A Methodology to Generate Efficient Neural Networks for Classification of Scientific Datasets

Ria Patel*, Ariel Keller Rorabaugh*, Paula Olaya*, Silvina Caino-Lores*, Georgia Channing*,
Catherine Schuman*, Osamu Miyashita†, Florence Tama‡, Michela Taufer*

*Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA
†Center for Comp. Science, RIKEN, Kobe, Japan
‡Nagoya University & Center for Comp. Science, RIKEN, Nagoya & Kobe, Japan

2022 IEEE 18th International Conference on e-Science (e-Science) | 978-1-6654-6124-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ESCIENCE55777.2022.00052

***Abstract*—Neural networks (NNs) are increasingly utilized in high-throughput scientific workflows. In this context, NN efficiency is essential for successful workflow management. We use a multi-objective Neural Architecture Search (NAS), NSGA-Net, to search for highly accurate NNs while optimizing for efficient use of computational resources by minimizing FLoating-point Operations Per Second (FLOPS). We define a domain-agnostic methodology to generate NNs with the support of NSGA-Net, select promising NNs that balance accuracy and FLOPS usage, and refine a subset of NNs in order to curate networks suitable for efficient data analysis. We apply this methodology to a protein diffraction use case. Preliminary results show NNs that efficiently classify conformation of proteins with a final accuracy of 97.7% or higher and using only 187 FLOPS.**

***Index Terms*—NSGA-Net, neural architecture search, NAS**

I. Motivation and Contributions

Modern high-throughput scientific workflows produce, manage, and analyze increasingly large amounts of data on high performance computing (HPC) resources. Neural networks (NNs) are a data analysis method that is currently being integrated in many scientific workflows across multiple domains because of their accuracy [1], [2]. The challenges associated with the management of scientific workflows require NNs to be not only accurate but also efficient in the use of HPC resources. In this work, we explore how neural architecture search (NAS), traditionally used to automate the process of finding near-optimal NN models, can be leveraged to design NNs suitable for efficient data analysis. Specifically, we define a domain-agnostic methodology to generate, select, and refine NNs with the support of NSGA-Net (Non-dominated Sorting Genetic Algorithm Network) [3], a NAS implementation. NSGA-Net incorporates a multi-objective optimization (MOO) evolutionary algorithm, distinguishing it from other NAS implementations which are usually single-objective. We optimize for accuracy by minimizing error and for efficiency by minimizing inference FLoating-point Operations Per Second (FLOPS). Note that FLOPS usage per second is a measure of NN efficiency, thus the lower the FLOPS value, the better. Preliminary results with our methodology demonstrate accuracy and efficiency on a use case of classification of protein diffraction patterns.

II. Domain-agnostic Methodology

Our methodology is composed of three steps in which we generate, select, and refine candidate NNs automatically designed by the NAS to meet both accuracy and efficiency objectives. These steps are domain-agnostic and can be applied to any learning problem since they are based exclusively on analysis of the NNs' relative performance and efficiency.

Generation of candidate NN pool. To find NNs with maximum accuracy and efficiency, we use NSGA-Net's "macro search." NSGA-Net implements a MOO algorithm, and its macro search looks for an NN topology optimized for specified objectives. We choose error and FLOPS as the objectives to minimize. We run NSGA-Net for 30 generations with a population and offspring size of 25, and with each NN trained for 15 epochs. At the end of the run, NSGA-Net stores in an output file the topology of each NN explored during the run, as well as its accuracy and FLOPS. This constitutes the pool of candidate NNs.

Selection of best NNs. Usually, the NNs with high accuracy show a larger and more sophisticated architecture than NNs with lower performance. These NNs have a larger number of layers and larger number of neurons per layer, which implies a higher arithmetic intensity during inference that translates into a higher FLOP count. Consequently, accuracy and FLOPS are conflicting objectives in which a trade-off must be studied. Using the output accuracy and FLOPS for the explored NNs, we construct a Pareto frontier to analyze the trade-off between maximizing accuracy and minimizing FLOPS. The NNs in the Pareto-optimal set are the best-performing and more efficient NNs from the pool of NNs generated by NSGA-Net.

Refinement of best NNs. In order to find the best possible NN we further refine a subset of the NNs in the Pareto-optimal set. We establish a FLOPS cutoff point that lies below the elbow region of the Pareto frontier. We train all NNs from the Pareto frontier that are below that cutoff for 500 epochs. This long and targeted training allows us to achieve accuracy that is comparable to the elbow region of the Pareto frontier but with a fraction of the FLOPS.

III. Case Study: Classification of Protein Diffraction Patterns

We use our methodology to obtain an NN able to classify the structure (i.e., conformation) of a protein from the protein's diffraction pattern. Diffraction patterns are 2D images generated by X-ray free electron laser (XFEL) experiments [4].

978-1-6654-6124-5/22/$31.00 ©2022 IEEE
DOI 10.1109/eScience55777.2022.00052

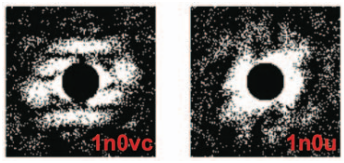

Fig. 1. Example of protein diffraction images for each conformation.

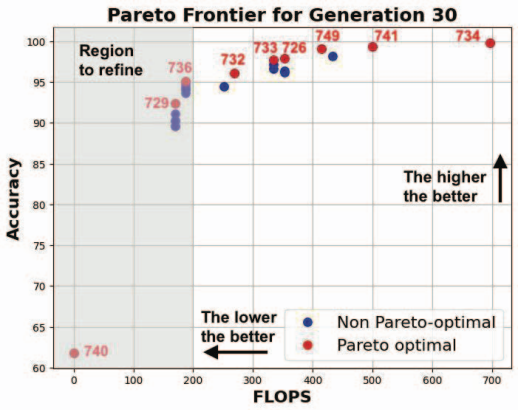

Fig. 2. Pareto frontier of generation 30 NN architectures, indicating the region to refine, which contains Pareto-optimal NNs 740, 729, and 736.

Structural properties of the protein—such as its conformation and orientation—have been previously studied in diffraction patterns through machine learning methods like KNN [5]. However, the performance of these methods has been evaluated purely in terms of accuracy, and computational efficiency has not been addressed. The availability of efficient and accurate NNs for diffraction pattern datasets has the potential to accelerate applications like 3D reconstruction [6]. The dataset consists of 79,384 protein diffraction pattern images, half of which represent the conformation *1n0u* and the other half represents the conformation *1n0vc*, as shown in Fig. 1.

Generation of candidate NN pool. We apply NSGA-Net to this dataset to observe and analyze the best generated NN architecture that satisfies the two desired objectives: minimized compute FLOPS usage during inference and maximized classification accuracy. Interfaces with NSGA-Net were adapted to support the protein diffraction dataset. We run NSGA-Net for 30 generations on the University of Tennessee Knoxville's Tellico cluster, an IBM Power9 (128 GB RAM) system with 2 NVIDIA Volta V100 GPUs. NSGA-Net completed the full run in 166.5 hours. We ran NSGA-Net with population and offspring size of 25, and as a result, 750 NNs were generated across 30 generations.

Selection and refinement of candidate NNs. Figure 2 shows the accuracy and FLOPS of the candidate NN pool after 30 generations, and highlights the Pareto-optimal set in red. We establish a cutoff FLOPS value of 200, so we select three NNs for refinement (i.e., models 740, 729 and 736). Table I shows a summary of the performance of these candidate models. Model 740 is extremely efficient and uses only 0.0161 FLOPS because it consists of a very small architecture with only one node. However, these structural characteristics also cause it to yield very low accuracy even after substantial training. Models 729 and 736 gain 5.3 and 3.5 percentage points in accuracy after full training for 500 epochs, which makes their accuracy comparable to the accuracy shown by NNs that double their FLOPS count.

TABLE I
FLOATING POINT OPERATIONS PER SECOND (FLOPS), VALIDATION ACCURACY BEFORE AND AFTER REFINEMENT, AND INFERENCE TIME FOR CANDIDATE NN MODELS 740, 729, AND 736.

NN model	FLOPS	Accuracy before refinement	Accuracy after refinement	Inference time (s)
740	0.016	57.1	57.1	0.407
729	169.125	92.4	97.7	1.161
736	187.125	94.6	98.1	1.250

IV. Conclusions and Relevance for eScience

We present a methodology to leverage NAS to obtain accurate and efficient NNs. These NNs can be used in scientific workflows and applications in need of efficient methods for high-frequency analysis. The methodology simplifies the process of developing a computationally efficient NN model for a given dataset, and it can be applied to data coming from any scientific domain.

Acknowledgment

This research was supported by the National Science Foundation (NSF) under grant numbers 1741057, 1740990, 1741040, 1841758 and 2223704; the Joint Directed Research Development (JDRD) program at UTK; the Oak Ridge Leadership Computing Facility under allocation CSC427; and IBM through a Shared University Research Award.

References

[1] M. Taufer, T. Estrada, and T. Johnston, "A Survey Of Algorithms For Transforming Molecular Dynamics Data Into Metadata For In Situ Analytics Based On Machine Learning Methods," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 378, p. 20190063, 03 2020.

[2] A. Keller Rorabaugh, S. Caíno-Lores, T. Johnston, and M. Taufer, "Building High-Throughput Neural Architecture Search Workflows via a Decoupled Fitness Prediction Engine," *IEEE Transactions on Parallel and Distributed Systems*, vol. 33, no. 11, pp. 2913–2926, 2022.

[3] Z. Lu, I. Whalen, V. Boddeti, Y. Dhebar, K. Deb, E. Goodman, and W. Banzhaf, "NSGA-Net: Neural Architecture Search Using Multi-Objective Genetic Algorithm," in *Proceedings of the Genetic and Evolutionary Computation Conference*, ser. GECCO '19. New York, NY, USA: Association for Computing Machinery, 2019, p. 419–427.

[4] A. Tokuhisa, S. Jonic, F. Tama, and O. Miyashita, "Hybrid Approach for Structural Modeling of Biological Systems from X-Ray Free Electron Laser Diffraction Patterns," *Journal of Structural Biology*, vol. 194, no. 3, pp. 325–336, 2016.

[5] P. Olaya, S. Caino-Lores, V. Lama, R. Patel, A. Rorabaugh, O. Miyashita, F. Tama, and M. Taufer, "Identifying Structural Properties of Proteins from X-ray Free Electron Laser Diffraction Patterns," in *18th IEEE International Conference on e-Science*, Salt Lake City, Utah, 2022.

[6] M. Nakano, O. Miyashita, S. Jonic, C. Song, D. Nam, Y. Joti, and F. Tama, "Three-Dimensional Reconstruction for Coherent Diffraction Patterns Obtained by XFEL," *Journal of synchrotron radiation*, vol. 24, pp. 727–737, Jul 2017.