, (c) ρ_{Si}, (d) surface area, and (e) pore volume, for zeolites with $k_H > 0$. A value of zero indicates zeolites without externally accessible channels for a probe of radius 1.2 Å. From (a) to (e), the bin widths are 0.089, 0.089, 0.091, 14.96, and 0.002 for PCOD zeolites and 0.68, 0.58, 0.74, 127.39, and 0.018 for IZA zeolites.

17574 | *J. Mater. Chem. A*, 2023, **11**, 17570–17580

This journal is © The Royal Society of Chemistry 2023

resulting correlations are largely noisy and non-predictive (see Fig. 3 of ref. 8 as an example). However, to provide a baseline to compare with results obtained with 3D ConvNets, we trained a MLP model and an XGBoost model to predict $\ln k_H$, the logarithmic Henry's constant for the adsorption of n-octadecane. The MLP achieved a value of r^2 of 0.783 and an MSE of 35.7, which corresponds to an error of $\sim$28.5 kJ mol^{-1} in the free energy of adsorption, ΔG_{ads}. The XGBoost model performs better, with $r^2 = 0.841$ and MSE = 26.2. The performance of the two models is significantly better than can be expected from the broad scatter plots of individual geometric descriptors.

3.3 Performance and optimization of ZeoNet

3.3.1 Comparing binary occupancy grids and distance grids. Among different volumetric representations, intuitively, energy grids would be expected to contain the most physical information, as they are widely used to speed up atomistic simulations. However, computing an energy grid involves calculating the interactions of a probe atom with all framework atoms and is thus computationally rather expensive. We therefore investigate two alternative volumetric representations that are easier to calculate, including binary occupancy grids that have been used by Lin's group[21] and distance grids implemented by Zeo++ (see Computational Details for the calculation of both grid representations). All four 3D ConvNet models were trained using both representations with an input shape of $100 \times 100 \times 100$. As shown in Table 4, distance grids outperform binary occupancy grids in almost all cases, with the only exception being VGG16 at a grid resolution of 1 Å. When using the default grid resolution in Zeo++, $\Delta d = 0.15$ Å, the values of r^2 for distance grids exceed those for occupancy grids by 0.014–0.046, while MSE is lower by 2.2–7.6. The largest difference is found for AlexNet, which also shows the worst performance for both representations, with $r^2 < 0.68$ and MSE > 54, while the deeper VGG16 model and the more modern architectures, ResNet18 and DenseNet121, exhibit a dramatic improvement, with $r^2 > 0.83$ and MSE < 28. Also included in Table 4 are the results obtained with the two representations down-sampled to a grid resolution of 1 Å (while keeping the same input grid dimension). The resulting coarser, but larger volumetric grids show even more pronounced improvements than achieved by the more modern 3D ConvNet architectures. The r^2 values are larger than 0.91 and MSE lower than 13.5 in all

cases, with a much smaller difference between the two representations. It is apparent that a large enough input volume is critical to ensure good performance, presumably due to the long-chain hydrocarbon molecule selected for the target application, which requires spatial learning of larger patches of the materials structure. As the input volume becomes more limited, the performance of the simplest AlexNet model suffers the most.

3.3.2 Effect of input volume and grid resolution. Following the observation that the size of the input volume greatly influences the performance of 3D ConvNets, in this section, the effect of input volume was systematically studied. We focus on the distance grid representation and vary the grid resolution from 0.15 to 1 Å while keeping its shape at $100 \times 100 \times 100$. Consequently, the distance grids represent a cubic input volume with a linear dimension, L, ranging from 15 to 100 Å (see Fig. 3). Fig. 4 (numerical data can be found in ESI Tables S5–S7†) shows how the performance metrics vary with input volume: As L increases, r^2 increases and MSE decreases sharply, by 0.09–0.27 and 15.5–45.0, respectively, until the model stabilizes roughly at $L \sim 45$ Å for AlexNet and VGG16 and at $L \sim 30$ Å for ResNet18 and DenseNet121. Above $L \sim 45$ Å, the performance of ResNet18 and DenseNet121 decreases slightly, by about 0.01 (r^2) and 1.9 (MSE) at $L = 100$ Å, indicating a potential loss of details due to the lower grid resolutions. The degradation is more significant for VGG16, with r^2 decreasing by 0.04 and MSE increasing by 7.1, although this observation may be an idiosyncrasy of the specific training runs (see ESI Fig. S6†). In contrast, the performance of AlexNet continues to improve, albeit slightly, up to the largest input volume tested, $L = 100$ Å. The relatively smaller depth of AlexNet might be limiting its ability to learn larger-scale features, leading to its lower accuracy than that of the ResNet models at all input volumes/grid resolutions.

Given the relatively similar performance, it is useful to compare the training speeds of the four 3D ConvNet models. With GPU acceleration using Nvidia RTX 2080TI, the ratio of training times per epoch is roughly 1 : 2 : 1.5 : 2.5 for AlexNet,

Table 4 Validation accuracy for binary occupancy grids and distance grids of different resolutions, Δd, in Å. The grid dimension is 100^3 in all cases

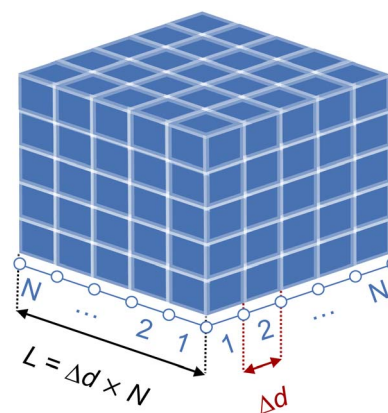|	Binary grids				Distance grids			
	$\Delta d = 0.15$		$\Delta d = 1$		$\Delta d = 0.15$		$\Delta d = 1$	
Model	r^2	MSE	r^2	MSE	r^2	MSE	r^2	MSE
AlexNet	0.630	61.6	0.937	10.4	0.676	54.0	**0.946**	**9.0**
VGG16	0.836	27.4	**0.942**	**9.6**	0.851	24.9	0.919	13.5
ResNet18	0.840	26.7	0.953	7.8	0.879	20.1	**0.961**	**6.5**
DenseNet121	0.867	22.1	0.957	7.2	0.881	19.9	**0.968**	**5.3**

Fig. 3 Illustration of a 3D volumetric grid with a shape of $N \times N \times N$ and a spatial resolution of Δd (in Å). $N\Delta d$ gives the linear dimension, L (in Å), of the input volume represented by the grid.
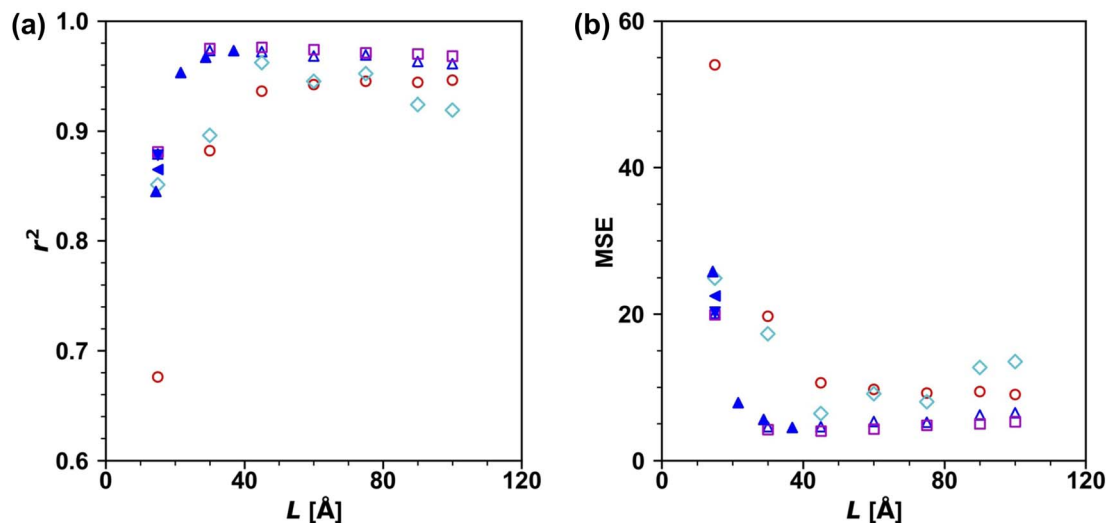
Fig. 4 The performance of 3D ConvNets as a function of the linear dimension of the input volume, showing r^2 (a) and MSE (b) for AlexNet (red circles), VGG16 (cyan diamonds), ResNet18 (open blue up triangles), DenseNet121 (magenta squares) with a fixed input tensor shape of $100 \times 100 \times 100$, for ResNet18 with a fixed grid resolution of 0.45 Å (filled blue up triangles), and for ResNet18 with a grid resolution of 0.3 Å and an input tensor shape of $50 \times 50 \times 50$ (blue down triangles) or with a grid resolution of 0.6 Å and an input tensor shape of $25 \times 25 \times 25$ (blue left triangles).

VGG16, ResNet18, and DenseNet121. It is also worth noting that the much larger models, VGG16 and DenseNet121, can only afford a batch size of 4 with the 11GB GPU memory on Nvidia RTX 2080TI. The training processes are thus much noisier and are completely unstable for DenseNet121. As a result, all training runs for DenseNet121 were run on Nvidia A100 with 40GB GPU memory using a batch size of 8. Considering these technical characteristics, ResNet18 provides the best balance in terms of model accuracy and computational efficiency and is therefore chosen to be the focus of subsequent studies.

Using ResNet18, the effect of input volume was examined at a fixed grid resolution of 0.45 Å. This set of data is shown as filled up triangles in Fig. 4, which largely fall onto the same trend line as the previous test with different grid resolutions but a fixed input grid shape. Differences become larger with smaller input volumes (those with $L < 30$ Å): comparing $L = 14.4$ Å and $\Delta d = 0.45$ Å and $L = 15$ Å and $\Delta d = 0.15$ Å, the r^2 and MSE values for the latter are better by 0.03 and 5.7, respectively. To compare model performance at exactly the same input volume, two additional tests were performed using a grid resolution of 0.3 Å and an input shape of 50^3 or a grid resolution of 0.6 Å and an input shape of 25^3, for an input volume with $L = 15$ Å. As shown in Fig. 4, the performance of grid resolutions of 0.15 and 0.3 is almost indistinguishable, while the grid resolution of 0.6 Å is slightly worse.

3.3.3 Optimization of hyperparameters. Here, the performance of ResNet18 was further optimized by tuning the size of the mini batches, optimizer, and learning rate. Given the comparisons in the previous section, a grid resolution of 0.45 Å and an input shape of 100^3 are considered nearly optimal and therefore used without change during the hyperparameter optimization process. Four optimizers were tested, including Adam,[33] Adagrad,[34] RMSprop[35] and vanilla stochastic gradient descent (SGD).[36] Table 5 summarizes the results obtained with

the different hyperparameters. First, the effect of learning rate was examined with a batch size of 64 (c.f., last three rows), the largest that can fit into the GPU memory of Nvidia A100. Next, the batch size was varied from 64 to 4, while the learning rate, according to the commonly used heuristic, was halved with every halving of batch size, resulting in a learning rate of 0.00025 for a batch size of 4 and a learning rate of 0.004 for a batch size of 64. Overall, the training of ResNet18 is largely insensitive to batch sizes and learning rates, achieving nearly identical results with all hyperparameters when the batch size is larger than 8. At the two smallest batch sizes, 4 and 8, the performance is slightly worse with the Adam or Adagrad optimizers. Adam is the best optimizer for this system, slightly outperforming the other three across combinations of batch sizes and learning rates. The best model was obtained using a batch size of 64 and a learning rate of 0.004, which achieved a r^2 coefficient of 0.974 and an MSE of 4.4 on the validation set. Very similar performance metrics ($r^2 = 0.973$ and MSE $= 4.4$) were found for the test set, indicating a good model generalization.

Table 5 Validation accuracy for ResNet18 trained with different optimizers, batch sizes, and learning rates

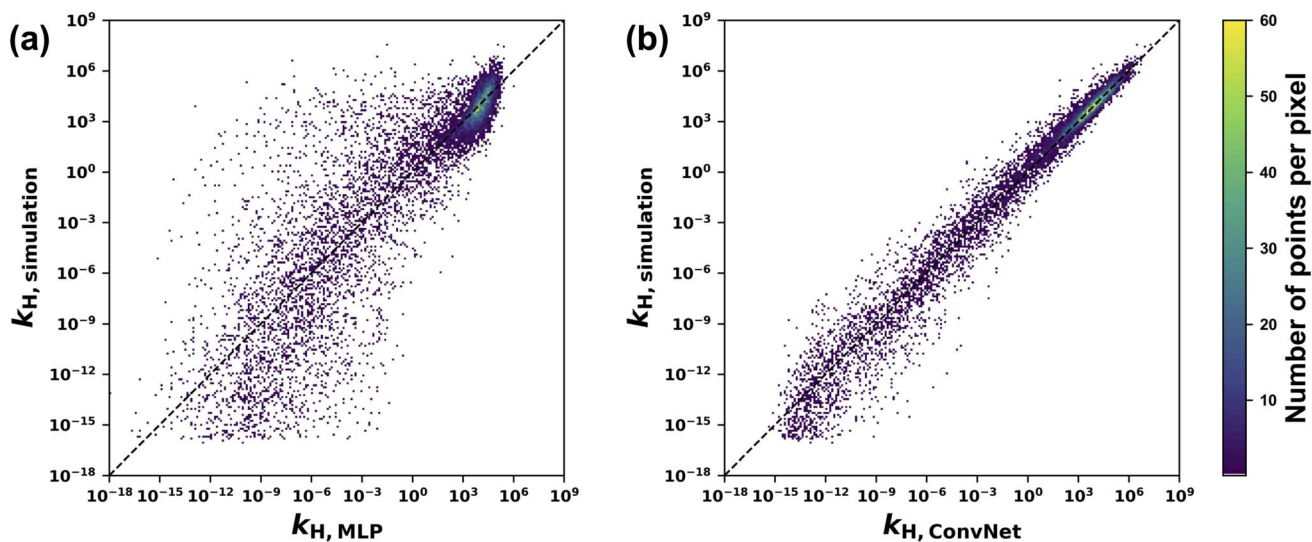| Batch size & learning rate	Adam		Adagrad		RMSprop		SGD	
	r^2	MSE	r^2	MSE	r^2	MSE	r^2	MSE
4 & 0.00025	0.968	5.3	0.935	10.9	0.971	4.8	0.972	4.6
8 & 0.0005	0.967	5.5	0.952	7.9	0.966	5.6	0.966	5.6
16 & 0.001	0.972	4.6	0.967	5.5	0.971	4.8	0.967	5.4
32 & 0.002	0.973	4.4	0.960	6.6	0.970	5.0	0.970	4.9
64 & 0.004	**0.974**	**4.4**	0.966	5.6	0.966	5.7	0.966	5.7
64 & 0.002	0.973	4.5	0.966	5.7	0.968	5.4	0.969	5.2
64 & 0.001	0.972	4.7	0.964	6.0	0.968	5.3	0.965	5.8

Fig. 5 Scatter plot of the testing performance, comparing Henry's constants for *n*-octadecane adsorption (unit: mol kg^{-1} MPa^{-1}) predicted by MLP (a) and ResNet18 (b) and target values from Monte Carlo simulations. Color indicates the number of points per pixel.

3.4 Analysis

3.4.1 Break-down of model performance. To gain a better understanding of model performance, a scatter plot was constructed to compare the target Henry's constants for *n*-octadecane adsorption from Monte Carlo (MC) simulations with the values predicted by the best ResNet18 model. As shown in Fig. 5, the predictions from ResNet18 cluster nicely around the parity line, although they are substantially more accurate for zeolites with larger values of k_H (*i.e.*, stronger adsorption). For $k_H < 1$ mol kg^{-1} MPa^{-1}, the correlation is visibly noisier. It is worth noting

that the ResNet18 model has a mean-squared error of 4.4 in ln k_H (Table 1), or 10.0 kJ mol^{-1} in ΔG_{ads}, but as k_H scales exponentially with ΔG_{ads}, even relatively small free energy differences manifest as large differences in Fig. 5. To quantify the distribution of prediction errors, k_H is grouped into nine classes and the resulting confusion matrix is shown in Fig. 6. Both figures show that the majority of zeolites have $k_H > 1$ mol kg^{-1} MPa^{-1} and these materials were predicted very well by the ResNet18 model. As k_H decreases, the prediction becomes less accurate and, interestingly, seems to be *slightly* positively biased (while still ranking near the bottom). Fig. 6 further demonstrates the generalizability of the trained model when they are applied to zeolites for which simulations predicted $k_H = 0$. These materials were excluded from the supervised learning
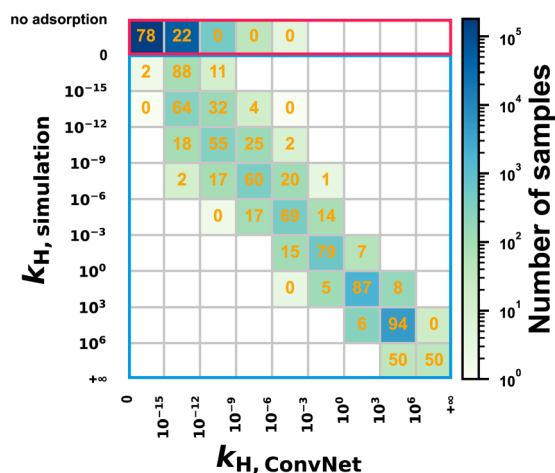

Fig. 6 Normalized confusion matrix showing the percentage of zeolites from a given target group (k_H for *n*-octadecane adsorption from Monte Carlo simulations; *y* axis) in different predicted groups of k_H using the best ResNet18 model. The background color of each grid indicates the number of zeolites represented. For comparison, the ResNet18 model was also applied to the zeolites with $k_H = 0$ that were not included in the training set. The predictions for these unseen zeolites are shown in the top row.
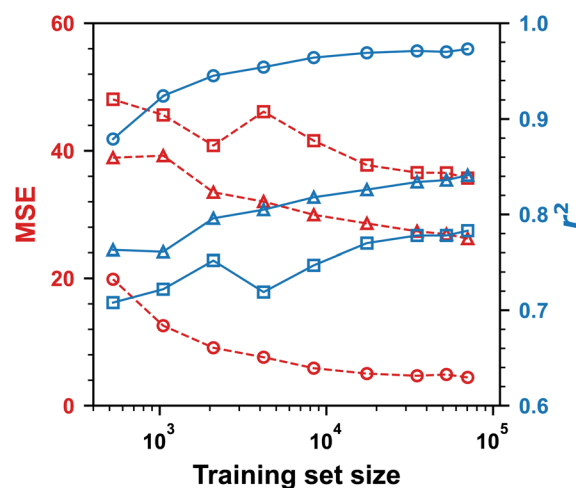

Fig. 7 MSE and r^2 values as a function of training set size for the best ResNet18 model and hyperparameters (circles), MLP (squares), and XGBoost (triangles). Each model was trained from scratch but evaluated on the same test set that consists of 10 000 samples.
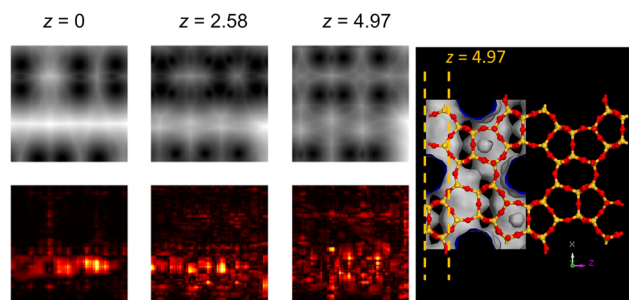
Fig. 8 Saliency maps based on the best ResNet18 model, showing three slices of the MFI zeolite at $z = 0$, 2.58, and 4.97 Å. The top row shows the distance grids and the bottom row shows the corresponding saliency maps. The right side illustrates the structure with the solvent accessible surface shown in white.

process as our loss function uses $\ln k_H$. For these zeolites, our ResNet18 model correctly predicted very small values of k_H and a similar small positive bias. As noted earlier, the dataset was generated from stochastic simulations of a finite length (9×10^4 MC steps; see ref. 8), where the Widom insertion MC moves used to compute Henry's constants[37] becomes much more difficult in weakly adsorbing zeolites, thus leading to larger uncertainties and potential under-predictions. To test this hypothesis, two zeolites with the largest ResNet18 over-predictions were selected, AEN-1 and PCOD-8314562, which had $k_H = 0$ and 3×10^{-15} mol kg^{-1} MPa^{-1} from previous MC simulations. Extending the simulations to 9×10^5 MC steps yielded $k_H = (5 \pm 45) \times 10^{-84}$ and $(7 \pm 14) \times 10^{-13}$ mol kg^{-1} MPa^{-1}. The new MC results indeed moved in the positive direction, although still much smaller than the ResNet18

predictions of $k_H = 3 \times 10^{-5}$ and 9×10^{-8} mol kg^{-1} MPa^{-1}. Examining these zeolites with the worst-case errors suggests that their pore diameters are barely large enough to fit linear alkanes (*e.g.*, AEN-1 has PLD = 3 and LCD = 3.93 Å) and increasing the pore sizes even slightly may lead to significant increases in k_H (within the rigid-zeolite assumption). We thus speculate that the spatial resolution of the ResNet18 model, while optimized for the prediction accuracy and efficiency over the entire dataset, may not be adequate to resolve the cutoff pore diameters.

3.4.2 Effect of training set size. To investigate how many training samples are needed to achieve good model performance, the best ResNet18 model was retrained from scratch using the optimal hyperparameters but with decreasing amounts of training data. These tests maintained the 7 : 2 training/validation split and used the same test set that consists of 10 000 samples. As summarized in Fig. 7, the model performance remains relatively unchanged as the number of training samples decreases from 70 416 to 17 500. Empirically, the minimum training set size for this adsorption system to achieve optimal results appears to be 10 000, below which the model performance degrades sharply. With 1050 training samples, the MSE in $\ln k_H$ increases to above 15 and r^2 drops below 0.93. Nonetheless, these values are still better than the best performance of the MLP and XGBoost models. Fig. 7 also shows the effect of training set size on the MLP and XGBoost models trained on high-level geometric features. r^2 decreases from 0.783 to about 0.7 (MLP) and from 0.841 to 0.76 (XGBoost), while MSE increases from 35.7 to above 45 (MLP) and from 26.2 to nearly 40 (XGBoost).

3.4.3 Feature visualization and attribution. To probe what is learned by a 3D ConvNet, one may ask two types of questions:
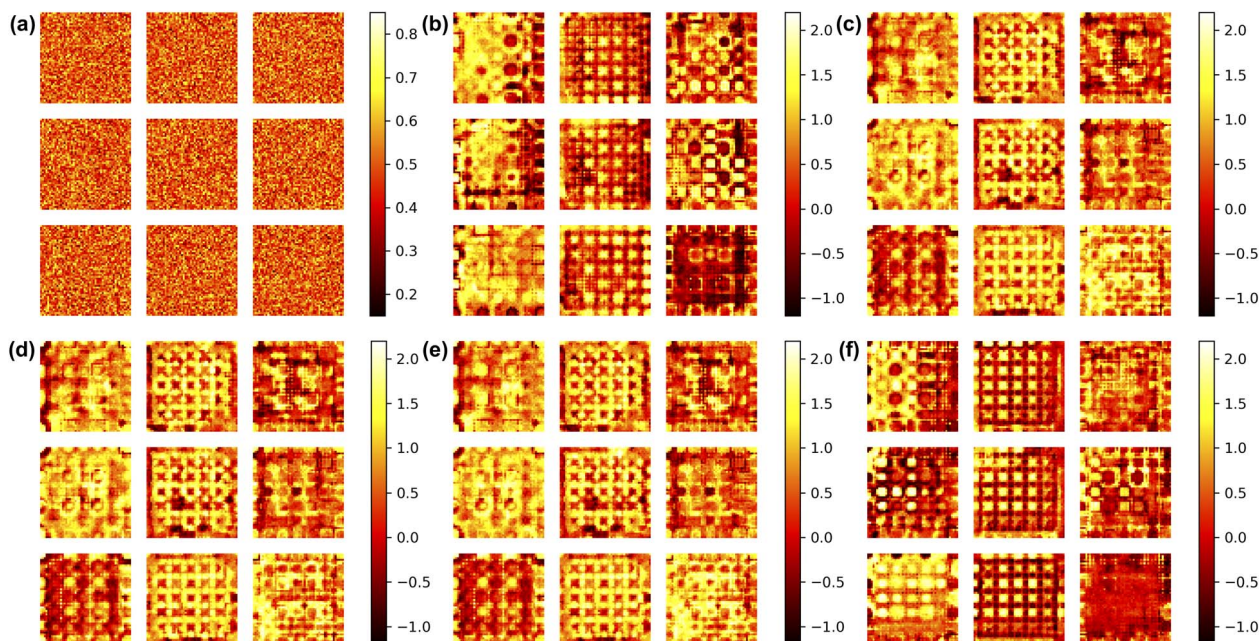

Fig. 9 Feature visualization showing (a) a random input and (b)–(f) input grids that strongly activate a particular feature map in the last Conv layer of the best ResNet18 model. In each subplot, the top, center, and bottom rows show 2D slices at *x*, *y*, and *z* = 1 (left), 25 (middle), and 50 (right).

17578 | *J. Mater. Chem. A*, 2023, **11**, 17570–17580

This journal is © The Royal Society of Chemistry 2023

(1) given a zeolite structure, what role different regions of the material play in directing the ConvNet to make its prediction; and (2) what features the ConvNet activates most in order to make predictions. Fig. 8 shows the saliency maps for the MFI zeolite based on the best ResNet18 model. Saliency maps are a feature attribution technique that assigns an importance value to each grid location as the gradient of the model output with respect to the input grid value.[38] The resulting 3D gradient fields thus characterize how much local changes of each grid influence the model prediction. By comparing with the corresponding distance grids, we found that, interestingly, the ResNet18 identifies the accessible pore volume and primarily relies on those regions to make its predictions. To answer the second question, one can look for the types of input structures that strongly activate a specified feature map, which can be obtained by solving an optimization problem starting from a random input (Fig. 9a). Here, we focus on feature maps in the last Conv layer of the best ResNet18 model as they represent higher-level features that may be more relatable. As shown in Fig. 9, the ConvNet appears to rely mostly on channels of different sizes and shapes (cylindrical *vs.* rectangular) to characterize zeolite structures. Finally, it is also worth noting that these visualizations are for structures that activate strongly, but not maximally the given feature map, as we found that pushing the optimization to convergence often yield unrealistic structures, *i.e.*, those with rapidly changing or even nearly discontinuous distance values, which may be due to strided convolutions and pooling operations.

4 Conclusions

In this work, we developed the ZeoNet representation learning framework to predict the adsorption of long-chain hydrocarbon molecules in all-silica zeolites using 3D ConvNets with volumetric representations. Using the logarithms of Henry's constants, $\ln k_H$, for *n*-octadecane adsorption as the target property, we performed a comprehensive evaluation of different ConvNet architectures and optimization of the grid representations and training hyperparameters. With almost all ConvNet models, it was found that distance grids, which contain the distances from each grid point to the nearest solvent accessible surface formed by zeolite framework atoms, outperformed binary occupancy grids that were used successfully for adsorption of small molecules (Table 4). Using the distance grid representation, we compared 3D variants of four popular ConvNet architectures: AlexNet, VGG16, ResNet18, and DenseNet121 (Table 1). These models all outperform a benchmark multi-layer perceptron trained on common geometric descriptors including pore-limiting diameters, largest-cavity diameters, surface areas, pore volume, and framework atom densities, which achieved a mean-squared error (MSE) of 35.7 and a correlation coefficient of $r^2 = 0.783$. The best prediction accuracy was obtained using DenseNet121, which reached $r^2 = 0.977$ and MSE = 3.8, corresponding to an error of 9.3 kJ mol^{-1} in adsorption free energy. AlexNet consistently underperformed modern ConvNets, with $r^2 = 0.944$ and MSE = 9.2. ResNet18 was found to provide the best balance between

expressiveness and efficiency, reaching an accuracy of $r^2 = 0.973$ and MSE = 4.4 but with a 70% faster training speed and a 75% reduction in memory requirements than DenseNet121. All 3D ConvNet models require a minimum input volume to obtain good performance, with AlexNet and VGG16 reaching a performance plateau at a linear dimension $L > 45$ Å and ResNet18 and DenseNet121 relatively stable between $L = 30$ and 100 Å (Fig. 4). The performance depends less sensitively on grid resolution, with a small benefit at $\Delta d = 0.30$–0.45 Å.

Analysis of the model performance (Fig. 5 and 6) reveals that ZeoNet is exceptionally accurate for zeolites with strong adsorption ($k_H > 1$ mol kg^{-1} MPa^{-1}) and slightly over-predicts compared to simulation results for weakly-adsorbing zeolites, which we argue may in fact be partly due to inadequate sampling by the grand-canonical Monte Carlo simulations for the more challenging adsorption systems. In addition, saliency maps suggest that the ConvNets mostly rely on the accessible pore volume to make predictions (Fig. 8) and visualization of feature maps further indicates that geometric primitives such as channels of different sizes and shapes are features learned by the ConvNets (Fig. 9). Finally, experiments with different training set and test set sizes suggest that a minimum of 10 000 samples are needed to reach peak accuracy and a minimum of 5000 – 10 000 samples are needed to obtain a precise estimate of performance metrics (three decimal digits in r^2 and one in MSE). These results provide benchmark quality data and comprehensive guidelines for using 3D ConvNets to model porous materials. ZeoNet and the associated dataset and software code provide a foundation for developing and comparing methods in future research efforts.

Data Availability Statement

The code[39] and datasets[8] for ZeoNet can be found at **https://r.bai.group/ZeoNet**

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

This journal is © The Royal Society of Chemistry 2023

J. Mater. Chem. A, 2023, **11**, 17570–17580 | 17579

Notes and references

1 B. Smit and T. L. Maesen, *Chem. Rev.*, 2008, **108**, 4125–4184.

2 Y. Li and J. Yu, *Nat. Rev. Mater.*, 2021, **6**, 1156–1174.

3 C. Baerlocher and L. McCusker, *Database of Zeolite Structures*, 2022, http://www.iza-structure.org/databases/.

4 R. Pophale, P. A. Cheeseman and M. W. Deem, *Phys. Chem. Chem. Phys.*, 2011, **13**, 12407–12412.

5 P. Pascual, P. Ungerer, B. Tavitian, P. Pernot and A. Boutin, *Phys. Chem. Chem. Phys.*, 2003, **5**, 3684–3693.

6 D. Dubbeldam, S. Calero, T. Vlugt, R. Krishna, T. L. Maesen and B. Smit, *J. Phys. Chem. B*, 2004, **108**, 12301–12313.

7 P. Bai, M. Tsapatsis and J. I. Siepmann, *J. Phys. Chem. C*, 2013, **117**, 24375–24387.

8 P. Bai, M. Y. Jeon, L. Ren, C. Knight, M. W. Deem, M. Tsapatsis and J. I. Siepmann, *Nat. Commun.*, 2015, **6**, 1–9.

9 Y. G. Chung, P. Bai, M. Haranczyk, K. T. Leperi, P. Li, H. Zhang, T. C. Wang, T. Duerinck, F. You, J. T. Hupp, *et al.*, *Chem. Mater.*, 2017, **29**, 6315–6328.

10 H. Fang, A. Kulkarni, P. Kamakoti, R. Awati, P. I. Ravikovitch and D. S. Sholl, *Chem. Mater.*, 2016, **28**, 3887–3896.

11 J. Kim, M. Abouelnasr, L.-C. Lin and B. Smit, *J. Am. Chem. Soc.*, 2013, **135**, 7545–7552.

12 L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, *et al.*, *Nat. Mater.*, 2012, **11**, 633–641.

13 P. Bai, M. Tsapatsis and J. I. Siepmann, *Langmuir*, 2012, **28**, 15566–15576.

14 J. Vamathevan, D. Clark, P. Czodrowski, I. Dunham, E. Ferran, G. Lee, B. Li, A. Madabhushi, P. Shah, M. Spitzer, *et al.*, *Nat. Rev. Drug Discovery*, 2019, **18**, 463–477.

15 R. Gaillac, S. Chibani and F.-X. Coudert, *Chem. Mater.*, 2020, **32**, 2653–2663.

16 R. Anderson, A. Biong and D. A. Gómez-Gualdrón, *J. Chem. Theory Comput.*, 2020, **16**, 1271–1283.

17 S. M. Auerbach, K. A. Carrado and P. K. Dutta, *Handbook of Zeolite Science and Technology*, CRC press, 2003.

18 M. E. Davis, *Nature*, 2002, **417**, 813–821.

19 B. J. Bucior, N. S. Bobbitt, T. Islamoglu, S. Goswami, A. Gopalan, T. Yildirim, O. K. Farha, N. Bagheri and R. Q. Snurr, *Mol. Syst. Des. Eng.*, 2019, **4**, 162–174.

20 Z. Li, B. J. Bucior, H. Chen, M. Haranczyk, J. I. Siepmann and R. Q. Snurr, *J. Chem. Phys.*, 2021, **155**, 014701.

21 E. H. Cho and L.-C. Lin, *J. Phys. Chem. Lett.*, 2021, **12**, 2279–2285.

22 T.-H. Hung, Z.-X. Xu, D.-Y. Kang and L.-C. Lin, *J. Phys. Chem. C*, 2022, **126**, 2813–2822.

23 S. Lee, B. Kim and J. Kim, *J. Mater. Chem. A*, 2019, **7**, 2709–2716.

24 B. Kim, S. Lee and J. Kim, *Sci. Adv.*, 2020, **6**, eaax9324.

25 A. Krizhevsky, I. Sutskever and G. E. Hinton, in *Advances in Neural Information Processing Systems* 25, 2012.

26 K. Simonyan and A. Zisserman, *arXiv*, preprint, arXiv:1409.1556, 2014.

27 K. He, X. Zhang, S. Ren and J. Sun, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

28 G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.

29 P. Bai, D. H. Olson, M. Tsapatsis and J. I. Siepmann, *ChemPhysChem*, 2014, **15**, 2225–2229.

30 T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Microporous Mesoporous Mater.*, 2012, **149**, 134–141.

31 H. Kataoka, T. Wakamiya, K. Hara and Y. Satoh, *arXiv*, preprint, arXiv:2004.04968, 2020.

32 K. He, X. Zhang, S. Ren and J. Sun, *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV*, 2016, vol. 14, pp. 630–645.

33 D. P. Kingma and J. Ba, *arXiv*, preprint, arXiv:1412.6980, 2014.

34 J. Duchi, E. Hazan and Y. Singer, *J. Mach. Learn. Res.*, 2011, **12**, 2121–2159.

35 T. Tieleman, G. Hinton *et al.*, *Coursera: Neural Networks for Machine Learning*, 2012, vol. 4, pp. 26–31.

36 H. Robbins and S. Monro, *Ann. Math. Stat.*, 1951, 400–407.

37 D. Frenkel and B. Smit, *Understanding Molecular Simulation: from Algorithms to Applications*, Elsevier, 2001, vol. 1.

38 K. Simonyan, A. Vedaldi and A. Zisserman, *arXiv*, preprint, arXiv:1312.6034, 2013.

39 Y. Liu, G. Perez, Z. Cheng, A. Sun, S. Hoover, W. Fan, S. Maji and P. Bai, *ZeoNet*, 2023, https://r.bai.group/ZeoNet.