FINETUNA: fine-tuning accelerated molecular simulations

Joseph Musielewicz, Xiaoxiao Wang, Tian Tian and Zachary Ulissi
Department of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States of America
1 These authors contributed equally to this work.
* Author to whom any correspondence should be addressed.
E-mail: zulissi@andrew.cmu.edu

Keywords: active learning, DFT, fine tuning, graph potential energy surface

Supplementary material for this article is available online

Abstract

Progress towards the energy breakthroughs needed to combat climate change can be significantly accelerated through the efficient simulation of atomistic systems. However, simulation techniques based on first principles, such as density functional theory (DFT), are limited in their practical use due to their high computational expense. Machine learning approaches have the potential to approximate DFT in a computationally efficient manner, which could dramatically increase the impact of computational simulations on real-world problems. However, they are limited by their accuracy and the cost of generating labeled data. Here, we present an online active learning framework for accelerating the simulation of atomic systems efficiently and accurately by incorporating prior physical information learned by large-scale pre-trained graph neural network models from the Open Catalyst Project. Accelerating these simulations enables useful data to be generated more cheaply, allowing better models to be trained and more atomistic systems to be screened. We also present a method of comparing local optimization techniques on the basis of both their speed and accuracy. Experiments on 30 benchmark adsorbate-catalyst systems show that our method of transfer learning to incorporate prior information from pre-trained models accelerates simulations by reducing the number of DFT calculations by 91%, while meeting an accuracy threshold of 0.02 eV 93% of the time. Finally, we demonstrate a technique for leveraging the interactive functionality built in to Vienna ab initio Simulation Package (VASP) to efficiently compute single point calculations within our online active learning framework without the significant startup costs. This allows VASP to work in tandem with our framework while requiring 75% fewer self-consistent cycles than conventional single point calculations. The online active learning implementation, and examples using the VASP interactive code, are available in the open source FINETUNA package on Github.

1. Introduction

Global population growth and climate change have greatly raised clean energy demand. Heterogeneous catalysis plays a crucial role in the development of renewable energy sources and sustainable chemical production processes [1, 2]. Examples include hydrogen generation from a renewable source, carbon dioxide conversion into liquid fuel, and ammonia synthesis for fertilization [3–8]. Catalyst discovery for each application is time consuming due to the enormous design space. The numerous combinations of compositions are impossible to be searched fully even with high throughput experimentation [9]. Alternatively, based on Brønsted–Evans–Polanyi relationships, computational screening uses adsorption energy as a descriptor of catalyst activity and allows a larger number of systems to be screened [10]. Generating accurate adsorption energies for a variety of catalyst systems is a rate limiting step. Adsorption energies are usually obtained using computational modeling methods like density functional theory (DFT) via geometric optimization [11]. Starting with an initial atomistic structure, in each step, the energy and forces are evaluated by DFT code. The structure is updated by an optimizer iteratively, until a local minimum
energy is found. Some common choices of optimizers include the conjugate gradient (CG) and Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithms. In this process, a series of DFT single point calculations are performed, which requires extensive computing resources.

With broad excitement for machine learning (ML) in the molecular simulation community, there is a push to apply ML to accelerate local relaxations and predict adsorption energies [12, 13]. Garijo del Rio et al developed Gaussian process minimizer (GPMin), which builds a Gaussian process (GP) model of the potential energy surface for fast convergence [14]. Given the nature of computational complexity scaling of GPs, GPMin is limited by the number of atoms in the atomistic systems. The Behler–Parrinello neural network paved the way for neural network based machine learning potential (MLP) applications in atomistic simulations [15]. Over the past decade, many MLPs have been developed to substitute expensive DFT calculations. For example, ANI-1 potential and its extensions are accurate and transferable models for organic molecules [16–19]. Recently, a number of graph neural network (GNN) architectures, e.g. DimeNet++ [20], DimeNet++ [21], SpinConv [22], and GemNet-T [23], have been developed and have shown increasing accuracy in energy and forces predictions. GNNs typically require a training data set consisting of a variety of catalyst systems and their adsorption energies, which is expensive to generate with DFT. The current state-of-the-art catalyst data set, the Open Catalyst Project 2020 (OC20) data set, is composed of over 1.2 million DFT relaxations. Using OC20 as a benchmark, the current most accurate general replacement for DFT is GemNet-T, a GNN [24].

In the case where a comprehensive training data set is not available, active learning strategies have been utilized to train MLPs with limited labeled data [25–32]. With an active learning framework, at each step the configuration is evaluated by a MLP surrogate model as long as the uncertainty of the prediction is within a threshold. Otherwise, a DFT single point calculation is triggered, and the model is retrained with the new configuration. Vandermause et al built a GP based model that can be trained on-the-fly for molecular dynamic simulations [29, 30]. Yang et al and Shuaibi et al demonstrated online and offline active learning applications to accelerate simulation tasks [31, 32]. However, these methods usually start from scratch, meaning the surrogate model has no prior information of the systems.

In this work, we propose an active learning scheme that leverages the prior information to accelerate geometry optimizations [33]. More specifically, we use a GemNet-T model pretrained on OC20 data set, and we retrain only a small portion of the model for the new optimization tasks. The acceleration is achieved by the active learning framework, where it queries sparsely from the potential energy surface and avoids unnecessary DFT calculations. To evaluate the performance of the proposed scheme, 30 catalyst systems from OC20 data are randomly chosen as benchmark systems, and we compare local relaxations on the benchmark systems using different optimization methods. We show that the fine-tuning process improves the prediction performance of a large GNN model on each individual system. Our best performing active learning strategy reduces DFT single point calculations by 91%, with 93% of the final relaxed energies are close to or lower than the reference DFT relaxed energy. We also discuss an efficient implementation of the interactive mode of the Vienna ab initio Simulation Package (VASP) calculator in the Atomistic Simulation Environment (ASE) Python package, which speeds up the calculation process by reducing overhead costs and the number self-consistent field (SCF) cycles [34–39].

2. Methods

2.1. Online learner framework

The online active learning workflow, seen in figure 1(a), starts by using a pre-trained MLP as a surrogate model to predict the energy and forces of an initial atomistic structure. Then, the optimizer uses the forces, as the gradients of energy, to update the atomic positions to reduce the overall energy of the system. The energy and forces of the new structure are again predicted by the MLP. With each MLP prediction, a querying criteria is used to accept or reject the prediction. If the prediction is accepted, the optimization continues as normal. If it is rejected, then the parent calculator (a role performed by VASP in this work) is queried. The parent calculator is used to compute the ground-truth energy and forces for the system in its current state, and this calculation is sent to the optimizer instead of the surrogate prediction. Additionally, the parent-calculated data point, consisting of positions labeled with the energy and forces, is added to a set of parent-data. It is immediately used to fine-tune the MLP, to make future predictions on the atomistic system more accurate. Since the BFGS optimizer is a second order optimizer with a numerical estimate of the hessian, we also reset its hessian estimate to use only the set of parent-data, so that it does not retain a hessian estimate which relies on a previous version of the MLP.

The active learning approach serves as an alternative to traditional numerical optimization approaches used for adsorbate-catalyst systems, such as BFGS and CG. While it still makes use of a numerical optimizer, it accelerates optimization by substituting a number of the expensive parent calculations with ML model
inference calculations. We expect that using a pre-trained ML model will allow it to incorporate the prior information learned by the ML model to further accelerate this process. However, a more obvious and perhaps simpler method of incorporating this prior information into the optimization is to simply perform an optimization on the initial adsorbate-catalyst structure, with the pre-trained model acting as the force calculator until it predicts it is fully optimized, and then to check that final structure with DFT. If the structure is not fully optimized, then finish the optimization using DFT as the force calculator. We refer to this method as the GemNet Warm Start method of optimization, and it is one of the benchmarks by which we must measure the active learning approach against.

2.2. Fine-tuning
The advantage of this framework over a more standard optimization approach, and over other active learning frameworks, is the incorporation of prior information learned by the pre-trained MLP. As seen in figure 1(b), we use a GemNet-T model, pre-trained on OC20, as the basis for the MLP. GNNs have recently shown great success at approximating DFT calculations for adsorbate-catalyst systems, and GemNet-T is currently state-of-the-art, according to the Open Catalyst Project (OCP) leaderboard. The GemNet-T architecture takes atomistic structures as input and encodes atoms as nodes and information about the interactions between atoms as edges. Edges encode three kinds of interaction information: two-body distance information, three-body angular information, and four-body dihedral information. This graphical
representation is passed through three interaction blocks, which allows information to flow between nodes and allows GemNet-T to learn a generalizable representation of atomistic structures. Each interaction block passes its graph representation to an output block, which learns to transform the representation into an energy and force contribution, which are summed to create the overall energy and force prediction [23].

Making an energy and force prediction with GemNet-T is inexpensive relative to DFT, on the order of seconds on a CPU core compared to hours on a CPU core [24]. Still, GemNet-T lacks the accuracy required to reliably find a structure with a local minimum energy, and it is expensive to train from scratch, so we fine-tune it on-the-fly using the parent data from the relaxation. Fine-tuning starts with the pre-trained GemNet-T model, and updates it by freezing the interaction blocks while training the final three GemNet-T output blocks using the AdamW optimizer, with similar settings to the ones used by the authors of the GemNet-T paper [23]. The goal of our training approach is to preserve the underlying physical information learned by the GemNet-T model while improving the model's accuracy for the specific atomistic system it is helping to optimize. We found that tuning the output blocks while freezing the interaction blocks resulted in the most consistent improvement to GemNet-T's prediction accuracy on later structures in the optimization. This process of fine-tuning large pre-trained models to 'transfer' their knowledge between domains has a track record of success in other domains of ML [40]. Although in this work we employ this strategy for an extremely low data regime to improve specific predictions on-the-fly, instead of using a larger data set to transfer on more general tasks.

In our on-the-fly training approach, we retrain the model sequentially on each new data point queried from the parent calculator. We take an incremental approach to fine-tuning, where the previous state of the model is retrained on only the most recent point for 400 epochs. We intend for this incremental approach to tune the model to match the most recent state of the atomistic system as it changes throughout the optimization. We use an initial learning rate of 0.0003 and a loss-based learning rate scheduler to reduce the learning rate by a factor of 0.9 when the loss fails to improve for at least three training steps. In this work, the fine-tuned state of the model is only used during the optimization of a specific system, and is not reused for other systems. However the model could be preserved for use in other active learning optimizations if similar systems needed to be optimized. Details on the effects of fine-tuning the GemNet model can be found in the supplementary information.

### 2.3. Querying criteria

The querying criteria is a key decision within the active learning framework which affects both its performance and accuracy. We assess three strategies for querying within our active learning framework, as seen in figure 1(c). Relaxed-only querying occurs only when the MLP predicts the forces on the atomic structure to be below the relaxed threshold, based on the definition used by the OCP paper [24]. Regardless of the other querying criteria used, in all three approaches we find that querying each point predicted to be below the relaxed threshold is necessary for convergence of the framework. K-steps querying occurs whenever the MLP has taken a number of optimization steps, k, since the last query. Uncertainty querying occurs whenever a measure of the MLP's uncertainty climbs above some heuristic threshold. Uncertainty metrics are commonly used in other active learning approaches, they are often derived from kernel methods and are used to ensure the MLP does not make unsafe predictions [29]. However we expect the prior information encoded in the pre-trained MLP to help alleviate this concern, potentially allowing relaxed-only or K-steps querying to be potentially effective. We also cannot use a kernel method approach to measuring uncertainty because our MLP is a graph neural network. Therefore we use the disagreement between the members of an ensemble of GemNet-T models to compute the uncertainty [41].

To obtain disagreement between models we first pre-train five GemNet-T models from different randomly initialized weights. Then we create ensembles of two, three, or five models. In addition to using distinct randomly initialized weights, we unfreeze and fine-tune a different output block in each member of the ensemble. Empirically this strategy led to better outcomes in terms of both accuracy and speed when querying based on uncertainty, and we suspect this is because it further differentiates the models. Finally we compute the standard deviation in the force predictions of the ensemble as a surrogate for uncertainty. We chose these ensemble sizes based on the computational cost of pre-training and fine-tuning each model. Since the goal of this method is to reduce the overall computational cost of optimization, a higher fine-tuning cost requires a greater reduction in the average number of parent calls (or possibly a superior accuracy) to be justified. All else being equal we prefer smaller ensembles, making the ensemble of two models the most attractive. We discuss the topic of relative speed-up during runtime further in section 3.3.

### 2.4. Experiment setup

To test each online active learning method we perform geometric optimizations on a set of 30 atomistic systems, and then compare their performance in terms of cost and accuracy metrics to some baseline
optimization methods. We sample the set of 30 atomistic systems from the OC20 validation set. We use systems only from the set of out-of-domain materials for both adsorbates and metal surfaces. This means each of these systems consist of an adsorbate placed above a metal substrate, and the pretrained GemNet-T model we use was never explicitly trained on any systems with the same adsorbate molecules or metal substrate. These 30 systems can be found in the FINETUNA GitHub repository with Google Colab ASE examples [33, 42]. We used VASP as our parent DFT calculator, with the settings specified by the OCP [24].

We compare each method to CG, an optimization method with a native implementation in VASP. We choose this for the baseline since it is the default optimization method for VASP, and it was used to generate all of the relaxation data for the OCP [24] data sets. We perform geometric optimizations using the following techniques for comparison to ASE CG and to each other:

- VASP CG: VASP with built-in conjugate gradient optimization.
- ASE BFGS: VASP ASE calculator with ASE BFGS optimization.
- GemNet Warm Start: Pretrained GemNet-T calculator with ASE BFGS optimization until relaxed, followed by ASE VASP calculator with ASE BFGS optimization from that point.
- ASE GPMin: An optimization strategy built in to ASE which uses VASP as a parent calculator and trains a Gaussian process on-the-fly to smooth out the potential energy surface of VASP [14].
- Relax-only Online Learner: Online learner with VASP as the parent calculator and GemNet-T as the MLP. Querying the parent calculator and retraining the MLP only when it predicts a relaxed structure.
- K-steps Online Learner: Same as Relax-only Online Learner while also querying whenever a number, k, steps have been taken since the last query.
- Uncertainty Online Learner: Same as Relax-only Online Learner while also querying whenever the prediction uncertainty metric rises above some threshold.

We introduce a framework for comparing geometric optimization methods in terms of speed and accuracy. As a proxy for speed we compute the ratio of the number of DFT singlepoint calculations (DFT calls) made during the optimization of a given system to the number of DFT calls made when using ASE CG to optimize that system. As a proxy for accuracy we compute the fraction of optimizations which satisfy two conditions. First, the parent calculator (VASP) calculates that the final point is relaxed, meaning the maximum force on any atom is \( F_{\text{max}} \leq 0.03 \text{ eV Å}^{-1} \). And second, the parent calculator (VASP) calculates the energy of final point is less than or equal to the energy of the final point of the ASE CG optimization, plus a 0.02 eV threshold for error. We choose the \( F_{\text{max}} \) threshold of 0.03 eV Å\(^{-1}\) to match the OCP criteria for relaxation [24]. We choose the 0.02 eV threshold for energy heuristically, based on experience working with these calculations, and the distribution of final energy differences between VASP CG and ASE BFGS. We expect 0.02 eV to be a reasonable guess for the energy threshold, but it could easily be changed depending on one’s expectation for similarity between optimization strategies.

3. Results and discussion

3.1. Mapping trajectories with principal component analysis (PCA)

The relaxation process of a randomly selected test system is illustrated in figure 2(a). In the GemNet Warm Start and active learning methods, the steps before the first DFT point are identical because they rely solely on the pretrained GemNet-T predictions. The active learning strategy shown in this figure is the K-steps, and a more detailed comparison of different querying strategies can be found in section 3.2. Figure 2(b) shows the DFT energy relative to the relaxed energy from VASP CG at each DFT call. In the GemNet Warm Start and active learning methods, DFT relaxation starts from an energy 0.18 eV above the relaxed energy, suggesting that the pretrained GemNet-T model has relaxed the structure close to the final structure. In the region near the local minimum (energy less than 0.2 eV), active learning converges much faster than other methods. Therefore, the overall acceleration of the relaxation process is a combined effect of the large pretrained GNN model and the active learning framework.

To visualize and qualitatively compare relaxation trajectories from each method, we perform PCA. We adapt the B2 descriptor from Vandermause et al [30] to represent each atomic configuration in the trajectory. The descriptors are vectors which we linearly transform into principle components which describe the majority of the variation in the positions. The first two components are plotted in figure 2. It is clear that ASE BFGS, GemNet Warm Start, and the online active learner take similar relaxation routes as they possess similar curves. This is reasonable because all three of these methods use the same BFGS algorithm as their underlying optimizer. VASP CG and ASE GPMin deviate further from the BFGS-based trajectories, but they
end up at the same relaxed configuration in PCA space. This suggests that the choice of optimizer affects the path optimization takes.

3.2. Comparing optimization methods

We find the K-steps, single GemNet online learner to be a significant improvement over other methods in terms of speed, without sacrificing accuracy. In figure 3 we see that it performs the set of optimizations with an average of ten times fewer parent calls when compared to the baseline, and at the same overall accuracy threshold as its underlying optimizer, BFGS. The addition of the K-steps querying criteria results in only a slight improvement over the relax-only querying strategy alone. This is due to the majority of parent calls being caused by relaxation queries, with only one or two calls at the beginning of the optimization being caused by K-steps. Occasionally the K-steps querying results in some speed-up, but in most cases has little impact on the optimization, leading to a small difference. The use of ensembling generally hurt the accuracy of the online learner strategy, and the use of uncertainty querying fails to work as expected. There is no clear correlation between ensemble size and performance for the uncertainty learners. While the $3 \times$ ensemble outperforms all other methods in speed, larger and smaller ensembles both underperform when compared to the single learners. The behavior of the uncertainty learner is too inconsistent to reliably make up for the increased time to fine-tune the larger ensembles. But, it does seem possible that with more thorough hyperparameter optimization the uncertainty querying strategy could outperform the simpler k-steps strategy. Regardless, based on these results we favor the use of single model strategies, along with the K-steps querying strategy, going forward.

Looking more closely at the distribution of accuracy and speed, we compare the distribution of outcomes of the K-steps online active learning strategy to the other methods in figure 4. Here we see the online active learning consistently improves speed when compared to all other methods, with rare exceptions where it is narrowly beaten by ASE GPMin or GemNet Warm Start techniques on a particular system. It always beats traditional numerical methods like BFGS and CG, which is expected. In terms of accuracy, the online active learner often finds the same energy outcome as VASP CG and the other methods. However, it is more likely to find a lower minimum energy than any of the others. We speculate that this could be due to the noise introduced by the regular fine-tuning of the GemNet-T model. Since the optimization algorithm is searching for a local minimum over a changing function, it may be occasionally forced out of shallow local minima, and therefore more likely to find deeper local minima. Regardless, the online active learner was just as likely as ASE BFGS and ASE GPMin to reach a local minimum below the threshold set by VASP CG which
Figure 3. Accuracy versus speed outcomes on the test set of 30 systems for the different optimization strategies. The accuracy metric is the fraction of test outcomes with an energy less than or equal to the base-case optimizer (ASE CG) relaxed energy for that system plus a margin of error of 0.02 eV. The speed metric is the average ratio of parent (DFT) calls made by the optimizer to parent calls made by the base-case optimizer.

makes it equivalent to ASE BFGS and ASE GPMin by this accuracy metric, but with consistently superior speed-up.

### 3.3. Optimizing run-time performance of online learners

The speed-up of online learners by means of parent DFT calls is not essentially the run-time speed-up in real-world applications, due to two factors: (a) steps to reach electronic convergence vary in each single point calculation, and (b) the GNN training time is non-negligible. In this section we discuss our software engineering efforts to address the above issues for performance optimization and take advantage of built-in interactive features of VASP.

Like other DFT packages, VASP uses an iterative scheme to solve the Kohn–Sham equation, making the time consumed by DFT calculations proportional to the number of SCF loops. In this respect, using the standard ASE VASP calculator has two drawbacks: more SCF loops to achieve convergence due to random wavefunction initialization, and system overhead caused by start/stop of the VASP program and file I/O. Here we design a new calculator interface for VASP by leveraging its interactive mode (VASPInteractive) to avoid shutting down the VASP program, and reduce the of SCF loops by an average of ~75% when compared with standard ASE VASP calculator. Details of the implementation can be found in the supplementary information.

We compare three different implementations of calculator interfaces to VASP, as shown in figure 5(a), to study the impact of SCF loops on run-time performance:

(a) Standard ASE VASP (M1): cold-start VASP process on each single point.
(b) ASE VASP with wavefunction cache (M2): use local file (WA VECAR) to store and pass wavefunction between single points.
(c) VASPInteractive (M3): stream-based calculator maintaining a long-running VASP process.
Figure 4. Comparison of all 30 systems with different optimization strategies. (a) The top plot shows the number of DFT single points calculations required for each strategy as a percentage of VASP CG. The bottom plot shows the relaxed energy difference with respect to VASP CG. The relaxed energy from different strategies is similar if it is within 0.02 eV of VASP CG reference energy. (b) Visualizations of the reference relaxed structures from VASP CG.

Figure 5(b) shows the total SCF loops compared with VASP CG, when choosing different optimization strategies (ASE BFGS, ASE GPMin or online learner) and VASP interfaces (M1–M3). We find that independent of the optimization strategy, ASE VASP with wavefunction cache (M2) and VASPInteractive (M3) reduce the total SCF loops compared with ASE VASP by roughly 50% and 75% on average, respectively. In both M2 and M3 the wavefunction information is passed between subsequent DFT single points, however M2 requires more steps for converging the orbitals when read from a locally-cached WAVECAR file. As a result, the choice of VASP interface has significant impact on performance: using ASE VASP (M1 or M2) as a calculator, ASE BFGS and ASE GPMin optimizers may have computational cost similar or even greater than VASP CG, despite reduction of parent DFT calls. On the other hand, the total SCF reduction when combining online learner and VASPInteractive is consistent with its reduction of parent DFT calls as shown in figure 3. We note that all of these metrics are dependent on precise computer hardware/architecture, and should only be viewed as rough estimates.

Further run-time performance optimization can be achieved by moving the ML training workload from CPU to GPU device. For the best fine-tuning strategy in section 3.2 (K-steps, single GemNet, 400 epochs), single training steps requires $O(10^2) \text{s}$ on a 32-core CPU, similar to DFT single point on the same CPU architecture ($O(10^3) \text{s} \sim O(10^5) \text{s}$). On the other hand, the training cost drastically drops to $O(10^0) \text{s} \sim O(10^1) \text{s}$ on a GPU device. Figure 5(c) shows rough walltime speed-ups of the whole relaxation process as compared with VASP CG, when choosing different optimization strategies and VASP interfaces, but precise speedups are dependent on precise computer hardware. When taking full advantage of VASP’s interactive capabilities in the online learner, simply by switching training from CPU to GPU, we can improve the mean walltime speed-up from $\sim5 \times$ to $\sim10 \times$, approaching the ideal speed-up by means of parent DFT calls as shown in figure 3. We note that all of this is possible only due to the much-appreciated functionality of VASP to run interactively.
Figure 5. Software engineering approaches for optimizing online learner run-time performance. (a) Schematic comparison between three implementations (M1–M3) of DFT calculators used in active learning framework. (b) Total DFT SCF loops of different optimization strategies and VASP calculator interfaces (M1–M3) compared with that of VASP CG. (c) Walltime speed-up of the whole relaxation process for different optimization strategies and VASP calculator interfaces (M1–M3) compared with that of VASP CG.

4. Conclusion

We have developed a comprehensive approach for performing and evaluating the acceleration of atomic simulations. We applied this approach to the local optimization of adsorbate-catalyst structures using VASP. We have shown that a transfer learning technique can be used to incorporate prior information from large-scale OCP graph models into an active learning framework. We also provided a method to evaluate its performance in terms of speed and accuracy. Using this method, we have shown that our active learning framework provides significant speed increases over other techniques for geometric optimizations of atomic structures, while maintaining equivalent accuracy, even without the use of an uncertainty metric. Further, we have demonstrated a method to run the VASP code in an interactive mode to mitigate the real-world computational cost of performing inefficient single point calculations during the optimization.

There are several obvious areas in which to expand this work. The first is to use the active learning framework on distinct atomic simulation tasks, to determine what limitations exist in terms of transferring the physical information learned by the pre-trained OCP model. The nudged elastic band method of finding
transition states might be an especially good use case for this method, since the fine-tuned model could be reused for each new relaxation [45, 46]. The second is to test more sophisticated and generalizable methods of fine-tuning on small amounts of data. The field of MLPs moves very quickly, and the speed-up we have demonstrated can be easily improved by incorporating the next better model. However, the fine-tuning approach must be adapted to each new architecture, either by manually testing which components of each model ought to be fine-tuned, or by implementing more generalized methods of adapting new model architectures. In other domains of ML there exists evidence that inserting adapter layers may be more effective for fine-tuning in very low data regimes [47]. Finally, it seems the performance of the online learning framework could be improved with more sophisticated querying strategies. Either with more reliable uncertainty metrics, or with optimizations to how those metrics are used to determine when to query. Reinforcement learning strategies have also shown some success in learning robust querying policies [48].

In addition to these areas of exploration, improvements or alternatives to the online active learning method of accelerating simulations described in this work should be compared using the framework of accuracy and speed we have described. This should result in faster and more useful techniques which can be adopted more confidently in practice. Methods for accelerating geometric optimizations performed by VASP involving single point calculations should also benefit from the efficiency gains provided by the VASPInteractive code. Future dataset generation and screening tasks normally performed using VASP should consider using online active learning and VASPInteractive in tandem to reduce computational costs and increase throughput. We also recommend that GPU-enabled machines be used to achieve the full acceleration enabled by the online active learning, but CPU-only machines will still achieve significant speed-up.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: https://github.com/ulissigroup/finetuna_manuscript.

Acknowledgments

We acknowledge the support from the U.S. Department of Energy, Office of Science, Basic Energy Sciences Award #DE-SC0019441, and Office of Energy Efficiency a Renewable Energy under Grant #DE-0008822. We also acknowledge support from the Swiss National Science Foundation. We also thank A J Medford (Georgia Tech), Andrew Peterson (Brown), Muhammed Shuaibi and Adeesh Kolluru (CMU), as well as C Lawrence Zitnick, Aditya Grover, Anuroop Sriram, Janice Lan, Nima Shoghi, and Siddharth Goyal (FAIR) for their insightful discussions. This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

ORCID iDs

Joseph Musielewicz  https://orcid.org/0000-0002-5296-9177
Xiaoxiao Wang  https://orcid.org/0000-0001-8587-8610
Zachary Ulissi  https://orcid.org/0000-0002-9401-4918

References

[1] Friend C M and Xu B 2017 Heterogeneous catalysis: a central science for a sustainable future Acc. Chem. Res. 50 517–21
[2] Liu H, Wei L, Liu F, Pei Z, Shi J, Wang Z J, He D and Chen Y 2019 Homogeneous, heterogeneous and biological catalysts for electrochemical N2 reduction toward NH3 under ambient conditions ACS Catal. 9 5245–57
[3] Yuranov I, Autissier N, Sordakis K, Dalebrook A F, Grasemann M, Orava V, Cendula P, Gubler L and Laurenczy G 2018 Heterogeneous catalytic reactor for hydrogen production from formic acid and its use in polymer electrolyte fuel cells ACS Sustain. Chem. Eng. 6 6635–43
[4] Ye R P et al 2019 CO2 hydrogenation to high-value products via heterogeneous catalysis Nat. Commun. 10 1–15
[5] De S, Dokania A, Ramirez A and Gascon J 2020 Advances in the design of heterogeneous catalysts and thermocatalytic processes for CO2 utilization ACS Catal. 10 14417–85
[6] Foster S L, Bakovic S I, Duda R D, Maheshwari S, Milton R D, Minter E D, Janik M J, Renner J N and Greenlee L F 2018 Catalysts for nitrogen reduction to ammonia Nat. Catal. 1 490–500
[7] Kobayashi Y, Tang Y, Kageyama T, Yamashita H, Masuda N, Hosokawa S and Kageyama H 2017 Titanium-based hydrides as heterogeneous catalysts for ammonia synthesis J. Am. Chem. Soc. 139 37
[8] Marakatti V S and Gaigneaux E M 2020 Recent advances in heterogeneous catalysis for ammonia synthesis ChemCatChem 12 5838–57
[9] McCullough K, Williams T, Mingle K, Jamshidi P and Lauterbach J 2020 High-throughput experimentation meets artificial intelligence: a new pathway to catalyst discovery Phys. Chem. Chem. Phys. 22 11174–96
[10] Bilaard T, Norskov J K, Dahl S, Matthiesen J, Christensen C H and Sehested J 2004 The Brønsted–Evans–Polanyi relation and the volcano curve in heterogeneous catalysis J. Catal. 224 206–17
[11] Kohn W, Becke A D and Parr R G 1996 Density functional theory of electronic structure J. Phys. Chem. 100 12974–80
[12] Goldsmith B R, Esterruizen J, Liu J X, Bartel C J and Sutton C 2018 Machine learning for heterogeneous catalyst design and discovery AIChE J. 64 2311–23
[13] Williams T, McCullough K and Lauterbach J A 2020 Enabling catalyst discovery through machine learning and high-throughput experimentation Chem. Mater. 32 157–65
[14] Del Rio E G, Mortensen J P and Jacobsen K W 2019 Local Bayesian optimizer for atomic structures Phys. Rev. B 100 104103
[15] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces Phys. Rev. Lett. 98 146401
[16] Smith J S, Iaseo O and Roitberg A E 2017 ANI-1: An extensible neural network potential with DFT accuracy at force field computational cost Chem. Sci. 8 3192–203
[17] Gao X, Ramezanghorbani F, Iaseo O, Smith J S and Roitberg A E 2020 TorchANI: a free and open source PyTorch-based deep learning implementation of the ANI neural network potentials J. Chem. Inf. Model. 60 5408–15
[18] Devereux C, Smith J S, Davis K K, Barros K, Zubatyuk R, Iaseo O and Roitberg A E 2020 Extending the applicability of the ANI deep learning molecular potential to sulfur and halogens J. Chem. Theory Comput. 16 4192–202
[19] Smith J S, Zubatyuk R, Neben G, Lubbens N, Barros K, Roitberg A E, Iaseo O and Tretiak S 2020 The ANI-1ccx and ANI-1x data sets, coupled-cluster and density functional theory properties for molecules Sci. Data 7 1–10
[20] Klicpera J, Grob J and Gümennüen S 2020 Directional message passing for molecular graphs (arXiv:2003.03123)
[21] Klicpera J, Sznit J and Gümennüen S 2020 Fast and uncertainty-aware directional message passing for non-equilibrium molecules (arXiv:2011.14115)
[22] Shuaibi M, Kolluru A, Das A, Grover A, Siriram A, Ulissi Z and Zitnick C L 2021 Rotation invariant graph neural networks using spin convolutions (arXiv:2106.09575)
[23] Gasteiger J, Becker F and Gümennüen S 2021 GemNet: Universal directional graph neural networks for molecules (arXiv:2106.08903)
[24] Chanussot L et al 2021 Open catalyst 2020 (OC20) dataset and community challenges ACS Catal. 11 6059–72
[25] Wang W, Yang T, Harris W H and Gómez-Bombarelli R 2020 Active learning and neural network potentials accelerate molecular screening of ether-based solvate ionic liquids J. Phys. Chem. Lett. 14 5755–63
[26] Lookman T, Balachandran P V, Xue D and Yuan R 2019 Active learning in materials science with emphasis on adaptive sampling using uncertainties for targeted design npj Comput. Mater. 5 1–17
[27] Tran K and Ulissi Z W 2018 Active learning across intermetallics to guide discovery of electrocatalysts for CO2 reduction and H2 evolution Nat. Catal. 1 696–703
[28] Zhong M et al 2020 Accelerated discovery of CO2 electrocatalysts using active machine learning Nature 581 178–83
[29] Vandermause J, Xie Y, Lim J S and Kozinsky B 2021 Active learning of reactive Bayesian force fields: application to heterogeneous hydrogen-platinum catalysis dynamics (arXiv:2106.01949)
[30] Vandermause J, Torrisi S B, Batzner S, Xie Y, Sun L, Kolkop A M and Kozinsky B 2020 On-the-fly active learning of interpretable Bayesian force fields for atomistic rare events npj Comput. Mater. 6 20
[31] Yang Y, Jiménez-Negrón O A and Kitchen J R 2021 Machine-learning accelerated geometry optimization in molecular simulation J. Chem. Phys. 154 234704
[32] Shuaibi M, Sivakumar S, Chen R Q and Ulissi Z W 2021 Enabling robust offline active learning for machine learning potentials using simple physics-based priors Mach. Learn.: Sci. Technol. 2 025007
[33] Musielwicz J, Wang X, Ulissi Z, Chen R Q, Adams M, Tian T and Shuaibi M 2022 FINETUNA: fine-tuning accelerated molecular simulations (available at: https://github.com/ulissigroup/finetuna) (Accessed 29 March 2022)
[34] Kresse G and Furthmüller J 1996 Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set Comput. Mater. Sci. 6 15–50
[35] Kresse G 1995 Ab initio molecular dynamics for liquid metals J. Non-Cryst. Solids 192–193 222–9
[36] Kresse G and Hafner J 1994 Ab initio molecular-dynamics simulation of the liquid-metalamorphous-semiconductor transition in germanium Phys. Rev. B 49 14251–69
[37] Kresse G and Furthmüller J 1996 Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set Phys. Rev. B 54 11169–86
[38] Hjorth Larsen A et al 2017 The atomic simulation environment—a Python library for working with atoms J. Phys.: Condens. Matter 29 273002
[39] Tian T and Musielwicz J 2022 VASPInteractive: interactive VASP calculator (available at: https://github.com/ulissigroup/vasp-interactive) (Accessed 3 April 2022)
[40] Sun C, Qu X, Xu Y and Huang X 2019 How to fine-tune BERT for text classification? Chinese Computational Linguistics (Lecture Notes in Computer Science vol 11856) pp 194–206
[41] Settles B 2009 Active learning literature survey Computer Sciences Technical Report vol 1648 (Madison, WI: University of Wisconsin)
[42] Musielwicz J and Wang X 2022 FINETUNA: fine-tuning accelerated molecular simulations manuscript (available at: https://github.com/ulissigroup/finetuna_manuscript) (Accessed 4 May 2022)
[43] Kresse G and Furthmüller J 1996 Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set Phys. Rev. B 54 11169–86
[44] Kresse G et al 2022 NELMDL: non-self consistent steps in electronic minimization from The VASP Manual (available at: www.vasp.at/wiki/index.php/NELMDL) (Accessed 3 April 2022)
[45] Garrido Torres J A, Jennings P C, Hansen M H, Boes J R and Bligaard T 2019 Low-scaling algorithm for nudged elastic band calculations using a surrogate machine learning model Phys. Rev. Lett. 122 156001
[46] Peterson A A 2016 Acceleration of saddle-point searches with machine learning J. Chem. Phys. 145 074106
[47] Houlsby N, Giurgiu A, Jastrzębski S, Morrone B, de Laroussilhe Q, Gesmundo A, Attariyan M and Gelly S 2019 Parameter-efficient transfer learning for NLP 36th Int. Conf. on Machine Learning, ICML 2019 vol 2019 pp 4944–53
[48] Fang M, Li Y and Cohn T 2017 Learning how to active learn: A deep reinforcement learning approach Proc. 2017 Conf. on Empirical Methods in Natural Language Processing (Copenhagen, Denmark) (Association for Computational Linguistics) pp 595–605