

# Scalable Reinforcement-Learning-Based Neural Architecture Search for Cancer Deep Learning Research

Prasanna Balaprakash\*  
pbalapra@anl.gov  
Argonne National Laboratory

Romain Egele\*  
regele@anl.gov  
Argonne National Laboratory

Misha Salim  
msalim@anl.gov  
Argonne National Laboratory

Stefan Wild  
wild@anl.gov  
Argonne National Laboratory

Venkatram Vishwanath  
venkat@anl.gov  
Argonne National Laboratory

Fangfang Xia  
fangfang@anl.gov  
Argonne National Laboratory

Tom Brettin  
brettin@anl.gov  
Argonne National Laboratory

Rick Stevens  
stevens@anl.gov  
Argonne National Laboratory

## ABSTRACT

Cancer is a complex disease, the understanding and treatment of which are being aided through increases in the volume of collected data and in the scale of deployed computing power. Consequently, there is a growing need for the development of data-driven and, in particular, deep learning methods for various tasks such as cancer diagnosis, detection, prognosis, and prediction. Despite recent successes, however, designing high-performing deep learning models for nonimage and nontext cancer data is a time-consuming, trial-and-error, manual task that requires both cancer domain and deep learning expertise. To that end, we develop a reinforcement-learning-based neural architecture search to automate deep-learning-based predictive model development for a class of representative cancer data. We develop custom building blocks that allow domain experts to incorporate the cancer-data-specific characteristics. We show that our approach discovers deep neural network architectures that have significantly fewer trainable parameters, shorter training time, and accuracy similar to or higher than those of manually designed architectures. We study and demonstrate the scalability of our approach on up to 1,024 Intel Knights Landing nodes of the Theta supercomputer at the Argonne Leadership Computing Facility.

## KEYWORDS

cancer, deep learning, neural architecture search, reinforcement learning

## 1 INTRODUCTION

Cancer is a disease that drastically alters the normal biological function of cells and damages the health of an individual. Cancer is estimated to be the second leading cause of death globally and was responsible for 9.6 million deaths in 2018 [10]. A thorough understanding of cancer remains elusive because of challenges due to the variety of cancer types, heterogeneity within a cancer type, structural variation in cancer-causing genes, complex metabolic pathways, and nontrivial drug-tumor interactions [31, 48, 57, 64, 67].

Recently, as a result of coordinated data management initiatives, the cancer research community increasingly has access to a large volume of data. This has led to a number of promising large-scale, data-driven cancer research efforts. In particular, machine learning (ML) methods have been employed for tasks such as identifying cancer cell patterns; modeling complex relationships between drugs and cancer cells; and predicting cancer types.

With the sharp increases in available data and computing power, considerable attention has been devoted to deep learning (DL) approaches. The first wave of success in applying DL for cancer stems from adapting the convolutional neural network (CNN) and recurrent neural network (RNN) architectures that were developed for image and text data. For example, CNNs have been used for cancer cell detection from images, and RNNs and its variants have been used for analyzing clinical reports.

These adaptations are possible because of the underlying regular grid nature of image and text data [24]. For example, images share the spatial correlation properties, and convolution operations designed to extract features from natural images can be generalized for detecting cancer cells in images with relatively minor modifications. However, designing deep neural networks (DNNs) for nonimage and nontext data remains underdeveloped in cancer research. Several cancer predictive modeling tasks deal with *tabular data* comprising an output and multidimensional inputs. For example, in the drug response problem, DNNs can be used to model a complex nonlinear relationship between the properties of drugs and tumors in order to predict treatment response [84]. Here, the properties of drugs and tumors cannot easily be expressed as images and text and cast into classical CNN and RNN architectures. Consequently, cancer researchers and DL experts resort to manual trial-and-error methods to design DNNs. Tabular data types are diverse; consequently, designing DNNs with shared patterns such as CNNs and RNNs is not meaningful unless further assumptions about the data are made. Fully connected DNNs are used for many modeling tasks with tabular data. However, they can potentially lead to unsatisfactory performance because they can have large numbers of parameters, overfitting issues, and difficult optimization landscape with low-performing local optima [33]. Moreover, tabular data often is obtained from multiple sources and modes, where combining certain inputs using *problem-specific* domain knowledge

\*Both authors contributed equally to this research.

can lead to better features and physically meaningful and robust models, thus preventing the design of effective architectures similar to CNNs and RNNs.

Automated machine learning (AutoML) [1, 4, 22, 37, 88] automates the development of ML models by searching over appropriate components and their hyperparameters for preprocessing, feature engineering, and model selection to maximize a user-defined metric. AutoML has been shown to reduce the amount of human effort and time required for a number of traditional ML model development tasks. Although DL reduces the need for feature engineering, extraction, and selection tasks, finding the right DNN architecture and its hyperparameters is crucial for predictive accuracy. Even on image and text data, DNNs obtained by using AutoML approaches have outperformed manually engineered DNNs that took several years of development [59, 86, 88].

AutoML approaches for DNNs can be broadly classified into hyperparameter search and neural architecture search (NAS). Hyperparameter search approaches try to find the best values for the hyperparameters for a fixed neural architecture. Examples include random search [20], Bayesian optimization [21, 42, 70], bandit-based methods [44, 70], metaheuristics [51, 54], and population-based training [38] approaches. NAS methods search over model descriptions of neural network specifications. Examples include discrete search-space traversal [49, 56], reinforcement learning (RL) [13, 61, 88], and evolutionary algorithms [34, 73, 75, 81].

We focus on developing scalable neural-network-based RL for NAS, which offers several potential advantages. First, RL is a first-order method that leverages gradients. Second, RL-based NAS construction is based on a Markov decision process: decisions that are made to construct a given layer depend on the decisions that were made on the previous layers. This exploits the inherent structure of DNNs, which are characterized by hierarchical data flow computations; While traditional RL methods pose several challenges [76] such as exploration-exploitation tradeoff, sample inefficiency, and long-term credit assignment, recent developments [35, 68] in the field are promising.

Although hyperparameter search work has been done on cancer data [83], to our knowledge scalable RL-based NAS has not been applied to cancer predictive modeling tasks. An online bibliography of NAS [6] and a recent NAS survey paper [32] did not list any cancer-related articles. The reasons may be twofold. First, AutoML, and in particular NAS, is still in its infancy. Most of the existing work in NAS focuses on image classification tasks on benchmark data sets. Since convolutions and recurrent cells form the basic building blocks for CNNs and RNNs, respectively, the problem of defining the search space for CNN and RNN architectures has become relatively easy [32]. However, no such generalized building block exists for nonimage and nontext data. Second, large-scale NAS and hyperparameter search require high-performance computing (HPC) resources and appropriate software infrastructure [17]. These requirements are attributed to the fact that architecture evaluations (training and validation) are computationally expensive and parallel evaluation of multiple architectures on multiple compute nodes through scalable search methods is critical to finding DNNs with high accuracy in short computation time. We note that the time needed for NAS can be more than the training time of a manually designed network. However, designing the network by manually

intensive trial-and-error approaches can take days to weeks even for ML experts [37].

We develop a scalable RL-based NAS infrastructure to automate DL-based predictive model development for a class of cancer data. The contributions of the paper are as follows:

- We develop a DL NAS search space with new types of components that take into account characteristics specific to cancer data. These include multidrug and cell line inputs, 1D convolution for traversing large drug descriptors, and nodes that facilitate weight sharing between drug descriptors.
- We demonstrate a scalable RL-based NAS on 1,024 Intel Knights Landing (KNL) nodes of Theta, a leadership-class HPC system, for cancer DL using a multiagent and multiworker approach.
- We scale asynchronous and synchronous proximal policy optimization, a state-of-the-art RL approach for NAS. Of particular importance is the convergence analysis of search methods at scale. We demonstrate that RL-based NAS achieves high accuracy on architectures because of the search strategy and not by pure chance as in random search.
- We show that the scalable RL-based NAS can be used to generate multiple accurate DNN architectures that have significantly fewer training parameters, shorter training time, and accuracy similar to or higher than those of manually designed networks.
- We implement our approach as a neural architecture search module within DeepHyper [14, 15], an open-source software package, that can be readily deployed on leadership-class machines for cancer DL research.

## 2 PROBLEM SETS AND MANUALLY DESIGNED DEEP NEURAL NETWORKS

We focus on a set of DL-based predictive modeling problem sets from the CANcer Distributed Learning Environment (CANDLE) project [5] that comprises data sets and manually designed DNNs for drug response; RAS gene family pathways; and treatment strategy at molecular, cellular, and population scales. Within these problem sets, we target three benchmarks, which represent a class of predictive modeling problems that seek to predict drug response based on molecular features of tumor cells and drug descriptors. An overview of these open-source [2] benchmarks (i.e., data set plus manually designed DNN) is given below.

### 2.1 Predicting tumor cell line response (Combo)

In the Combo benchmark [3], recent paired drug screening results from the National Cancer Institute (NCI) are used to model drug synergy and understand how drug combinations interact with tumor molecular features. Given drug screening results on NCI60 cell lines available at the NCI-ALMANAC database, Combo’s goal is to build a DNN that can predict the growth percentage from the cell line molecular features and the descriptors of drug pairs. The manually designed DNN comprises three input layers: one for cell expression (of dimension  $d = 942$ ) and two for drug descriptors ( $d = 3,820$ ). The two input layers for the drug pairs are connected by a shared submodel of three dense layers ( $d = 1,000$ ). The cell expression layer is connected to a submodel of three dense layers each with  $d = 1,000$ . The outputs of these submodels are concatenated and connected to three dense layers each with  $d = 1,000$ . The scalar

output layer is used to predict the percent growth for a given drug concentration. The training and validation input data are given as matrices of sizes  $248,650$  (number of data points)  $\times$   $4,762$  (total input size) and  $62,164 \times 4,762$ , respectively. The training and validation output data are matrices of size  $248,650 \times 1$  and  $62,164 \times 1$ , respectively.

## 2.2 Predicting tumor dose response across multiple data sources (Uno)

The Uno benchmark [9] integrates cancer drug screening data from 2.5 million samples across six research centers to examine study biases and to build a unified drug response model. The associated manually designed DNN has four input layers: a cell RNA sequence layer ( $d=942$ ), a dose layer ( $d=1$ ), a drug descriptor layer ( $d=5,270$ ), and a drug fingerprints layer ( $d=2,048$ ). It has three feature-encoding submodels for cell RNA sequence, drug descriptor, and drug fingerprints. Each submodel is composed of three hidden layers, each with  $d=1,000$ . The last layer for each of the submodels is connected to the concatenation layer along with the dose layer. This is connected to three hidden layers each with  $d=1,000$ . The scalar output layer is used to predict tumor dose. We used the single drug paclitaxel, a simplified indicator, for this study. The training and validation input data are given as matrices of sizes  $9,588 \times 8,261$  and  $2,397 \times 8,261$ , respectively. The training and validation output data are given as matrices of sizes  $9,588 \times 1$  and  $2,397 \times 1$ , respectively.

## 2.3 Classifying RNA-seq gene expressions (NT3)

The NT3 benchmark [8] classifies tumors from normal tissue by tracking gene-expression-level tumor signatures. The associated manually designed DNN has an input layer for RNA sequence gene expression ( $d=60,483$ ). This is connected to a 1D convolutional layer of 128 filters with kernel size 20 and a maximum pooling layer of size 1. This is followed by a 1D convolutional layer of 128 filters with kernel size 10 and a maximum pooling layer of size 10. The output of the pooling layer is flattened and given to the dense layer of size 200 and a dropout layer with 0.1%. This is followed by a dense layer of size 20 and a dropout layer with 0.1%. The output layer of size 2 for the two classes with softmax activation is used to predict the tissue type. The training and validation input data are given as matrices of sizes  $1,120 \times 60,483$  and  $280 \times 60,483$ , respectively. The training and validation output data are given as matrices of sizes  $1,120 \times 1$  and  $280 \times 1$ , respectively.

## 3 RL-BASED NAS

NAS comprises (1) a search space that defines a set of feasible architectures, (2) a search strategy to search over the defined search space, and (3) a reward estimation strategy that describes how to evaluate the quality of a given neural architecture.

### 3.1 Search space

We describe the search space of a neural architecture using a graph structure. The basic building block is a set of nodes  $\mathcal{N}$  with possible choices; typically these choices are nonordinal (i.e., values that cannot be ordered in a numeric scale). For example,  $\{\text{Dense}(10, \text{sig}), \text{Dense}(50, \text{relu}), \text{Dropout}(0.5)\}$  respectively represent a

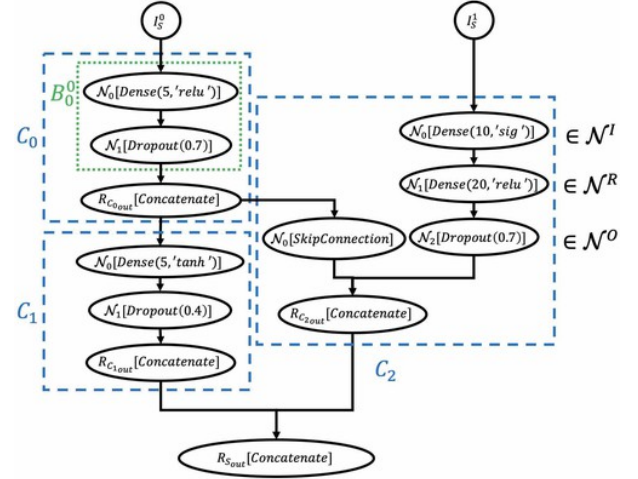


Figure 1: Example search space for NAS

dense layer with 10 units and sigmoid activation, a dense layer with 50 units with relu activation, and a layer with 50% dropout. A block  $B$  is a directed acyclic graph:  $(\mathcal{N} = (\mathcal{N}^I, \mathcal{N}^O, \mathcal{N}^R), \mathcal{R}_\mathcal{N})$ , where the set of nodes is differentiated by input nodes  $\mathcal{N}^I$ , intermediate nodes  $\mathcal{N}^R$ , and output nodes  $\mathcal{N}^O$  and where  $\mathcal{R}_\mathcal{N} \subseteq (\mathcal{N}^I \cup \mathcal{N}^R) \times (\mathcal{N}^R \cup \mathcal{N}^O)$  is a set of binary relations<sup>1</sup> that describe the connections among nodes in  $\mathcal{N}$ . A cell  $C_i$  consists of a set of  $L_i$  blocks  $\{B_i^0, \dots, B_i^{L_i-1}\}$  and a rule  $R_{C_i \text{ out}}$  to create the output of  $C_i$ . The structure  $S$  is given by the set  $\{(I_S^0, \dots, I_S^{P-1}), (C_0, \dots, C_{K-1}), R_{S \text{ out}}\}$ , where  $(I_S^0, \dots, I_S^{P-1})$  is a tuple of  $P$  inputs,  $(C_0, \dots, C_{K-1})$  is a tuple of  $K$  cells, and  $R_{S \text{ out}}$  is a rule to create the output of  $S$ . Users can define cell-specific blocks and block-specific input, intermediate, and output nodes.

Figure 1 shows a sample search space. The structure  $S$  is made up of three cells ( $C_0, C_1, C_2$ ). Cell  $C_0$  has one block  $B_0$  that has one input ( $\mathcal{N}_0$ ) and one output node ( $\mathcal{N}_1$ ). The rule to create the output is concatenation. Since there is only one block, the output from  $\mathcal{N}_1$  is the output layer for  $C_0$ ; cell  $C_1$  is similar to cell  $C_0$ . Since  $\mathcal{N}_0$  is a dense layer, the output of  $C_0$  is connected as an input to the  $C_1$ .

The search space definition that we have differs in two ways from existing chain-structured neural networks, multibranch networks, and cells blocks for designing CNNs and RNNs. The first is the flexibility to define multiple input layers  $(I_S^0, \dots, I_S^{J-1})$  (e.g., to support cell expression, drug descriptors in Combo; and RNA sequence, dose, drug descriptor, and drug fingerprints in Uno) and a cell for each of them. The second is the node types. By default, each node is a VariableNode, which is characterized by a set of possible choices. In addition, we define two types of nodes. ConstantNode, with a particular operation, is excluded from the search space but will be used in neural architecture construction. This allows for domain knowledge encoding—for example, if we want the dose value in Uno in every block, we can define a constant node for every block and connect them to the dose input layer. MirrorNode is used to reuse an existing node. For example, in Combo, drug1.descriptors and

<sup>1</sup>Without loss of generality, this can be extended to multiple intermediate nodes.

drug2.descriptors share the same submodel for feature encoding. To support such shared submodel construction, we define a cell with variable nodes for drug1.descriptors and a cell with mirror nodes drug2.descriptors. Consequently, the mirror nodes are not part of the specified search space.

Using the search space formalism, we define the search spaces for Uno, Combo, and NT3. We consider a small and a large search space for each of Combo and Uno. For NT3, we define only a small search space because the baseline DNN obtains 98% accuracy on the validation data.

**3.1.1 Combo.** We define a VariableNode as a node consisting of options representing identity operation, a dense layer with  $x$  units and activation function  $y$  (Dense( $x, y$ )), and a dropout layer Dropout( $r$ ) where  $r$  is a fraction of input units to drop (e.g., Identity, Dense(100, *relu*), Dense(100, *tanh*), Dense(100, *sigmoid*), Dropout(0.05), Dense(500, *relu*), Dense(500, *tanh*), Dense(500, *sigmoid*), Dropout(0.1), Dense(1000, *relu*), Dense(1000, *tanh*), Dense(1000, *sigmoid*), Dropout(0.2)). We refer to this VariableNode as MLP\_Node, where MLP stands for multilayered perceptron.

For the small Combo search space, we define cells  $C_0$ ,  $C_1$ , and  $C_2$ . Cell  $C_0$  receives input from three input layers, cell expression, drug 1 descriptors, and drug 2 descriptors, and has three blocks,  $\langle C_0, B_0 \rangle$ ,  $\langle C_0, B_1 \rangle$ , and  $\langle C_0, B_2 \rangle$ . The block  $\langle C_0, B_0 \rangle$  receives input from the cell expression layer and comprises three MLP\_Nodes connected sequentially in a feed-forward manner. The block  $\langle C_0, B_1 \rangle$  receives input from drug 1 descriptors and has three MLP\_Nodes similar to  $\langle C_0, B_0 \rangle$ . The block  $\langle C_0, B_2 \rangle$  receives input from drug 2 descriptors but has three Mirror\_Nodes that reuse the MLP\_Nodes of  $\langle C_0, B_1 \rangle$  to enable sharing the same submodel between drug 1 descriptors and drug 2 descriptors. The output from  $C_0$  is used as input to the cell  $C_1$  that contains two blocks,  $\langle C_1, B_0 \rangle$ ,  $\langle C_1, B_1 \rangle$ . The former has three MLP\_Nodes with feed-forward connectivity.  $\langle C_1, B_1 \rangle$  has one VariableNode with a Connect operation that includes options to create skip-connections (i.e., Null, Cell expression, Drug 1 descriptors, Drug 2 descriptors, Cell 1 output, Inputs, Cell expression & Drug 1 descriptors, Cell expression & Drug 2 descriptors, Drug 1 & 2 descriptors).

The output from  $C_1$  is used as input to the cell  $C_2$  that has one block  $\langle C_2, B_0 \rangle$  with three MLP\_Nodes with feed-forward connectivity. The Concatenate operation is used to combine the outputs from  $C_0$ ,  $C_1$ , and  $C_2$  to form the final output. The size of the architecture space is  $\approx 2.0968 \times 10^{14}$ .

For the large search space, we replicate  $C_1$  8 times. For  $C_i$  for  $i \in [1, \dots, 8]$ , we update the set of Connect operations by adding outputs of  $C_1, \dots, C_{i-1}$  (i.e., outputs of previous cells). The size of the architecture space is  $\approx 2.987 \times 10^{44}$ .

**3.1.2 Uno.** For the small search space, we define two cells,  $C_0$  and  $C_1$ . The cell  $C_0$  has four blocks,  $\langle C_0, B_0 \rangle$ ,  $\langle C_0, B_1 \rangle$ ,  $\langle C_0, B_2 \rangle$ , and  $\langle C_0, B_3 \rangle$ , that take cell rna-seq, dose, drug descriptors, and drug fingerprints as input, respectively. Each block has three MLP\_Nodes that are connected sequentially. The output rule of  $C_0$  is Concatenate. The cell  $C_1$  has one block  $\langle C_1, B_0 \rangle$  that takes the  $C_0$  output as input and has five nodes:  $\langle C_1, B_0, N_0 \rangle$ ,  $\langle C_1, B_0, N_1 \rangle$ ,  $\langle C_1, B_0, N_2 \rangle$ ,  $\langle C_1, B_0, N_3 \rangle$ , and  $\langle C_1, B_0, N_4 \rangle$ . The nodes  $\langle C_1, B_0, N_2 \rangle$  and  $\langle C_1, B_0, N_4 \rangle$  are ConstantNodes with the operation Add (i.e.,

elementwise addition for tensors). The other three are sequential MLP\_Nodes. The five nodes are connected sequentially, and  $\langle C_1, B_0, N_0 \rangle$  and  $\langle C_1, B_0, N_2 \rangle$  are connected to  $\langle C_1, B_0, N_2 \rangle$  and  $\langle C_1, B_0, N_4 \rangle$ , respectively. The size of the architecture space is  $\approx 2.3298 \times 10^{13}$ .

For the large search space, we have nine cells. The cell  $C_0$  is the same as the one we used for the small search space. Each cell  $C_i$ , for  $i \in [1, 8]$ , has two blocks. The block  $\langle C_i, B_0 \rangle$  has one MLP\_Node. The block  $\langle C_i, B_1 \rangle$  has one VariableNode with the following set of Connect operations to create skip connections: Null, all combinations of inputs (i.e., 15 possibilities), all outputs of previous cells, and all  $N_0$  of previous cells except  $C_0$ . Each  $C_i$  takes as input the output of the cell  $C_{i-1}$  for  $i \in [1, \dots, 8]$ . The size of the architecture space is  $\approx 5.7408 \times 10^{29}$ .

**3.1.3 NT3.** We define five types of nodes: Conv\_Node, Act\_Node, Pool\_Node, Dense\_Node, and Dropout\_Node. The Conv\_Node has the following options, where  $x$  in Conv1D( $x$ ) is the filter size. Here, the number of filters and the stride are set to 8 and 1, respectively: Identity, Conv1D(3), Conv1D(4), Conv1D(5), Conv1D(6). The Act\_Node has the following options, where  $x$  in Activation( $x$ ) is a specific type of activation function: Identity, Activation(*relu*), Activation(*tanh*), Activation(*sigmoid*). The Pool\_Node has the following options, where  $x$  in MaxPooling1D( $x$ ) represents the pooling size: Identity, MaxPooling1D(3), MaxPooling1D(4), MaxPooling1D(5), MaxPooling1D(6). The Dense\_Node has the following options: Identity, Dense(10), Dense(50), Dense(100), Dense(200), Dense(250), Dense(500), Dense(750), Dense(1000).

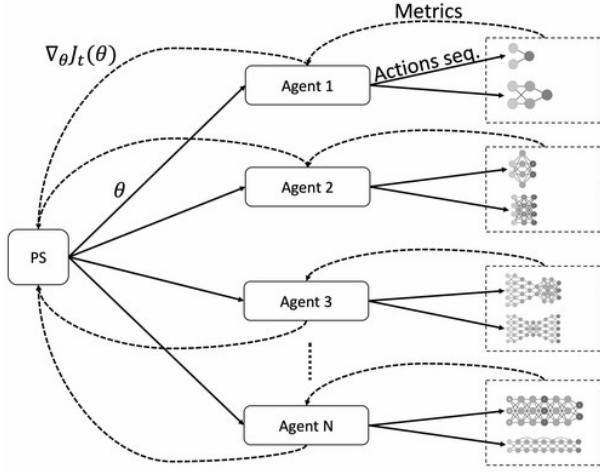
The Drop\_Node has the following options: Identity, Dropout(0.5), Dropout(0.4), Dropout(0.3), Dropout(0.2), Dropout(0.1), Dropout(0.05).

For the small search space, we define four cells:  $C_0$ ,  $C_1$ ,  $C_2$ , and  $C_3$ . Each cell  $C_i$  has one block  $\langle C_i, B_0 \rangle$ , which takes the output of the previous cell as input except for the first block  $\langle C_0, B_0 \rangle$ , which takes RNA-seq gene expression as input. This is followed by CONV\_Node, ACT\_Nodes, and POOL\_Node, which are connected sequentially. The blocks  $\langle C_0, B_0 \rangle$  and  $\langle C_1, B_0 \rangle$  have three sequentially connected VariableNodes: Conv\_Node, Act\_Nodes, and Pool\_Node. The blocks  $\langle C_2, B_0 \rangle$  and  $\langle C_3, B_0 \rangle$  have three sequentially connected VariableNodes: Dense\_Node, Act\_Node, and Drop\_Node. The size of the architecture space is  $6.3504 \times 10^8$ .

## 3.2 Search strategy

Different approaches have been developed to explore the space of neural architectures described by graphs. These approaches include random search, Bayesian optimization, evolutionary methods, RL, and other gradient-based methods. We focus on RL-based NAS, where an agent generates a neural architecture, trains the generated neural architecture on training data, and computes an accuracy metric on validation data. The agent receives a positive (negative) reward when the validation accuracy of the generated architecture increases (decreases). The goal of the agent is to learn to generate neural architectures that result in high validation accuracy by maximizing the agent’s reward.

Policy gradient methods have emerged as a promising optimization approach for leveraging DL for RL problems [76, 77]. These methods alternate between sampling and optimization using a loss



**Figure 2: Synchronous and asynchronous manager-worker configuration for scaling the RL-based NAS**

function of the form

$$J_t(\theta) = \hat{\mathbb{E}}_t[\log \pi_\theta(a_t|s_t)\hat{A}_t], \quad (1)$$

where  $\pi_\theta(a_t|s_t)$  is a stochastic policy given by the action probabilities of a neural network (parameterized by  $\theta$ ) that, for given a state  $s_t$ , performs an action  $a_t$ ;  $\hat{A}_t$  is the advantage function at time step  $t$  that measures goodness of the sampled actions from  $\pi_\theta$ ; and  $\hat{\mathbb{E}}_t$  denotes the empirical average over a finite batch of sampled actions. The gradient of  $J_t(\theta)$  is used in a gradient ascent scheme to update the neural network parameters  $\theta$  to generate actions with high rewards.

Actor-critic methods [35, 76] improve the stability and convergence of policy gradient methods by using a separate critic to estimate the value of each state that serves as a state-dependent baseline. The critic is typically a neural network that progressively learns to predict the estimate of the reward given the current state  $s_t$ . The difference between the rewards collected at the current state from the policy network  $\pi_\theta(a_t|s_t)$  and the estimate of the reward from the critic is used to compute the advantage. When the reward of the policy network is better (worse) than the estimate of a critic, the advantage function will be positive (negative), and the policy network parameters  $\theta$  will be updated by using the gradient and the advantage function value.

Proximal policy optimization (PPO) is a policy gradient method for RL [68] that uses a loss function of the form

$$J_t(\theta) = \hat{\mathbb{E}}_t[\min(r_t(\theta)\hat{A}_t, \text{clip}(r_t(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)], \quad (2)$$

where  $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_{\text{old}}}(a_t|s_t)}$  is the ratio of action probabilities under the new and old policies; the clip/median operator ensures that the ratio is in the interval  $[1 - \epsilon, 1 + \epsilon]$ ; and  $\epsilon \in (0, 1)$  is a hyperparameter (typically set to 0.1 or 0.2). The clipping operation prevents the sample-based stochastic gradient estimator of  $\nabla_\theta J_t(\theta)$  from making extreme updates to  $\theta$ .

We used two algorithmic approaches to scale up the NAS: synchronous advantage actor-critic (A2C) and asynchronous advantage actor-critic (A3C). Both approaches use a manager-worker

distributed learning paradigm as shown in Fig. 2. In A2C, all  $N$  agents start with the same policy network. At each step  $t$ , agent  $i$  generates  $M$  neural architectures, evaluates them in parallel (training and validation), and computes the gradient estimate  $\nabla_\theta J_t(\theta)$  using the PPO method. Once the parameter server (PS) receives the PPO gradients from the  $N$  agents, it averages the gradients and sends the result to each agent. The parameters of the policy network for each agent are updated by using the averaged gradient. The A3C method is similar to the A2C method except that an agent  $i$  sends the PPO gradients to the PS, which does not wait for the gradients from all the agents before computing the average. Instead, the PS computes the average from a set of recently received gradients and sends it to the agent  $i$ .

The synchronous update of A2C guarantees that the gradient updates to the PS are coordinated. A drawback of this approach is that the agents must wait until all  $M * N$  tasks have completed in a given iteration. Given a wide range of training times for the generated networks, A2C will underutilize nodes and limit parallel scalability. On the other hand, A3C increases the node utilization at the expense of gradient staleness due to the asynchronous gradient updates, a staleness that grows with the number of agents. While several works address synchronous and asynchronous updates in large batch supervised learning, studies in RL settings are limited, and none exists for the RL-based NAS methods studied here.

### 3.3 Reward estimation strategy

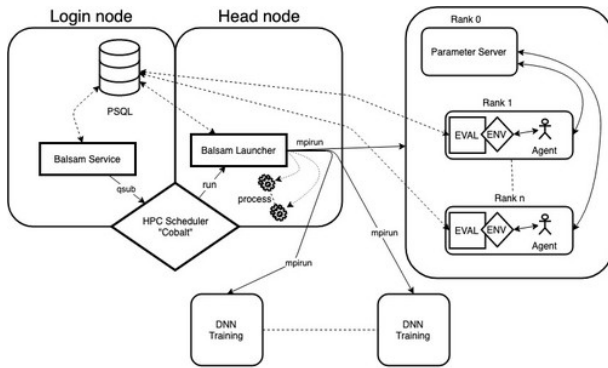
Crucial to the effectiveness of RL-based NAS is the way in which rewards are estimated for the agent-generated architectures. A naive approach of training each architecture from scratch on the full training data is computationally expensive even at scale and can require thousands of single-GPU days [32]. A common approach to overcome this challenge is low-fidelity training, where the rewards are estimated by using a smaller number of training epochs [87], a subset of original data [43], a smaller proxy network for the original network [89], and a smaller proxy data set [27]. In this paper, we use a smaller number of training epochs, a subset of the full training data, and timeout strategies to reduce the training time required for estimating the reward for architectures generated by NAS agents.

For the image data sets, research has showed that low-fidelity training can introduce a bias in reward estimation, which requires a gradual increase in fidelity as the search progresses [44]. Whether this is the case for nonimage and nontext cancer data is not clear, however. Moreover, the impact of low-fidelity training on NAS at scale is not well understood. As we show next, at scale the RL-agent behavior (and consequently the generated architectures) exhibits different characteristics based on the fidelity level employed.

## 4 SOFTWARE DESCRIPTION

Our open source software comprises three Python subpackages: benchmark, a collection of representative NAS test problems; evaluator, a model evaluation interface with several execution backends; and search, a suite of parallel NAS implementations that are implemented as distributed-memory mpi4py applications, where each MPI rank represents an RL agent.

As new architectures are generated by the RL agents, the corresponding reward estimation tasks are submitted via the evaluator



**Figure 3: Distributed NAS architecture.** The Balsam service runs on a designated node, providing a Django interface to a PostgreSQL database and interfacing with the local batch scheduler for automated job submission. The launcher is a pilot job that runs on the allocated resources and launches tasks from the database. The multiagent search runs as a single MPI application, and each agent submits model evaluation tasks through the Balsam service API. As these evaluations are added, the launcher continually executes them by dispatch onto idle worker nodes.

interface. The evaluator exposes a three-function API, that generically supports parallel asynchronous search methods. In the context of NAS, `add_eval_batch` submits new reward estimation tasks, while `get_finished_evals` is a nonblocking call that fetches newly completed reward estimations. This API enforces a complete separation of concerns between the search and the backend for parallel evaluation of generated architectures. Moreover, a variety of evaluator backends, ranging from lightweight threads to massively parallel jobs using a workflow system, allow a single search code to scale from toy models on a laptop to large DNNs running across leadership-class HPC resources.

We used the DeepHyper [15] software module on Theta, our target HPC platform, to dispatch reward estimation tasks to Balsam [66], a workflow manager enabling high-throughput, asynchronous task launching and monitoring for supercomputing platforms. Each agent exploited DeepHyper’s evaluation cache and leveraged this to avoid repeating reward estimation tasks. A global cache of evaluated architectures is not maintained because that would nullify the benefit of agent-specific random weight initialization. Balsam’s performance monitoring capabilities are used to infer utilization as the fraction of allocated compute nodes actively running evaluation tasks at any given time  $t$ ; the maximum value of 1.0 indicates that all worker nodes are busy evaluating configurations.

A schematic view of the NAS-Balsam infrastructure for parallel NAS is shown in Fig. 3. The BalsamEvaluator queries a Balsam PostgreSQL database through the Django model API. Each NAS agent interacts with an environment that contains a BalsamEvaluator; therefore each agent has a separate database connection. The Balsam launcher, in turn, pulls new reward estimation tasks and launches them onto idle nodes using a pilot-job mechanism. The launcher

monitors ongoing tasks for completion status and signals successful evaluations back to the BalsamEvaluator.

For the implementations of A3C and A2C, we interfaced our NAS software with OpenAI Baselines [29], open-source software with a set of high-performing state-of-the-art RL methods. We developed an API for the RL methods in OpenAI Baselines so that we can leverage any new updates and RL methods that become available in the package. We followed the same interface as in OpenAI Gym [23] to create a NAS environment that encapsulates the Evaluator interface of Balsam to submit jobs for reward estimation.

The interface to specify the graph search space comprises support for structure, cell, block, variable node, constant node, mirror node, and operation. These are implemented as Python objects that allow the search space module to be extensible for different applications. The different choices for a given variable node are specified by using the `add_op` method. These choices can be any set of Dense or Connect operations; the former creates a Keras [26] dense layer and the latter creates skip connections. After a neural architecture is generated, the corresponding Keras model is created automatically for training and inference.

The analytics module of the software can be used to analyze the data obtained from NAS. This module parses the logs from the NAS to extract the reward trajectory over time and to find the best architectures, worker utilization from the Balsam database, and number of unique architectures evaluated.

## 5 EXPERIMENTAL RESULTS

For the NAS search, we used Theta, a 4,392-node, 11.69-petaflop Cray XC40-based supercomputer at the Argonne Leadership Computing Facility (ALCF). Each node of Theta is a 64-core Intel Knights Landing (KNL) processor with 16 GB of in-package memory, 192 GB of DDR4 memory, and a 128 GB SSD. The compute nodes are interconnected by using an Aries fabric with a file system capacity of 10 PB.

The reward estimation for a given architecture uses only a single KNL node (no distributed learning) with the number of training epochs set to 1 and timeout set to 10 minutes.

The reward estimation for a generated architecture comprises two stages: training and validation. For Combo, the training is performed by using only 10% of the training data. For Uno and NT3, since the data sizes are smaller, the full training data are used. The reward is computed by evaluating the trained model on the validation data set. For Combo and Uno, we use  $R^2$  value as the reward; for NT3, we use classification accuracy (ACC). While we focus on accuracy in this paper, other metrics can be specified, such as model size, training time, and inference time for a fixed accuracy using a custom reward function. To increase the exploration among agents, we used random weight initialization in the DNN training using agent-specific seeds during the reward estimation. Consequently, different agents generating the same architecture can have different rewards. For the policy network and value networks, we used a single-layer LSTM with 32 units and trained them with epochs=4, clip=0.2, and learning\_rate=0.001, respectively.

Once the NAS search was completed on Theta, we selected the top 50 DNN architectures from the search based on the estimated reward values. We performed post-training, where we trained the



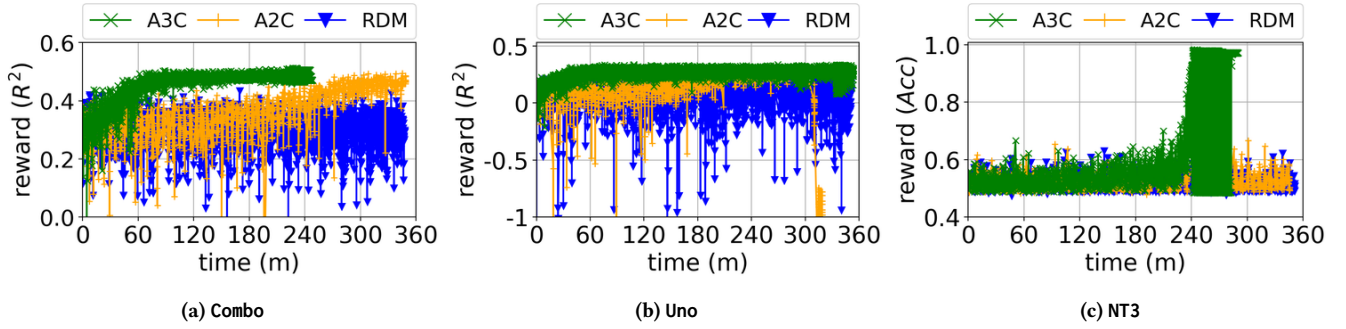


Figure 4: Search trajectory showing reward over time for A3C, A2C, and RDM on the small search space

DNNs for a larger number of epochs (20 for all experiments), without a timeout, and on the full training data. Running post-training on the KNL nodes was slower; therefore we used Cooley, a GPU cluster at the ALCF. Cooley has 126 compute nodes; each node has 12 CPU cores, one NVIDIA Tesla K80 dual-GPU card, with 24 GB of GPU memory and 384 GB of DDR3 CPU memory. The manually designed Combo network took 2215.13 seconds and 705.26 seconds for training on KNL and K80 GPUs, respectively. We ran 50 models on 25 GPUs with two model trainings per K80 dual-GPU card, one per GPU. For both the reward estimation and post-training, we used the Adam optimizer with a default learning rate of 0.001. The batch size was set to 256, 32, and 20 for Combo, Uno, and NT3, respectively. The same values were used in the manually designed networks.

We evaluated the generated architectures after post-training with respect to three metrics: **accuracy ratio** ( $R^2/R_b^2$  or  $ACC/ACC_b$ ), given by the ratio of  $R^2$  ( $ACC$ ) of a NAS-generated architecture and the manually designed network for Combo and Uno (NT3); **trainable parameters ratio** ( $P_b/P$ ), given by the ratio of number of trainable parameters of the manually designed network and the given NAS-generated architecture (this metric helped us evaluate the ability of NAS to build smaller networks, which have better generalization and fewer overfitting issues compared with larger networks); and **training time ratio** ( $T_b/T$ ), given by the ratio of the post-training time (on a single NVIDIA Tesla K80 GPU) of the manually designed network and the given NAS-generated architecture (this metric allows us to evaluate the ability of the NAS to find faster-to-train networks, which are useful for hyperparameter search and subsequent training with additional data).

The Theta environment consists of Cray Python 3.6.1, TensorFlow 1.13.1 [11]. Based on ALCF recommendations, we used the following environment variable settings to increase the performance of TensorFlow: `KMP_BLOCKTIME='0'`, `KMP_AFFINITY='granularity=fine,compact,1,0'`, and `intra_op_parallelism_threads=62`. The Cooley environment consists of Intel Python 3.6.5, Tensorflow-GPU 1.13.1.

## 5.1 Evaluation of the search strategy

In this section, we show that in spite of gradient staleness issue, A3C has a faster learning capability and a better system utilization than does A2C; synchronized gradient updates and the consequent node idleness adversely affect the efficacy of A2C.

We evaluated the learning and convergence capabilities of A3C and A2C by comparing them with random search (RDM), where agents perform actions at random and will not compute and synchronize gradients with the parameter server. This comparison was to ensure that the search space was the same as A3C, A2C, and RDM and allowed us to evaluate the search capabilities of A3C and A2C with all other settings remaining constant. We used 256 Theta nodes for A3C, A2C, and RDM with 21 agents and 11 workers per agent<sup>2</sup>: 21 agent nodes, 231 worker nodes, 1 Balsam workflow node, and 3 unused nodes.

Figure 4 shows rewards obtained over time for A3C, A2C, and RDM. We observe that A3C outperforms A2C and RDM with respect to both time and rewards obtained. A3C exhibits a faster learning trajectory than does A2C and reaches a higher reward in a shorter wall-clock time. A3C reaches reward values of 0.5 and 0.4 in approximately 70 and 35 minutes for Combo and Uno, respectively, after which the increase in the reward values is small. On Combo and NT3, A3C ends in 250 and 285 minutes, respectively, because all the agents generate the same architecture for which the agent-specific cache returns the same reward value. We detected this and stopped the search since it could not proceed in a meaningful way. On Uno, A3C generates different architectures and does not end before the wall-clock time limit. A2C shows a slower learning trajectory; it eventually reaches the reward values of A3C on Combo and Uno, but its reward value on NT3 is poor. As expected, RDM shows neither learning capability nor the ability to reach higher reward values. On NT3, we found an oscillatory behavior with A3C toward the end. After finding higher rewards, A3C did not converge as expected. After a closer examination of the architectures and their reward values, we found that although the agents are producing similar architectures, the reward estimation is sensitive to a random initializer with one training epoch and a batch size of 20. Consequently, the same architecture produced by two different agents had significantly different rewards (e.g., 1.0 and 0.4).

Figure 5 shows the utilization over time for A3C, A2C, and RDM on the small search space. The utilization of the RDM on Combo stays at 1.0 in the initial search stages, but after that it averages 0.75. Although RDM lends itself to an entirely asynchronous search, the estimation of  $M$  rewards per agent was blocking in our implementation. This per-agent synchronization, combined with variability

<sup>2</sup>We want to set number of agents  $\approx 2 \times$  number of workers: 21 agents and 11 workers satisfy the constraint with minimal unused nodes. We requested 256 instead of 253 to get 6 hours of running time.

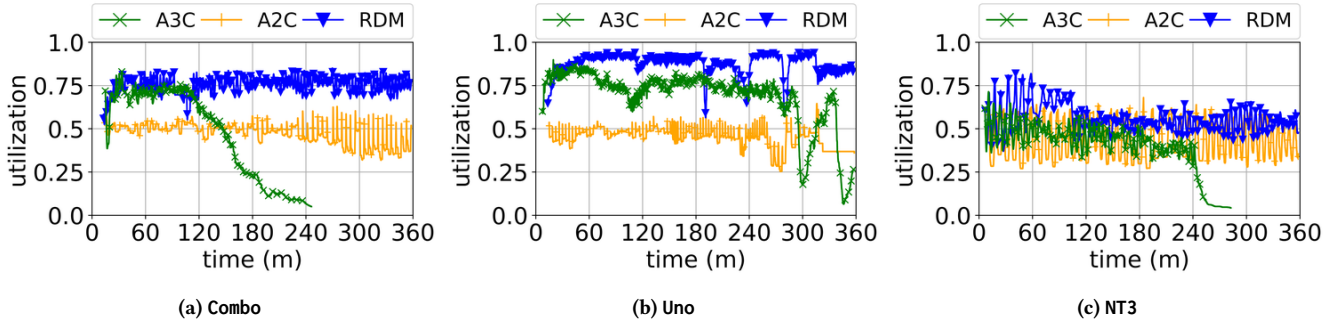


Figure 5: Utilization for A3C, A2C, and RDM on the small search space

of the reward estimation times, leads to suboptimal utilization. The utilization of A3C is similar to that of RDM until 100 minutes, after which there is a steady decrease due to an increase in the caching effect; this is just a manifestation of the convergence of A3C, which stops after 160 minutes.

On Uno, the utilization of RDM becomes high, with an average of 0.9. This is because randomly sampled DNNs in this space have a smaller variance of reward estimation times. The utilization of A3C is similar to that of RDM in the beginning of the search, but it decreases after 50 minutes because it learns to generate architectures that have a shorter training time with higher rewards. On NT3, the utilizations of RDM and A2C are similar to that of Combo but with even lower values because per batch several architectures have a shorter reward estimation time. The utilization of A2C shows a sawtooth shape; because of the synchronous nature, at the start of each batch the utilization goes to 1, then drops off and becomes zero when all agents finish their batch evaluation.

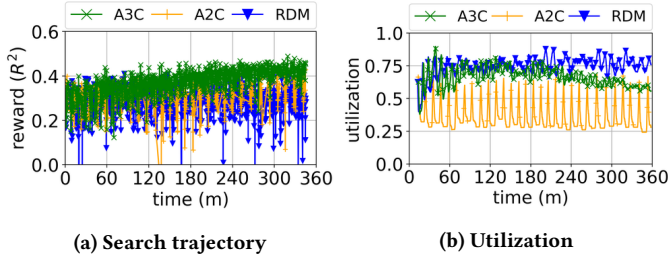


Figure 6: Results on Combo with the large search space

Figure 6 shows the search trajectory and utilization of A3C on Combo with the large search space. We observe that A3C finds higher rewards faster than do A2C and RDM. The utilization of A3C is similar to that of RDM (75% average) until 200 minutes, after which there is a gradual decrease because of the caching effect. Nevertheless, the search did not converge and stop as it did in the small search space.

## 5.2 Comparison of A3C-generated networks with manually designed networks

Here, we show that A3C discovers architectures that have significantly fewer trainable parameters, shorter training time, and accuracy similar to or higher than those of manually designed architectures.

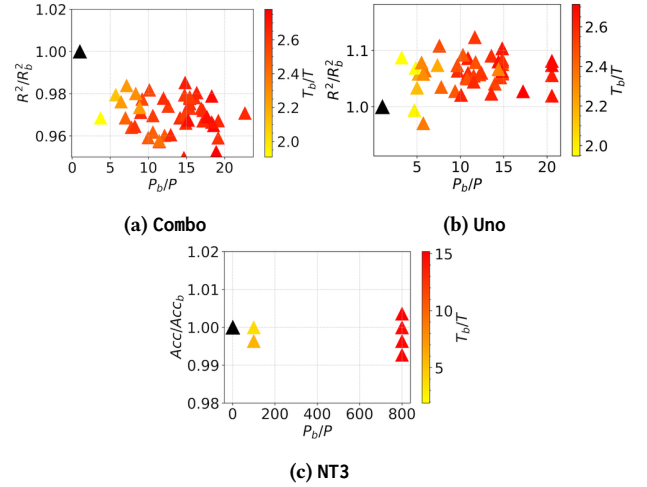


Figure 7: Post-training results on the top 50 A3C architectures from the small search space run on 256 nodes. Accuracy ratios ( $R^2/R_b^2$ ,  $Acc/Acc_b$ )  $> 1.0$  indicate a A3C-generated architecture outperforming the manually designed network. Trainable parameter ratios ( $P_b/P$ )  $> 1.0$  indicate that a A3C-generated architecture has fewer trainable parameters than the manually designed network. Training time ratios ( $T_b/T$ )  $> 1.0$  indicate that a A3C-generated architecture is faster than the manually designed network.

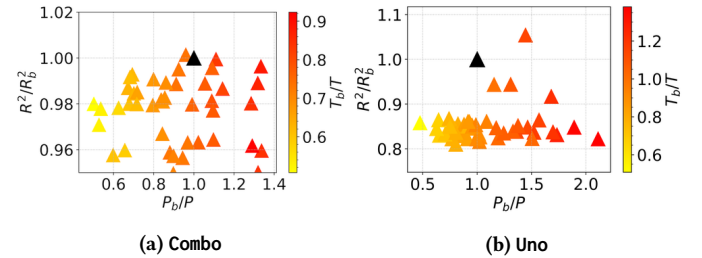


Figure 8: Post-training results on top-50 A3C architectures from the large search space run on 256 nodes

Figure 7 shows the post-training results of A3C on the 50 best architectures selected based on the estimated reward during the NAS. From the accuracy perspective, on Combo, five A3C-generated



architectures obtain  $R^2$  values that are competitive with the manually designed network ( $R^2/R_b^2 > 0.98$ ); on Uno, more than forty A3C-generated architectures obtain  $R^2$  values that are better than the manually designed network value ( $R^2/R_b^2 > 1.0$ ); on NT3, three A3C-generated architectures obtain accuracy values that are higher than that of the manually designed network ( $ACC/ACC_b > 1.0$ ). From the trainable parameter ratio viewpoint, A3C-generated architectures completely outperform the manually designed networks on all three data sets. On Combo, A3C-generated architectures have 5x to 15x fewer trainable parameters than the manually designed network has; on Uno this is between 2x to 20x; on NT3, A3C-generated architectures have up to 800x fewer parameters than the manually designed network has. The significant reduction in the number of trainable parameters is reflected in the training time ratio, where we observed up to 2.5x speedup for Combo and Uno and up to 20x for NT3.

Figure 8 shows the post-training results of A3C with the large search space on Combo and Uno. On Combo, use of the large search space allowed A3C to generate a number of architectures with accuracy values higher than those generated with the small search space. Among them, five architectures obtained  $R^2/R_b^2 > 0.99$ ; one was better than the manually designed network. The large search space increases the number of training parameters and training time significantly. Nevertheless, on Uno, we found that the larger search space decreases the accuracy values significantly, which can be attributed to the overparameterization given the relatively small amount of data, and additional improvement in accuracy was not observed after a certain number of epochs.

### 5.3 Scaling A3C on Combo with large search space

In this section, we demonstrate that increasing the number of agents and keeping the number of workers to a smaller value in A3C result in better scalability and improvement in accuracy.

We ran A3C on Combo with the large search space on 512 and 1,024 KNL nodes.<sup>3</sup> We studied two approaches to scaling. In the first approach, called *worker scaling*, we fixed the number of agents at 21 and varied the number of workers per agent. For 512 and 1,024 nodes, we tested 23 and 47 workers per agent, respectively. In the second approach, called *agent scaling*, we fixed the number of workers per agent at 11 and increased the number of agents. For 512 and 1,024 nodes, we used 42 and 85 agents, respectively.

The utilization of A3C is shown in Fig. 9. We observe that scaling the number of agents is more efficient than is scaling the number of workers per agent. In particular, the utilization values of agent scaling, 512(a) and 1,024(a), are similar to those measured at 256 nodes; there is no significant loss in utilization by going to higher node counts. On the other hand, utilization suffers as the number of workers per agent is increased. The reason is that the worker evaluations are batch synchronous and the increase in workers results in an increase in the number of idle nodes within a batch. The decreasing overall trend in utilization can be attributed to the increased cache effect.

Figure 10 shows the post-training results of the 50 best architectures from the 512 and 1,024-node agent scaling experiments.

<sup>3</sup>We did not use more than 1,024 nodes in this experiment because of a system policy limiting the total number of concurrent application launches to 1,000.

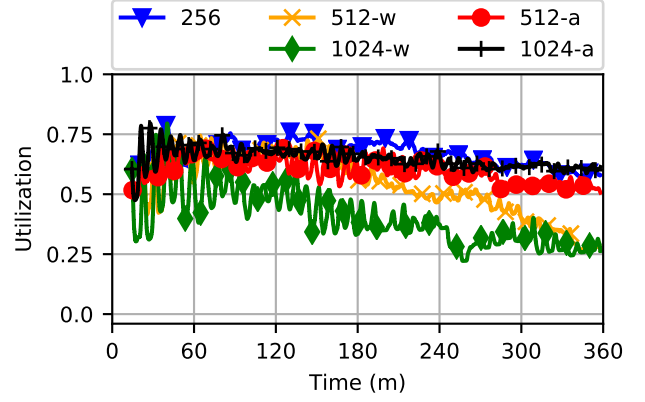


Figure 9: Utilization of A3C on Combo with the large search space run on 512 and 1,024 nodes with agent and worker scaling; 256 nodes are used as reference.

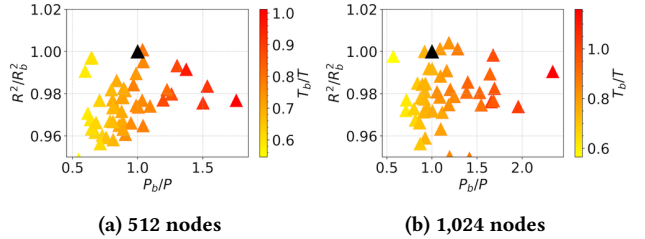


Figure 10: Post-training results of A3C on Combo with the large search space run on 512 and 1,024 nodes with agent scaling

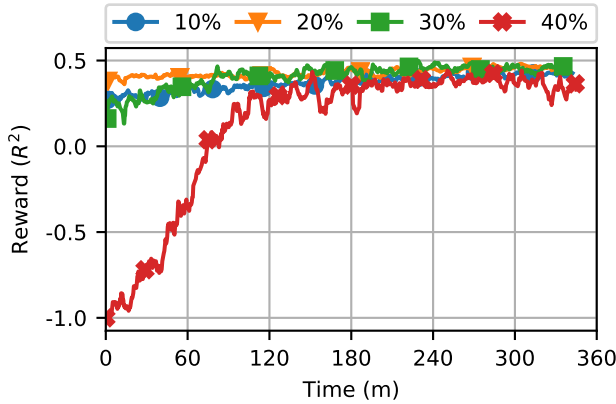
Compared with the 256-node experimental results (see Fig. 8a), both the 512-node and 1,024-node experiments result in network architectures that have better accuracy, fewer trainable parameters, and lower training time. In particular, scaling on 1,024 nodes results in nine networks with  $R^2/R_b^2 > 0.99$ ; among them four networks were better than the manually designed network. These networks have as few as 50% fewer parameters than the manually designed network has. An increase in the number of nodes and agents results in higher exploration of the architecture space, which eventually increases the chances of finding a diverse range of architectures without sacrificing accuracy.

### 5.4 Impact of fidelity in reward estimation

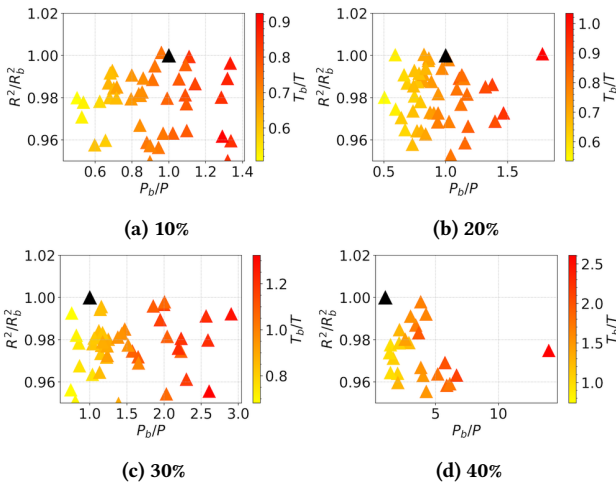
Here, we show that, at scale, the fidelity of the reward estimation affects agent learning behavior in different ways and can generate diverse architectures.

We analyzed the impact of fidelity in the reward estimation strategy by increasing the training data size in A3C from the default of 10% to 20%, 30%, and 40% on Combo. We ran the experiments on 256 nodes and used the default values for the training epochs and the timeout.

Figure 11 shows the search trajectory of A3C. We can observe that on 10%, 20%, and 30% training data, A3C generates architectures with high rewards within 80 minutes. With 40% training data, the improvement in the reward is slow. The reason is that the large



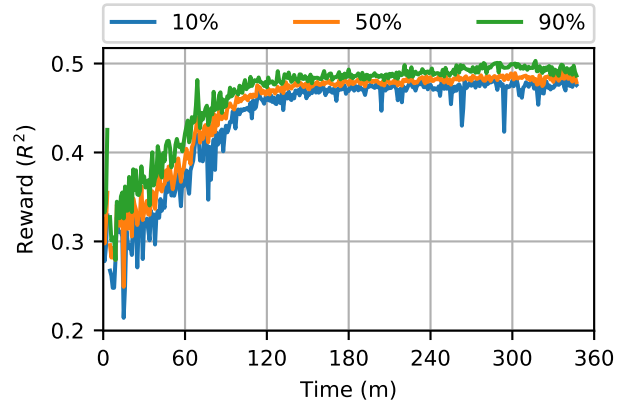
**Figure 11: Rewards over time obtained by A3C for Combo with the large search space on 256 nodes with different training data sizes**



**Figure 12: Post-training results of A3C on Combo with a large search space run on 256 nodes**

number of architectures generated by A3C cannot complete training before the timeout. Consequently, it takes 80 minutes to reach reward values greater than 0. Nevertheless, it slowly learns to generate architectures that can be trained within the timeout—within 160 minutes, A3C with 40% of the training data reaches the reward values found by A3C with less training data.

The post-training results are shown in Fig. 12. As we increase the training data size in the reward estimation, we can observe a trend in which the best architectures generated by A3C have fewer trainable parameters and shorter post-training time. In the 10% case, the training time in reward estimation is not a bottleneck. Consequently, the agents generate networks with fewer trainable parameters to increase the reward, and the post-training time of the best architectures often exceeds that of the manually designed network. Increasing the training data size to 20% results in the best architectures that have smaller trainable parameters and longer post-training time than the manually designed network has. We found that the agents that can achieve faster rewards by using fewer parameters update the parameter server and bias the search. In the



**Figure 13: Statistics of the A3C search trajectory computed over 10 replications on Combo with small search space**

30% case, the training time affects the best architectures. Consequently, several best architectures have fewer trainable parameters and shorter post-training time than the 10% and 20% cases have. In the 40% case, the training time in the reward estimation becomes a bottleneck. As a result, the agent learns to maximize the reward by generating architectures with faster training time in the reward estimation by using fewer trainable parameters.

## 5.5 Impact of randomness in A3C

The A3C strategy that we used in NAS is a randomized method. The randomness stems from several sources, including random weight initialization of the neural networks, asynchronicity, and stochastic gradient descent for reward estimation. Here, we analyze the impact of randomness on the search trajectory of A3C. We repeated A3C 10 times on the Combo benchmark with the small search space. The results are shown in Fig. 13. Given a time stamp, we compute 10%, 50% (median), and 90% quantiles from 10 values—this removes both the best and worst values (outliers) for a given time stamp. In the beginning of the search, the differences in the search trajectory of A3C are noticeable, where the quantiles of the reward range between 0.2 and 0.4. Nevertheless, the variations become smaller as the search progresses. At the end of the search, all the quantile values are close to 0.5, indicating that the search trajectories of different replications are similar and the randomness does not have a significant impact on the search trajectory of A3C.

## 5.6 Summary of the best A3C architectures

Table 1 summarizes the results of the best A3C-generated architectures with respect to the manually designed networks on three data sets. On Combo, the accuracy of the best A3C-generated architecture is slightly better than that of the manually designed network. However, it has 7.3x fewer trainable parameters and 2.5x faster training time. On Uno, the best A3C-generated architecture outperforms the manually designed network with respect to all three factors: trainable parameters, training time, and accuracy. The best NAS architecture obtained an  $R^2$  value of 0.729 with 11.5x fewer trainable parameters and 2.5x faster training time than that of the manually

**Table 1: Summary of best architectures found by A3C**

	Trainable Parameters	Training Time (s)	$R^2$ or $ACC$
Combo			
manually designed	13,772,001	705.26	0.926
<b>A3C-best</b>	<b>1,883,301</b>	<b>283.00</b>	<b>0.93</b>
Uno			
manually designed	19,274,001	164.94	0.649
<b>A3C-best</b>	<b>1,670,401</b>	<b>63.53</b>	<b>0.729</b>
NT3			
manually designed	96,777,878	247.63	0.986
<b>A3C-best</b>	<b>120,968</b>	<b>16.65</b>	<b>0.989</b>

designed network. On NT3, the best A3C-generated network obtained 98% accuracy (similar to the manually designed network), but it has 800x fewer trainable parameters and 14.8x faster training time. Moreover, whereas the manual design of networks for these data sets took days to weeks, a NAS run with our scalable open-source software will take only six hours of wall-clock time for similar data sets.

## 6 RELATED WORK

We refer the reader to [6, 17, 32] for a detailed exposition on NAS related work. Here, to highlight our contributions, we discuss related work across five dimensions.

**Application:** A majority of the NAS literature focuses on the automatic construction of CNNs for image classification tasks applied primarily to benchmark data sets such as CIFAR and ImageNet. This is followed by recurrent neural nets for text classification tasks on benchmark data sets. Application of new domain applications beyond standard benchmark tasks is still in its infancy [32]. Recent examples include language modeling [88], music modeling [62], image restoration [74], and network compression [12] tasks. While there exist several prior works on DL for cancer data, we believe our work is the first application of NAS for cancer predictive modeling tasks.

**Search space:** Two key elements define the search space: primitives and the architecture template. Existing works have used convolution with different numbers of filters, kernel size, and strides; pooling such as average and maximum depthwise separable convolutions; dilated convolutions; RNN/LSTM cells; and a layer with a number of units for fully connected networks. Motivated by the requirements of the cancer data, we introduced new types of primitives such as multiple input layers, variable nodes, fixed nodes, and mirror nodes, which allow us to explicitly incorporate cancer domain knowledge in the search space. Existing architecture templates range from simple chain-structured skip connections to cell-based architectures [32]. The NAS search space that we designed is not specific to a single template, and it can handle all three template types. More important, we can define templates that enable a search over cells specific to the cancer data.

**Search method:** Different search methods have been used to navigate the search space, such as random search [20, 45, 69], Bayesian optimization [21, 30, 41, 42, 65, 70, 80, 82], evolutionary methods [25, 28, 46, 47, 50, 53, 59, 72, 74, 78, 86], gradient-based methods, and reinforcement learning [12, 13, 16, 18, 19, 28, 36, 72,

85, 88]. Currently, there is no clear winner (for example, see Real et al. [63]); most likely a single method may never outperform all other methods on all data sets under all possible settings (as a consequence of the "no free lunch" theorem). Therefore, we need to understand the strengths and limitations of these methods based on the data set. We compared A3C and A2C methods at scale and analyzed their convergence on nontext and nonimage data sets. We showed that A3C, despite gradient staleness due to asynchronous gradient update, can find high-performing DNNs in a short computation time. We evaluated the efficacy of the RL-based NAS at scale with respect to accuracy, training time, parameters, and reward estimation fidelity.

**RL-based NAS scalability:** In [88], RL-based NAS was scaled on 800 GPUs using 20 parameter servers, 100 agents, and 8 workers per agent. This approach was run for three to four weeks. In [89], a single RL agent generated 450 networks and used 450 GPUs for concurrent training across four days. In both these works, the primary goal was to demonstrate that the NAS can outperform manually designed networks on image and text classification tasks. We demonstrated RL-based NAS experiments on up to 1,024 KNL nodes and for a much shorter wall-clock time of 6 hours on cancer data.

**Open source software:** AutoKeras [39] is an open source automated machine learning package that uses Bayesian optimization and network morphism for NAS. The scalability of the package is limited because it is designed to run on single node with multiple GPUs that can evaluate few architectures in parallel. Microsoft's Neural Network Intelligence [7] is a open source AutoML package designed primarily to tune the hyperparameters of a fixed DNN architecture by using different types of search methods; it lacks capabilities for architecture templates. Ray [55] is an open source high-performance distributed execution framework that has modules for RL and hyperparameter search but does not have support for NAS. AMLA [40] is a framework for implementing and deploying AutoML neural network generation algorithms. Nevertheless, it has not been demonstrated on benchmark applications or at scale. TPOT [58] optimizes scikit-learn [60], a library of classical machine learning algorithms, using evolutionary algorithms, but it does not have support for NAS. Our package differs from the existing ones with respect to customized NAS search space for cancer data and scalability.

## 7 CONCLUSION AND FUTURE WORK

We developed scalable RL-based NAS to automate DNN model development for a class of cancer data. We designed a NAS search space that takes into account characteristics specific to nonimage and nontext cancer data. We scaled the proximal policy optimization, a state-of-the-art RL approach, using a manager-worker approach on up to 1,024 Intel Knights Landing nodes and evaluated its efficacy using cancer DL benchmarks. We demonstrated the efficacy of this method at scale and showed that the asynchronous actor critic method (A3C) outperforms its synchronous and random variants. The results showed that A3C can discover architectures that have significantly fewer training parameters, shorter training time, and accuracy similar to or higher than those of manually designed architectures. We experimented with the volume of training data in reward estimation, analyzed the impact of fidelity in reward

estimation on the agent learning capabilities, and showed that it can be used to discover different types of network architectures.

Our future work will include applying NAS on a broader class of cancer data, conducting an architecture search for transformer and attention-based networks [52, 79], adapting NAS for multiple objectives, developing adaptive reward estimation approaches, developing multiparameter servers to improve scalability, integrating hyperparameter search approaches, and comparing our approach with extremely scalable evolutionary approaches such as MEN-NDL [59, 86] and Bayesian optimization methods [71]. Our NAS framework is designed to be flexible for developing surrogate DNN models for tabular data. We will explore NAS for reduced-order modeling in scientific application areas such as climate and fluid dynamics simulations.

NAS has the potential to accelerate cancer deep learning research. A scalable open-source NAS package such as ours can allow cancer researchers to automate neural architecture discovery using HPC resources and to experiment with diverse DNN architectures. This will be of paramount importance in order to tackle cancer as we incorporate more diverse and complex data sets.

## ACKNOWLEDGMENT

This material is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Advanced Scientific Computing Research, under Contract DE-AC02-06CH11357. This research used resources of the Argonne Leadership Computing Facility, which is a DOE Office of Science User Facility.

## REFERENCES

- [1] [n. d.]. AutoML Workshops. <https://www.ml4aad.org/automl/>
- [2] [n. d.]. CANDLE Exascale Computing Program Application. <https://github.com/ECP-CANDLE/Benchmarks>
- [3] [n. d.]. Combo Benchmark. <https://github.com/ECP-CANDLE/Benchmarks/tree/master/Pilot1/Combo>
- [4] [n. d.]. An End-to-End AutoML Solution for Tabular Data at KaggleDays. <https://ai.googleblog.com/2019/05/an-end-to-end-automl-solution-for.html>
- [5] [n. d.]. Exascale Deep Learning and Simulation Enabled Precision Medicine for Cancer. <https://candle.cels.anl.gov>
- [6] [n. d.]. Literature on Neural Architecture Search. <https://www.ml4aad.org/automl/literature-on-neural-architecture-search/>
- [7] [n. d.]. Neural Network Intelligence. <https://github.com/Microsoft/nni>
- [8] [n. d.]. NT3 Benchmark. <https://github.com/ECP-CANDLE/Benchmarks/tree/master/Pilot1/NT3>
- [9] [n. d.]. Uno Benchmark. <https://github.com/ECP-CANDLE/Benchmarks/tree/master/Pilot1/Uno>
- [10] [n. d.]. World Health Organization: Cancer key facts. <https://www.who.int/news-room/fact-sheets/detail/cancer>
- [11] Martin Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. TensorFlow: A system for large-scale machine learning. In *OSDI*, Vol. 16. 265–283.
- [12] Anubhav Ashok, Nicholas Rhinehart, Fares Beainy, and Kris M Kitani. 2017. N2n learning: Network to network compression via policy gradient reinforcement learning. *arXiv preprint 1709.06030* (2017).
- [13] Bowen Baker, Otkrist Gupta, Nikhil Naik, and Ramesh Raskar. 2016. Designing neural network architectures using reinforcement learning. *arXiv preprint 1611.02167* (2016).
- [14] P. Balaprakash, R. Egele, M. Salim, V. Vishwanath, and S. M. Wild. 2018. DeepHyper: Scalable automated machine learning package. <https://github.com/deephyper/deephyper>
- [15] Prasanna Balaprakash, Michael Salim, Thomas Uram, Venkat Vishwanath, and Stefan Wild. 2018. DeepHyper: Asynchronous Hyperparameter Search for Deep Neural Networks. In *HiPC 2018: 25th edition of the IEEE International Conference on High Performance Computing, Data, and Analytics*.
- [16] Irwan Bello, Barret Zoph, Vijay Vasudevan, and Quoc V Le. 2017. Neural optimizer search with reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, Vol. 70. JMLR. org, 459–468.
- [17] T. Ben-Nun and T. Hoefler. 2018. Demystifying Parallel and Distributed Deep Learning: An In-Depth Concurrency Analysis. *CoRR* abs/1802.09941 (Feb. 2018).
- [18] Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. 2018. Understanding and simplifying one-shot architecture search. In *International Conference on Machine Learning*. 549–558.
- [19] Gabriel Bender, Pieter-Jan Kindermans, Barret Zoph, Vijay Vasudevan, and Quoc Le. 2018. Understanding and Simplifying One-Shot Architecture Search. In *Proceedings of the 35th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Jennifer Dy and Andreas Krause (Eds.), Vol. 80. PMLR, 550–559.
- [20] James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* 13, Feb (2012), 281–305.
- [21] James Bergstra, Dan Yamins, and David D Cox. 2013. Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. In *Proceedings of the 12th Python in Science Conference*. 13–20.
- [22] James Bergstra, Daniel Yamins, and David Daniel Cox. 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. (2013).
- [23] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba. 2016. OpenAI Gym. *arXiv:arXiv:1606.01540*
- [24] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. 2017. Geometric deep learning: Going beyond euclidean data. *IEEE Signal Processing Magazine* 34, 4 (2017), 18–42.
- [25] Yukang Chen, Qian Zhang, Chang Huang, Lisen Mu, Gaofeng Meng, and Xing-gang Wang. 2018. Reinforced Evolutionary Neural Architecture Search. *arXiv preprint 1808.00193* (2018).
- [26] François Chollet et al. 2017. Keras (2015).
- [27] Patryk Chrabaszcz, Ilya Loshchilov, and Frank Hutter. 2017. A downsampled variant of ImageNet as an alternative to the CIFAR datasets. *arXiv preprint 1707.08819* (2017).
- [28] Xiangxiang Chu, Bo Zhang, Hailong Ma, Ruijun Xu, Jixiang Li, and Qingyuan Li. 2019. Fast, Accurate and Lightweight Super-Resolution with Neural Architecture Search. *arXiv preprint 1901.07261* (2019).
- [29] Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, Yuhuai Wu, and Peter Zhokhov. 2017. OpenAI baselines. <https://github.com/openai/baselines>.
- [30] Georgi Dikov, Patrick van der Smagt, and Justin Bayer. 2019. Bayesian Learning of Neural Network Architectures. *arXiv preprint 1901.04436* (2019).
- [31] Jesse R Dixon, Jie Xu, Vishnu Dileep, Ye Zhan, Fan Song, Victoria T Le, Galip Gürkan Yardımcı, Abhijit Chakraborty, Darrin V Bann, Yanli Wang, et al. 2018. Integrative detection and analysis of structural variation in cancer genomes. *Nature Genetics* 50, 10 (2018), 1388.
- [32] Thomas Elsken, Jan Hendrik Metzen, and Frank Hutter. 2018. Neural architecture search: A survey. *arXiv preprint 1808.05377* (2018).
- [33] Manuel Fernández-Delgado, Eva Cernadas, Senén Barro, and Dinani Amorim. 2014. Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research* 15, 1 (2014), 3133–3181.
- [34] Dario Floreano, Peter Dür, and Claudio Mattiussi. 2008. Neuroevolution: From architectures to learning. *Evolutionary Intelligence* 1, 1 (2008), 47–62.
- [35] Ivo Grondman, Lucian Busoniu, Gabriel AD Lopes, and Robert Babuska. 2012. A survey of actor-critic reinforcement learning: SStandard and natural policy gradients. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 1291–1307.
- [36] Minghao Guo, Zhao Zhong, Wei Wu, Dahua Lin, and Junjie Yan. 2018. IRLAS: Inverse Reinforcement Learning for Architecture Search. *arXiv preprint 1812.05285* (2018).
- [37] F. Hutter, L. Kotthoff, and J. Vanschoren (Eds.). 2019. *Automated Machine Learning: Methods, Systems, Challenges*. Springer International Publishing.
- [38] Max Jaderberg, Valentin Dalibard, Simon Osindero, Wojciech M Czarnecki, Jeff Donahue, Ali Razavi, Oriol Vinyals, Tim Green, Iain Dunning, Karen Simonyan, et al. 2017. Population Based Training of Neural Networks. *arXiv preprint 1711.09846* (2017).
- [39] Haifeng Jin, Qingquan Song, and Xia Hu. 2018. Efficient neural architecture search with network morphism. *arXiv preprint 1806.10282* (2018).
- [40] Purushotham Kamath, Abhishek Singh, and Debo Dutta. [n. d.]. AMLA: An AutoML framework for Neural Network Design. ([n. d.]).
- [41] Kirthevasan Kandasamy, Willie Neiswanger, Jeff Schneider, Barnabas Poczos, and Eric P Xing. 2018. Neural architecture search with Bayesian optimisation and optimal transport. In *Advances in Neural Information Processing Systems*. 2020–2029.
- [42] A. Klein, S. Falkner, N. Mansur, and F. Hutter. 2017. RoBO: A flexible and robust Bayesian Optimization framework in Python. In *NeurIPS 2017 Bayesian Optimization Workshop*.
- [43] Aaron Klein, Stefan Falkner, Jost Tobias Springenberg, and Frank Hutter. 2016. Learning curve prediction with Bayesian neural networks. (2016).

- [44] Lisha Li, Kevin Jamieson, Giulia DeSalvo, Afshin Rostamizadeh, and Ameet Talwalkar. 2016. Hyperband: Bandit-based configuration evaluation for hyperparameter optimization. (2016).
- [45] Liam Li and Ameet Talwalkar. 2019. Random Search and Reproducibility for Neural Architecture Search. *arXiv preprint 1902.07638* (2019).
- [46] Jason Liang, Elliot Meyerson, Babak Hodjat, Dan Fink, Karl Mutch, and Risto Miikkulainen. 2019. Evolutionary Neural AutoML for Deep Learning. *arXiv preprint 1902.06827* (2019).
- [47] Jason Liang, Elliot Meyerson, and Risto Miikkulainen. 2018. Evolutionary architecture search for deep multitask networks. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 466–473.
- [48] Shaoping Ling, Zheng Hu, Zuyu Yang, Fang Yang, Yawei Li, Pei Lin, Ke Chen, Lili Dong, Lihua Cao, Yong Tao, et al. 2015. Extremely high genetic diversity in a single tumor points to prevalence of non-Darwinian cell evolution. *Proceedings of the National Academy of Sciences* 112, 47 (2015), E6496–E6505.
- [49] Chenxi Liu, Barret Zoph, Jonathon Shlens, Wei Hua, Li-Jia Li, Li-Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. 2017. Progressive neural architecture search. *arXiv preprint 1712.00559* (2017).
- [50] Pablo Ribalta Lorenzo and Jakub Nalepa. 2018. Memetic evolution of deep neural networks. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 505–512.
- [51] Pablo Ribalta Lorenzo, Jakub Nalepa, Luciano Sanchez Ramos, and José Ranilla Pastor. 2017. Hyper-parameter selection in deep neural networks using parallel particle swarm optimization. In *Proceedings of the Genetic and Evolutionary Computation Conference Companion*. ACM, 1864–1871.
- [52] Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. *arXiv preprint 1508.04025* (2015).
- [53] Krzysztof Maziarsz, Andrey Khorlin, Quentin de Laroussilhe, and Andrea Gesmundo. 2018. Evolutionary-Neural Hybrid Agents for Architecture Search. *arXiv preprint 1811.09828* (2018).
- [54] Risto Miikkulainen, Jason Liang, Elliot Meyerson, Aditya Rawal, Dan Fink, Olivier Francon, Bala Raju, Arshak Navruzyan, Nigel Duffy, and Babak Hodjat. 2017. Evolving deep neural networks. *arXiv preprint 1703.00548* (2017).
- [55] Philipp Moritz, Robert Nishihara, Stephanie Wang, Alexey Tumanov, Richard Liaw, Eric Liang, Melih Elibol, Zongheng Yang, William Paul, Michael I Jordan, et al. 2018. Ray: A Distributed Framework for Emerging {AI} Applications. In *13th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 18)*. 561–577.
- [56] Renato Negrinho and Geoff Gordon. 2017. Deeparchitect: Automatically designing and training deep architectures. *arXiv preprint 1704.08792* (2017).
- [57] Michail Nikolaou, Athanasia Pavlopoulou, Alexandros G Georgakilas, and Efthymios Kyrodimos. 2018. The challenge of drug resistance in cancer treatment: A current overview. *Clinical & Experimental Metastasis* 35, 4 (2018), 309–318.
- [58] Randal S. Olson, Nathan Bartley, Ryan J. Urbanowicz, and Jason H. Moore. 2016. Evaluation of a Tree-based Pipeline Optimization Tool for Automating Data Science. In *Proceedings of the Genetic and Evolutionary Computation Conference 2016 (GECCO '16)*. ACM, New York, NY, USA, 485–492. <https://doi.org/10.1145/2908812.2908918>
- [59] Robert M Patton, J Travis Johnston, Steven R Young, Catherine D Schuman, Don D March, Thomas E Potok, Derek C Rose, Seung-Hwan Lim, Thomas P Karnowski, Maxim A Ziatdinov, et al. 2018. 167-PFlops deep learning for electron microscopy: From learning physics to atomic manipulation. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage, and Analysis*. IEEE Press, 50.
- [60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2011. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* 12 (2011), 2825–2830.
- [61] Hieu Pham, Melody Y Guan, Barret Zoph, Quoc V Le, and Jeff Dean. 2018. Efficient Neural Architecture Search via Parameter Sharing. *arXiv preprint 1802.03268* (2018).
- [62] Aditya Rawal and Risto Miikkulainen. 2018. From nodes to networks: Evolving recurrent neural networks. *arXiv preprint 1803.04439* (2018).
- [63] Esteban Real, Alok Aggarwal, Yanping Huang, and Quoc V Le. 2018. Regularized evolution for image classifier architecture search. *arXiv preprint 1802.01548* (2018).
- [64] Ed Reznik, Augustin Luna, Bülent Arman Aksoy, Eric Minwei Liu, Konnor La, Irina Ostrovskaya, Chad J Creighton, A Ari Hakimi, and Chris Sander. 2018. A landscape of metabolic variation across tumor types. *Cell Systems* 6, 3 (2018), 301–313.
- [65] Raanan Y Rohekar, Shami Nisimov, Yaniv Gurwicz, Guy Koren, and Gal Novik. 2018. Constructing Deep Neural Networks by Bayesian Network Structure Learning. In *Advances in Neural Information Processing Systems*. 3051–3062.
- [66] Michael A. Salim, Thomas D. Uram, Taylor Childers, Prasanna Balaprakash, Venkatram Vishwanath, and Michael E. Papka. 2018. Balsam: Automated Scheduling and Execution of Dynamic, Data-Intensive Workflows. In *PyHPC 2018: Proceedings of the 8th Workshop on Python for High-Performance and Scientific Computing*.
- [67] Francisco Sanchez-Vega, Marco Mina, Joshua Armenia, Walid K Chatila, Augustin Luna, Konnor C La, Sofia Dimitriadou, David L Liu, Havish S Kantheti, Sadeq Saghafinia, et al. 2018. Oncogenic signaling pathways in the cancer genome atlas. *Cell* 173, 2 (2018), 321–337.
- [68] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
- [69] Christian Sciuto, Kaicheng Yu, Martin Jaggi, Claudiu Musat, and Mathieu Salzmann. 2019. Evaluating the Search Phase of Neural Architecture Search. *arXiv preprint 1902.08142* (2019).
- [70] Jasper Snoek, Hugo Larochelle, and Ryan P Adams. 2012. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems*. 2951–2959.
- [71] Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Mostofa Patwary, Mr Prabhat, and Ryan Adams. 2015. Scalable Bayesian optimization using deep neural networks. In *International Conference on Machine Learning*. 2171–2180.
- [72] Kenneth O. Stanley, Jeff Clune, Joel Lehman, and Risto Miikkulainen. 2019. Designing neural networks through neuroevolution. *Nature Machine Intelligence* 1, 1 (2019), 24–35. <https://doi.org/10.1038/s42256-018-0006-z>
- [73] Kenneth O Stanley, David B D’Ambrosio, and Jason Gauci. 2009. A hypercube-based encoding for evolving large-scale neural networks. *Artificial Life* 15, 2 (2009), 185–212.
- [74] Masanori Suganuma, Mete Ozay, and Takayuki Okatani. 2018. Exploiting the potential of standard convolutional autoencoders for image restoration by evolutionary search. *arXiv preprint 1803.00370* (2018).
- [75] Masanori Suganuma, Shinichi Shirakawa, and Tomoharu Nagao. 2017. A genetic programming approach to designing convolutional neural network architectures. In *Proceedings of the Genetic and Evolutionary Computation Conference*. ACM, 497–504.
- [76] Richard S. Sutton and Andrew G. Barto. 2018. *Reinforcement learning: An introduction*. MIT Press.
- [77] Richard S. Sutton, David A. McAllester, Satinder P. Singh, and Yishay Mansour. 2000. Policy gradient methods for reinforcement learning with function approximation. In *Advances in neural information processing systems*. 1057–1063.
- [78] Gerard Jacques van Wyk and Anna Sergeevna Bosman. 2018. Evolutionary Neural Architecture Search for Image Restoration. *arXiv preprint 1812.05866* (2018).
- [79] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in Neural Information Processing Systems*. 5998–6008.
- [80] Jiazhao Wang, Jason Xu, and Xuejun Wang. 2018. Combination of hyperband and Bayesian optimization for hyperparameter optimization in deep learning. *arXiv preprint 1801.01596* (2018).
- [81] Daan Wierstra, Faustino J Gomez, and Jürgen Schmidhuber. 2005. Modeling systems with internal state using Evolino. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*. ACM, 1795–1802.
- [82] Martin Wistuba. 2017. Bayesian Optimization Combined with Incremental Evaluation for Neural Network Architecture Optimization. *Proceedings of the International Workshop on Automatic Selection, Configuration and Composition of Machine Learning Algorithms* (2017).
- [83] Justin M. Wozniak, Rajeev Jain, Prasanna Balaprakash, Jonathan Ozik, Nicholas T. Collier, John Bauer, Fangfang Xia, Thomas S. Brettin, Rick Stevens, Jamaludin Mohd-Yusof, Cristina Garcia-Cardona, Brian Van Essen, and Matthew Baughman. 2018. CANDLE/Supervisor: A workflow framework for machine learning applied to cancer research. *BMC Bioinformatics* 19-S, 18 (2018), 59–69. <https://doi.org/10.1186/s12859-018-2508-4>
- [84] Fangfang Xia, Maulik Shukla, Thomas Brettin, Cristina Garcia-Cardona, Judith Cohn, Jonathan E Allen, Sergei Maslov, Susan L. Holbeck, James H Doroshov, Yvonne A Evrard, et al. 2018. Predicting tumor cell line response to drug pairs with deep learning. *BMC Bioinformatics* 19, 18 (2018), 486.
- [85] Sirui Xie, Hehui Zheng, Chunxiao Liu, and Liang Lin. 2018. SNAS: Stochastic neural architecture search. *arXiv preprint 1812.09926* (2018).
- [86] Steven R Young, Derek C Rose, Travis Johnston, William T Heller, Thomas P Karnowski, Thomas E Potok, Robert M Patton, Gabriel Perdue, and Jonathan Miller. 2017. Evolving deep networks using HPC. In *Proceedings of the Machine Learning on HPC Environments*. ACM.
- [87] Arber Zela, Aaron Klein, Stefan Falkner, and Frank Hutter. 2018. Towards automated deep learning: Efficient joint neural architecture and hyperparameter search. *arXiv preprint 1807.06906* (2018).
- [88] Barret Zoph and Quoc V Le. 2016. Neural architecture search with reinforcement learning. *arXiv preprint 1611.01578* (2016).
- [89] Barret Zoph, Vijay Vasudevan, Jonathon Shlens, and Quoc V Le. 2018. Learning transferable architectures for scalable image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 8697–8710.



The submitted manuscript has been created by UChicago Argonne, LLC, Operator of Argonne National Laboratory ("Argonne"). Argonne, a U.S. Department of Energy Office of Science laboratory, is operated under Contract No. DE-AC02-06CH11357. The U.S. Government retains for itself, and others acting on its behalf, a paid-up nonexclusive, irrevocable worldwide license in said article to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan. <http://energy.gov/downloads/doe-public-access-plan>