

# Graph Convolution Neural Network for predicting Photo-excited State Properties of Indole C<sub>8</sub>H<sub>7</sub>N Absorption Spectra

Khai Nguyen<sup>†</sup>, Salsabil Abou-Hatab<sup>\*</sup>, Pratip Chakraborty<sup>‡</sup>, and Spiridoula Matsika<sup>\*</sup>

<sup>\*</sup>Department of Chemistry, <sup>†</sup>Department of Computer & Information Sciences, Temple University

<sup>‡</sup>Department of Chemistry and Chemical Biology, Rutgers University-New Brunswick

{khainguyen, salsabil.abou-hatab, smatsika}@temple.edu, pratip.chakraborty@rutgers.edu

**Abstract**—Absorption spectroscopy is a very important tool in chemistry which predicts which part of the electromagnetic radiation is absorbed by a molecule, giving a unique signature for each molecule. All photo-responsive biological phenomena (i.e., photosynthesis, photocatalysis, UV absorption of DNA leading to skin cancer, and fluorescence in biological systems) are initiated by absorption of light, so many recent efforts have been shifted towards accurately modeling the spectra of molecules, where the collective vibrational effects have been identified to significantly affect the spectra shape in electronic spectroscopy. With the growing promise of deep learning in predicting properties in photo-excited molecules, this research explores the feasibility of applying the SchNet deep learning model to predict excited states properties of a molecule with added complexity outside the popular QM9 and MD17 datasets. We have found the graph convolutional structure of SchNet to account for the molecular dynamics requirements and reproduce the spectra with significantly less time and computational resources.

## I. INTRODUCTION

Deep learning has been successfully used to address clustering, classification, and detection tasks in the domain of computer vision, natural language processing, and even computer hardware. Some recent research has geared the focus towards applications of deep learning algorithms in the chemistry and other natural sciences space, especially for molecules where graph neural networks in [14], [28] have been tested against the *MoleculeNet* benchmark dataset [27]. In our work, we chose indole C<sub>8</sub>H<sub>7</sub>N due to its use in the n-Cyanoindole fluorescent probes in spectroscopy, and picked SchNet as our model due to its good benchmarking results on QM9 and MD17, as well as its established documentation. QM9 is a dataset of 131k small organic molecules with up to nine heavy atoms and MD17 is a collection of eight molecular dynamics simulations for small organic molecules. SchNet uses graph convolutional neural networks (GCNN) that maps a molecule into an atom embedding space dependent on the atom types, then models atomic interactions with continuous-filter convolutional (*cfconv*) layers produced from filter-generating sub-networks, and training subsequently is performed by reducing the neural network loss function. In our work, the original loss function of SchNet is extended with appropriate trade-offs to cover the loss terms for the excited state energies and dipole moments. The absorption spectra generated by SchNet are compared with ones produced by Quantum Mechanics/classical Molecular Dynamics (QM/MD) simulations to confirm our model’s predictive capability.

To produce the absorption spectra, the energies and oscillator strengths are needed. Under resonance condition, i.e., the energy of the photons of the incident light equals to the energy gap between two electronic states, single- or multiphoton excitations can take place to one or more excited states if an oscillating dipole is induced. The oscillator strength,  $f_{ij}$  between electronic states  $i$  and  $j$ , can be calculated from the transition dipole moment,  $\mu_{ij}$ , of the respective electronic states as well as the energy difference,  $\Delta E_{ij}$ :

$$f_{ij} = \frac{2}{3} \Delta E_{ij} |\mu_{ij}|^2 \quad (1)$$

## II. DEEP LEARNING BACKGROUND

An overview of recent research in deep learning for quantum-mechanical molecular modeling and simulations was drawn out by [24] and [29]. For deep learning architectures used in molecular dynamics, some of the key requirements includes:

- Rotational invariance: the changes in positions of the system’s particles in the absence of external field(s) should not impact the model’s predictive ability [29].
- Symmetry: all relevant rigid space group symmetries and dynamic non-rigid symmetries should be incorporated [3] [4] [6].

Aside from the requirements above, some models also require the differential relationships of response properties [7]. For models that work with potential energy surfaces (PES), both the PES and its partial derivatives have to be smooth to perform geometry optimization [21].

Two main deep learning architectures that are currently adapted for molecular dynamics include: descriptor-based models and graph neural network (GNN)-based models [29]. For descriptor based models, molecular systems are represented by descriptor vectors, and choosing descriptors is seen as feature engineering. Once descriptors are defined they can act as direct inputs to a feed-forward artificial neural network, also known as a basic type neural network, for prediction of the target molecular property. However, applications of descriptor-based models are limited to relatively simple systems and have limited transferability between different molecular systems, which discourages their use for force field development. In addition, choosing good descriptors relevant to a particular property can be defined as a machine learning problem that requires further studies.

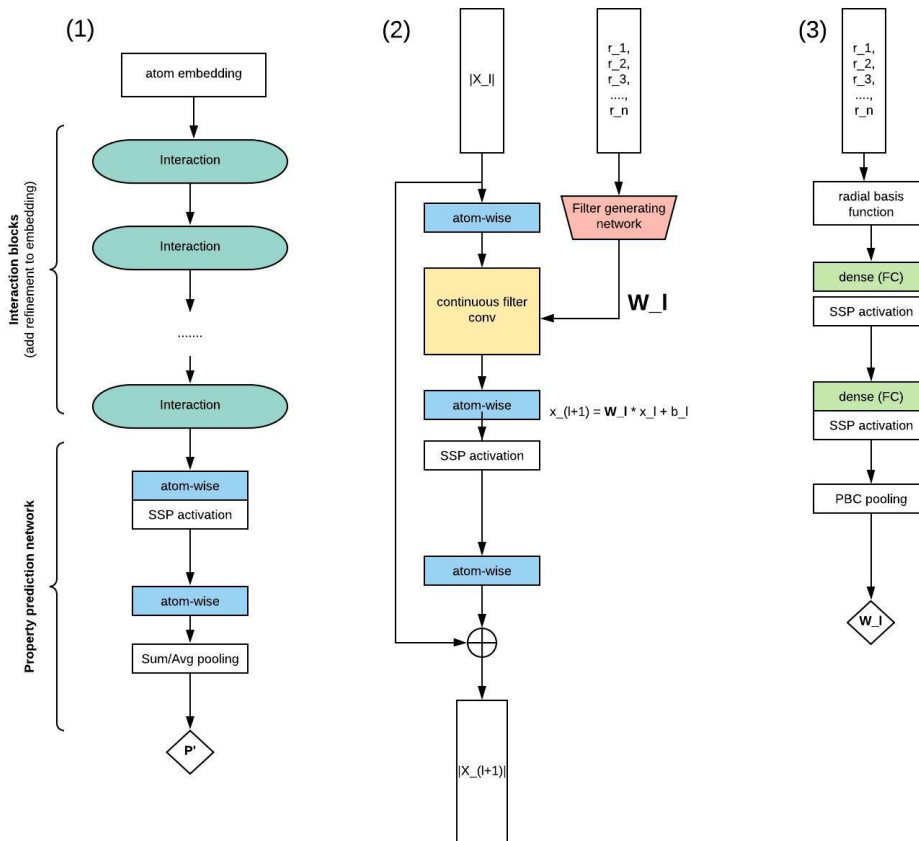


Fig. 1. Basic building blocks of a model predicting the property  $P'$  from the positions and atomic numbers of an atomistic system. The figure shows the (1) Full network, (2) Interaction block, (3) Filter generating network. Adapted from ref. [20].

On the other hand, GNN-based models treat the molecular system as a graph, where each particle is represented by a vector corresponding to a graph node and two-body interactions are modeled by the weighted edges between the nodes. Such graphical representation of molecules naturally captures rotational invariance for molecular systems and can be further engineered to account for other problem-specific requirements, namely graph convolutional neural networks [8] and message-passing neural networks (MPNNs) [11]. In addition, approaches that learn a representation directly from atom types and positions have also been developed such as Deep Tensor Neural Networks (DTNNs) [18] [17], and Gradient-domain machine learning (GDML) [6].

#### A. Graph Convolutional Neural Network

Given the requirements mentioned above, Graph Convolutional Neural Network is seen as an active area of research for deep learning in quantum chemistry. Graph convolutional networks for circular fingerprint [9] and molecular graph convolutions [15] are the works that allow the graphical representation of molecules of arbitrary size. For our specific work with indole, SchNet is chosen as our deep learning architecture due to its ability to incorporate fundamental symmetries of atomistic systems, rotational invariance and invariance to atom indexing. Compared to the well-known DTNN model, SchNet

adds on the continuous-filter convolution interpretation that was not available in the interaction layers of DTNN.

#### B. SchNet

1) *Continuous-filter convolutional layer*: SchNet can be defined as a spatial approach in GCNN which propagates information along the node neighborhoods in the graph [25]. The key feature of SchNet that allows it to outperform some of the previously developed models is its *cfconv* layers. Similar to Dynamic Filter Networks (DFN) [5], the function to generate filters of SchNet is modeled with a neural network. While DFN generates weights restricted to a grid structure, SchNet generalizes this to arbitrary positions and number of objects.

2) *Architecture*: The full SchNet architecture marked (1) in Fig. 1 first consists of initial embedding of atoms onto  $R^F$  based on their positions and atom type, where  $F$  is the dimension of the embedding space. The embeddings for  $n$  atoms at a layer  $l$  in our molecular system  $(x_1^l, \dots, x_n^l)$  with  $x_i^l \in R^F$  are first initialized randomly and then gradually optimized during training. Next comes the interaction blocks, each of which presented in detail in (2) of Fig. 1. Interaction blocks update the atomic embeddings that got passed through based on the molecular geometry  $R = (r_1, \dots, r_n)$ , or namely

re-position our atoms in the embedding space  $R^F$ . These blocks, or layers, are built from the concept of ResNet [13], which was tested with 101 convolutional layers on the CIFAR-10 image dataset of daily life objects. Building the interaction block in such way inherits the capability of ResNet for addressing vanishing/exploding gradients and improving network performance with considerably increased network depth. Inside each interaction block, there exists a *cfconv* generated by a filter generating network as depicted in (3) of Fig. 1, which is simply built-up from a stack of fully connected, or dense layers. Each atom  $i$  will have a filter  $W$  computed from a vector of distances from atom  $i$  to all other atoms in our molecule. Periodic boundary conditions are also incorporated into the filters via the cutoff value  $r_{cut}$ . Since SchNet embeds our molecular system onto  $R^F$ , the generated *cfconv* filter will then be applied on our molecular embedding in that space.

$$x_i^{l+1} = \sum_{j=0}^{n_{atoms}} x_j^l \circ W^l(\mathbf{r}_j - \mathbf{r}_i), \quad (2)$$

where  $W^l : R^3 \rightarrow R^F$  is the mapping function represented as the filter generating network that takes  $(\mathbf{r}_j - \mathbf{r}_i)$  as input. This use of pairwise distances serves the required rotational invariance for our model. “ $\circ$ ” is the element-wise multiplication. Finally, after going through the interaction blocks, the embeddings reach the atom-wise layers, which are fundamentally fully-connected layers that process the interaction between atoms and further fine-tune the embedding. Property prediction  $P^l$  comes at the end by summing or averaging the contributions from all the atoms.

For benchmarking the performance of SchNet, the evaluation was done on predicting energies for molecules in QM9 dataset as well as energies and forces for all molecular dynamics trajectories in MD17. A new dataset consisting of 645,000 conformations of various  $C_7O_2H_{10}$  isomers, ISO17, was also introduced in the evaluation<sup>1</sup>.

### III. MODELS AND METHODS

#### A. Datasets

Our dataset was generated from [10] and we extracted 2000 geometries of  $C_8H_7N$  from equilibrated molecular dynamics simulations of indole in water. The excitation energies, transition dipole moments, and oscillator strengths for each geometry were calculated using the CAM-B3LYP/6-31G(d) approach and the Gaussian software. The data was split into train / validation / test in 1500 / 300 / 200. We first planned to discard the few first tens geometries to ensure geometry stabilization in the simulation but eventually included the non-equilibrium conformations to our dataset to add random noises, which is suggested as a form of data augmentation [12], and to avoid overfitting to further improve the predictive ability of the network when perform on unseen data [22].

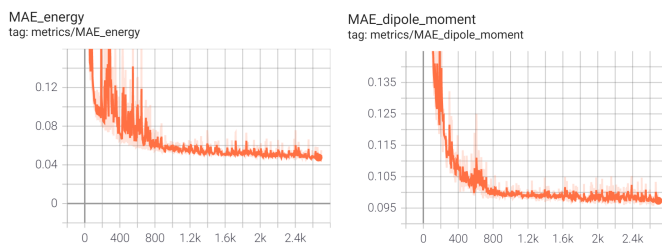


Fig. 2. Energy and dipole moment loss w.r.t epochs. An epoch is a pass of the whole training set through the model for gradient update.

#### B. Experiment

In this project, the first step was slicing the 2000ps trajectory into single geometries to produce the excited state properties with Gaussssian. The computation of the CAM-B3LYP/6-31G(d) reference calculations were then performed on Xeon(R) CPU E5-2670. Properties that are important to the absorption spectra are the energies, oscillator strengths and transition dipole moments, in which we hypothesizes using only the first and second excited states is sufficient to obtain a statistically converged spectrum. Next, to perform training using the SchNet deep learning model, ASE database format must be used [19]. Each database will include the  $C_8H_7N$  Cartesian geometries and the energy, oscillator strength, and dipole moment for each geometry. Since we chose to have a separate SchNet model for each excited state, two databases corresponding to  $S_1$  and  $S_2$  were prepared.

TABLE I  
Different test configurations of SchNet for prediction of energies and transition dipole moments of  $C_8H_7N$  compared using %MAE.

n_features	n_gauss	n_interactions	%MAE Energy	%MAE Dipole
30	25	7	11.0	24.0
64	25	10	2.15	23.1
<b>128</b>	<b>25</b>	<b>10</b>	<b>1.02</b>	<b>11.3</b>
128	20	10	1.68	24.9
128	20	7	2.02	30.0
128	25	15	2.02	32.0
128	25	7	6.66	25.3

The SchNet models for both excited state  $S_1$  and  $S_2$  are trained on the same configuration of deep learning parameters. We tested against different neural network configurations and saw the network architecture with 128-dimension embedding feature space, 25 gaussians for the filter resolution, and 10 interaction blocks works best for the molecule and dataset we have (Table I).

The loss function was customized for our specific problem as below

$$L_{SchNet} = \rho_{energy} \cdot L_{energy} + \rho_{dipole} \cdot L_{dipole} \quad (3)$$

where  $L_{energy}$  and  $L_{dipole}$  are the mean absolute error (MAE) of the actual and predicted value for each corresponding property. The trade-off terms  $\rho_{energy}$  and  $\rho_{dipole}$  should be chosen to make the loss of each property equally weighted in the total loss function. Mean Absolute Percentage Error

<sup>1</sup>QM9, MD17 and ISO17 are publicly available at <http://www.quantum-machine.org/datasets/>

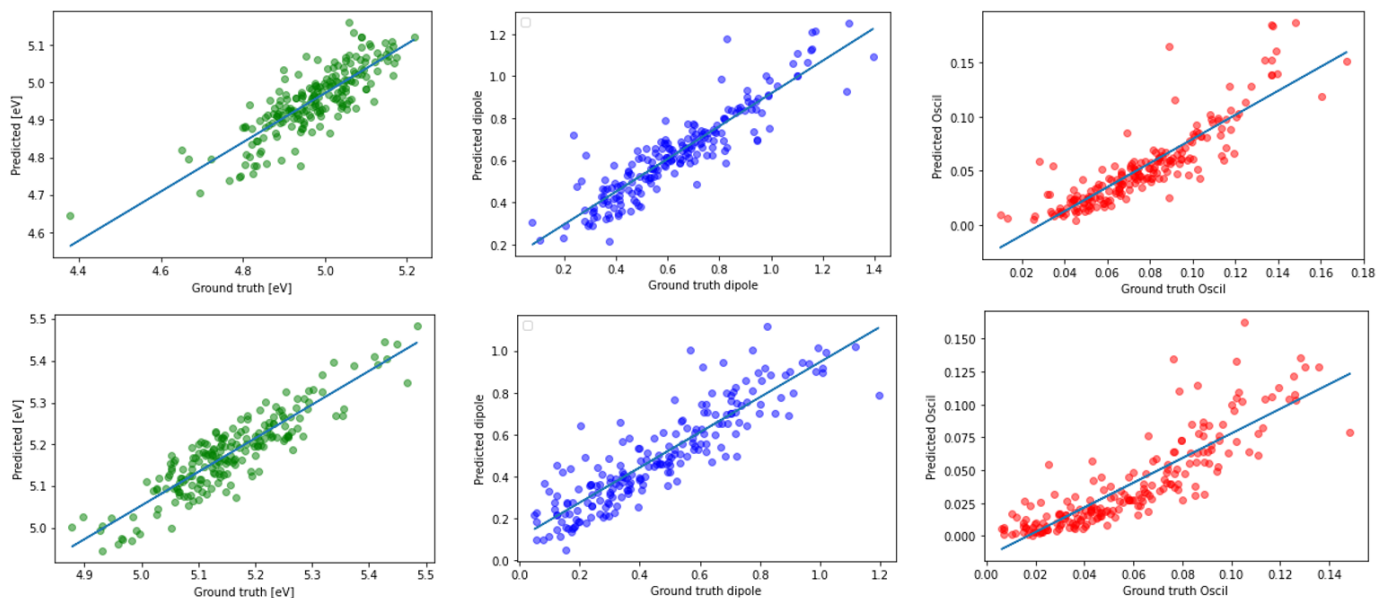


Fig. 3. Correlation graphs of energies ( $eV$ ), dipole moments ( $a.u.$ ), and oscillator strengths. The graphs on the top row corresponds to the  $S_1$  excited state and the bottom corresponds to  $S_2$ .

(%MAE) of each property was calculated to assess the model’s performance. Since oscillator strength is an intermediate variable based on equation 1, we decided to train the model to output the energy and transition dipole moment.

To determine the reasonable trade-off ratios for the two properties in our loss function, we first started with equal trade-offs and observed the changes in the loss for each property in the first tens of epochs as depicted in Fig. 2. The trade-offs of **0.1** and **0.9** were finally chosen for energy and dipole moment respectively. To enable the model to predict two properties, we have to define the two separate output modules, *Atomwise* for energy and *DipoleMoment* for dipole moment. Experiments were performed with ReLU versus Shifted-Soft Plus (SSP)  $ssp(x) = \ln(0.5e^x + 0.5)$  activation functions. Both yield relatively similar accuracy but results showed faster training time with ReLU. However, the SSP function allows creation of a smooth potential energy surface and has infinite order of continuity.

The popular Adam optimizer [23], which is a combined version of Momentum and RMSprop where past gradients are used to calculate current gradients [16], was chosen for updating the network weights. All deep learning training, validation and testing were done on a Tesla T4 GPU supported by Google Research.

## IV. RESULTS AND DISCUSSION

### A. Results

We normalized and plotted the ab initio reference and predicted absorption spectra to test the performance of our model. The spectrum computed using quantum chemical calculation shows a peak at around  $5.0 eV$  (Fig. 4) while the predicted one shows two peaks with the second lying around  $5.2 eV$ .

A comparison of actual and predicted values for the energy and dipole moment plotted on correlation graphs in Fig. 3

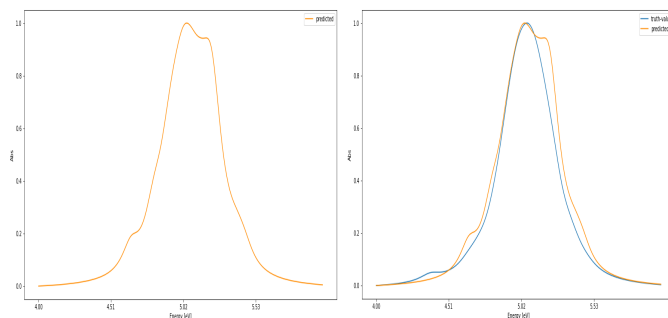


Fig. 4. Predicted absorption spectrum and comparison with reference spectrum. Spectra were produced from energies and oscillator strengths convoluted with Lorentzian functions.

TABLE II  
The %MAE of energy and dipole moment for our best model

States	%MAE Energy	%MAE Dipole
$S_1$	1.343	11.618
$S_2$	0.844	20.944

shows promising results since we have the slope of the line-of-best-fit close to 1. We further computed the corresponding graphs for oscillator strengths of excited states  $S_1$  and  $S_2$  and observed the two oscillator strengths cluster more in the 0.02 to 0.10 range. Our experiment also showed a good comparison in computation time for the two methods. Training for 3000 epochs for each deep learning model, without early stopping, takes 50 min and testing takes 2 min while the quantum calculations for 2000 geometries with Gaussian costs approximately 4320 minutes (3 days).

For our experiment setup, compared with previous works in [20] [26], our custom SchNet model requires up to 10 interaction blocks possibly due to the added complexity of our molecule, which shows the scalability of SchNet.

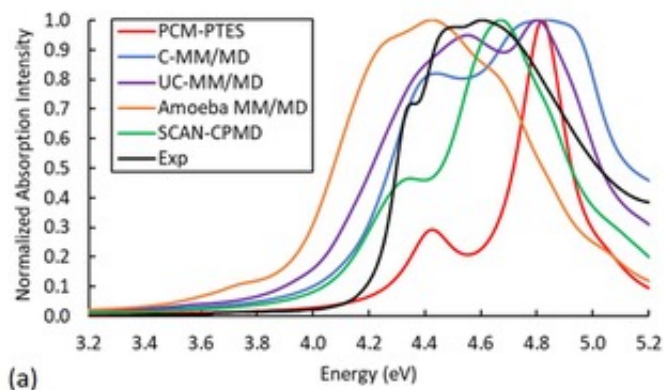


Fig. 5. Spectra generated from experiment and other MD methods. Adapted from ref. [1].

### B. Limitations

On the chemical molecular dynamics side, we further compare our produced spectra with the spectra from experiment and other molecular dynamics method. Experiment (Fig. 5) shows two peaks at around 4.4 eV and 4.6 eV. The fact our predicted result diverges from this can be credited to the training set that peaks at 5.0 eV generated from the CAM-B3LYP/6-31G(d) method, which is not as good as the methods shown in Figure 5. The structure of the spectrum is also not reproduced because of the CAM-B3LYP limitations. Interestingly, the predicted spectrum has more structure than the original. One also wants to note that spectra generated from energy and oscillator strength or transition dipole cannot tell which mode of vibration is responsible for the additional peaks/shoulders of the spectra.

On the deep learning side, with the %MAE values obtained above, we recommend that more data is needed for the training of our model. Also, one direct consequence of the embedding technique used in SchNet is it disregards any information about the environment around the atoms. If water molecules were to be added in the prediction, a separate sub-network for the waters has to be added. In addition, the  $cf_{conv}$  that exists for SchNet only uses shallow isotropic gaussian kernels to represent the weights. There exists a Deep Parametric Continuous Convolutional Neural Networks tested on Stanford large-scale 3D indoor scene dataset [2] that uses expressive deep networks to parameterize the continuous filters, which can be the scope for further studies [25].

## V. SUMMARY AND OUTLOOK

In this work we performed experimental work on the predictive capabilities of SchNet on a larger molecule outside of the QM9 and MD17 datasets. By fixing the loss function to account for the training of both excited state energies and transition dipole moment for the  $S_1$  and  $S_2$  states, and adjusting hyper-parameters of the GCNN SchNet architecture, we obtained a predicted spectrum much closely aligned with the one generated with quantum calculations. The core GCNN of SchNet is fundamental to account for both rotational invariance and symmetry required for deep learning in excited

state processes. Our work demonstrated a step closer to deep learning in computational chemistry and advancing research in photo-excited states processes in biology, chemistry and the medical industry. While results for our experiment of SchNet in indole showed promises, future outlook involves the careful re-design of suitable neural network architectures to model quantum interactions especially in molecules of larger sizes and complexity, and even in liquid phase environments, as well as doing prediction via training on molecule with resembling structures to support efforts towards transferability in the excited states.

### CODE AVAILABILITY

Currently all codes are held private under the Matsika’s lab. Please contact the authors for potential access.

### ACKNOWLEDGMENT

This work was supported by the College of Science and Technology Undergraduate Research Program (URP) and Temple University’s Honors Program. The computational results presented have been achieved using computational resources from the Matsika’s Lab. The main author thank Kiet Chu and Dr. Kai Zhang for helpful discussions regarding the output modules and penalty values for the loss function.

### REFERENCES

- [1] Salsabil Abou-Hatab, Vincenzo Carnevale, and Spiridoula Matsika. Modeling solvation effects on absorption and fluorescence spectra of indole in aqueous solution. *The Journal of Chemical Physics*, 154(6):064104, 2021.
- [2] Iro Armeni, Ozan Sener, Amir R. Zamir, Helen Jiang, Ioannis Brilakis, Martin Fischer, and Silvio Savarese. 3d semantic parsing of large-scale indoor spaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [3] Jörg Behler. Atom-centered symmetry functions for constructing high-dimensional neural network potentials. *The Journal of Chemical Physics*, 134(7):074106, 2011.
- [4] Jörg Behler, Sönke Lorenz, and Karsten Reuter. Representing molecule-surface interactions with symmetry-adapted neural networks. *The Journal of Chemical Physics*, 127(1):014705, 2007.
- [5] Bert De Brabandere, Xu Jia, Tinne Tuytelaars, and Luc Van Gool. Dynamic filter networks, 2016.
- [6] Stefan Chmiela, Huziel E. Sauceda, Igor Poltavsky, Klaus-Robert Müller, and Alexandre Tkatchenko. sgdml: Constructing accurate and data efficient molecular force fields using machine learning. *Computer Physics Communications*, 240:38–45, 2019.
- [7] Anders S. Christensen, Felix A. Faber, and O. Anatole von Lilienfeld. Operators in quantum machine learning: Response properties in chemical space. *The Journal of Chemical Physics*, 150(6):064105, 2019.
- [8] David K. Duvenaud, Dougal Maclaurin, Jorge Aguilera-Iparraguirre, Rafael Gómez-Bombarelli, Timothy Hirzel, Alán Aspuru-Guzik, and Ryan P. Adams. Convolutional networks on graphs for learning molecular fingerprints. In *NIPS*, pages 2224–2232, 2015.
- [9] David K Duvenaud, Dougal Maclaurin, Jorge Iparraguirre, Rafael Bombarelli, Timothy Hirzel, Alan Aspuru-Guzik, and Ryan P Adams. Convolutional networks on graphs for learning molecular fingerprints. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pages 2224–2232. Curran Associates, Inc., 2015.



- [10] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian<sup>16</sup> Revision C.01, 2016. Gaussian Inc. Wallingford CT.
- [11] Justin Gilmer, Samuel S. Schoenholz, Patrick F. Riley, Oriol Vinyals, and George E. Dahl. Neural message passing for quantum chemistry. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 1263–1272. PMLR, 06–11 Aug 2017.
- [12] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- [13] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [14] Katsuhiko Ishiguro, Shin ichi Maeda, and Masanori Koyama. Graph warp module: an auxiliary module for boosting the power of graph neural networks in molecular graph analysis. 2019.
- [15] Steven Kearnes, Kevin McCloskey, Marc Berndl, Vijay Pande, and Patrick Riley. Molecular graph convolutions: moving beyond fingerprints. *Journal of Computer-Aided Molecular Design*, 30(8):595–608, Aug 2016.
- [16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [17] Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett, editors. *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5–10, 2016, Barcelona, Spain*, 2016.
- [18] Kristof T. Schütt, Farhad Arbabzadah, Stefan Chmiela, Klaus R. Müller, and Alexandre Tkatchenko. Quantum-chemical insights from deep tensor neural networks. *Nature communications*, 8:13890–13890, Jan 2017. 28067221[pmid].
- [19] K. T. Schütt, P. Kessel, M. Gastegger, K. A. Nicoli, A. Tkatchenko, and K.-R. Müller. Schnetpack: A deep learning toolbox for atomistic systems. *Journal of Chemical Theory and Computation*, 15(1):448–455, 2019.
- [20] K. T. Schütt, H. E. Sauceda, P.-J. Kindermans, A. Tkatchenko, and K.-R. Müller. Schnet – a deep learning architecture for molecules and materials. *The Journal of Chemical Physics*, 148(24):241722, 2018.
- [21] Kristof T. Schütt, Pieter-Jan Kindermans, Huziel E. Sauceda, Stefan Chmiela, Alexandre Tkatchenko, and Klaus-Robert Müller. Schnet: A continuous-filter convolutional neural network for modeling quantum interactions, 2017.
- [22] Jocelyn Sietsma and Robert J.F. Dow. Creating artificial neural networks that generalize. *Neural Networks*, 4(1):67–79, 1991.
- [23] Derya Soydaner. A comparison of optimization algorithms for deep learning. *CoRR*, abs/2007.14166, 2020.
- [24] O. Anatole von Lilienfeld, Klaus-Robert Müller, and Alexandre Tkatchenko. Exploring chemical compound space with quantum-based machine learning. *NATURE REVIEWS CHEMISTRY*, 4(7):347–358, JUL 2020.
- [25] Shenlong Wang, Simon Suo, Wei-Chiu Ma, Andrei Pokrovsky, and Raquel Urtasun. Deep parametric continuous convolutional neural networks. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun 2018.
- [26] J. Westermayr and P. Marquetand. Deep learning for uv absorption spectra with schnarc: First steps toward transferability in chemical compound space. *The Journal of Chemical Physics*, 153(15):154112, 2020.
- [27] Zhenqin Wu, Bharath Ramsundar, Evan N. Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S. Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chem. Sci.*, 9:513–530, 2018.
- [28] Kevin Yang, Kyle Swanson, Wengong Jin, Connor Coley, Philipp Eiden, Hua Gao, Angel Guzman-Perez, Timothy Hopper, Brian Kelley, Miriam Mathea, Andrew Palmer, Volker Settels, Tommi Jaakkola, Klavs Jensen, and Regina Barzilay. Analyzing learned molecular representations for property prediction. *Journal of Chemical Information and Modeling*, 59(8):3370–3388, 2019. PMID: 31361484.
- [29] Jun Zhang, Yao-Kun Lei, Zhen Zhang, Junhan Chang, Maodong Li, Xu Han, Lijiang Yang, Yi Isaac Yang, and Yi Qin Gao. A perspective on deep learning for molecular modeling and simulations. *The Journal of Physical Chemistry A*, 124(34):6745–6763, 2020. PMID: 32786668.