OAK RIDGE National Laboratory

Ensembles of Networks Produced from Neural Architecture Search

Emily Herron
Bredesen Center, University of Tennessee

Steven R. Young, Thomas Potok
Oak Ridge National Laboratory

ORNL is managed by UT-Battelle, LLC
for the US Department of Energy

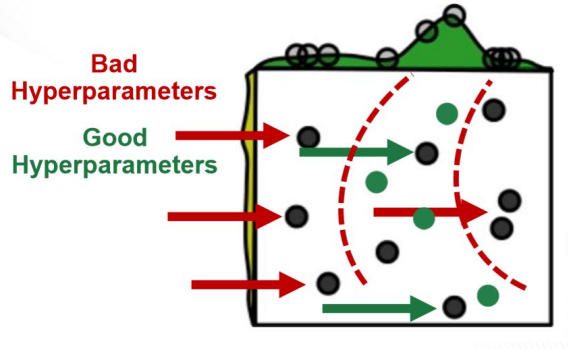
U.S. DEPARTMENT OF ENERGY

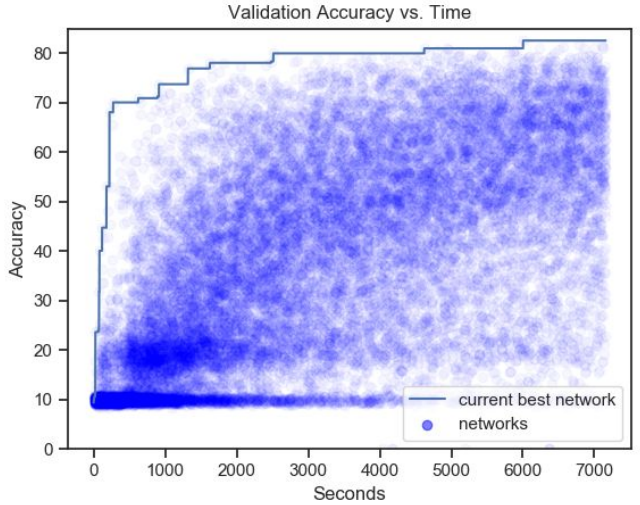
Outline

- Problem Overview
- Introduction
  - Neural Architecture Search
  - Neural Network Ensembles
- Methods
  - MENNDL
  - Ensembles
  - Experiments
  - Datasets
- Results
- Summary & Future Work

Problem Overview

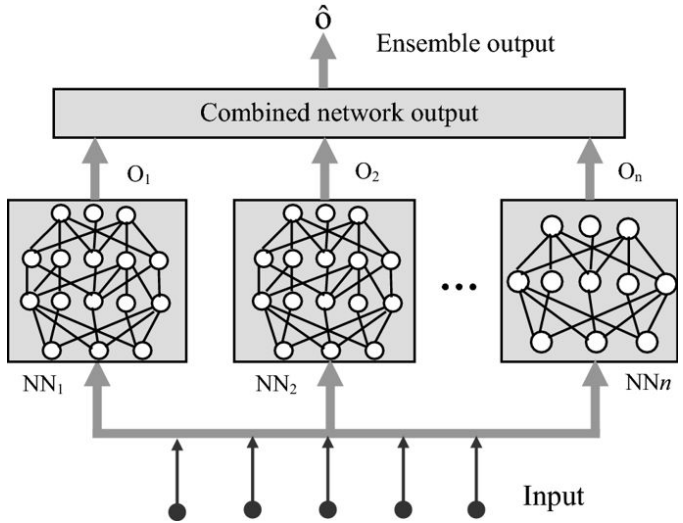
- Neural architecture search automatically designs neural networks for various challenges, focusing on single best network
- Single network with optimal performance may have limited knowledge of data distribution or over/under-fitted to training data.
- Many networks created & evaluated throughout NAS, providing opportunity to assemble network ensembles.
- Neural network ensembles combine outputs from neural networks with different parameters, offering improved prediction accuracies.





Problem Overview

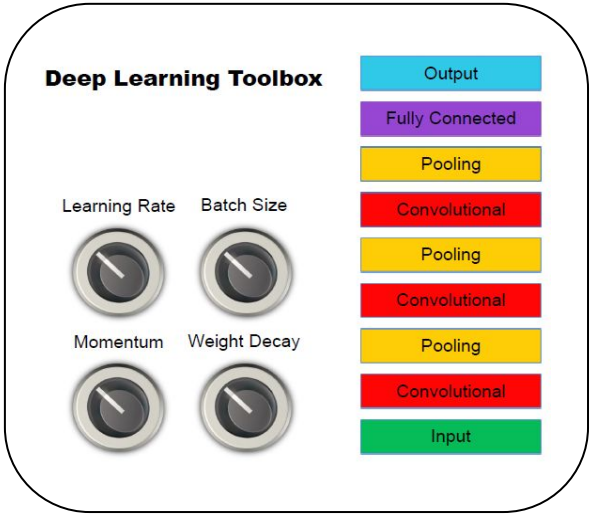
- We produced network ensembles with results from one or more runs of **M**ulti-node **E**volutionary **N**eural **N**etworks for **D**eep **L**earning (MENNDL)
- Two approaches considered and applied to two traditional image dataset benchmarks.
- We detail effects of ensembling networks from NAS method including:
  - Ensembles created from multiple instantiations of method.
  - Size of ensemble on performance.
  - Performance measured with accuracy & ensemble diversity.

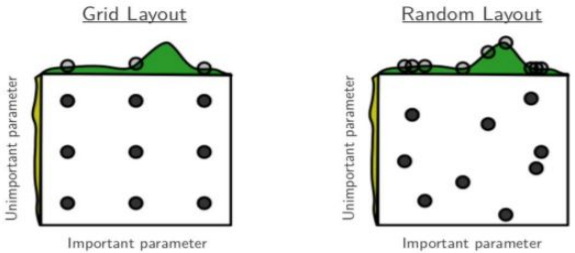


Alam, K, Siddique, N., Adeeli, H. A dynamic ensemble learning algorithm for neural networks, Neural Computing and Applications. Jul. 2019

Neural Architecture Search

- Features & learning capacity of deep convolutional neural networks (CNN) controlled by hyperparameters
- Tailoring architecture to data set is computationally expensive & time-consuming
- Hyperparameters traditionally selected manually or by grid or random search.
- We use MENNDL- evolutionary optimization approach to NAS
  - HPC framework that uses evolutionary algorithm to parallelize large-scale network evaluation.
  - Allows efficient hyperparameter search, by considering previous results, produces networks with increased accuracy & efficiency

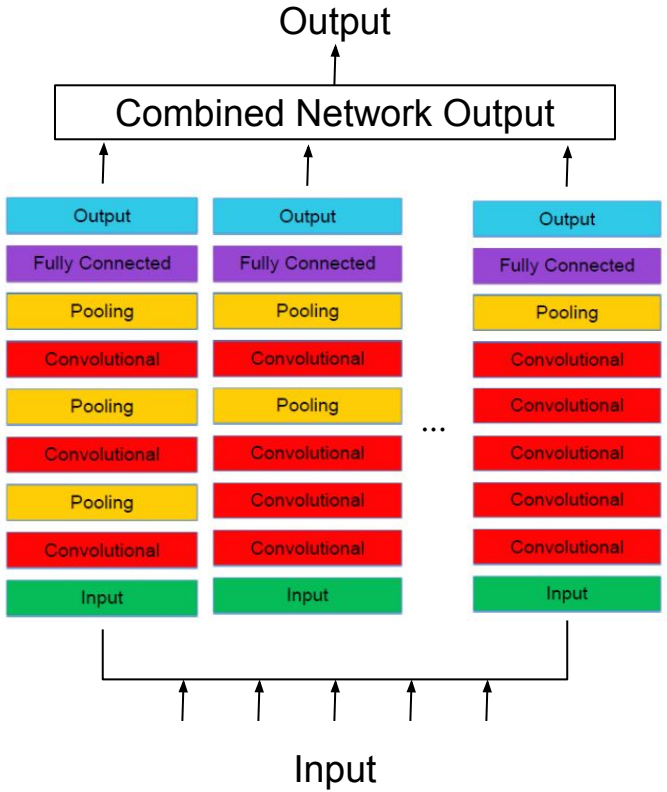




Bergstra, J, and Bengio, Y. Random Search for Hyperparameter Optimization, Journal of Machine Learning Research, Feb. 2012

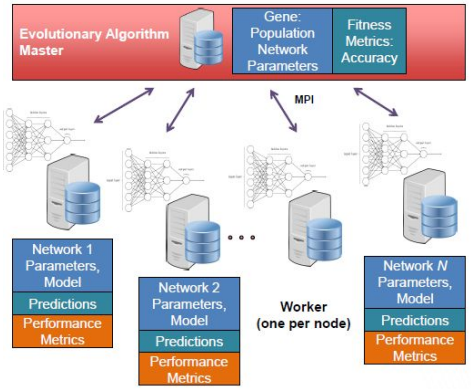
Neural Network Ensembles

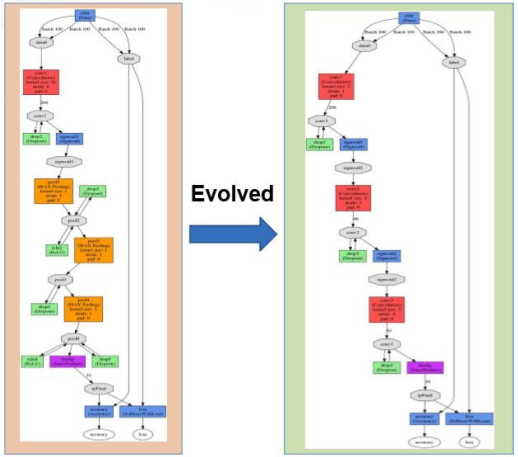
- Collection of neural networks trained on same task; results combined to produce model high generalization ability
- Successful deep learning models learn distribution of dataset; single models often overfit
- Network ensembles with different parameters and architectures learn varying aspects of training set
- Ensemble networks with randomly generated topologies, weights, or that learn random subsets of training data to encourage training error diversity.



MENNDL Overview

- **M**ulti-node **E**volutionary **N**eural **N**etworks for **D**eep **L**earning (MENNDL)- software framework that implements evolutionary algorithm for optimizing neural network topology & hyperparameters.
- Optimizes number of layers, layer type for each layer, and the corresponding layer hyperparameters.
- Utilizes asynchronous approach to evaluate the networks it generates in parallel in order to maximize utilization of leadership scale HPC



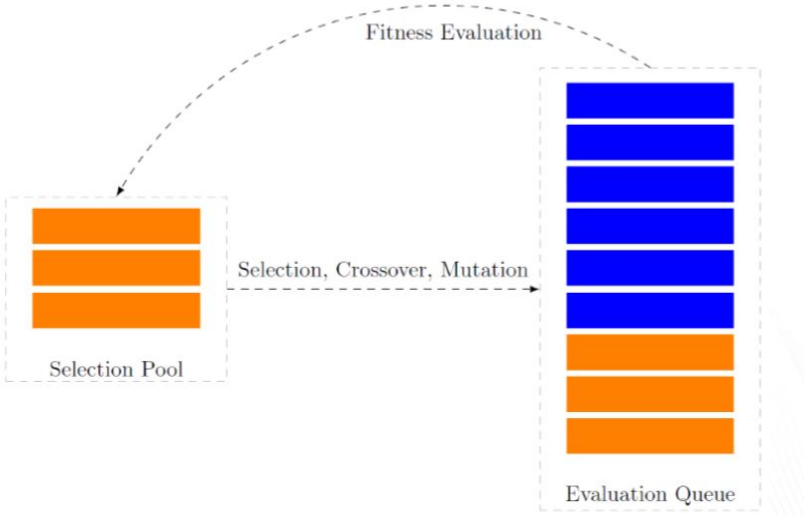


OAK RIDGE National Laboratory

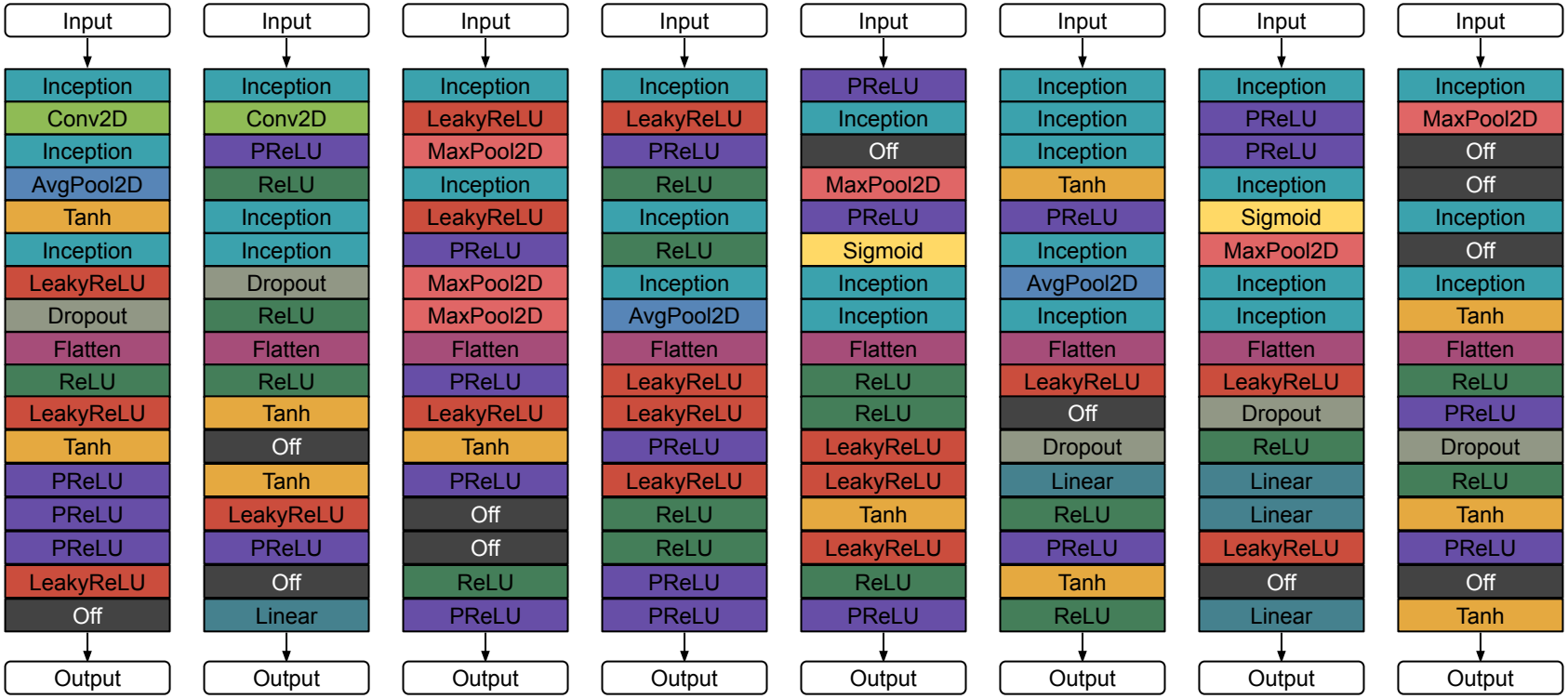
MENNDL

- Evolutionary algorithms mimic natural selection
  - Neural network population as individuals each with set of architectural hyperparameters or genes.
- Evolutionary algorithms are good for global search of spaces with many local minima.
- Basic evolutionary process:
  - Fitnesses of individuals in each generation evaluated (e.g. validation accuracy)
  - selection → mutation → crossover

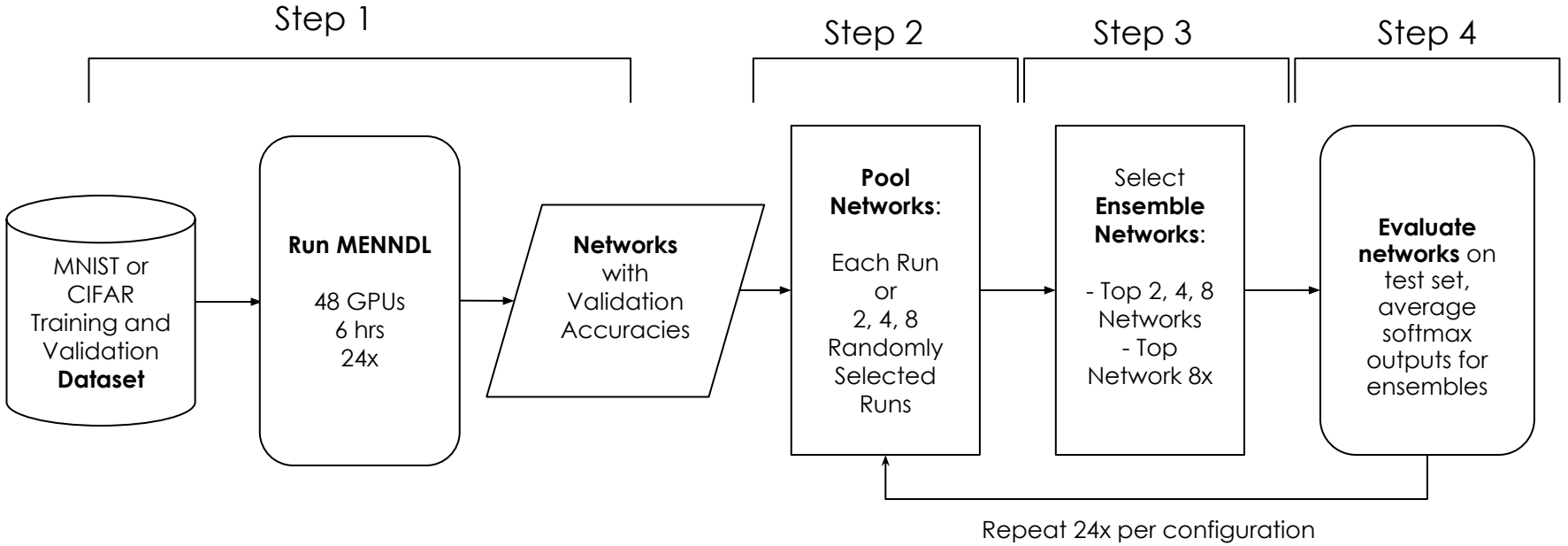
MENNDL Algorithm



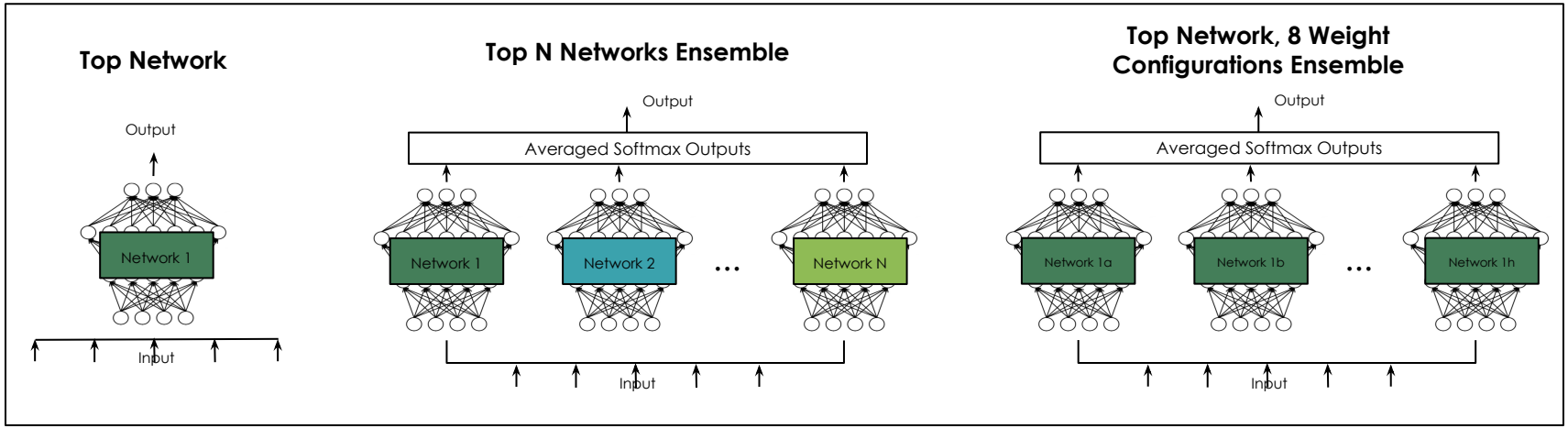
CIFAR-10 MENNDL Top Networks Runs 1-8

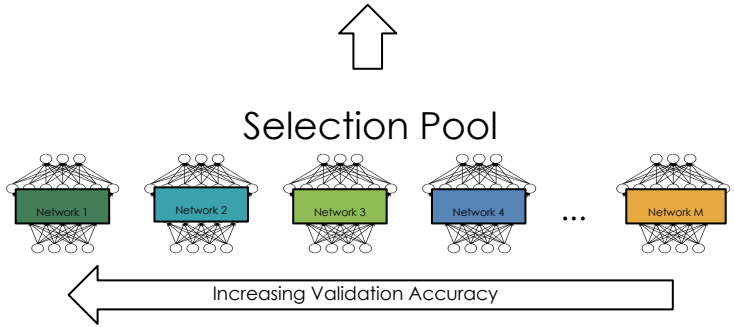


MENNDL Network Ensembles



Ensemble Configurations





OAK RIDGE National Laboratory

Measuring Ensemble Diversity

- Ensemble diversity measured by averaging total disagreement between predicted outputs for each sample
- Result is value that measures probability two networks in ensemble disagree with one another on given test sample.

Diversity Metric

$$d(p_i, p_j) = \frac{1}{m} \sum_{k=1}^{m} \psi(p_{ik}, p_{jk})$$

$$\psi(p_{ik}, p_{jk}) = \begin{cases} 0, & p_{ik} = p_{jk} \\ 1, & \text{otherwise} \end{cases}$$

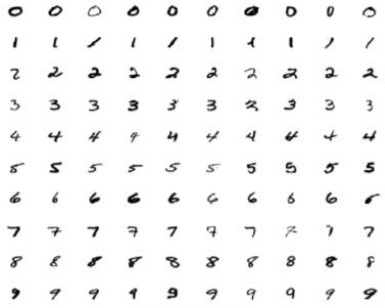
$$\frac{1}{n^2} \sum_{i=1}^{n} \sum_{j=1}^{n} d(p_i, p_j)$$

Experiment Details

- Each MENNDL run and ensemble experiment carried out on Summit supercomputer at Oak Ridge National Laboratory
- System has 4608 nodes, each with 2 IBM POWER9 CPUs and 6 NVIDIA Volta GPUs

Datasets

- CIFAR-10 dataset consists of 10 classes of 60,000 32 by 32 multicolor images
- MNIST consists of 70,000 28 by 28 grayscale images of handwritten digits ranging 0 to 9
- Dataset samples normalized; no data augmentation
- 10% random training samples as validation set per network.
- Individual networks trained with batch size of 64 on remaining training samples.
- Networks evaluated on validation set to obtain fitnesses for selection
- Ensemble accuracies based on test sets



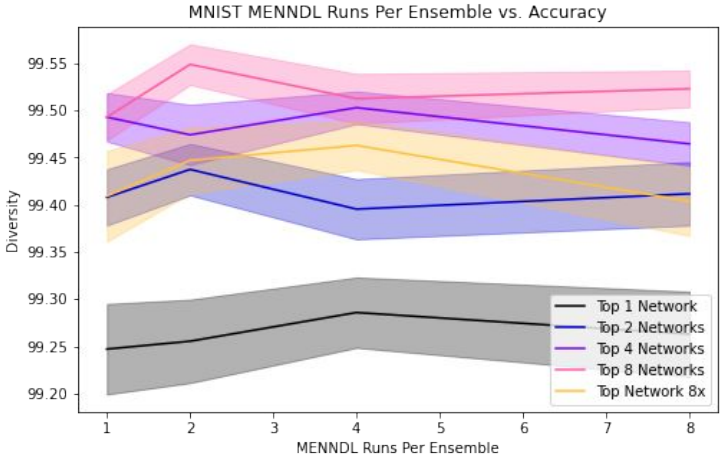
MNIST

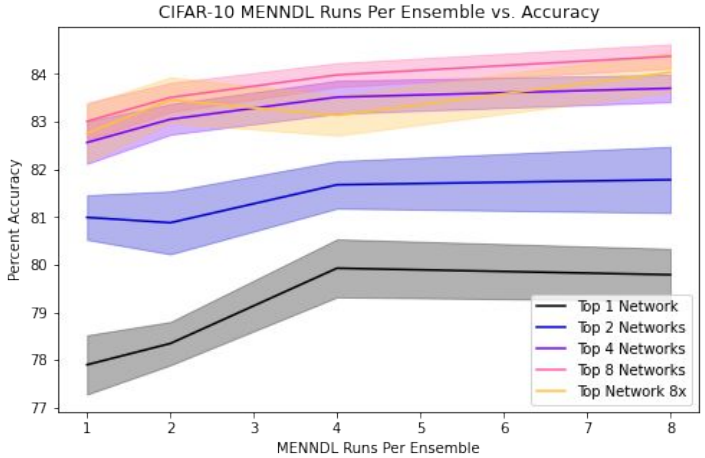
CIFAR

Results

MENNDL Run Statistics

Statistic	Dataset	
	MNIST	CIFAR-10
Total Networks	589.63±73.71	607.63±86.35
Generations	13.08±1.61	13.54±1.76
Best Network Fitness	99.33±0.10	78.47±1.26





- **Ensemble accuracies consistently higher when composed of more top networks**.
- Ensemble of top 2 networks offered significant accuracy improvements over individual networks.
- **Ensembles of top 2+ MENNDL runs improve upon generalizability of single best-performing network**.

Results

MNIST Mean Accuracy

MENNDL Runs	Ensemble Method				
	Top Network	Top 2 Networks	Top 4 Networks	Top 8 Networks	Top Network 8x
1	99.2471±0.1225	99.4079±0.0761	99.4929±0.0658	99.4929±0.0624	99.4092±0.1226
2	99.2554±0.1129	99.4375±0.0697	99.4742±0.0815	**99.5487±0.0550**	99.4471±0.0897
4	99.2858±0.0953	99.3954±0.0816	99.5029±0.0443	99.5125±0.0673	99.4629±0.0666
8	99.2629±0.1154	99.4117±0.0860	99.4646±0.0587	99.5229±0.0500	99.4038±0.0933

CIFAR-10 Mean Accuracy

MENNDL Runs	Ensemble Method				
	Top Network	Top 2 Networks	Top 4 Networks	Top 8 Networks	Top Network 8x
1	77.9025±1.5848	80.9925±1.2150	82.5629±1.1345	83.0067±0.9954	82.7583±1.5473
2	78.3483±1.1599	80.8808±1.6867	83.0500±0.8213	83.5075±0.7859	83.4538±1.2226
4	79.9271±1.5532	81.6767±1.2697	83.5146±0.8869	83.9796±0.6361	83.1325±1.0810
8	79.7904±1.3920	81.7825±1.7717	83.6996±0.7334	**84.3708±0.6521**	84.0350±1.0589

- CIFAR-10 ensembles tended to achieve **higher accuracies with larger pools of runs**
- **MNIST ensembles did not**, likely result of low error rates
- Misclassification rate lower with MNIST, little room to add functionally diverse networks to ensemble while maintaining high classification rates

Results

- **Ensembles of top 8 networks** yielded **diversities consistently higher** than 8 retrained copies of top network
- **Decreasing diversity with increasing pool size is an artifact of the diversity metric penalizing agreement even when all networks get the correct answer**

MNIST

MENNDL Runs	Ensemble Runs			
	Top 8 Networks		Top Network 8x	
	Diversity	Accuracy	Diversity	Accuracy
1	0.0610±0.0059	99.4954±0.0536	0.0060±0.0012	99.4083±0.1242
2	**0.0073±0.0007**	**99.5262±0.0569**	0.0055±0.0014	99.4429±0.1129
4	0.0069±0.0006	99.5262±0.0585	0.0058±0.0013	99.4154±0.1126
8	0.0066±0.0006	99.5212±0.0348	0.0060±0.0013	99.4421±0.0826

CIFAR-10

MENNDL Runs	Ensemble Runs			
	Top 8 Networks		Top Network 8x	
	Diversity	Accuracy	Diversity	Accuracy
1	**0.2118±0.0199**	83.0267±1.0399	0.1804±0.0197	82.6771±1.5128
2	0.2008±0.0117	83.7908±0.5655	0.1736±0.0161	82.9808±1.3632
4	0.1885±0.0138	83.8854±0.9317	0.1656±0.0152	83.5458±1.2053
8	0.1777±0.0158	**84.2163±0.6691**	0.1647±0.0158	83.9296±1.1975

Summary & Future Work

- Ensembles of multiple different networks produce better results than best network produced by search method, even when multiple copies of best network retrained several times
- Increased diversity of ensemble network structure produces increased diversity in network predictions, leading to improved ensemble performance.
- As we have demonstrated the diversity of network structures improves performance, we will look to explicitly leverage this by evolving ensembles of networks in NAS approach instead of creating ensemble in post-process, allowing NAS to explicitly identify networks that complement each other

OAK RIDGE National Laboratory

Questions?