2023 IEEE International Conference on Cluster Computing Workshops (CLUSTER Workshops) | 979-8-3503-7062-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/CLUSTERWorkshops61457.2023.00020

VINARCH: A Visual Analytics Interactive Tool for Neural Network Archaeology

Seoyoung An, Georgia Channing, Catherine Schuman, Michela Taufer
Department of Electrical Engineering and Computer Science, University of Tennessee, Knoxville, TN, USA

***Abstract*—Most neural networks (NNs) generated by neural architecture search (NAS) are discarded except for the final output to limit the memory usage on high performance computing (HPC) systems on which the search is performed. However, discarded NNs are vital for understanding the NAS structure's evolution and reproducibility. We design a visual interactive tool for NN archaeology that explores the evolution of NAS structures, finds matching subsequences in the structures, and visualizes NN similarities across NAS outputs, including discarded NNs. We demonstrate the capabilities of our tool to discover and visualize matching subsequences on a dataset of NNs generated by NSGA-Net, a genetic NAS.**

***Index Terms*—neural architecture search, high performance computing, genetic algorithm, NSGA-Net**

I. CHALLENGES WITH NEURAL NETWORK DESIGN

Neural network (NN) architecture design plays a significant role in deep learning because it determines the deep learning model's performance. Neural architecture search (NAS) automates the NN design [1], but it generates hundreds of NNs in the process. To limit memory usage on the HPC systems used for the search, most created NNs, except the final output, are usually discarded [1]. However, discarded NNs are vital for understanding how architecture evolves and its reproducibility. We address this challenge by introducing a visual interactive tool for neural network archaeology (VINARCH). Our tool enables users to explore NN datasets for architecture similarities, the first step towards understanding NN evolution and reproducibility. Our contributions are two-fold. First, we present VINARCH and its capabilities to analyze matching subsequences within NN architectures, calculate the distance between networks to quantify similarities, and summarize a wide range of validation accuracy statistics. Second, we use VINARCH with a dataset of evolving NNs generated by NSGA-Net, a genetic NAS, demonstrating VINARCH's capabilities to discover and visualize matching subsequences in the dataset.

II. DESIGNING OUR VISUAL ANALYTICS TOOL

We design VINARCH to integrate visualization and analytics features of NN architectures. VINARCH currently supports analyzing NNs created by NSGA-Net as nested binary arrays, but work in progress extends our tool to other NAS implementations. NSGA-Net is a well-known, open-source NAS that uses a genetic algorithm for NN searches [2]. Its goal is to reduce the error rate and required computational resources by optimizing prediction performance and computational complexity [2]. Figure 1 shows an example of NSGA-Net networks. The figure depicts an NN architecture as a graphic and binary array. The graphic and binary array are color-coded to show three connectivity phases within the NN structure. The three arrays (colored in red, blue, and orange) each represent a connectivity phase with up to four basic computational units [2]. Each phase shows the presence (1) or absence (0) of a connection between the basic computational units. VINARCH drops the brackets of the resultant NSGA-Net NN structures such that a binary array represents each NN and compares all NN structures pairwise. Going through each NN, VINARCH stores the starting index when binary arrays match. It stores the matching subsequence and sequence length in a summary CSV file, including variables such as the binary arrays of the matching subsequences, their length, starting index, NN index, count, and mean validation accuracy. Only subsequences longer than four are stored. VINARCH assigns a unique ID number to the unique NN structures. The analysis counts the number of unique NN structures with subsequences and calculates the mean validation accuracy across all networks with that subsequence. Using this data, VINARCH generates a set of visualization plots, including a scatter plot of subsequence statistics. To study similarities between networks, VINARCH supports the creation of heatmaps for the distances between the networks from selected subsequences. It offers two choices for distance calculation for distance metrics: Euclidean and city block distance calculation. Distance depicts the similarities and differences between these networks; a higher distance represents that the two networks are less similar, and a lower distance represents that they are more similar. VINARCH also generates an NN visualization of the structures for selected unique NN indices from the scatter plot in the ascending order of corresponding validation accuracy to get a deeper view and see the actual NN structures and their validation accuracy. Using our visualization, users can compare or contrast the NN structures and identify the matching subsequence structure starting from the scatter plot.

III. DEPLOYING OUR TOOL WITH A USE CASE

To demonstrate the capabilities of VINARCH's visualization and analytics, we use a dataset of 100 NNs generated by NSGA-Net's evolutionary algorithm [3] and trained for 13 to 25 epochs [4], [5]. We extract matching subsequences with their corresponding validation accuracy among all created

---

This research was supported by the National Science Foundation (NSF) under grant numbers 1841758 and 2223704 and by IBM through a Shared University Research Award.

979-8-3503-7062-1/23/$31.00 ©2023 IEEE

DOI 10.1109/CLUSTERWorkshops61457.2023.00020

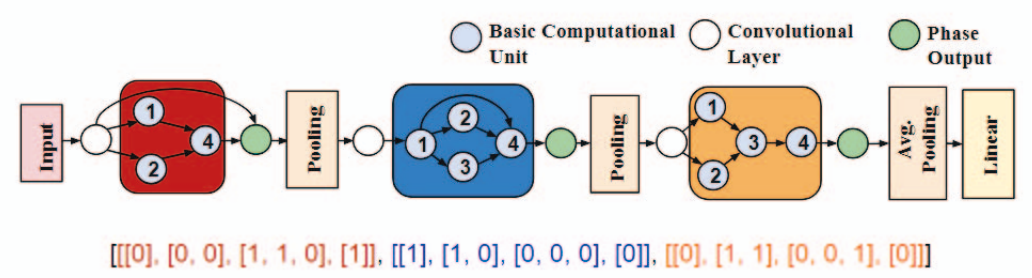

Fig. 1. Example of NSGA-Net's NN structure with its connectivity phases and basic computational units, both in graphic and array representation.

NNs; the subsequence lengths range from 5 to 20 binary digits. VINARCH assigns a unique ID number to the 97 unique NN structures from 0 to 96. Figure 2 shows our dataset's scatter plot of subsequence statistics using Bokeh library. The x-axis represents the sequence length of common binaries found, and the y-axis represents the mean validation accuracy. The size of each point represents the count of unique NN structures.

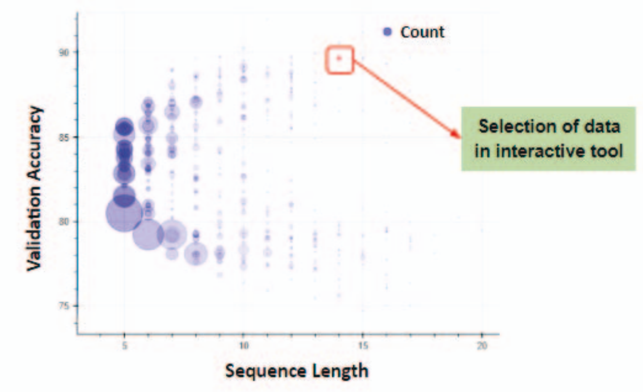

Fig. 2. Scatter plot of matching network subsequences statistics for our dataset.

In our demonstration, we select one point with 14 binary digits using the interactive features of our tool. Table I describes the statistics associated with the selected point (i.e., the matching subsequence, subsequence length, starting index, network indices, count, and mean validation accuracy). According to this data, the unique NN structures of 11, 29, 30, 43, and 49 contain the identical subsequence of length 14 starting at the first index. Figure 3 shows the structures of the selected NN indices for our use case in Table I. Yellow cells indicate 1, and purple cells indicate 0 of the binary array. VINARCH allows us to identify a matching subsequence across the five structures (outlined by dashed lines). Our tool also provides us with the validation accuracy for each sequence.

IV. Conclusions and Future Work

We present VINARCH to allow users to analyze NNs generated by NAS across multiple generations through various visualization tools. VINARCH can assist with studying the evolution and reproducibility of NAS-produced NN architectures. Future work includes expanding VINARCH to support a larger number of NAS, integrating new visual analytics capabilities, and studying reproducibility across larger NN databases.

TABLE I
Summary of Selected Point in Fig. 2.

Matching subsequence	[1, 0, 1, 1, 0, 1, 1, 0, 1, 0, 1, 0, 1, 0]
Sequence length	14
Starting index	0
Network indices	11, 29, 30, 43, 49
Count	5
Validation accuracy	89.66

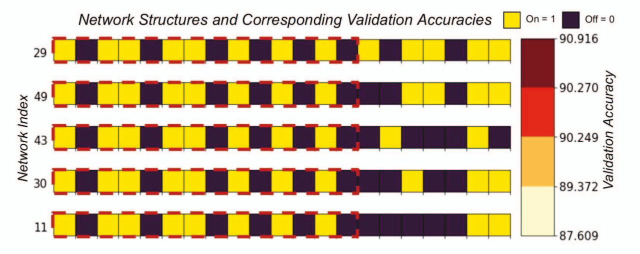

Fig. 3. Network structure of selected network indices in our dataset in ascending validation accuracy. The dashed line represents the matching subsequence.

References

[1] T. Elsken, J. H. Metzen, and F. Hutter, "Neural architecture search: A survey," *J. ML Research*, vol. 20, no. 1, pp. 1997–2017, 2019.

[2] 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 *Proc. of the Genetic and Evolutionary Computation CConference*, 2019, pp. 419–427.

[3] A. Keller Rorabaugh and et al., "Building high-throughput neural architecture search workflows via a decoupled fitness prediction engine," *IEEE Tran. on Parallel and Distr. Systems*, vol. 33, no. 11, pp. 2913–2926, 2022.

[4] G. Channing and et al., "Composable Workflow for Accelerating Neural Architecture Search Using In Situ Analytics for Protein Characterization," in *Proc. of the 52nd Int. Conference on Parallel Processing (ICPP)*, 2023.

[5] R. Patel and et al., "A Methodology to Generate Efficient Neural Networks for Classification of Scientific Datasets," in *Proc. of the IEEE 18th Int. Conference on e-Science (e-Science)*, 2022.