From Conventional Machine Learning to AutoML

Ziqiao Weng
School of Information and Electrical Engineering, Zhejiang University City College,
No. 48 Huzhou Street, Hangzhou, PRC, 310015
31502301@stu.zucc.edu.cn

Abstract. Machine Learning has enabled conspicuous progress over the past decade on various areas, such as image analysis, computer vision, natural language processing, quantitative finance, etc. This paper presents an overview of Machine Learning algorithms from conventional Machine Learning to Deep Learning then to Automated Machine Learning (AutoML). In the first two sections, a few classical and distinct algorithms will be discussed in the same structure: 1) the principle of algorithm, 2) specific applications or improvements, and 3) contributions or limitations. At the last section, the state-of-the-art AutoML algorithms will be briefly introduced along with their innovation, which achieve unexpected higher efficiency. It turns out that there is a growing interest in AutoML methods. In the future, more and more novel neural architectures will be proposed and applications to real-world scenario will be expanded.

1. Introduction
Machine Learning, which serves as the core of data science as well as artificial intelligence, is being studied and explored by more and more people, including students and researchers. This is a very appealing area, often combined with other areas such as statistics and biology, and in the recent decade Machine Learning has been rapidly and widely used throughout computer science and beyond. Machine Learning is applied in computer vision, natural language processing, quantitative finance and many other applications. Thus, it has given us a glimpse of the great potential of Artificial Intelligence technology, which contributes to this article.

What is Machine Learning (ML)? In the classical textbook Machine Learning [1], the author Tom Mitchell gave the following definition: a computer program is said to learn from experience with respect to some class of tasks and performance measure, if its performance at tasks in the class, as measured by the performance measure, improves with experience. In a word, the research on Machine Learning seeks to improve the performance of computer systems automatically through experience. According to the information provided by the training data and the different ways of feedback, Machine Learning can be mainly divided into Supervised Learning, Unsupervised Learning, Semi-Supervised Learning and Reinforcement Learning.

This passage will focus on Supervised Learning and Unsupervised Learning, especially on classification and regression problems. Several typical algorithms or models will be discussed according in chronological order from conventional ML algorithms to Artificial Neural Network to state-of-the-art Automated Machine Learning algorithms. Each algorithm will be demonstrated in the same structure: 1) the principle of algorithm, 2) specific applications or improvements, and 3)
contributions or limitations. Thus, the ultimate goal of this article is to help other learners to get better understanding of Machine Learning and the areas in which different ML algorithms can handle.

2. Conventional Machine Learning algorithm
In the past 20 years of golden development, conventional Machine Learning community was not too much hype "smart" or "cognitive", but rather focused on the introduction of statistics, and data analysis and processing, with Supervised Learning and Unsupervised Learning So, this chapter will illustrate both of them with their distinct algorithms.

2.1. Supervised Learning
Supervised Learning aims to get an optimal model under a certain criterion, through training the existing training sample with known input denoted by \( X \) (with feature) and the corresponding output denoted by \( Y \) (with label). The goal is to obtain an optimal model through training existing samples, where the known inputs correspond to outputs, so that the unknown data can be classified by this model. Typical methods of Supervised Learning are K-Nearest Neighbor (KNN) and Support Vector Machine (SVM).

2.1.1. K-Nearest Neighbor. K-Nearest Neighbor (KNN) is one of the most fundamental and simplest classification methods. It is usually used as the prior choices for a classification study. It is still widely used in facial recognition, movie/music recommendation, detection of diseases, etc., when the distribution of the data is unknown [2]. This method only determines the category of the sample to be subdivided according to the category of the nearest one or several samples. The dissimilarities between samples are measured by Euclidean distance [3].

Usually, there are six public datasets, from the Machine Learning repository at University of California at Irvine (UCI), are used for several K-Nearest Neighbor based algorithm to test, including Glass, Iris, Heart, Wine, Diabetes, and Aust. Guo and other researchers [4] proposed a KNN model-based approach that achieved competitive performance on the data sets above compared with previous C5.0 and KNN method, with higher classification accuracy than the standard KNN by reducing the dependency on \( K \). KNN based algorithm can also address other real world problem. Zhang’s group [5] proposed a multi-label lazy learning approach named ML-KNN that could predict the label sets of unseen examples according to labeled training data. The experiment on yeast gene functional analysis, natural scene classification and automatic web page categorization revealed that ML-KNN outperforms some widely-used multi-label learning algorithms such as BoosTexter, Adtboost and Rank-SVM.

The significant advantage of the KNN method is that it is easy to understand and to implement [4][6]. New data can be added directly into the data set without retraining [1]. It is suitable for rare event classification and multi-classification problems.

However, an obvious disadvantage of KNN is the time complexity of making predictions [7]. For the dataset with large sample capacity, the calculation amount is relatively large. So, as a lazy learning method, KNN is difficult to be used in many applications, like dynamic web mining [4]. Besides, the dependency on the selection of a proper \( K \) affects its efficiency [4]. At last, the simple Euclidean distances used in most KNN classifiers cannot reveal any statistical regularities in a large training examples [3].

2.1.2. Support Vector Machine. In Machine Learning, Support Vector Machine (SVM) [8] is a Supervised Learning model that analyzes data for classification or regression. SVM is trying to construct an optimal high-dimensional hyper-plane, which is able to separate the non-linear data from the low-dimensional plane that are difficult to distinguish. Usually, the kernel function, originally proposed by Aizerman et al [9], is applied to solve nonlinear classification problems. SVM has been applied to various tasks such as object recognition and classification of images.
There are lots of applications of SVM for the forecasting of hydrology areas, such as rainfall and runoff forecasting and soil moisture estimation [10]. Among applications in the field of surface water hydrology, Anirudh and Umesh [11] investigated on using SVM to predict the occurrence of rainy days and they achieved satisfactory performance by using Gaussian kernel functions. Besides, for the prediction of ground water hydrology. Jin et al. [12] proposed a dynamic algorithm based on the Least Squares Support Vector Machine (LS-SVM) for predicting groundwater level, which had a positive influence on ground water resources management.

On the task of text-independent speaker identification, the SVM classifier also achieves excellent performance on the TIMIT corpus [13], compared with Naive Bayes (NB), decision tree C4.5 and Multi-Layer Perceptron (MLP). The final result has revealed that the SVM trained with polynomial kernel performs best on speaker identification tasks, as shown in the Figure. 1.

![Figure 1. The IR average of SVM, NB, C4.5 and MLP classifiers for all the data sets in TIMIT corpus, where the IR refers to the identification rate. Indeed, the SVM provided remarkably outperforms NB, C4.5 and MLP algorithms.](image)

SVM is able to produce accurate and robust classifiers, even dealing with non-linearly separable input data [10]. However, apart from its advantages, SVM also has some deficiencies. It is difficult to apply SVM algorithm in a data set with large scale and its time complexity is high [14]. In addition, SVM is sensitive to missing data and the selection of kernel functions and hyper parameters are time-consuming [10]. The dependence on the previous data as support vectors may result in poor model extrapolation [10].

2.2. Unsupervised Learning
Unsupervised Learning is where the inputs do not correspond to outputs, which means the data has not been classified or categorized. The goal is to learn the unknown data from the unlabeled data.

Unsupervised Learning algorithms can be included into two categories: clustering and association. Clustering problems are looking for an unsupervised algorithm that is able to discover optimal model by exploring similarities or dissimilarities between individual data points in a data set. Association problems, after clustering, are trying to find certain associations (relationships, dependencies) in large sets of data items.

2.2.1. K-means. Simply speaking, K-means clustering is an algorithm to classify or to group objects based on attributes/features, into \( K \) number of groups where \( K \) is a positive integer. Thus, the purpose of K-means clustering is to group the data by minimizing the mean squared distances between data and the associated cluster center.

Among a number of variants to K-means algorithms, Lloyd's algorithm is one of the most popular iterative heuristics scheme for finding a locally minimal solution [15]. Lloyd's algorithm repeatedly looks for the center of each set in the partition and then re-partitions the input under the simple
observation that the optimal location of a center is at the center of the associated cluster. Kanungo’s group [15] presented an efficient implementation of Lloyd’s algorithm named filtering algorithm and practice it on both synthetically generated and real data sets. And they found that the algorithm becomes more efficient as the separation of cluster increases.

The basic K-means algorithm has been extended into various variants with different ways. Dunn [16][17] proposes Fuzzy C-means algorithm where each data point can be grouped to multiple clusters with a relationship value. Kaufman and Rousseveu propose K-medoid [16][18] which replaces the mean of data by using the median of data. Huang [19] extends the K-modes algorithm to cluster data by using modes of the data and a simple matching dissimilarity approach to categorize objects. Then Huang updated it into K-prototypes algorithm which combines the K-means and K-modes processes to cluster data using mixed numeric and categorical values [19].

The K-means algorithm is easy to understand and is widely used, but it has two major defects:
1. As an input parameter, an improper $K$ may lead to non-ideal results [19]. However, in many cases, the estimation of $K$ is very difficult. Some variants of K-means like ISODATA include a procedure. In order to find a best $K$, some variants of K-means like ISODATA are proposed at the sacrifice of performance [19][20].
2. The sensitivity of K-means to the initial cluster centers makes clustering results in different random seed points are totally different. The K-means++ algorithm that minimizes the average squared distance between samples in the same cluster is used to optimize this problem [21].

3. Deep Learning
The most rudimentary version of Deep Learning is Artificial Neural Networks, a subfield of Machine Learning that attempts to mimic the human brain and automatically extract data features through more complex structures.

Before the development of Deep Learning, although Machine Learning had been developed for decades, there were still many problems that could not be solved by Artificial Intelligence, such as image classification, natural language processing, computer vision etc. This is because conventional Machine Learning techniques are limited in their ability to process raw or large amounts of data. The emergence of Deep Learning solved some problems in these fields.

Deep Learning is a representation learning method that converts primitive data into more abstract expressions through non-linear models. With enough combinations of transformations, very complex functions can also be learned. Deep Learning can discover complex structures in big data by using Back-Propagation (BP) algorithm [22], which can indicate the machine how to obtain errors from previous layers and updates the internal parameters or weights of each layer. The ideal parameters can be used to construct a certain expressible formula. With the boost of big data and the emergence of high-performance GPU, more complex network models are becoming possible, which has also promoted the further development of Deep Learning.

After years of exploration and research, many deep neural network models have been produced, of which Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN) are two typical models.

3.1. CNN
Among various deep neural network structures, Convolutional Neural Network (CNN) is most commonly applied to object recognition field, which is deep, feed-forward artificial neural networks.

The earliest convolutional network is LeNet [23], which is applied to recognize and classify handwritten digits, and is used in American Banks to read about 10% of checks in North America. LeNet laid the foundation of modern convolutional neural network and reveal its typical structure that consists of convolutional layers, pooling (sub-sampling) layers and fully connected layers. However, due to the shortage of large-scale well-established samples and low performance of computer hardware, LeNet was not effective in dealing with complex problems.
In recent years, the appearance of large-scale labelled data like ImageNet and the rapid improvement of GPU computing performance have led to a rapid spurt of research on convolutional neural network, such as AlexNet [24], ZF Net [25], GoogLeNet [26], ResNets [27], DenseNet [28], etc.

3.1.1. AlexNet. AlexNet consists of eight layers, including five convolutional layers and three fully connected layers. It uses the technique of Dropout, Local Response Normalization (LRN) and ReLU activation function [24]. AlexNet's great success established the dominance of Deep Learning in computer vision and also promoted the expansion of Deep Learning in the fields of speech recognition, natural language processing, and Reinforcement Learning.

Before the advent of AlexNet, the best published results of top-1 as well as top-5 test set error rates on ILSVRC 2010 competition were 45.7% and 25.7%. However, AlexNet achieved the state of the art performance at that time with the error rates of 37.5% and 17.0%. AlexNet also achieved best results on ILSVRC 2012, with 15.3% top-5 test error, which outperforms the second-best network with 26.2% top-5 test error.

AlexNet has considerable applications. Ding’s group [29] investigated the pre-trained AlexNet model for feature extraction of high resolution remote sensing image took from Google image in Ludian area of Yunnan Province with 0.3-meter spatial resolution and combined it with spectral feature and GLCM texture respectively. The results showed that the deep feature of AlexNet can obviously get higher classification and extraction accuracy, comparing to spectral and texture features. Ballester and other researchers [30] studied the understanding ability of deep neural network when learning to classify abstract representations (sketches) based on previous experience of learning more realistic subjects (photos). They used AlexNet trained over the ImageNet data set on TU-Berlin sketch data set [31], containing 20,000 sketches of everyday subjects with 250 categories. They found that AlexNet could achieve higher classification accuracy as categories have more well classified training images. Jiao’s group [32] improved AlexNet restructure for rapeseed crop remote sensing identification. They analyzed the cultivated area of rape crop in Jingmen, Hubei province from and the result showed the model achieved 2.39% error rate of actual area and estimated area.

The improvements and deficiencies of AlexNet can be summarized as follows: 1) ReLU nonlinearity is successfully used as the activation function to solve the gradient diffusion problem when the network was deep [24]. But because it is a nonlinear excitation combined with max-pooling, it may lead to misclassification and leakage for high-brightening and low-brightening objects [29]; 2) two GPUs are used for parallel accelerated training, which greatly reduces the training time [24]; 3) using data augmentation, regularization and dropout method to reduce overfitting and enhance the generalization ability [24]; and 4) AlexNet is unable to transfer the learned concepts from realistic objects to abstract and simple representations (sketch) and to compete with humans unless it trains with much more labelled images [31].

3.1.2. DenseNet. DenseNet has absorbed the best parts of ResNet and done more innovative work on it, further improving the performance of network. DenseNet is a convolutional neural network with dense connections. In this network, there is a direct connection between any two layers. In other words, the input of each layer of the network is the output of all previous layers, and the feature map will be passed to all the subsequent layers directly for output [28].

DenseNet can be applied in different areas. Kamal’s group [33] proposed a DenseNet pipeline based on the DenseNet-201 architecture for source camera model identification and post-processing detection. They used the strategy, which was proved to be very robust, that concatenated 3 patch-sizes as output features to secondary network to make the prediction and used Empirical Mode Decomposition (EMD) to augment the data. This model achieved the state-of-the-art performance: an overall accuracy of 98.37% on Forensic Camera-Model Identification Dataset and an overall test accuracy of 92% on Dresden Image Database for Camera Model Identification. Zhou [34] presented a weakly supervised adaptive DenseNet model for classifying diseases and identifying abnormalities to automatically diagnose and localize chest diseases. The results proved the effectiveness of their model.
which outperformed the previous well-established models on the ChestX-ray14 data set, a largest public chest radiography data set. Tao’s group [35] proposed a novel depth-width-reinforced DNN based on DenseNet for the Very High Resolution Remote Sensing (VHRRS) per-pixel classification. The proposed method solved problems of VHRRS per-pixel classification when the amount of labelled training samples is insufficient by strengthening the gradients as the depth of network increases and reducing overfitting and redundancy of too many parameters. It achieved competitive classification accuracy on the BJ02, GF02, geoeye and quickbird satellites images.

This original structure gives DenseNet distinct advantages [28][35]: 1) the vanishing-gradient problem has been alleviated by applying dense connectivity; 2) the feature propagation has been strengthened and the reuse of feature has been enhanced; and 3) the number of parameters has been reduced substantially that contributes to high efficiency.

4. Automated Machine Learning

Traditionally, the term Automated Machine Learning (AutoML) is used to describe automated methods for model selection or hyper parameter optimization, which includes many types of algorithms, such as random forests, neural network, etc. Recently, researches of AutoML focus on neural network architecture search algorithm, usually using reinforcement learning or evolutionary algorithms, which means we can use the neural network to design a more complex and optimal neural network. The automatically searched architectures have been used in tasks such as object detection and image classification, and have achieved unexpected competitive performances on these filed.

Negrinho and Gordon [36] propose a framework to automatically design and train deep neural network models. It shows a search space specification language combined with evaluation algorithms to easily represent complex search spaces and automatically compile models to computational graphs, thus to release experts from adjusting architectures laboriously. The experiments conducted on CIFAR-10, using different model search algorithms, such as Monte Carlo tree search(MCTS) and sequential model-based optimization(SMBO), show that this framework is effective for discovering expressive models.

Zoph et al. [37] propose the NASNet model, which extends the Neural Architecture Search (NAS) framework [38] by using proximal policy optimization (PPO) rather than policy gradient optimization. The NASNet contains a controller Recurrent Neural Network (RNN) that samples child networks in search space with different architectures and get validation accuracy from child network to update the controller so that a best architecture will be generated. As a result, the NASNet achieves better performances (82.7% accuracy on validation data set) on ImageNet than previous networks and the best architecture learned from CIFAR-10 is transferable to ImageNet image classification and COCO object detection.

However, the Neural Architecture Search (NAS) faces two problems: high computational consumption and instability. In order to solve these two problems, Chen’s group [39] proposes the Reinforced Evolutionary Neural Architecture Search (RENAS) that extends NAS with an evolutionary method. The RENAS employs the reinforced mutation controller, which takes mutation actions from effects of slight modifications, to help the exploration of the architecture space more stable and efficient.

Cai et al. [40] propose a new framework, called Efficient Architecture Search (EAS), where a reinforcement learning agent is employed as the meta-controller to explore the architecture space by network transformation operations and reusing weights. The experiment based on image benchmark datasets (CIFAR-10 and SVHN) shows that even with limited computational resources (5 GPUs), EAS can design highly competitive networks. In order to solve the limitation of layer-level architecture modifications, Cai’s group [41] then proposes the path-level network transformation to modify the path topology, further proposing a bidirectional tree structured reinforcement learning meta-controller, which leads to a simple but effective multi-branch structures.

Liu et al. [42] propose Differentiable Architecture Search (DARTS) model that regards the search space as to be continuous rather than discrete. Thus, the architecture can be optimized by gradient
descent rather than inefficient black-box search method by searching in a continuous architecture space. The experiments on image classification and language modelling tasks show that DARTS performs better than other non-differentiable search algorithms on CIFAR-10 and PTB. Besides, it is transferable, which means the architecture learned from CIFAR-10 and PTB can be transferred to ImageNet and WikiText-2, respectively.

Barreiro et al. [43] propose the Net-Net AutoML approach for the prediction of Biological Ecosystem Networks (BENs), complex webs of biological species. This work uses for the first-time Shannon entropy information measures (Shk) to quantify both the structure of complex biological systems to be predicted and ML algorithms to be tested. Mohr et al. [44] present a new AutoML approach named ML-Plan based on hierarchical planning and it achieves highly competitive compared with other well-established approaches, such as Auto-WEKA, TPOT and auto-sklearn.

5. Conclusion
This paper presented an overview of the development and distinct algorithms of Machine Learning, consisting of three sections: 1) conventional Machine Learning, 2) Deep Learning, and 3) Auto Machine Learning (AutoML). From the first section, we can see that some conventional ML algorithms are still active in classification and forecasting areas. From the second section, we can conclude that Deep Learning methods, including CNN and RNN, are more widely used and perform much better than conventional methods in image classification and natural language processing. In the third section, AutoML significantly improves the efficiency of Machine Learning and has achieved considerable successes in recent years. In the future, more researchers will focus on this area to design more efficient algorithms and accelerate their applications to solve real-world issues. Moreover, the increasing requirement of hardware for strong computing power could also be the research emphasis in both academic and industry field.

References
[1] Mitchell T M. Machine learning. [M]. China Machine Press, 2003.
[2] Peterson L. K-nearest neighbor[J]. Scholarpedia, 2009, 4(2):1883.
[3] Weinberger K Q, Saul L K. Distance Metric Learning for Large Margin Nearest Neighbor Classification[J]. Journal of Machine Learning Research, 2009, 10(1):207-244.
[4] Suja R N, Deka P C. Support vector machine applications in the field of hydrology: A review[J]. Applied Soft Computing Journal, 2014, 19(6):372-386.
[5] Guo G, Wang H, Bell D, et al. KNN Model-Based Approach in Classification[C]// Otm Confederated International Conferences "on the Move To Meaningful Internet Systems. Springer, Berlin, Heidelberg, 2003:986-996.
[6] Zhang M L, Zhou Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7):2038-2048.
[7] Bramer M. Principles of Data Mining[M]. MIT Press, 2001.
[8] Elkan C. Nearest Neighbor Classification[J]. Survey Paper, 2011, 27(5):622 - 627.
[9] Cortes C, Vapnik V. Support-vector networks[J]. Machine learning, 1995, 20(3): 273-297.
[10] Aizerman M A. Theoretical foundations of the potential function method in pattern recognition learning[J]. Automation and remote control, 1964, 25: 821-837.
[11] Sujay R N, Deka P C. Support vector machine applications in the field of hydrology: A review[J]. Applied Soft Computing Journal, 2014, 19(6):372-386.
[12] V. Anirudh, C.K. Umesh, Classification of rainy days using SVM, in: Proceedings of Symposium at HYDRO, Norfolk, Virginia, USA, 2007.
[13] Liu J, Chang J X, Zhang W G. Groundwater Level Dynamic Prediction Based on Chaos Optimization and Support Vector Machine[C]// International Conference on Genetic & Evolutionary Computing. IEEE Computer Society, 2009:39-43.
[14] Tong S, Koller D. Support vector machine active learning with applications to text classification[M]. JMLR.org, 2002.
[14] Suliman A, Nazri N, Othman M, et al. Artificial neural network and support vector machine in flood forecasting: A review[C]// 2013.
[15] Kanungo T, Mount D M, Netanyahu N S, et al. An efficient k-means clustering algorithm: analysis and implementation[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2002, 24(7):881-892.
[16] Jain A K. Data Clustering: 50 Years Beyond K-means[M]// Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2008:3-4.
[17] Dunn J C. A fuzzy relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters[J]. Journal of Cybernetics, 1973, 3(3):32-57.
[18] Gentle J E. Finding Groups in Data: An Introduction to Cluster Analysis. by L. Kaufman; P. J. Rousseuw[J]. Wiley, 2005.
[19] Huang Z. Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values[J]. Data Mining & Knowledge Discovery, 1998, 2(3):283-304.
[20] Ball G H, Hall D J. A clustering technique for summarizing multivariate data.[J]. Systems Research & Behavioral Science, 2010, 12(2):153-155.
[21] Arthur D, Vassilvitskii S. k-Means++: the advantages of careful seeding, in: SODA ‘07[J]. Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2007, 11(6):1027-1035.
[22] Rumelhart, David E, Hinton, et al. Learning representations by back-propagating errors[J]. 1986, 323(6088):399-421.
[23] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11):2278-2324.
[24] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc. 2012:1097-1105.
[25] Zeiler M D, Fergus R. Visualizing and understanding convolutional networks (2013)[J]. arXiv preprint arXiv:1311.2901, 2013.
[26] Szegedy C. Robust and fast model fitting by adaptive sampling: U.S. Patent 8,756,175[P]. 2014-6-17.
[27] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.
[28] Huang G, Liu Z, Maaten L V D, et al. Densely Connected Convolutional Networks[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2017:2261-2269.
[29] Ding L, Li H, Hu C, et al. AlexNet Feature Extraction and Multi-Kernel Learning for Object-Oriented Classification [J]. ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2018: 277-281.
[30] Ballester P, Araujo R M. On the performance of GoogLeNet and AlexNet applied to sketches[C]// Thirtieth AAAI Conference on Artificial Intelligence. AAAI Press, 2016:1124-1128.
[31] Eitz M, Hays J, Alexa M. How do humans sketch objects?[M]. ACM, 2012.
[32] Jiao J, Zhang F, Zhang L. Remote Sensing Estimation of Rape Planting Area Based on Improved AlexNet Model[J]. Computer Measurement & Control, 2018.
[33] Kamal U, Rafi A M, Hoque R, et al. Application of DenseNet in Camera Model Identification and Post-Processing Detection[J]. 2018.
[34] Zhou, Bo & Li, Yuemeng & Wang, Jiangcong. (2018). A Weakly Supervised Adaptive DenseNet for Classifying Thoracic Diseases and Identifying Abnormalities. Remote Sens. 2018, 10, 779.
[36] Negrinho R, Gordon G. DeepArchitect: Automatically Designing and Training Deep Architectures[J]. 2017.
[37] Zoph B, Vasudevan V, Shlens J, et al. Learning Transferable Architectures for Scalable Image Recognition[J]. 2017.
[38] Zoph B, Le Q V. Neural Architecture Search with Reinforcement Learning[J]. 2016.
[39] Chen Y, Zhang Q, Huang C, et al. Reinforced Evolutionary Neural Architecture Search[J]. 2018.
[40] Cai H, Chen T, Zhang W, et al. Efficient Architecture Search by Network Transformation[C]// AAAI Conference on Artificial Intelligence. 2018.
[41] Cai H, Yang J, Zhang W, et al. Path-Level Network Transformation for Efficient Architecture Search[J]. 2018.
[42] Liu H, Simonyan K, Yang Y. DARTS: Differentiable Architecture Search[J]. 2018.
[43] Barreiro E, Munteanu C R, Cruzmonteagudo M, et al. Net-Net Auto Machine Learning (AutoML) Prediction of Complex Ecosystems[J]. Scientific Reports, 2018, 8.
[44] Mohr F, Wever M, Höllerman E. ML-Plan: Automated machine learning via hierarchical planning[J]. Machine Learning, 2018, 107(8-10):1495-1515.