International Journal of Neural Systems, Vol. XX, No. X (2020) 1–16
© World Scientific Publishing Company

A Multiobjective Evolutionary Approach Based on Graph-in-graph for Neural Architecture Search of Convolutional Neural Networks

Yu Xue[†]
1. School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China
2. Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science and Technology, Nanjing, China
E-mail: xueyu@nuist.edu.cn

Pengcheng Jiang[†]
School of Computer and Software, Nanjing University of Information Science and Technology, Nanjing, China
E-mail: pcjiang@nuist.edu.cn

Ferrante Neri*
COL Laboratory, School of Computer Science, University of Nottingham, Nottingham, UK
E-mail:ferrante.neri@nottingham.ac.uk

Jiayu Liang
Tianjin Key Laboratory of Autonomous Intelligent Technology and System, Tiangong University, Tianjin, China
Email: yyliang2012@hotmail.com

With the development of deep learning, the design of an appropriate network structure becomes fundamental. In recent years, the successful practice of Neural Architecture Search (NAS) has indicated that an automated design of the network structure can efficiently replace the design performed by human experts. Most NAS algorithms make the assumption that the overall structure of the network is linear and focus solely on accuracy to assess the performance of candidate networks.
This paper introduces a novel NAS algorithm based on a multi-objective modeling of the network design problem to design accurate Convolutional Neural Networks (CNNs) with a small structure. The proposed algorithm makes use of a graph-based representation of the solutions which enables a high flexibility in the automatic design. Furthermore, the proposed algorithm includes novel ad-hoc crossover and mutation operators. We also propose a mechanism to accelerate the evaluation of the candidate solutions.
Experimental results demonstrate that the proposed NAS approach can design accurate neural networks with limited size.

Keywords: Deep Learning, Neural Architecture Search, Multi-objective Optimization, Genetic Algorithm.

1. Introduction

Convolution neural networks (CNNs) have achieved remarkable results in solving many problems, such as image classification [16] and image segmentation [40]. CNNs are very efficient at obtaining features from 1-D sequences of data, 2-D images,

†: These authors contributed equally to this work and should be considered *co-first* authors

and 3-D images. The features extracted from 1-D sequences of sound data by 1-D convolution neural networks can be used to extract voiceprint features [96; 48]. The features extracted from 2-D image data by convolution neural network can be used for image content recognition, prediction, and segmentation. The features in 3-D space in 3-D image data (mostly medical image data) can be extracted by 3-D convolution kernels, which is very useful in predicting diseases and identifying lesions [35; 34]. In addition, video data with time attributes can also be classified by 3-D convolutional neural networks [21; 73].

Among the plethora of real-world applications of CNNs, some modern examples representing the state-of-the-art in the field of neural systems are to analyse the electroencephalogram signals to diagnose seizures [2; 46; 42] or depression [3]. A neural system based on multiple CNNs is proposed in Ref.[52] to control epileptic seizures. Other studies propose CNNs to diagnose epilepsy in infants [7] and children [43] by classifying electroencephalogram signals. CNNs have been also successfully used to classify medical images to diagnose Parkinson's disease [54; 12] and detect pupils[72]. Another popular application domain for CNNs is civil engineering. Some examples of application include damage detection in concrete structures[39] and roads[53]. Some other examples are about vibration-based structural state identification[95] and effect of wind on structures[59]. In addition, CNNs can be combined with other technologies to be applied in more fields. In Ref. [56] CNNs are combined with Long Short Term Memory to accurately predict the remaining useful life of components, thus helping to make an optimal decision for maintenance management.

There have been many classical network structures, such as Alexnet [36], VGG [74], GoogLenet [84], Inception-V4 [83], Inception-Resnet [83], Resnet [31], Densenet [33], etc., which appear to perform well in image classification and image segmentation. However, due to high complexity, it is impractical to use these CNNs on mobile platforms since they would require an excessive amount of computational resources thus leading to an unreasonable waiting time, memory overflow, and high energy consumption. Therefore, some new lightweight network structures for mobile platforms have been proposed, such as MobileNet [68], ShuffelNet [51], MnasNet [85], EfficientNet [86], Xception [19], etc. All the network structures mentioned above are the result of (human) expert design.

In recent years, Neural Architecture Search (NAS) methods[17], that automatically search the network architectures, are progressively becoming more popular to design CNNs. Most NAS methods are to search the blocks or cells which are consist of convolution kernels with different sizes (such as 3×3, 5×5, etc.) and the position of pool layers [50; 82; 80]. Moreover, in MUXConv [49] and Shufflenet [51], it is pointed out that the generalization performance of the network can be improved by channel multiplexing, spatial multiplexing, and channel shuffling, and then the accuracy of recognition can be improved. The majority of the NAS methods in the literature perform the automatic design by using accuracy as the sole objective of the targets. However, operational efficiency is also an extremely important aspect of the functioning of the network, especially in mobile applications.

In order to simultaneously address accuracy and computational cost, unlike the other studies in the literature, we propose an encoding mechanism with multi-objective evaluate mechanism of the problem where besides the accuracy of the CNN also the number of network parameters is taken into consideration [65; 66; 75; 88; 69; 70].

Existing NAS methods design and limit the search space and search domain to reduce the time complexity of the optimization problem. An usual strategy consists of defining some building blocks which are defined by a human expert. This study proposes a graph-based flexible representation that supports a higher level of automatism of the design process. Furthermore, the proposed method relates to the concept of regularized evolutionary algorithm[67; 62] in that the approaches aim at reducing the computational overhead (e.g. memory employment) by performing an action on the optimization algorithm.

The remainder of this paper is organised in the following way. Section 2 provides the background about NAS methods, encoding mechanism and evaluation of candidate network architectures. Section 3 provides the details of the proposed NAS method. Section 4 provides the numerical results of this study.

2. Related Work: Neural Architecture Search

The majority of NAS methods can be categorised according to their search logic:

- Gradient-based methods [47; 92; 15];
- Reinforcement Learning (RL) [98; 28; 9];
- Evolutionary Algorithms (EA) [79; 50; 91; 63],[82].

This list does not mean to be exhaustive since other methods not belonging to any of the categories above exist, such as Monte Carlo Tree search[90]. The various NAS methods belonging to each category above present advantages and disadvantages. Specifically, RL-based algorithms require a large computational time to perform the automatic design, even on median-scale datasets, such as cifar10 and cifar100 [37]. Unlike RL-based algorithms, gradient-based algorithms are usually very fast. Besides, their search logic leads to obtaining a local optimum problem which may have a much poorer performance than the desired optimal design. Moreover, the gradient-based search algorithm needs to construct a super network in advance, which should contain as much search space as possible. The construction of this super network requires substantial human intervention of an expert, see Ref.[15; 25]. Although EAs are not theoretically guaranteed to converge to the global optimum of problem, they are able to overcome the local optima. Also, they do not require a super network. Thus, EAs are often considered a viable compromise for NAS since they are relatively fast and can be applied to NAS without human intervention or prior knowledge of the problem. One pioneering example is in Ref.[94]. It is worthwhile remarking that there exist other search strategies integrated in NAS methods such as Ref.[57], Ref.[18], and Ref.[55].

This paper focuses on EAs for NAS. In the following subsections, some context is provided around the two major challenges of this approach: encoding mechanism and evaluation of the candidate solutions.

2.1. *Encoding of NAS*

The encoding of candidate network architectures for NAS methods are broadly divided into two categories [38]: direct encoding and indirect encoding. Indirect encoding was often used in early works on NAS usually referred to as *Neuroevolution*, see Ref.[71], which is similar to NAS. Neuroevolution uses evolutionary computation to optimize the structure and parameters of neural networks at the same time [4; 27; 30; 26; 1], and many researchers still work on it[76; 77; 64; 32; 8]. However, due to the limitations of equipment at that time, the neuroevolution can only be performed on small networks. Furthermore, due to the very large number of parameters in fully connected networks, direct encoding cannot be used to represent the whole network. Therefore, a lot of effort is made to find simple ways (*i.e.*,indirect encoding) to represent the connections and weight parameters of neurons. Thus, indirect encoding is a popular strategy to simplify the search space. These search purposes determine that search space is difficult to represent with direct encoding, so indirect encoding is needed to simplify the encoding and early researchers used indirect encoding to represent individuals.

In recent years, most of the NAS studies have been conducted on neural networks that albeit complex, can be naturally schematised as interconnected blocks. This is the case, besides the CNNs, of Generative Adversarial Networks (GANs) [28], and Recurrent Neural Networks (RNNs) [47]. For networks of these types, direct encoding is an easy and natural option. For example, CNNs contain convolution blocks, pooling blocks, batch normalization operations, and sometimes activation functions. These blocks are often represented by a few parameters. Convolution blocks can be fully represented by the number of convolution cores, the size of the convolution cores, stride, padding, dilation and groups (in fact, some parameters can be directly ignored based on the actual search strategy and purpose). In most cases, pooling blocks, batch normalization operations and activation functions do not even require parameters for special representations, and they just need the position in the structure to represent the modules.

For each block's position in the structure, there exist two encoding mechanisms

- linear structure [81], that is the sequential (linear) arrangement of all blocks or units composed of blocks;
- graph structure [91], that is a planar (graph) arrangement of interconnected blocks.

Although formally a linear structure is a special graph structure (a sequence is a special graph), we emphasise the distinction since the two encoding mechanisms correspond to two significantly different implementations.

Adjacency matrices are better suited for dense graph structures [58], since sparse structures can waste a lot of space in adjacency matrices. Sparse structures are better represented by adjacency tables (or adjacency lists). While adjacency matrices are matrices of '0' and '1' to indicate connection or without connection between nodes, adjacency tables are lists that indicate the for each node which nodes are linked to it. The latter allows a compact representation of large sparse networks.

The main advantage of a linear structure is its simplicity compared to that of graph structure. Besides, linear structures cannot represent all the networks. In some cases, like the example in Fig. 1, a linear structure would yield an ambiguous representation of a neural network.

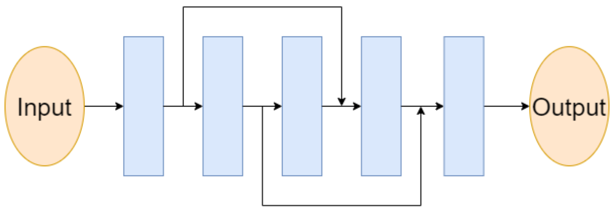

Fig. 1. An example of architecture that cannot be represented by a linear structure. Blue blocks are modules in CNNs. This architecture has two skip connections, so it can't be represented by a linear structure.

2.2. *Evaluation of NAS*

To evaluate a candidate structure, the general practice is to train the network and calculate its accuracy, see Ref.[80].

Since the training time of the network is very time-consuming, there are many ways to reduce the total time of the evaluation phase. There are two ways to reduce the total time: foresight and early closure. Foresight methods make use of models to predict the performance of the training network. Some researchers use the performance during training to predict the future performance. For example, MetaQNN [10] gives the first 25% of the historical data of the Stochastic Gradient Descent (SGD) training curve to the time series model for prediction and estimates the final accuracy of the network structure. Some researchers use other models, such as random forest, Bayes methods or other models to predict the possible representations of particular network architectures. The reason why they use this method is that the structures searched for by the same NAS method often have a great deal of similarity, and when encoded, it is possible to work out whether the network is good or not from the encoding directly. For instance, PNAS [45] uses the model to predict the top-1 accuracy of candidate networks. Ref.[78] proposes an end-to-end offline performance predictor based on the random forest to accelerate the evaluation.

Early closure is another way to reduce the total time of the evaluation phase. This type of approach reduces overall time through targeted evaluations. For example, many researchers used subsets of the dataset for training [99; 89], so that the time of training each network will decrease. Also, Ref.[89] uses a strategy to identify the required structure in advance. In ChamNet [20], only 300 high-accuracy (or other indicators) samples with different efficiency are selected for each training. Another approach is to keep the good structure and weight so that the new structure requires fewer times to train. There are three specific implementations of this approach: weight sharing, One-Shot method, and weight inheritance. The weight sharing method, which is mostly used in NAS based on gradient, makes use of shared weights from a super network to accelerate the training process, see Ref.[47; 92; 15; 98; 50]. The one-shot method consists of adding components to a small network or deleting components from a large network [41; 22; 29; 11]. The weight inheritance method is mostly used in NAS based on EAs [63; 23; 14; 24]. This method requires that the candidate networks of the entire search space have similar structures. Most of the network structures found by NAS based on EAs meet this condition.

Figure 2 illustrates the weight inheritance method. In the upper part of the figure, two parent solutions with a crossover point (indicated as a diamond) are depicted. The first parent solution is composed of the sequences G_1 and $G2$ (representing the network structure) with the corresponding weights W_1 and W_2. Analogously, the second parent solution is composed of G_3 and $G4$ with the weights W_3 and W_4. In the lower-left part of the figure, the standard crossover is illustrated. The sequences G_2 and

G_4 are swapped over and four sets of corresponding weights W_5, W_6, W_7 and W_8 are randomly initialized, thus generating new networks (indicated with a darker colour). In the lower right part of the figure, the weight inheritance method is illustrated. When the crossover occurs, the offspring solutions inherit the weights of the parent (the weights of that portion of the network). Thus, the first offspring solution is composed G_1 and G_4 with the weights W_1 and W_4 while the second solution is composed of G_3 and G_2 with the weights W_3 and W_2.

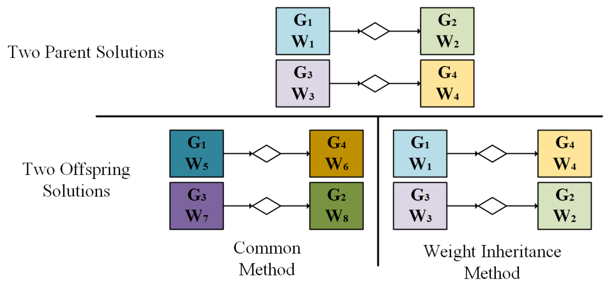

Fig. 2. Comparison between basic crossover (with random initialization of the weights) and crossover with weight inheritance method.

3. The Proposed Approach: MOGIG-Net

In this section, we introduce the framework of the proposed NAS algorithm, namely Multi-Objective Graph-in-graph Network (MOGIG-Net) whose flowchart is shown in Fig. 3.

This section firstly introduces the overall framework of the proposed algorithm and then describes the encoding mechanism, crossover, mutation, decoding method, evaluation, and environment selection in details.

3.1. *Overall Description of the MOGIG-Net Framework*

Fig. 4 displays the structure of the whole algorithm. First, the initial population is obtained through random initialization (line 1), and then the fitness evaluation of the initial population is calculated (line 2-3).

After the initialization, the algorithm makes use of generation cycles to process the population (line 4-15). New individuals are generated through crossover and mutation. The new individuals are selected for the survival of the fittest by evaluating the fitness values for each objective (line 12-13). Finally, individual sets with better performance on multiple objectives are obtained.

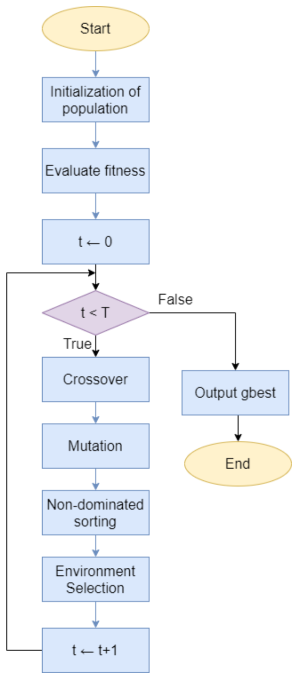

Fig. 3. Flowchart of the of the MOGIG-Net framework

Input: The population size p, the maximal generation number T, the crossover probability μ, the mutation probability ν, the maximal nodes number in each block (M_{min}, M_{max}), the maximal blocks number (N_{min}, N_{max}).
Output: Collection of individuals on the pareto frontier meeting the minimization goals.

```
1:  $P_0$ ←Initialize a population with the size of $p$ by using the proposed
    encoding strategy in Algorithm 6;
2:  Convert all genes in $P_0$ to models and evaluate the fitness of the mod-
    els by method 12, and record the fitness of each corresponding indi-
    vidual;
3:  Record the fingerprint and fitness value of each individual.
4:  $t \leftarrow 0$
5:  while $t < T$ do
6:    $Q \leftarrow \phi$
7:    if the length of $Q < P$ then
8:      Two individuals were randomly selected and will cross and mutate
        by the method of Algorithm 7 and 8, and then two offspring will be
        generated.
9:      Record the fingerprint of each offspring, and add the two offspring
        into population $Q$.
10:   end if
11:   $P_t \leftarrow P_t \cup Q$
12:   Convert genes to models and evaluate the fitness of individuals in
      $Q$ by method 12 to control sequence, and record the fitness of each
      individual;
13:   Sort $P_t$ by non-dominated sorting algorithm, and retain $P$ individ-
      uals with better performance, and delete other individuals with poor
      performance, then we will get $P_{t+1}$;
14:   $t \leftarrow t + 1$
15: end while
    Return: $P_{T-1}$
```

Fig. 4. Framework of the MOGIG-Net algorithm

In NAS problems, the evaluation phase is by far the computationally most expensive as it requires the training of the candidate network structure. In order to avoid the re-evaluation of the same architectures/structures, we keep an archive of visited solutions with their objective function value. If a solution is re-visited the archived objective function values are used.

3.2. *Encoding Mechanism of MoGIG-Net*

In this study, we use a graph structure to encode the architecture of the network. We propose the encoding of a CNN in a chromosome divided into blocks linked by separators. To understand the proposed encoding, let us remark that CNNs are composed of blocks, three of them being essential and named 1) convolution; 2) pooling; 3) fully connection. The chromosome representing the CNN is described as follows:

$$\mathrm{CB}_1\text{-}\mathrm{CB}_2\text{-}\cdots\text{-}\mathrm{CB}_n\text{-S-P}$$

where each CB_j is a convolution block, S represents the structure how the convolution block are interlinked and P describes the presence of pooling layers in the CNN.

The convolution block CB_j is a sequence of separators and binary numbers. The $'1'$ indicates a link between neurons while $'0'$ indicates the dismiss a connection. A convolution block containing m neurons is represented by a sequence of $\frac{m(m-1)}{2}$ binary numbers grouped in sub-blocks of $1, 2, \ldots m-1$ binary numbers. Each sub-block is separated by a dot. This sequence of binary numbers is the adjacency matrix associated with the convolution block. More specifically, each sub-block contains the information of a column of the adjacency matrix. Fig. 5 provides an example of the proposed encoding for $m = 5$. On the top of the figure, the encoding used in this study is shown. Below the chromosome, the corresponding adjacency matrix and table are displayed. It may be noticed that each block of the chromosome contains the columns of the adjacency matrix. At the bottom of Fig. 5, the corresponding network structure is represented. Also, like CB_j representing the connections in each block, the binary numbers in S is also the same representation as CB_j, and thus represents the connection between blocks.

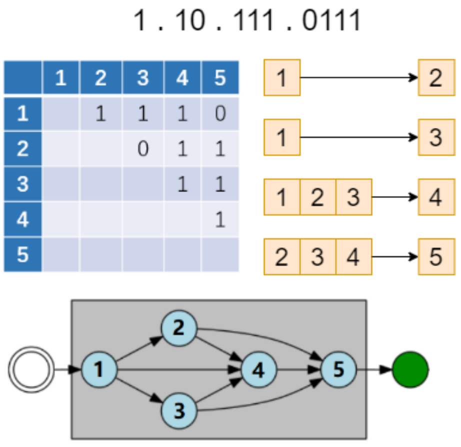

Fig. 5. Encoding of a convolution block CB_j (part of the chromosome) of the candidate CNN. The corresponding adjacency matrix and table are displayed as well as the graph of the encoded network. Convolution blocks formed by these binary blocks are the components of the CNN

The structure S is also a sequence of binary numbers which has $\frac{n(n-1)}{2}$ bits. The 1 indicates a link between two convolution blocks CB_i and CB_j while 0 indicates the dismiss of connections between blocks. The sequence S is also divided into sub-blocks composing the columns of the adjacency matrix that describes the topology of the interconnections among convolution blocks. The sequence P is composed of n binary numbers, one for each convolution block CB_j composing the CNN. The sequence P can be seen as a binary sequence Z where $z_j = 1$ represents the presence of pooling layers (P_j in Fig. 9) pointing to CB_{j+1} while $z_j = 0$ represents the absence of a pooling layer pointing to CB_{j+1}. The last binary number z_n indicates the presence or the absence of a pooling layer between CB_n and the fully connected layer FC.

Fig. 6 provides the implementation details of the encoding mechanism in the context of the initialization of the population to be processed by MOGIG-Net.

The chromosome code only contains the topological structure before the fully connected layer. The connection mode between the components of each individual is determined at the beginning of the algorithm (residual connection[31] and close connection[33]).

Let us indicate with α the maximum number of cells of and with β the maximum number of blocks of the CNN. The longest possible code to search for contains L bits, and L is calculated by the following formula.

$$L = \frac{\alpha(\alpha-1)}{2} + \frac{\beta(\beta-1)}{2} + \alpha$$

The search space contains up to 2^L possible candidate networks.

```
Input: The limit of nodes number in each block ($M_{min}, M_{max}$), the limit of blocks number ($N_{min}, N_{max}$), the maximal pooling blocks number $K$.
Output: One chromosome

1: Generate a random number $n, n \in (N_{min}, N_{max})$;
2: $flag \leftarrow 0$
3: $gene \leftarrow$ empty string
4: while $flag < n$ do
5:   Generate a random number $m, m \in (M_{min}, M_{max})$;
6:   Generate a random sequence $s$ of $\frac{m(m-1)}{2}$ binary numbers and allocate them with "." separators in CB$_{flag}$
7:   Make sure that the sequence represents a connected graph, see Fig. 10.
8:   $gene \leftarrow gene{+}s{+}'-'$, the '-' in this paper is the separator between genes of blocks and pools
9:   $flag \leftarrow flag + 1$
10: end while
11: Generate a random sequence $s$ of $\frac{n(n-1)}{2}$ numbers and allocate them with "-" separators in S.
12: Make sure that the series can represent an oriented connected graph, see Fig. 10.
13: Generate a random sequence $p$ of $n$ binary numbers and allocate them in P
14: $gene \leftarrow gene + s +' -' + p$
    Return: chromosome
```

Fig. 6. MOGIG-Net Encoding Strategy and Initialization

3.3. *Crossover and Mutation*

Due to the encoding mechanism proposed in this paper, an ad-hoc crossover operator is here proposed to ensure that the offspring solutions meaningfully represent structures of neural networks[13]. Furthermore, a meaningful chromosome must represent a connected graph.

The proposed crossover operator combines two chromosomes I and II by selecting randomly some blocks from the first and then filling the missing gaps with the genotype of the second to ensure that the offspring is meaningful. Fig. 7 provides the implementation details of the crossover.

For the chromosome I, two separators are randomly selected. Then the number of separators n between the two selected separators is calculated (line 6). Then, two separators in the chromosome II are selected while the number of separators between these two separators is ensured to be also n (line 7). Finally, the genes between the two separators are exchanged (line 8).

The mutation operation, outlined in Fig. 8, consists of the random flip from 0 to 1 or from 1 to 0 of a gene (except for the position of separator). Although the location of mutation changes is limited, the fact is that only small connection changes will affect all the input feature maps after this.

```
Input: Two parents, $p_1$ and $p_2$, probability of crossover $\mu \in (0, 1)$.
Output: Two offspring, $q_1$ and $q_2$.

1: Generate a random number $flag$;
2: if $flag > \mu$ then
3:   if num of separators in $p_1$ > num of separators in $p_2$ then
4:     $p_1, p_2 \leftarrow p_2, p_1$
5:   end if
6:   Select two different positions of separator randomly, $l_1$ and $l_2$, in $p_1$, (suppose $l_1 < l_2$);
7:   Select two different positions of separator randomly, $l_3$ and $l_4$, in $p_2$, and make sure that the num of separators in $p_1[l_1 : l_2]$ is the same as the num of separators in $p_2[l_3 : l_4]$;
8:   Exchange the parts $p_1[l_1 : l_2]$ and $p_2[l_3 : l_4]$, then get two offspring, $q_1$ and $q_2$;
9: else
10:   $q_1 \leftarrow p_1$;
11:   $q_2 \leftarrow p_2$;
12: end if
    Return: $q_1$ and $q_2$.
```

Fig. 7. MOGIG-Net Crossover

```
Input: One individual, $p$, probability of mutation $\nu \in (0, 1)$, bits to change, $n$.
Output: One individual, $q$.

1: Generate a random number $flag$;
2: if $flag > \nu$ then
3:   $q \leftarrow$ modify $n$ different bits in $q$;
4: else
5:   $q \leftarrow p$;
6: end if
    Return: $q$.
```

Fig. 8. MOGIG-Net Mutation

3.4. *Decoding of MOGIG-Net*

Fig. 9 represents the construction of the CNN from its chromosome. At first, the CB$_j$ (in blue) are decoded. If the CB$_j$ is the same as that in the corresponding position in its parents, the module is copied from its parents. Otherwise, the module is generated according to the procedure illustrated in Fig. 5. Then, the S is decoded and the corresponding connection is represented by an input array of each block (such as the two red arrows pointing to CB$_3$). Finally, P is decoded and the corresponding position in each input array of each block is wrapped by an adaptive pooling (like the right sub-figure in 9). The connection method in the detailed structure depends on the method which we choose before the algorithm. If we use the residual structure, we add the connection directly. If we use the dense structure, we adjust the channel and merge it by using the 1x1 convolution kernels to a unitize the channel number.

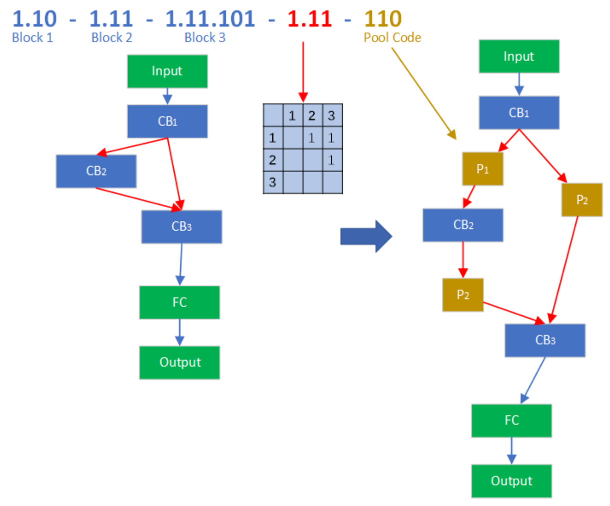

Fig. 9. Construction of a CNN from its chromosome: The blocks or connections are decided by the part of encoding in the same colour. The green squares represent fixed structures. FC means a fully connected layer. P means a pooling layer. Blocks are built in Fig. 5

We also implemented a mechanism to handle missing connections within and among blocks when an adjacency list is generated. Let us consider at first the nodes within a block. If the generated solution contains a node which has inputs and no outputs, then a link between the node and the output node of the block is created. If the generated solution contains a node which has outputs and no outputs, then a link from the input node is generated. If a node has neither inputs nor outputs, then the node is removed. The same reasoning is performed about the connectivity among blocks where each node represents a block while input and output blocks of the CNN are considered instead of input and output nodes of the block. Figure 10 describes this mechanism by showing the three possible scenarios where node 3 has only inputs (left), has only outputs (centre), has neither inputs nor outputs.

During the construction of a CNN from its chromosome, the skip connections (in blocks and between blocks), which need the sizes of the input and output to be the same, are fundamental to achieve a graph structure network. However, convolution and pooling operations can both change the size of the image. This characteristic of CNNs makes difficult to unify the input size of each part in the graph structure network. Therefore, we maintain the size consistency in the input and output of each block or search unit, so that the size reduction is completely controlled by the pooling layer. It is simple to keep the image size unchanged in the convolution block, only by adjusting the super parameters of the convolution kernel and avoiding the use of a pooling layer.

The formula for calculating the size of input and output is given in Eq.(1), where X_{out} and X_{in} are the size of the input and output, p is the number of padding around the input, d is the offset of two adjacent points of the dilated convolution, k is the size of the convolution kernel, and s is the step size of the convolution operation. Therefore, the size is controlled by means of the convolution kernel.

$$X_{out} = \left\lfloor \frac{X_{in} + 2 \times p - d \times (k-1) - 1}{s} + 1 \right\rfloor \quad (1)$$

Furthermore, since maintaining the consistency of image size outside the convolution block (i.e., the macro structure) another countermeasure has been adopted. We also encode the reduced position of the size (but did not add into the genes) as the reduction of the size does not affect the use of convolution kernel, see Fig. 6 line 11.

We chose adaptive pooling, which is different from the traditional pooling operation. This operation can dynamically create pooled cores according to the input and output requirements, and it has been used in the last layer of many existing models [31; 33]. The step size of the adaptive pooling layer can be obtained by Eq (2)

$$stride = \lfloor \frac{size_{in}}{size_{out}} \rfloor \quad (2)$$

where $size_{in}$ is the size of input feature map and $size_{out}$ is the size of output feature map. The size of pool $size_{pool}$ is then calculated in Eq (3)

$$size_{pool} = size_{in} - stride * (size_{out} - 1) \quad (3)$$

On the basis of these two formulas, we can adjust the kernel of the adaptive pooling layer from the size of the input feature map and the size of the desired output feature map. As shown in Fig. 11, the two pool layers before the block and FC marked in

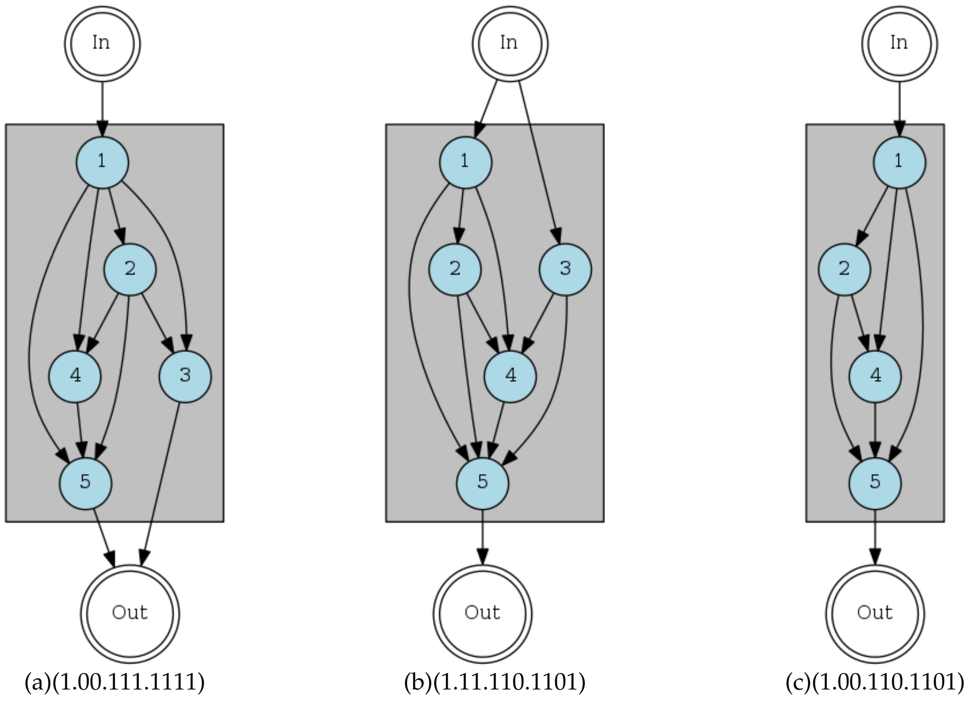

Fig. 10. Three scenarios to guarantee connected CNN blocks. In the left encoding, node 3 would have only inputs. Thus, an output link is generated to guarantee connectivity. In the central encoding, node 3 would have only outputs. Thus, an input link is generated to guarantee connectivity. In the right encoding, node 3 would be isolated. Thus, the node is removed from the graph.

bold because we choose to add adaptive pools before them. In this way, we can control the size of input and output in each layer by controlling the position of adaptive pooling. The location of adaptive pooling and the combination of these channels are referred to as the detailed structure of the individual and are recorded separately.

3.5. *Evaluation and Environment Selection*

We divide the training sets D into two parts, 80% of which are real training sets D_{train}, and the rest are validation sets D_{valid}. When the new population of offspring solutions is generated, their performance must be assessed to select the population undergoing the following generation. The networks composing the new population undergo training by means of the training set D_{train}. When the change range is below a pre-arranged threshold, the learning rate is adjusted accordingly. If the learning rate adjustment is less than a prearranged value, the training will be stopped.
In our approach, we use weight inheritance to speed up the search. Since our crossover operation can ensure that most of the modules of the network remain unchanged, the weight of the model constructed by the child will directly inherit the weight from the model of the parent. This method, like weight sharing, can make the network model obtain a relatively high accuracy rate at the early stage of evolution. In this way, we only need to continue training at a relatively small learning rate to achieve the best performance of each network.

After the training, the accuracy $q.acc$ (that is the error rate) of the network is assessed by means of the validation set D_{valid}. Furthermore, the model size in terms of the number of parameters $q.params$ is also calculated. Both the scores $q.acc$ and $q.params$ characterise the quality of the candidate CNN. The non-dominated sorting [50] is used to select among parent and offspring solutions the population undergoing the following generation, which often used to evaluate the quality of two solutions in the process of multi-objective optimization[93]. The condition for one individual to dominate another is to have a performance not worse than the other according to all objective and

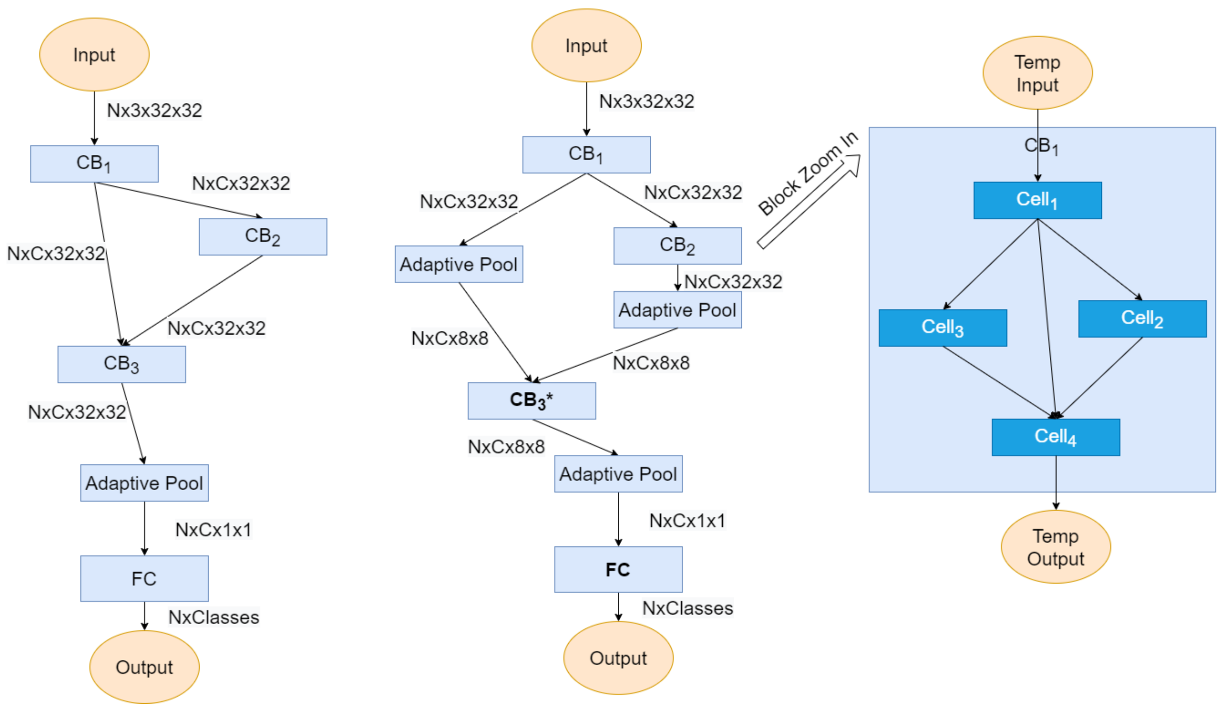

Fig. 11. The left subgraph is the macro structure without pool layers. After executing line 11 of the algorithm in Fig. 6, the adaptive pooling is added at the specified location (center subgraph). The right subgraph is micro structure. Each block includes some convolution cells, and each cell is consist of 3x3 and 1x1 convolution kernels, which do not change the size of input.

to outperform it according to at least one objective.

Fig. 12 provides the implementation details of evaluation and selection mechanisms.

3.6. *Limitations of MOGIG-Net and Countermeasures*

Without a prior knowledge on the problem, each connection has initially the same probability to be set as 0 or as 1. Thus, on average initialized solutions contain approximately half of the skip connections, many of them being unnecessary. These skip connections can cause a slow down of the network training. Thus, the search efficiency of our method is rather low in the early stages. However, the method of weight inheritance accelerates the search and partially mitigates this limitation. Already from the second generation of the population, we observed a large number of excellent structures in the population and its parameters are retained along with the encoding, which makes the training process overall efficient and yields high-performance candidate solutions.

4. Experiments

This section displays the results of the proposed MOGIG-Net on two popular datasets and compares its performance with that of seventeen NAS methods previously proposed in the literature.

```
Input: The population $P_t$, the training set $D_{train}$, the validation set $D_{valid}$
Output: The new population $P_{t+1}$

1:  for all individual $q$ in population $P$ do
2:    Check the database of fingerprint
3:    if fingerprint of $q$ is in the database then
4:      Get $q.acc$ and $q.params$ from database;
5:    else
6:      $cnn$ ← Generate the network with $q$;
7:      train $cnn$ on $D_{train}$ until the loss and accuracy don't change significantly;
8:      $q.acc$ ← the rate of accuracy assessed on the valid set;
9:      $q.params$ ← the number of parameters contained in the model $cnn$ itself;
10:   end if
11:   Update individual $q$ in population $P$;
12: end for
13: Do non-dominated sorting [50] and select half of the individuals who were better at multiple goals from $P_{t+1}$.
    Return: $P_{t+1}$
```

Fig. 12. MOGIG-Net Multi-objective Evaluation and Selection

The popular datasets considered in this study are Cifar-10 and cifar-100 proposed by the Canadian Institute for Advanced Research [37]. These two datasets are often used to verify the performance of network models. Each dataset comprises 60000 images, including 50000 in the training set and 10000 in the test set. Each image is a 3-channel colour image, and the height and the width are both 32. There are 10 categories in cifar-10 and 100 categories in cifar-100. Both cifar-10 and cifar-100 come from a larger dataset of 80 million small images. Therefore, to a certain extent, cifar-10 and cifar-100 can illustrate the predictive ability of the model.

Table 1 displays the results of MOGIG-Net and twenty-one NAS competitors on cifar-10 and cifar-100. The listed methods are divided into three design categories: NAS human design, single-objective approaches and multi-objective approaches. For each NAS method considered in this study, the reference to its original implementation. For each method we report the result of the objectives in the proposed model, that is the accuracy $q.acc$ expressed in terms to percentage error for Cifar-10 and Cifar-100 and the complexity $q.param$ expressed in million of parameters of the network designed by the corresponding NAS method. We may observe that the proposed MOGIG-Net can efficiently detect networks which combine a relatively low number of parameters and a low percentage error. For example, none of the seventeen competitor NAS methods can achieve an error rate of 14.38% on Cifar-100 with only 3.7 million parameters. With respect to NSGA-Net [50], that is a recent NAS method considered the state-of-the-art in the field, the proposed MOGIG-Net designed networks with a comparable performance notwithstanding a lower number of parameters (approximately 10% fewer parameters).

Figures 13 and 14 display the solutions in the objective space considered in this study detected by the proposed MOGIG-Net and its competitor. To enhance the readability of the figures, we present a zoom around the non-dominated solutions.

We noticed that when the network structure is relatively large, the number of pooling in the detailed structure greatly affects the required training time and the memory space. When the number of pooling is small and the network structure is large, the size of intermediate variables is very large and the training time is very long. The results in this study have been detected after two weeks of calculation.

Experimental results show that for networks with similar structures, the accuracy of large models is higher than that of small models, including our method. The reason of this phenomenon is that the increase in the number of parameters appears to improve the generalization capability of the model. Therefore, the maximum accuracy that can be achieved with large models is higher than that of smaller models.

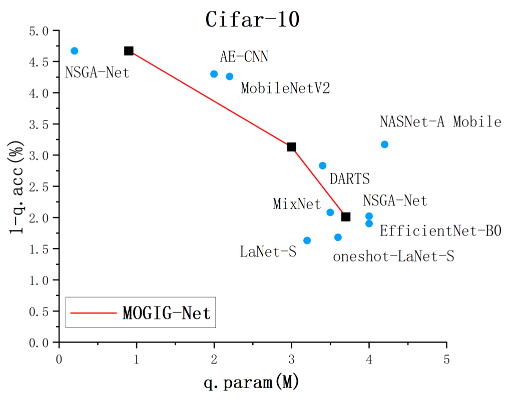

Fig. 13. Solutions detected by MOGIG-Net and its competitors represented in the objective space (Cifar-10)

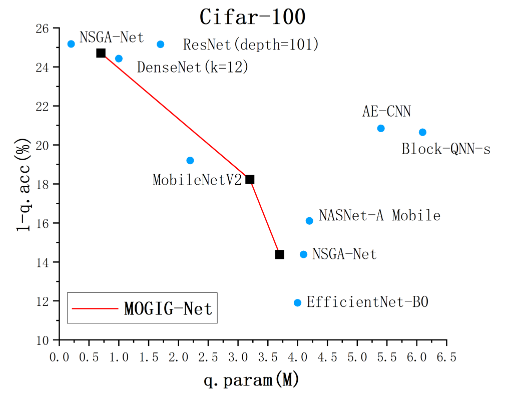

Fig. 14. Solutions detected by MOGIG-Net and its competitors represented in the objective space (Cifar-100)

The results in Fig. 13 and Fig. 14 show that MixNet and MobileNetV2 display excellent performance. However, MixNet and MobileNetV2, unlike the proposed MOGIG-Net are human-designed networks with a predefined purpose. Thus, the performance of the methods cannot be directly compared. Also, LaNet produced a solution that dominates the MOGIG-Net solution for cifar-10. We suspect that this may be because LaNet tends to select large models with high accuracy, and we come to this conclusion because some of the networks in the search space, like LaNet-L and oneshot-LaNet-L, seems to be large.

However, when the network structure is relatively small, with the increase of total computation times (Multiply-Adds operations), the generalization performance of the network is also improving. Consequently, the next step of this work will be to transform the network structure and/or to determine the number of pooling which is randomly added to the network structure, instead of randomly generating several pooling layers and inserting them into random locations.

Table 1. Results on Cifar-10 and Cifar-100 datasets [37] of the proposed MOGIG-Net against twenty-one NAS methods. The percentage error "Error Rate (%)" and number of parameters expressed in million pf parameters "Params(M)" are reported.

Name	Params(M)	Error Rate(%)	
		Cifar-10	Cifar-100
Human Design			
DenseNet(k=12)[33]	1.0	5.24	24.42
ResNet(depth=101)[31]	1.7	6.43	25.16
ResNet(depth=1202)[31]	10.2	7.93	27.82
MobileNetV2[68]	2.2	4.26	19.20
NASNet-A Mobile[99]	4.2	3.17	16.10
EfficientNet-B0[86]	4.0	1.90	11.90
MixNet[87]	3.5	2.08	-
DARTS[47]	3.4	2.83	-
VGG[74]	20.1	6.66	28.05
NIN[44]	-	8.81	35.68
Single-Objective Approaches			
Genetic CNN[91]	-	7.10	29.05
Block-QNN[97]	39.8	3.50	-
Block-QNN-s[97]	6.1	4.38	20.65
LaNet-S[90]	3.2	1.63	-
LaNet-L[90]	44.1	0.99	-
oneshot-LaNet-S[90]	3.6	1.68	-
oneshot-LaNet-L[90]	45.3	1.20	-
Large-scale Evolution[63]	5.4	5.40	
	40.4		23.00
MetaQNN[9]	-	6.92	27.14
AE-CNN[80]	2.0	4.30	
	5.4		20.85
Multi-Objective Approaches			
NSGA-Net[50]	0.2	4.67	
	4.0	2.02	
	0.2		25.17
	4.1		14.38
MOGIG-Net	0.9	4.67	-
	3.0	3.13	-
	3.7	2.01	-
	0.7	-	24.71
	3.2	-	18.23
	3.7	-	14.38

Since numerical results indicate that the proposed MOGIG-Net is able to design excellent CNNs, a future direction of our research will include the extension of the encoding strategy to other ingenious neural systems recently proposed in the literature, such as Enhanced Probabilistic Neural Network[5], Neural Dynamic Classification Algorithm[61], Dynamic Ensemble Learning Algorithm [6], and Finite Element Machine for Fast Learning[60]

5. Conclusion

This paper proposes a NAS method to design CNNs with high performance in terms of accuracy and a limited impact on the computational resources.

The proposed algorithm indicated with MOGIG-Net makes use of a novel block logic based on adjacency list to compose the network structure. The encoding mechanism proposed in this paper can naturally represent the structure of any graph. Moreover, MOGIG-Net employs ad-hoc crossover and mutation operators which are designed to explore the search space and identify potential candidate structures. At last, the proposed network encoding enables that the parent structures can be effectively and naturally transferred to the offspring during the crossover process. The proposed approach overcomes the limitation of classical NAS approaches based on Evolutionary Algorithms which require a search in a large space and an overhead due to multiple re-training sessions. Numerical results on two popular datasets Cifar-10 and Cifar-100 show that MOGIG-Net can exceed most existing network structures.

This paper confirms that multi-objective optimization modelling is a promising direction of research in the field of NAS. Future research will consider further objectives and strategies to reduce the computational cost of the training by e.g. limiting the number of skip connections in the first generation.

Acknowledgment

This work was partially supported by the National Natural Science Foundation of China (61876089, 61876185, 61902281), the opening Project of Jiangsu Key Laboratory of Data Science and Smart Software (No.2019DS301), the Natural Science Foundation of

Jiangsu Province (BK20141005), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (14KJB520025).

References

1. H. A. Abbass, Pareto neuro-evolution: Constructing ensemble of neural networks using multi-objective optimization, *IEEE Congress on Evolutionary Computation*, **3** 2003, pp. 2074–2080.
2. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan and H. Adeli, Deep convolutional neural network for the automated detection and diagnosis of seizure using eeg signals, *Computers in biology and medicine* **100** (2018) 270–278.
3. U. R. Acharya, S. L. Oh, Y. Hagiwara, J. H. Tan, H. Adeli and D. P. Subha, Automated eeg-based screening of depression using deep convolutional neural network, *Computer methods and programs in biomedicine* **161** (2018) 103–113.
4. A. Agogino, K. Stanley and R. Miikkulainen, Online interactive neuro-evolution, *Neural Processing Letters* **11**(1) (2000) 29–38.
5. M. Ahmadlou and H. Adeli, Enhanced probabilistic neural network with local decision circles: A robust classifier, *Integrated Computer-Aided Engineering* **17**(3) (2010) 197–210.
6. K. M. R. Alam, N. Siddique and H. Adeli, A dynamic ensemble learning algorithm for neural networks, *Neural Computing and Applications* **32**(12) (2020) 8675–8690.
7. A. H. Ansari, P. J. Cherian, A. Caicedo, G. Naulaers, M. De Vos and S. Van Huffel, Neonatal seizure detection using deep convolutional neural networks, *International journal of neural systems* **29**(04) (2019) p. 1850011.
8. A. Asseman, N. Antoine and A. S. Ozcan, Accelerating deep neuroevolution on distributed fpgas for reinforcement learning problems, *ACM Journal on Emerging Technologies in Computing Systems (JETC)* **17**(2) (2021) 1–17.
9. B. Baker, O. Gupta, N. Naik and R. Raskar, Designing neural network architectures using reinforcement learning, *International Conference on Learning Representations*, 2017.
10. B. Baker, O. Gupta, R. Raskar and N. Naik, Accelerating neural architecture search using performance prediction, *International Conference on Learning Representations*, 2018.
11. G. Bender, P.-J. Kindermans, B. Zoph, V. Vasudevan and Q. Le, Understanding and simplifying one-shot architecture search, *International Conference on Machine Learning*, 2018, pp. 550–559.
12. S. Bhat, U. R. Acharya, Y. Hagiwara, N. Dadmehr and H. Adeli, Parkinson's disease: Cause factors, measurable indicators, and early diagnosis, *Computers in biology and medicine* **102** (2018) 234–241.
13. C. Blum, R. Chiong, M. Clerc, K. De Jong, Z. Michalewicz, F. Neri and T. Weise, Evolutionary optimization, *Variants of evolutionary algorithms for real-world applications*, (Springer, 2012), pp. 1–29.
14. H. Cai, J. Yang, W. Zhang, S. Han and Y. Yu, Path-level network transformation for efficient architecture search, *PMLR International Conference on Machine Learning*, 2018, pp. 678–687.
15. H. Cai, L. Zhu and S. Han, Proxylessnas: Direct neural architecture search on target task and hardware, *International Conference on Learning Representations*, 2018.
16. X. Cao, J. Yao, Z. Xu and D. Meng, Hyperspectral image classification with convolutional neural network and active learning, *IEEE Transactions on Geoscience and Remote Sensing* **58**(7) (2020) 4604–4616.
17. F. Charte, A. J. Rivera, F. Martínez and M. J. del Jesus, Evoaaa: An evolutionary methodology for automated neural autoencoder architecture search, *Integrated Computer-Aided Engineering* (Preprint) (2020) 1–21.
18. L.-C. Chen, M. Collins, Y. Zhu, G. Papandreou, B. Zoph, F. Schroff, H. Adam and J. Shlens, Searching for efficient multi-scale architectures for dense image prediction, *Advances in neural information processing systems*, 2018, pp. 8699–8710.
19. F. Chollet, Xception: Deep learning with depthwise separable convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
20. X. Dai, P. Zhang, B. Wu, H. Yin, F. Sun, Y. Wang, M. Dukhan, Y. Hu, Y. Wu, Y. Jia *et al.*, Chamnet: Towards efficient network design through platform-aware model adaptation, *Proceedings of the IEEE Conference on computer vision and pattern recognition*, 2019, pp. 11398–11407.
21. A. Diba, A. Pazandeh and L. Van Gool, Efficient two-stream motion and appearance 3d cnns for video classification, *Proceedings*, 2016, pp. 1–4.
22. X. Dong and Y. Yang, One-shot neural architecture search via self-evaluated template network, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 3681–3690.
23. T. Elsken, J. H. Metzen and F. Hutter, Efficient multi-objective neural architecture search via lamarckian evolution, *International Conference on Learning Representations*, 2018.
24. T. Elsken, J.-H. Metzen and F. Hutter, Simple and efficient architecture search for convolutional neural networks, *International Conference on Learning Representations*, 2018.
25. J. Fang, Y. Sun, Q. Zhang, Y. Li, W. Liu and X. Wang, Densely connected search space for more flexible neural architecture search, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10628–10637.
26. F. Gomez, J. Schmidhuber and R. Miikkulainen, Efficient non-linear control through neuroevolution, *European Conference on Machine Learning*, (Springer,

2006), pp. 654–662.

27. F. J. Gomez and R. Miikkulainen, Solving non-markovian control tasks with neuroevolution, *IJCAI*, **99** 1999, pp. 1356–1361.

28. X. Gong, S. Chang, Y. Jiang and Z. Wang, Autogan: Neural architecture search for generative adversarial networks, *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 3224–3234.

29. Z. Guo, X. Zhang, H. Mu, W. Heng, Z. Liu, Y. Wei and J. Sun, Single path one-shot neural architecture search with uniform sampling, *European Conference on Computer Vision*, (Springer, 2020), pp. 544–560.

30. M. Hausknecht, J. Lehman, R. Miikkulainen and P. Stone, A neuroevolution approach to general atari game playing, *IEEE Transactions on Computational Intelligence and AI in Games* **6**(4) (2014) 355–366.

31. K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

32. O. M. Hooman, M. M. Al-Rifaie and M. A. Nicolaou, Deep neuroevolution: Training deep neural networks for false alarm detection in intensive care units, *2018 26th IEEE European Signal Processing Conference (EUSIPCO)*, 2018, pp. 1157–1161.

33. G. Huang, Z. Liu, L. Van Der Maaten and K. Q. Weinberger, Densely connected convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.

34. H. Jiang, F. Gao, X. Xu, F. Huang and S. Zhu, Attentive and ensemble 3d dual path networks for pulmonary nodules classification, *Neurocomputing* **398** (2020) 422–430.

35. N. Karthikeyan and R. Sukanesh, Cloud based emergency health care information service in india, *Journal of medical systems* **36**(6) (2012) 4031–4036.

36. A. Krizhevsky, One weird trick for parallelizing convolutional neural networks, tech. report (2014).

37. A. Krizhevsky, G. Hinton and others, Learning multiple layers of features from tiny images, tech. report, University of Toronto (2009).

38. H. Kwasnicka and M. Paradowski, Efficiency aspects of neural network architecture evolution using direct and indirect encoding, *Adaptive and Natural Computing Algorithms*, (Springer, 2005), pp. 405–408.

39. S. Li, X. Zhao and G. Zhou, Automatic pixel-level multiple damage detection of concrete structure using fully convolutional network, *Computer-Aided Civil and Infrastructure Engineering* **34**(7) (2019) 616–634.

40. X. Li, Y. Jiang, M. Li and S. Yin, Lightweight attention convolutional neural network for retinal vessel image segmentation, *IEEE Transactions on Industrial Informatics* **17**(3) (2020) 1958–1967.

41. X. Li, C. Lin, C. Li, M. Sun, W. Wu, J. Yan and W. Ouyang, Improving one-shot nas by suppressing the posterior fading, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13836–13845.

42. Y. Li, Z. Yu, Y. Chen, C. Yang, Y. Li, X. Allen Li and B. Li, Automatic seizure detection using fully convolutional nested lstm, *International journal of neural systems* **30**(04) (2020) p. 2050019.

43. L.-C. Lin, C.-S. Ouyang, R.-C. Wu, R.-C. Yang and C.-T. Chiang, Alternative diagnosis of epilepsy in children without epileptiform discharges using deep convolutional neural networks, *International journal of neural systems* **30**(05) (2020) p. 1850060.

44. M. Lin, Q. Chen and S. Yan, Network in network, *International Conference on Learning Representations*, 2014.

45. C. Liu, B. Zoph, M. Neumann, J. Shlens, W. Hua, L.-J. Li, L. Fei-Fei, A. Yuille, J. Huang and K. Murphy, Progressive neural architecture search, *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 19–34.

46. G. Liu, W. Zhou and M. Geng, Automatic seizure detection based on s-transform and deep convolutional neural network, *International journal of neural systems* **30**(04) (2020) p. 1950024.

47. H. Liu, K. Simonyan and Y. Yang, Darts: Differentiable architecture search, *International Conference on Learning Representations*, 2018.

48. Z. Liu, Z. Wu, T. Li, J. Li and C. Shen, Gmm and cnn hybrid method for short utterance speaker recognition, *IEEE Transactions on Industrial informatics* **14**(7) (2018) 3244–3252.

49. Z. Lu, K. Deb and V. N. Boddeti, MUXConv: Information Multiplexing in Convolutional Neural Networks, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12044–12053.

50. Z. Lu, I. Whalen, V. Boddeti, Y. Dhebar, K. Deb, E. Goodman and W. Banzhaf, Nsga-net: neural architecture search using multi-objective genetic algorithm, *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019, pp. 419–427.

51. N. Ma, X. Zhang, H.-T. Zheng and J. Sun, Shufflenet v2: Practical guidelines for efficient cnn architecture design, *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 116–131.

52. Z. Ma, Reachability analysis of neural masses and seizure control based on combination convolutional neural network, *International journal of neural systems* **30**(01) (2020) p. 1950023.

53. K. Maeda, S. Takahashi, T. Ogawa and M. Haseyama, Convolutional sparse coding-based deep random vector functional link network for distress classification of road structures, *Computer-Aided Civil and Infrastructure Engineering* **34**(8) (2019) 654–676.

54. O. M. Manzanera, S. K. Meles, K. L. Leenders, R. J. Renken, M. Pagani, D. Arnaldi, F. Nobili, J. Obeso, M. R. Oroz, S. Morbelli and N. M. Mauritis, Scaled subprofile modeling and convolutional neural networks for the identification of parkinson's disease in 3d nuclear imaging data, *International journal of neural systems* **29**(09) (2019) p. 1950010.

55. H. Mo, L. L. Custode and G. Iacca, Evolutionary neural architecture search for remaining useful life prediction, *Applied Soft Computing* **108** (2021) p. 107474.
56. H. Mo, F. Lucca, J. Malacarne and G. Iacca, Multi-head cnn-lstm with prediction error analysis for remaining useful life prediction, *2020 27th Conference of Open Innovations Association (FRUCT)*, (IEEE, 2020), pp. 164–171.
57. R. Negrinho, M. Gormley, G. J. Gordon, D. Patil, N. Le and D. Ferreira, Towards modular and programmable architecture search, *Advances in Neural Information Processing Systems*, 2019, pp. 13715–13725.
58. F. Neri, *Linear Algebra for Computational Sciences and Engineering*, second edn. (Springer, 2019).
59. B. K. Oh, B. Glisic, Y. Kim and H. S. Park, Convolutional neural network-based wind-induced response estimation model for tall buildings, *Computer-Aided Civil and Infrastructure Engineering* **34**(10) (2019) 843–858.
60. D. R. Pereira, M. A. Piteri, A. N. Souza, J. P. Papa and H. Adeli, Fema: a finite element machine for fast learning, *Neural Computing and Applications* **32**(10) (2020) 6393–6404.
61. M. H. Rafiei and H. Adeli, A new neural dynamic classification algorithm, *IEEE transactions on neural networks and learning systems* **28**(12) (2017) 3074–3083.
62. E. Real, A. Aggarwal, Y. Huang and Q. V. Le, Regularized evolution for image classifier architecture search, *Proceedings of the aaai conference on artificial intelligence*, **33**(01) 2019, pp. 4780–4789.
63. E. Real, S. Moore, A. Selle, S. Saxena, Y. L. Suematsu, J. Tan, Q. V. Le and A. Kurakin, Large-scale evolution of image classifiers, *International Conference on Machine Learning*, 2017, pp. 2902–2911.
64. S. Risi and K. O. Stanley, Deep neuroevolution of recurrent and discrete world models, *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019, pp. 456–462.
65. S. Rostami, F. Neri and M. Epitropakis, Progressive preference articulation for decision making in multi-objective optimisation problems, *Integrated Computer-Aided Engineering* **24**(4) (2017) 315–335.
66. S. Rostami, F. Neri and K. Gyaurski, On algorithmic descriptions and software implementations for multi-objective optimisation: A comparative study, *SN Computer Science* **1**(5) (2020) 1–23.
67. C. Saltori, S. Roy, N. Sebe and G. Iacca, Regularized evolutionary algorithm for dynamic neural topology search, *International Conference on Image Analysis and Processing*, (Springer, 2019), pp. 219–230.
68. M. Sandler, A. Howard, M. Zhu, A. Zhmoginov and L.-C. Chen, Mobilenetv2: Inverted residuals and linear bottlenecks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
69. K. C. Sarma and H. Adeli, Fuzzy discrete multicriteria cost optimization of steel structures, *Journal of structural engineering* **126**(11) (2000) 1339–1347.
70. K. C. Sarma and H. Adeli, Bilevel parallel genetic algorithms for optimization of large steel structures, *Computer-Aided Civil and Infrastructure Engineering* **16**(5) (2001) 295–304.
71. W. W. Seeley, J. M. Allman, D. A. Carlin, R. K. Crawford, M. N. Macedo, M. D. Greicius, S. J. Dearmond and B. L. Miller, Divergent social functioning in behavioral variant frontotemporal dementia and alzheimer disease: reciprocal networks and neuronal evolution, *Alzheimer Disease & Associated Disorders* **21**(4) (2007) S50–S57.
72. J. Shen, X. Xiong, Z. Xue and Y. Bian, A convolutional neural-network-based pedestrian counting model for various crowded scenes, *Computer-Aided Civil and Infrastructure Engineering* **34**(10) (2019) 897–914.
73. W. Shin, S.-J. Bu and S.-B. Cho, 3d-convolutional neural network with generative adversarial network and autoencoder for robust anomaly detection in video surveillance, *International Journal of Neural Systems* **30**(06) (2020) p. 2050034.
74. K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, *International Conference on Learning Representations*, 2015.
75. M. G. Soto and H. Adeli, Many-objective control optimization of high-rise building structures using replicator dynamics and neural dynamics model, *Structural and Multidisciplinary Optimization* **56**(6) (2017) 1521–1537.
76. K. O. Stanley, J. Clune, J. Lehman and R. Miikkulainen, Designing neural networks through neuroevolution, *Nature Machine Intelligence* **1**(1) (2019) 24–35.
77. F. P. Such, V. Madhavan, E. Conti, J. Lehman, K. O. Stanley and J. Clune, Deep neuroevolution: Genetic algorithms are a competitive alternative for training deep neural networks for reinforcement learning, *International Conference on Learning Representations*, 2019.
78. Y. Sun, H. Wang, B. Xue, Y. Jin, G. G. Yen and M. Zhang, Surrogate-assisted evolutionary deep learning using an end-to-end random forest-based performance predictor, *IEEE Transactions on Evolutionary Computation* **24**(2) (2019) 350–364.
79. Y. Sun, B. Xue, M. Zhang and G. G. Yen, A particle swarm optimization-based flexible convolutional autoencoder for image classification, *IEEE transactions on neural networks and learning systems* **30**(8) (2018) 2295–2309.
80. Y. Sun, B. Xue, M. Zhang and G. G. Yen, Completely automated CNN architecture design based on blocks, *IEEE transactions on neural networks and learning systems* **31**(4) (2019) 1242–1254, Publisher: IEEE.
81. Y. Sun, B. Xue, M. Zhang and G. G. Yen, Evolving deep convolutional neural networks for image classification, *IEEE Transactions on Evolutionary Computation* **24**(2) (2019) 394–407, Publisher: IEEE.
82. Y. Sun, B. Xue, M. Zhang, G. G. Yen and J. Lv, Au-

tomatically designing cnn architectures using the genetic algorithm for image classification, *IEEE Transactions on Cybernetics* (2020).

83. C. Szegedy, S. Ioffe, V. Vanhoucke and A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.
84. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, Going deeper with convolutions, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
85. M. Tan, B. Chen, R. Pang, V. Vasudevan, M. Sandler, A. Howard and Q. V. Le, Mnasnet: Platform-aware neural architecture search for mobile, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2820–2828.
86. M. Tan and Q. Le, Efficientnet: Rethinking model scaling for convolutional neural networks, *International Conference on Machine Learning*, 2019, pp. 6105–6114.
87. M. Tan and Q. V. Le, Mixconv: Mixed depthwise convolutional kernels, *30th British Machine Vision Conference 2019, BMVC 2019, Cardiff, UK, September 9-12, 2019*, (BMVA Press, 2019), p. 74.
88. K. Thurnhofer-Hemsi, E. López-Rubio, N. Roe-Vellve and M. A. Molina-Cabello, Multiobjective optimization of deep neural networks with combinations of lp-norm cost functions for 3d medical image super-resolution, *Integrated Computer-Aided Engineering* (Preprint) (2020) 1–19.
89. B. Wang, Y. Sun, B. Xue and M. Zhang, Evolving deep neural networks by multi-objective particle swarm optimization for image classification, *Proceedings of the Genetic and Evolutionary Computation Conference*, 2019, pp. 490–498.
90. L. Wang, S. Xie, T. Li, R. Fonseca and Y. Tian, Sample-efficient neural architecture search by learning actions for monte carlo tree search, *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2021) 1–1, to appear.
91. L. Xie and A. Yuille, Genetic cnn, *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1379–1388.
92. S. Xie, H. Zheng, C. Liu and L. Lin, Snas: stochastic neural architecture search, *International Conference on Learning Representations*, 2018.
93. Y. Xue, Y. Tang, X. Xu, J. Liang and F. Neri, Multi-objective feature selection with missing data in classification, *IEEE Transactions on Emerging Topics in Computational Intelligence* (2021).
94. X. Yao and M. M. Islam, Evolving artificial neural network ensembles, *IEEE Computational Intelligence Magazine* **3**(1) (2008) 31–42.
95. Y. Zhang, Y. Miyamori, S. Mikami and T. Saito, Vibration-based structural state identification by a 1-dimensional convolutional neural network, *Computer-Aided Civil and Infrastructure Engineering* **34**(9) (2019) 822–839.
96. D. Zhipeng, W. Jingcheng, X. Yumin, M. Qingmin and W. Xiaoming, Voiceprint recognition based on bp neural network and cnn, *Journal of Physics: Conference Series*, **1237**(3) 2019, p. 032032.
97. Z. Zhong, J. Yan and C.-L. Liu, Practical network blocks design with q-learning, *arXiv preprint arXiv:1708.05552* **1**(2) (2017) p. 5.
98. B. Zoph and Q. V. Le, Neural architecture search with reinforcement learning, *International Conference on Learning Representations*, 2017.
99. B. Zoph, V. Vasudevan, J. Shlens and Q. V. Le, Learning transferable architectures for scalable image recognition, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.