Network Traffic Classification Based on Deep Learning

Junwei Li[1,2], Zhisong Pan[1]*
1Institute of Command Control Engineering, Army Engineering University,
Nanjing 210007, P. R. China
[lijw@188.com, panzs@nuaa.edu.cn]
2Institute of Computer and Information Engineering, XinXiang University,
Xinxiang 453003, P. R. China
*Corresponding author: Zhisong Pan

Received August 3, 2019; revised November 20, 2019; revised February 2, 2020;
revised February 27, 2020; accepted March 18, 2020; published November 30, 2020

Abstract

As the network goes deep into all aspects of people's lives, the number and the complexity of network traffic is increasing, and traffic classification becomes more and more important. How to classify them effectively is an important prerequisite for network management and planning, and ensuring network security. With the continuous development of deep learning, more and more traffic classification begins to use it as the main method, which achieves better results than traditional classification methods. In this paper, we provide a comprehensive review of network traffic classification based on deep learning. Firstly, we introduce the research background and progress of network traffic classification. Then, we summarize and compare traffic classification based on deep learning such as stack autoencoder, one-dimensional convolution neural network, two-dimensional convolution neural network, three-dimensional convolution neural network, long short-term memory network and Deep Belief Networks. In addition, we compare traffic classification based on deep learning with other methods such as based on port number, deep packets detection and machine learning. Finally, the future research directions of network traffic classification based on deep learning are prospected.

Keywords: Traffic classification, deep learning, convolution neural network, stack auto encoder, long short-term memory network

http://doi.org/10.3837/tiis.2020.11.001 **ISSN : 1976-7277**

1. Introduction

With the rapid popularization and development of computer network, especially mobile internet, the antenna of network has penetrated into all aspects of people's lives. According to the "Statistical Report on the Development of Internet in China" issued by China Internet Network Information Center (CNNIC) on February 28, 2019, By the end of December 2018, the number of netizens had reached 829 million. There were 56.53 million new netizens in the year, with a penetration rate of 59.6% [1]. **Fig. 1** shows the trend of change in the last decade.

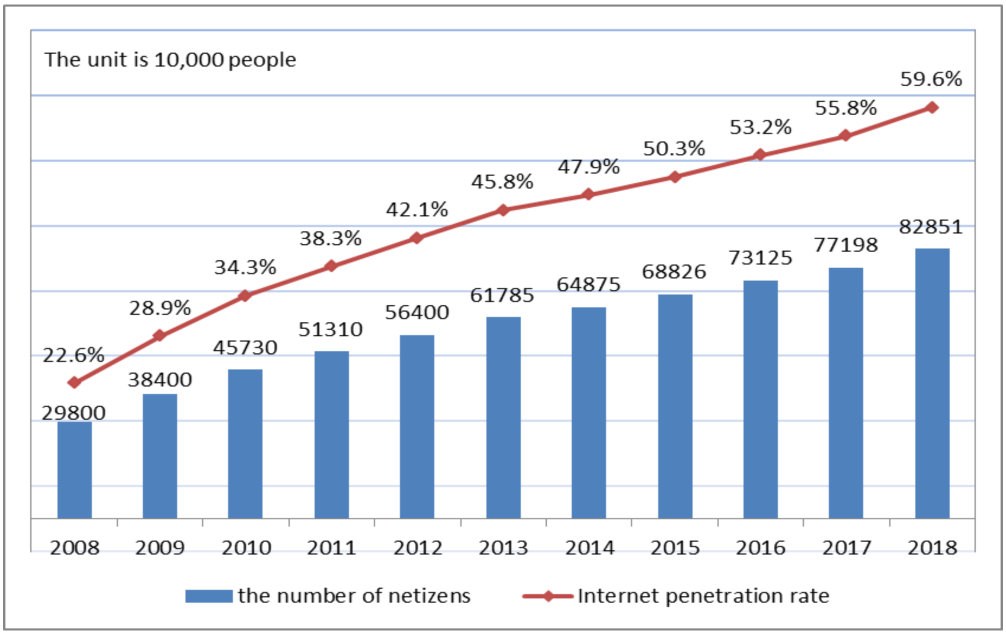

Fig. 1. Number of Internet users and Internet penetration

As a result, there are more and more network applications, which will generate a large number of different kinds of network traffic in the process of information communication and data transmission. Various kinds of network traffic bring great challenges to network management and security. Therefore, network traffic classification, identification of different kinds of traffic or encrypted traffic[2], detection of malicious traffic [3], discovery of network attacks or intrusions [4], and improvement of network operation efficiency become more and more important. They are prerequisites for network management, network operation and network security [5].

Research on network traffic classification has been developing continuously at home and abroad. From the traditional classification based on port number and deep packets inspection to the classification based on machine learning such as Support Vector Machine (SVM) and decision tree, various kinds of classification research emerge in endlessly. There are also some surveys about traffic classification, comparing and analyzing these classification methods.

With the rise of deep learning, classification methods based on convolution neural network (CNN) and recurrent neural (RNN) network have emerged. At present, there are many papers on network traffic classification, mostly based on traditional classification methods and machine learning classification methods. Also, there are some papers on improving the deep learning algorithm and improving the performance, even there are lots of papers about different deep learning algorithms. However, to the best of our knowledge, there is a lack of review papers of traffic classification based on deep learning. In view of this, we systematically summarizes and compares traffic classification based on deep learning such as stack autoencoder (SAE), one-dimensional convolution neural network (1D-CNN), two-dimensional convolution neural network (2D-CNN), three-dimensional convolution neural network (3D-CNN) and long short-term memory network (LSTM) from the research background, basic concepts, research progress, classification objects, classification methods and evaluation criteria. Emphasis is placed on the main ideas, preprocessing methods, using models, technical implementation and classification results of various methods. It provides an effective help for people to quickly understand the network traffic classification based on deep learning, to further study the network traffic classification method, to improve the classification performance and to expand the application scope of classification.

2. Research Objects and Evaluation Criteria

2.1 Classification Objects

When two hosts which follow the same network protocol communicate or transmit data, each layer of the network architecture will generate its own network traffic. According to the size of granularity, traffic can be divided into flow level, packet level, host level and session level, in which flow level can be divided into unidirectional flow and bidirectional flow [6]. When two hosts communicate with each other, the continuous data packets generated by the same pair of network services or applications form network flows. Usually a flow is uniquely identified or represented by a quintuple <source IP address, source port number, destination IP address, destination port number, transport layer protocol> [7]. These different granularity traffic is the object of traffic classification. First of all, traffic classification needs to determine the object of classification [8]. The descriptions of traffic classification objects at each level are listed in **Table 1**.

Table1. Traffic classification objects

Classification object	**Focus**	**Representation features**	**Extension of application**
Unidirectional flow level	Characteristics and arrival process of flow from the same direction	Duration of flow and Number of stream bytes	★★★★
Bidirectional flow level	Characteristics and arrival process of flow from the anti direction	Duration of flow and Number of stream bytes	★★★
Packet level	Characteristics and arrival process of packets	Distribution of grouping size and packets arrival time interval	★★
Host level	Connection mode between hosts	Connection Degree and Port Number of Communication Flow with Host	★
Session level	Characteristics and arrival process of sessions	Session bytes and session duration	★

2.2 Evaluation Criteria

To compare and analyze different kinds of classification methods, first of all, we should choose the right evaluation criteria. Classification accuracy and error rate are generally used to evaluate the classification results [9-10]. However, due to the uneven distribution of some samples or the consideration of multi-angle evaluation of classification results, performance metrics such as precision and recall will also be used [11]. The evaluation criterias are described as follows.

Accuracy = the number of samples correctly detected / the total number of samples, as shown in formula (1).

Given training set D= {(x_1,y_1) , (x_2,y_2), … , (x_m ,y_m)}, where x_i∈(x_1,x_m) is the sample to be classified while y_i∈(y_1,y_m) is its real classification, and f is the classifier, then the accuracy (ACC) can be expressed as follows.

$$ACC(f;D)=\frac{1}{m}\sum_{i=1}^{m}\prod\left(f\left(x_i=y_i\right)\right) \tag{1}$$

Let TP denote the number of samples labeled as a class and also belong to a class actually. Let FP denote the number of samples labeled as a class but not belong to the class actually. Let FN denote the number of samples not labeled as a class but belong to the class actually.Let TN denote the number of samples not labeled as a class and also not belong to the class. The precision (PRE) and recall (REC) can be described as formula (2) and formula separately(3).

$$PRE=\frac{TP}{TP+FP} \tag{2}$$

$$REC=\frac{TP}{TP+FN} \tag{3}$$

A good traffic classification method should have high accuracy, precision and recall at the same time [12]. In some cases, however, the accuracy and recall rates are contradictory. When you can't give consideration to both, you should choose which is more important according to the application scenario. When describing the coverage of test results, more emphasis can be placed on the recall rate. For example, screening cancer patients through physical examination should try to find all cancer patients as far as possible, so the recall rate is more important than the precision rate. But when describing the authenticity of test results, more emphasis is placed on precision. For example, most of the alarms of electric bicycle are just because that the electric bicycle is accidentally touched. If you look at the alarm every time you hear it, it will waste too much useless energy. Therefore, this scene should pay more attention to the precision.

3. Research Progress of Traffic Classification

Up to now, there have been many studies on traffic classification [13-15], which aim at different classification objects and methods as well as output classification results [16-17]. According to the order of occurrence and the adoption of technology, the traffic classification methods can be divided into three stages. Firstly, flow classification based on traditional methods [18], including classification based on port number [19], depth packets detection [20], etc. Secondly, traffic classification based on statistical features[21] or host behavior characteristics[22] and online learning[23], which are usually combined with machine learning. Thirdly, traffic classification based on deep learning mainly adopts deep neural network such as CNN[24] and RNN [25-26]. The overall research progress and contents are shown in **Fig. 2**.

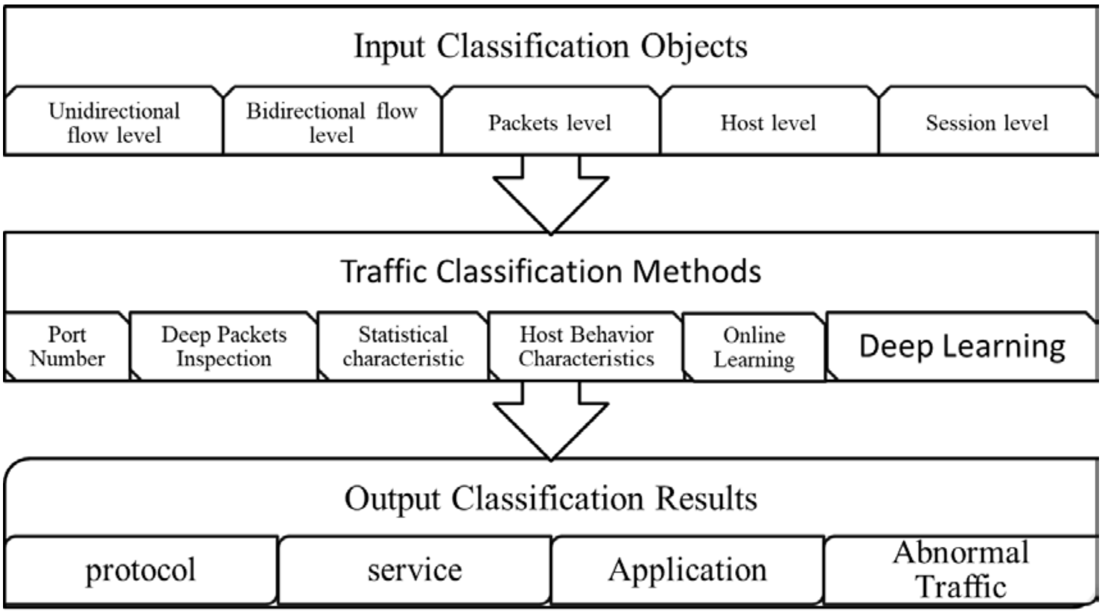
Input Classification Objects
Unidirectional flow level
Bidirectional flow level
Packets level
Host level
Session level
Traffic Classification Methods
Port Number
Deep Packets Inspection
Statistical characteristic
Host Behavior Characteristics
Online Learning
Deep Learning
Output Classification Results
protocol
service
Application
Abnormal Traffic

Fig. 2. Research progress of network traffic classification

3.1 Traffic Classification Based on Port Number

Internet Assigned Numbers Authority (IANA) publishes a list of common port numbers[27], assigning a fixed port number to each common application or service. Therefore, the protocol and application using the port number can be determined simply by querying the corresponding table of the port number through the port field of the packet header. The method of classification based on port number is simple and fast, and it is one of the main methods used in the early stage of traffic classification[28-29]. However, with the development of the Internet, the accuracy of this classification method is getting lower and lower. It makes the traffic classification based on port number ineffective. However, this method has not completely faded out of the stage of history, and usually appears as an important supplement to other classification methods.

3.2 Traffic Classification Based on Deep Packets Inspection

With the decline of classification accuracy of methods based on port number, people turn to other alternative methods, and Deep Packets Inspection based classification methods come out [30]. As can be seen from the name, this method not only detects the port number, but also takes the header and content of the packets as the detection object.

The classification based on deep packets detection has two prerequisites. On the one hand, the protocol for generating network traffic to be classified is known and key fingerprints can be extracted from it. On the other hand, the load of traffic packets is visible and can be compared with all fingerprints byte by byte [31-32]. Compared with the classification based on port number, the method based on deep packet detection is more accurate and reliable, and early recognition can be achieved by recognizing the first few data of the packet head.

3.3 Traffic Classification Based on Statistical Characteristics

In order to solve the defect of deep packet detection, a classification method based on statistical theory is proposed, which expands the classified objects from the data packets of the flow to the whole network flow. Therefore, this method is also called the deep flow detection based classification method [33].It considers that each flow generated by one application has its own statistical characteristics in terms of the time interval of grouping in the stream and the number of bits produced per second. It analyses and chooses the statistical characteristics of different streams, and then combines machine learning algorithms such as Support Vector Machine , Bayesian Network or Decision Tree to learn from the stream. In statistical features, we learn the classification method to distinguish the application of different flow, and achieve traffic classification [34].Unlike the deep packets inspecton, the statistical classification method takes the whole network flow as the research object, and does not need to pay attention to the contents of the data packets in the flow.

3.4 Traffic Classification Based on Host Behaviors

The classification based on host behaviors also realizes network traffic classification through machine learning. Unlike the classification based on statistical characteristics, it focuses on the behavior information of host communication. It is believed that all kinds of host applications will have different behavior patterns from other applications, which is the characteristics of the algorithm to learn. Its advantages, disadvantages and application occasions are roughly the same as those of classification based on statistical characteristics [35].

3.5 Traffic Classification Based on Online Learning

Although the traffic classification based on machine learning such as statistics and behavior have achieved good results, with the popularization of the network and the development of technology, especially the implementation of the national policy of speed-up and fee-reduction, the speed and scale of the network have been greatly increased, and the data generated are in a geometric explosion state. In this era of data explosion, the classification method based on traditional machine learning is challenged greatly [36]. The improvement of machine operation ability is slow, the training time of model becomes longer, and the classification accuracy decreases. In view of these requirements, a classification method based on online learning is proposed.

Unlike previous algorithms, it only processes one or a batch of data at a time according to the time series, and the processed data will not be computed any more, and only uses incremental data to learn the model. Compared with the previous batch processing methods, it runs faster and achieves higher efficiency. It is more suitable for processing large-scale time series data and identifying network traffic in real time and online.

4. Traffic Classification Based on Deep Learning

Machine learning algorithms such as SVM and decision tree belong to shallow algorithm in essence. In the case of limited data samples, the generalization ability of the algorithm is not high, and even can not express very complex non-linear functions [37]. Deep learning is a perceptron network with multiple hidden layers[38]. Because of the quantity of layers, it has a stronger ability to fit complex functions, and each layer can extract different features, and form higher-level features through combination. The ability of feature extraction is more reliable and accurate than that of manual extraction by experts.

Unlike classification method based on traditional machine learning depending on feature extraction artificially, classification based on deep learning enables the neural network to learn how to extract traffic features by itself, so as to achieve end-to-end traffic classification, i.e., the raw network traffic is input and output is the classification of traffics or corresponding services and applications, as shown in **Fig. 3**.

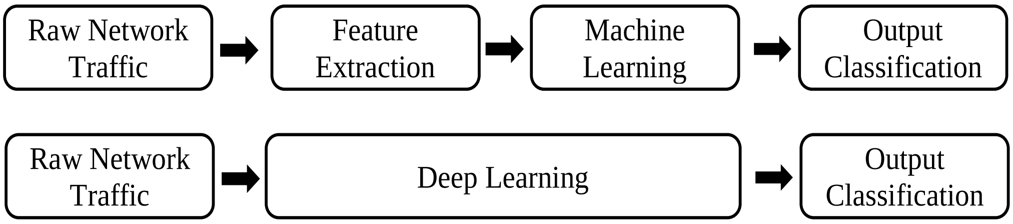

Fig. 3. Comparison traffic classification based on machine learning with deep learning

The first breakthrough in deep learning is CNN network, which is mainly used in image recognition, face recognition and audio and video data expressed in time-domain and frequency-domain. Later, RNN and LSTM network appeared, which are used in speech recognition, natural language processing and other fields, and good results have been achieved.

These implementation frameworks of traffic classification based on deep learning are basically the same. They can be divided into training and testing.

4.1 Traffic Classification Based on Stack Autoencoder

Wang Zhanyi proposed the application of deep learning in network traffic classification firstly, which was the originator of the application of deep learning technology to end-to-end network traffic classification. Later, many articles were inspired by it and carried out research, and proposed different methods [39]. But he did not elaborate on the details. He simply applied the Stack Autoencoder (SAE) model to traffic identification and classification, and gave the experimental results which were better than the classical methods directly. SAE is formed by stacking multiple autocoders (AE) together, and the model structure of AE is shown in **Fig. 4**.

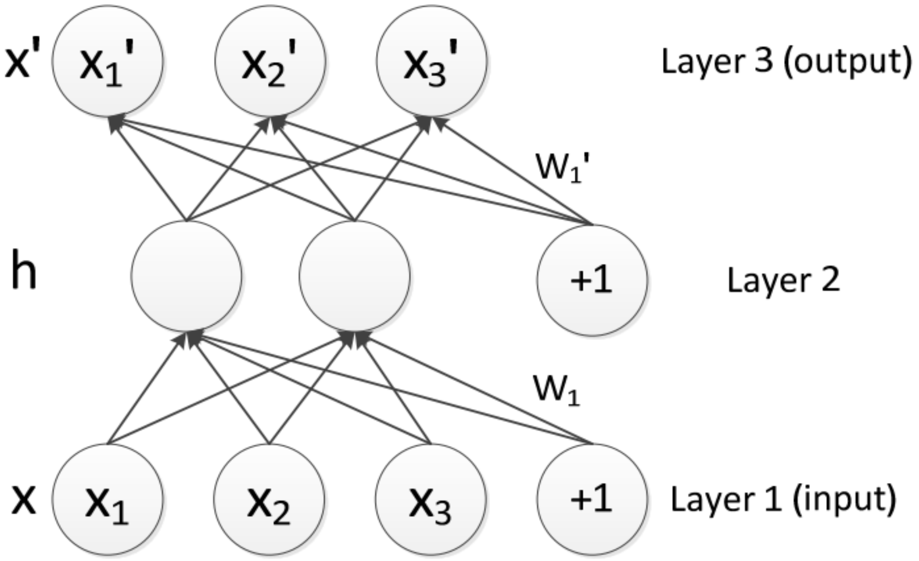

Fig. 4. AE model

AE can be regarded as a variant of multi-layer perceptron[40], which reconstructs input data through hidden layer to form output, and requires output data to be consistent with input data as far as possible, so as to find the potential relationship between them to form a model. It generally consists of input layer, hidden layer and output layer. The dimensions of input layer and output layer are identical, and the dimensions of hidden layer are usually smaller than those of input layer. Because it can accept unlabeled data as input, AE is an unsupervised learning model. Since labeled data is often more difficult to obtain than unlabeled data in the real world, this is an obvious advantage.

As shown in **Fig. 5**, SAE is a type of deep neural network, in which multiple AE are stacked and the hidden layer of the first AE is used as the input layer of the second one.

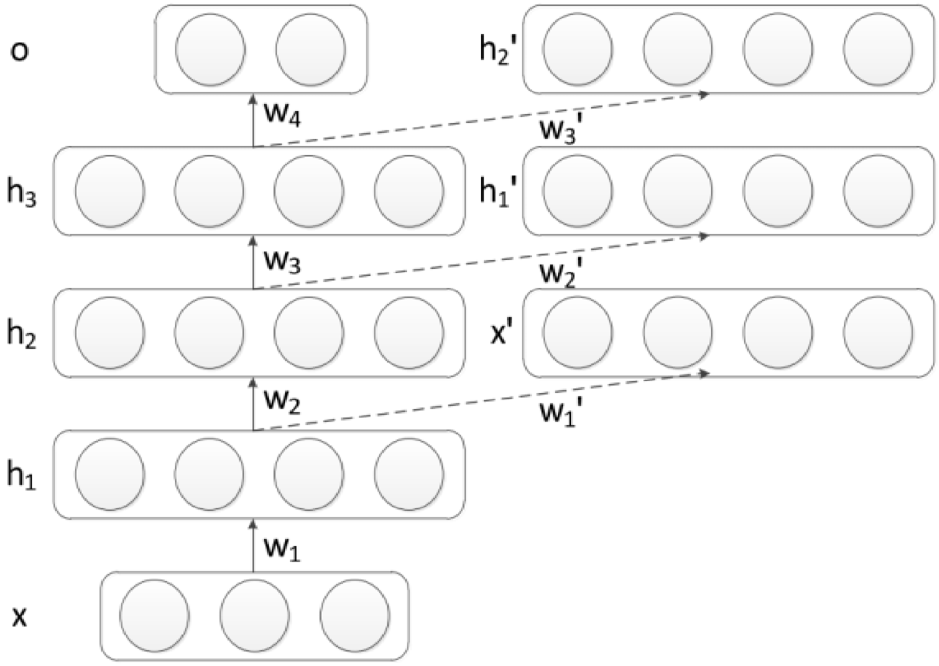

Fig. 5. SAE model

To the best of our knowledge, this is the first time that deep learning is used in traffic classification. Compared with machine learning based traffic classification method, it no longer needs manual feature selection, but achieves end-to-end traffic classification, that is, the input is the original traffic data, and the output is the result of traffic classification, which saves the workload greatly.

The classification results are as follows: the recognition rate of known traffic is more than 90%, and even 100% for individuals., reaching the level of practical application. And the recognition rate of unknown traffic is 83%, which is very high at that time.

Data preprocessing is to take the first 1000 bytes of each flow, normalize each byte from [0-255] to [0-1] and input it directly into SAE.

The data set contains a total of 300 thousand pieces of data collected from the real data of 360 company network.

The main idea is to regard the byte data of the flow as the pixels of a picture or the words of a document, and use S AE model to learn and implement classification.

4.2 Traffic Classification Based on Convolution Neural Network

Convolutional neural network developed earlier, and it was proposed by Yann LeCun, who won the 2018 Turing Award of the highest honor in computer field and is known as one of the troikas of artificial intelligence. The high performance of the network was verified in handwritten numeral recognition tasks. In the ImageNet Image Classification Competition, several champions have used this network structure with excellent performance. Compared with ordinary neural networks, it has three core improvements: local connection, weight sharing and pooling[41].

In artificial neural networks, every two nodes in different layers are connected, that is, full connection. As the complexity of fitting function increases, the number of neurons is increasing, and the number of connections is increasing geometrically, which leads to the extremely time-consuming training of the model. However, in convolutional neural networks, inspired by the local receptive field of biological nervous system, people begin to consider connecting only one node to some adjacent nodes of the previous layer instead of all nodes, which reduces the number of connections effectively. Weight sharing is that each neuron uses the same weight to connect with the previous neuron. It is also called filter or convolution core, which can extract a corresponding feature. In practical networks, multiple filters are often used to extract multiple features of input. This improved weight sharing reduces the number of parameters that need to be trained by multiple times. In order to further reduce the data scale, people have proposed pooling, which divides the feature map obtained by convolution into several small areas, and then replaces the original area with the maximum or average value, that is, the maximum or average pooling, so as to replace the whole area with one value in the original area. In addition to multiplying the amount of data and getting more representative features, pooling can also prevent the occurrence of over-fitting by dropping off part of the data, and even reduce the impact of noise and interference.

Traffic classification based on convolutional neural network is an end-to-end method, which can automatically select features from the input traffic data and obtain the weights and other parameters of the network model through training and learning. After the model is determined, traffic classification can be carried out. Since the raw traffic data does not meet the input requirements of convolutional neural network, all data need to be preprocessed first.

According to the different objects to be processed, convolutional neural networks can be divided into one-dimensional, two-dimensional and three-dimensional types. 1D-CNN is usually used to process temporal data such as speech and text. 2D-CNN is generally used to process images and audio and video data expressed in time-domain and frequency-domain, and 3D-CNN can process three-dimensional data such as video and stereo images. Most traffic classification methods based on convolutional neural networks use 2D-CNN [42], while a few use 1D-CNN [43] and 3D-CNN [44-45].

4.2.1 Traffic Classification Based on 2D-CNN

Main idea: The byte data of network traffic is converted into gray scale image in bytes, and the problem of traffic classification is transformed into the problem of picture classification.

Data Set: USTC-TFC 2016. Its data format is PCAP, a total of 3.71GB, including 10 kinds of malicious traffic collected by CTU and 10 kinds of normal traffic collected by professional simulation equipment BPS of IXIA company.

Data preprocessing: The original data of PCAB format is transformed into data of IDX format which can be directly processed by convolutional neural network, that is, the flow data is visualized, and then the convolutional neural network model is trained to classify the images. Firstly, cut the network flow into several traffic data of equal size, usually 1024 or 784 bytes as the standard. Then, considering the range of binary number that each traffic byte can represent, the pixel value of gray image is the same, which is 0-255, so the traffic data can be directly converted into pictures.

Through the gray scale image after transformation, we can see that the texture of transformed graphs from the same traffic is basically the same, and the ones from different traffic is very different. This provides a visual basis for further classification using 2D-CNN.

The model structure used is shown in **Fig. 6**, which is a general 2D-CNN .

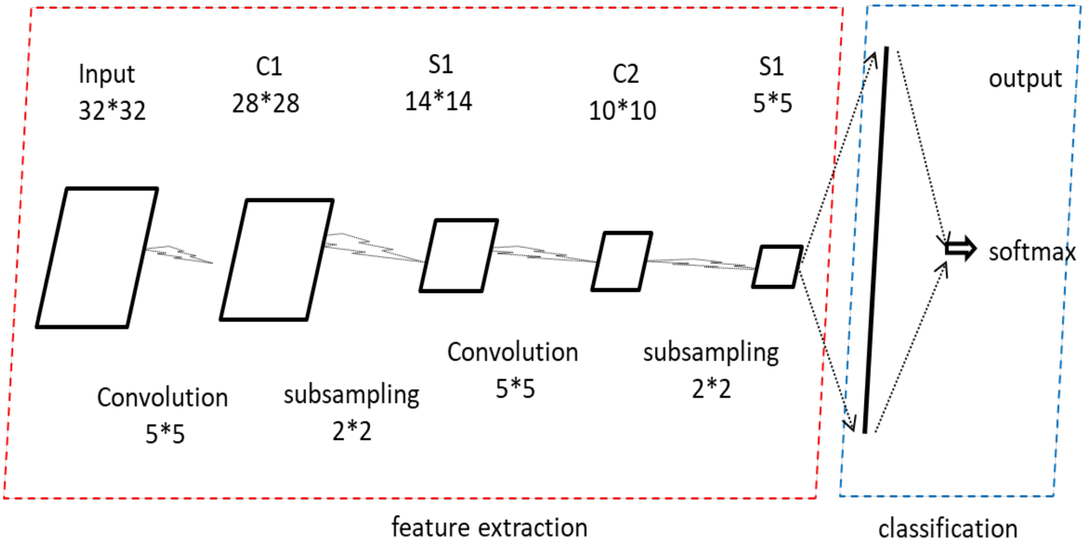

Fig. 6. Structure of 2D-CNN

The size of input data is 32*32, and the first convolution core size is 5*5. Then the size of data becomes from 28*28 to 14*14 through the pooling layer of 2*2. So far, the first convolution and pooling have been completed. Similarly, after the second convolution layer with a convolution core of 5*5, it becomes 10*10, and after the pooling of 2*2, it becomes 5*5. Finally, it flattens the whole connection layer into one-dimensional vectors, and classifies them through the softmax layer. This is an example based on the existing network structure, which can be changed according to the need in practical application.

Through the above steps, the classification and recognition of raw network traffic will be transformed into the training of 2D-CNN for classification by taking the traffic bytes as the gray value of the image, and the latter has a very high success rate in image recognition. Through building a series of links such as model, training and testing, traffic classification is realized by convolution neural network.

4.2.2 Traffic Classification Based on 1D-CNN

Considering that the byte data of traffic is linearly arranged, Wangwei et al. proposed using 1D-CNN to classify network traffic, and achieved better results than 2D-CNN.

Main idea: The byte data of traffic is regarded as characters in natural language, and the document classification technology in natural language processing is used to realize traffic classification.

Data Set: ISCX VPN-NONVPN DATASET. There are 12 kinds of data, including 6 kinds of routine encrypted traffic and 6 kinds of traffic encapsulated by VPN protocol.

Data preprocessing: The fixed length byte data of the intercepted traffic is directly used as input of 1D-CNN.

The overall structure of the model is similar to that of the 2D-CNN. The difference is that the input data, the convolution core and the output data are all one-dimensional, and the size of the model has changed, as shown in **Fig. 7**.

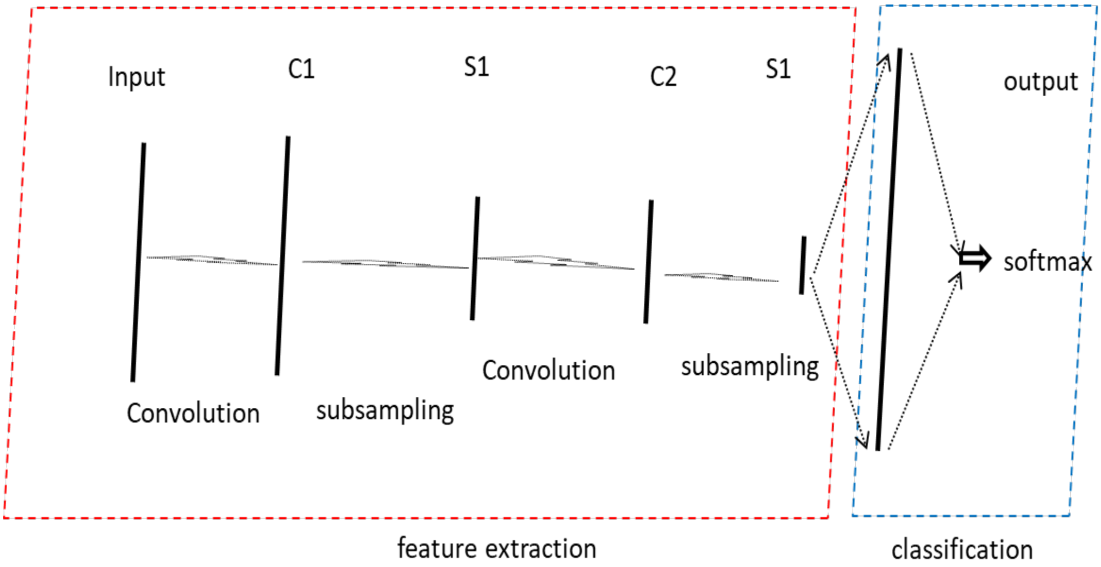

Fig. 7. Structure of 1D-CNN

Classification results: Different from using 2D-CNN to convert traffic into images, 1D-CNN classifies network traffic by analogy of bytes, frames, data packets and flow into letters, words, sentences and articles. It realizes the classification of network traffic by text categorization, and the classification results are better than 2D-CNN.

4.2.3 Traffic Classification Based on 3D-CNN

Considering that the network traffic data has the characteristics of time series, Chen Yexin and other researchers proposed using 3D-CNN to classify network traffic, in order to make better use of the time series characteristics which are not used by one-dimensional and two-dimensional neural networks. Compared with 1D-CNN and 2D-CNN, 3D-CNN is basically the same in overall thought and data flow, but the difference is data preprocessing, model structure and other aspects, the classification effect is also slightly improved.

The main idea is to convert the packet header data of traffic bytes into pictures, and add the dimension of time characteristics to form three-dimensional data as input, which is equivalent to converting traffic data into multi-frame gray images in video processing, splicing one frame to one large image, and learning the characteristics of these large images by using 3D-CNN model to improve learning efficiency.

Data Set: Five malicious traffic collected by CTU and five normal traffic collected by BPS, the professional simulation equipment of IXIA Company, are selected from USTC-TFC2016 to form a new data set, which contains 65,000 data streams. In order to verify the generalization performance of the model, tests are carried out on ISCX VPN-NONVPN DATASET, and the results are better than 1D-CNN and 2D-CNN.

Data preprocessing: Firstly, the raw network traffic is divided into different network flows, and the same number of packets are extracted from the front of each flow, which is the same as 1D-CNN and 2D-CNN, but the following steps are different with them.

Firstly, the same length of the front is reserved for each packet, and the sequence of packets in each flow is treated as a dimension. There is no such step in 1D-CNN and 2D-CNN.

Secondly, the header data in each packet is converted into two-dimensional data by one-hot encoding, and the time dimension is added to form three-dimensional data as input, instead of converting traffic bytes into two-dimensional gray image directly as input data of 2D-CNN. The pre-processed data is equivalent to multi-frame gray image in video processing. That is to say, this method is equivalent to splicing a frame-by-frame image into a large image, so that the model can learn the characteristics of these large images, so as to improve the learning efficiency and classification accuracy.

The model used: The network traffic classification system model based on 3D-CNN still consists of convolution layer, pooling layer, full connection layer and output layer, but the data scale and dimension have changede in each layer.

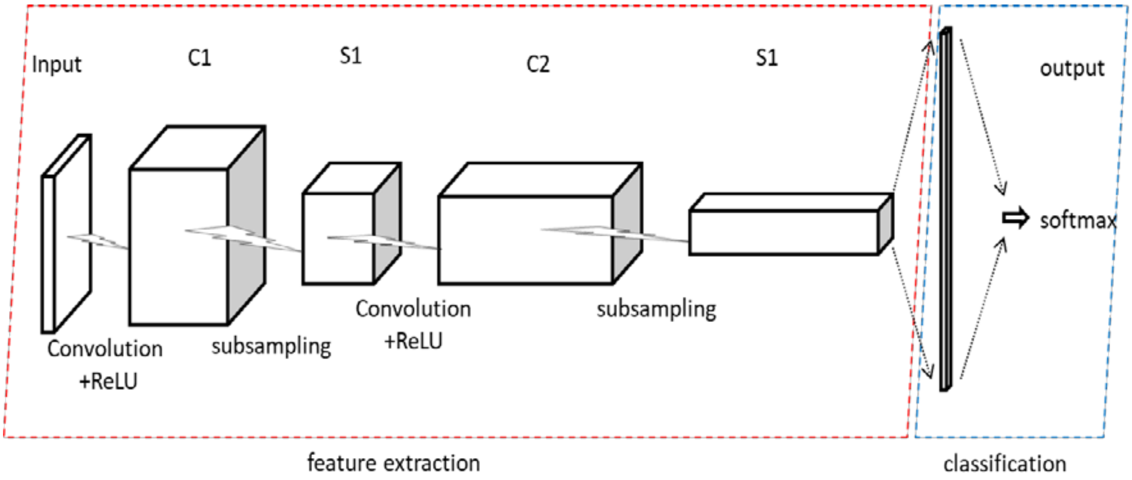

Fig. 8. Structure of 3D-CNN

As shown in **Fig. 8**, lightning arrows represent layers, while cubes and rectangles represent the output of each layer. The first and third layers are convolution layers, and the output is three-dimensional data. The second and fourth layers are pooling layers, and the output data is half the size of the input. The fifth and sixth layers are full connection layers, flattened and expanded to output one-dimensional data, and some data are dropped out randomly. Finally, the last layer is the soft max output layer, which judges the traffic type and outputs results

Classification results: Compared with 2D-CNN, 3D-CNN achieves higher precision and recall. Compared with the classification method using RNN to process time characteristics, 3D-CNN reduces the quantity of parameters used and the computational complexity obviously on the premise of guaranteeing accuracy.

4.2.4 Comparison of Classification Methods Based on CNN with Different Dimensions

It can be seen that although the main ideas and general processes of traffic classification based on CNN in different dimensions are basically the same, they are still different from each other at the specific implementation level. Generally speaking, there are four differences.

First, the specific ideas of each method are different, such as: some directly input traffic byte data, some convert traffic byte into image or video frame data, etc.

Secondly, different classification methods use different models in training and testing, such as SAE, CNN, RNN or LSTM, and among which CNN has 1D-CNN, 2D-CNN and 3D-CNN.

Thirdly, different models require different formats of input data, which makes the data preprocessing stage have different processing methods.

Fourthly, the data sets used by each classification method are different. There are data collected by themselves, such as the enterprise network data of 360 company, or datasets published on the Internet, such as ISCX VPN-NONVPN DATASET.

The following is a detailed comparison and analysis of the main methods of network traffic classification based on deep learning from several key aspects, such as main ideas, data sets, data preprocessing methods, training and testing models and classification results, as shown in **Table 2**.

Table 2. Comparison of classification based on CNN

Contrastive focus	**1D-CNN**	**2D-CNN**	**3D-CNN**
Main Idea	The byte data of traffic is regarded as characters in natural language, and the document classification technology in natural language processing is used to realize traffic classification	The byte data of network traffic is transformed into gray scale image in bytes, and the problem of traffic classification is transformed into the problem of image classification	The header data of traffic bytes is converted into images, and the time dimension is added to form 3-D data, which is equivalent to multi-frame gray image in video processing
DataSet	ISCX VPN-NONVPN DATASET	USTC-TFC2016	USTC-TFC2016 and□ISCX VPN-NONVPN DATASET
Verification method	10 fold cross validation	10 fold cross validation	5 fold cross validation
Data preprocessing	Direct use of traffic byte data	Converting traffic byte data into gray scale image	The traffic byte data is transformed into gray scale image, and the time characteristic is added to form 3-D video data.
Model Structure	1 dimension	2 dimension	3 dimension
Classification effect	All of them have achieved practical application effect. On the whole, 3D-CNN is better than 1D-CNN than 2D-CNN.		

4.3 Traffic Classification Based on Long Short-Term Memory

In addition to traffic classification method based on convolutional neural network, some researchers use the Recurrent Neural Network (RNN) or its variant long short-term memory network (LSTM) [46] to learn the temporal characteristics of network traffic, and then classify them.

In CNN, the signals of neurons cannot be transmitted on the same layer, but only on the upper layer, which leads to non-connection between the processing of samples at different time points. Therefore, it is difficult to learn changes in time series. However, natural language processing, speech recognition and so on need to focus on the processing of time series property [47], so recurrent neural network is generated, whose model structure is shown in **Fig.**

9.

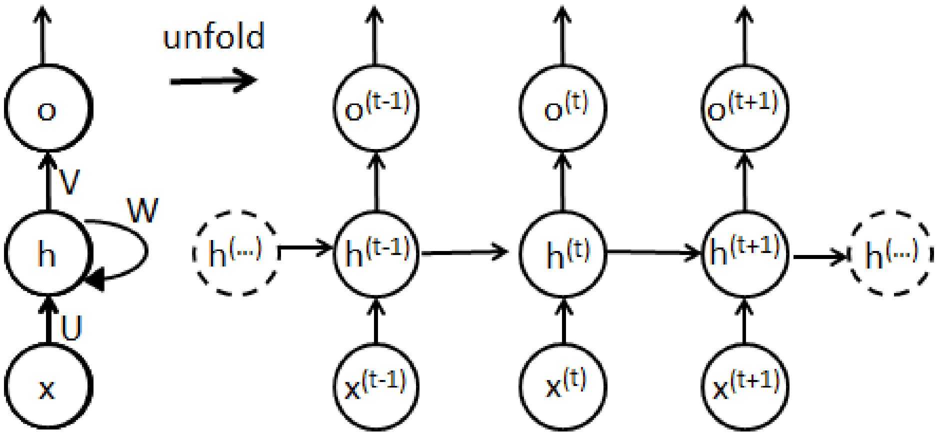

Fig. 9. Structure of recurrent neural network

However, RNN did not pay enough attention to temporal attributes for a long time. And its variant LSTM introduces forgetting gate to control whether the output of the previous several moments contributes to the output of this moment. That is, the output of RNN is only determined by the input of this moment, while the output of LSTM is also related to the previous inputs and controlled by forgetting gate. Therefore, its ability of learning temporal features is stronger than CNN [48]. For example, in learning the sentence "I love you!", RNN usually only uses the current information to judge the output information, ignoring the previous semantic information, while LSTM will judge all the information including the previous several information "I love you!" in a comprehensive way, so the output vector data is more accurate. [49]. In reality, LSTM is generally used to integrate the two networks and extract temporal feature information at different levels to achieve better classification results[50].

The main idea is that the raw traffic data is directly input, and LSTM, which is divided into two stages, is responsible for learning and extracting features at different levels, and then combining them for classification.

Data Set: ISCX2012 Data Set. It was published by the University of New Brunswick in Canada, consists of normal traffic and four malicious traffic, which contains seven days of data.

Data Processing: Firstly, LSTM is used to learn the time series characteristic data at the packet level, and the direction and size of each packet can be obtained to form a vector about the packet. Then, on the basis of the vector formed, we can learn the relationship between these packets, and form a vector about the network flow. Finally, two vectors are synthesized. The LSTM learning results of each stage can get more comprehensive attributes of network traffic, and then classify the output using softmax [51].

The model used is shown in **Fig. 10**.

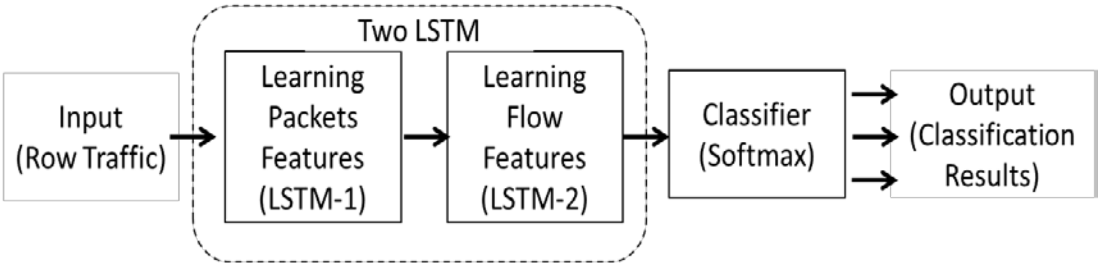
Two LSTM
Input (Row Traffic)
Learning Packets Features (LSTM-1)
Learning Flow Features (LSTM-2)
Classifier (Softmax)
Output (Classification Results)

Fig. 10. Traffic classification based on LSTM

Classification results: Using LSTM to classify traffic, the experimental data can also achieve the actual usable effect, but the network structure is more complex than SAE and CNN. And there are more parameters to learn while the computational complexity is higher.

4.4 Traffic Classification Based on Deep Belief Networks

The original Deep Belief Networks (DBN) model was proposed by Hinton [51] which can not only realize the automatic learning of characteristics but also learn the essential features that characterize the data and overcome the difficulty in the training through the method of layer-by-layer initialization. The components of the DBN are Restricted Boltzmann Machines. The process of training DBN is done layer by layer. Its structure is shown in **Fig. 11**.

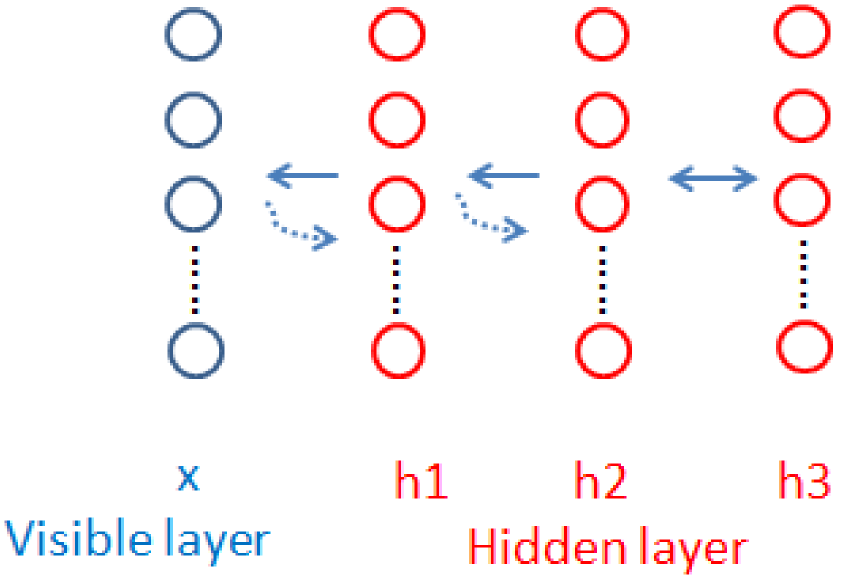

Fig. 11. Structure of DBN

In each layer, the data vector is used to infer the hidden layer, and this hidden layer is treated as the data vector of the next layer. The utilization of DBN technique for characteristic classification and recognition has obvious advantages.

In 2018, Hong Shao et al.[52] proposes proposes a network application classification model based on Deep Belief Networks and construct a DBN-based model suitable for network applications classification with the Tensorflow framework. the classification performances of this DBN-based model and the BP-based model are compared on the real data sets. The experimental results show that the applications classification model based on DBN has higher classification accuracy forP2P applications.

4.5 Comparison of Classification Methods Based on Deep Learning

In order to get the characteristics of various traffic classification methods based on deep learning more intuitively, we compares the various traffic classification methods based on SAE, CNN and LSTM from a higher level, as shown in **Table 3**.

Table 3. Comparison of Classification Methods Based on Various Types of Deep Learning

Classification Based on Deep Learning	**Supervisory/ Unsupervisory**	**Domain informatio n**	**Interception of Fixed-length Stream Data**	**Focus**	**Parameters to be learned**
Classification Based on SAE	Both are acceptable	No	No	Denoising and Dimension Reduction	Less
Classification Based on CNN	Supervisory	No	Yes	Spatial characteristics	Less
Classification Based on LSTM	Supervisory	Yes	No	Time characteristics	More

4.6 Comparison of Classification Based on Deep Learning and Others

According to the above analysis, compared with the traditional traffic classification methods based on features such as port number and deep packet detection, the traffic classification based on deep learning does not need to carry out port number detection and byte-by-byte comparison within the packet. It focuses on the characteristics of network traffic, and overcomes the shortcomings of not being able to deal with encrypted flow and unknown traffic.

Compared with the traffic classification based on machine learning such as statistical and host behavior characteristics, the traffic classification based on deep learning achieves an end-to-end classification method, which can automatically extract features by learning and get rid of the dependence on manual feature extraction.

Compared with the classification which needs a lot of calculation to get the result, the traffic classification based on deep learning can adopt the offline-training and online-identification method to further reduce the computational complexity, which can improve the efficiency of real-time classification. It is the most commonly used method for traffic classification at present. We use 1 for yes and 0 for No, and compare the characteristics of the five methods from four dimensions. As can be seen more intuitively from **Fig. 12**, the network traffic classification method based on deep learning can adopt the offline-training and online-identification method to further reduce the computational complexity, which can improve the efficiency of real-time classification. It is the most commonly used method for traffic classification at present. It is really a better way to be more intuitive with providing graph based comparison of every possible work.

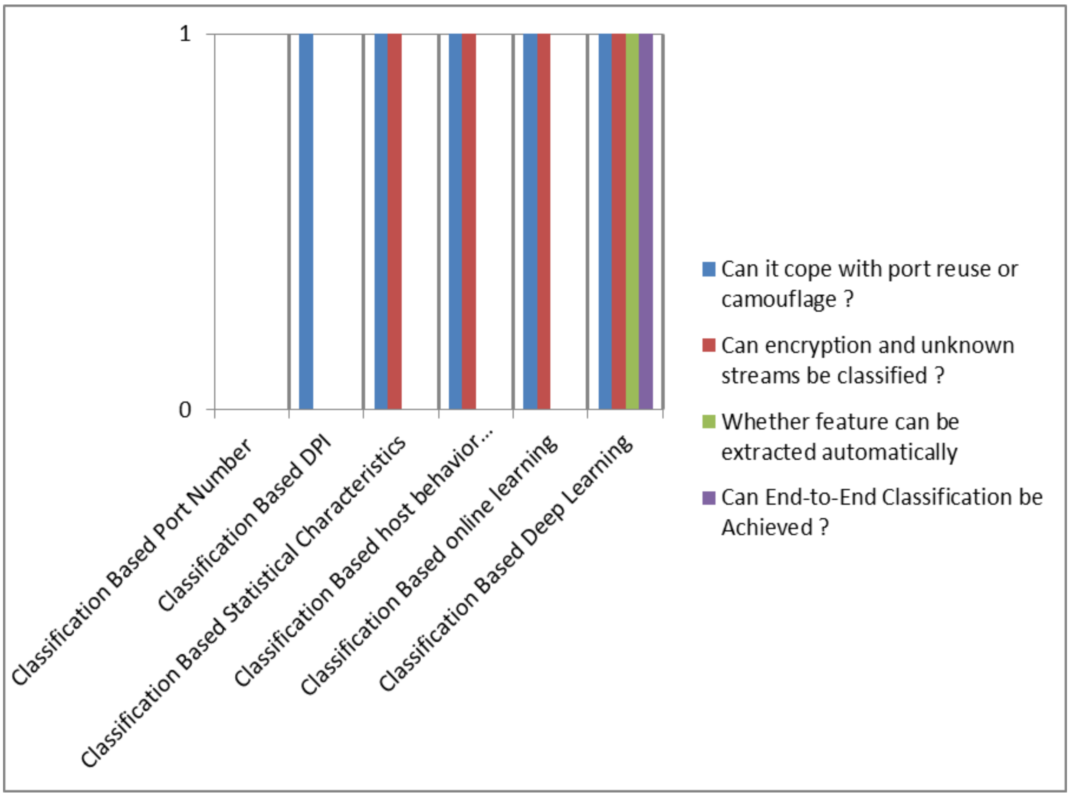

Fig. 12. Comparison of Classification Method Based on Deep Learning with Others

5. Problems and Prospects

The research of network traffic classification has entered a new stage of deep learning, and has made great progress. However, it also faces some problems and challenges, such as the lack of open data sets and the need to further improve the performance of the algorithm. The next step is to do further research in the following aspects.

(1) Forming an open, reliable and user-friendly dataset. In the field of network traffic classification, although there are open datasets such as DARPA1998 and KDD99, the development age is relatively long. Some studies believe that most of the traffic is not in line with the current environment. For convenience of research, many researchers grab network traffic by themselves, and use the existing filtering function of grabbing tools to label traffic classifications, partly for training, and the rest for testing. Most of these datasets are not open, and depend on the environment and the level of researchers'operation at that time. It is not conducive to comparison among different researchers. Therefore, the design and production of accurate, open and credible network traffic classification dataset is an important direction of future work.

(2) Combine with specific scenarios. In order to further improve the performance and efficiency of network traffic classification based on deep learning, in addition to efforts in data preprocessing, model training and algorithm improvement, another idea is to combine with specific application scenarios and make full use of the expertise in various fields. For example, in mobile internet, the traffic of mobile Internet is mostly based on HTTP protocol, and the difference between mobile traffic and computer network traffic is fully considered, and then

classified. In addition, the combination with the Internet of Things requires the use of its specialized domain knowledge for algorithm design, in order to improve the classification performance effectively. Also,most of the current researches on traffic classification suffers from a lack of future and open challenges sections that highlight the limitation and of current tools, solutions, and colored requirements for the future integrated attentions of such technologies in various areas like 5G, visual techniques, Drones, and smart cities.So Combine with specific scenarios such as 5G, visual techniques, Drones, and smart cities are also an important research direction.

(3) Unsupervised and semi-supervised methods are used for classification. It is good to train deep neural network with labeled data for classification, but in reality, labeled data is difficult to obtain. It is also an important research direction that how to use unsupervised and semi-supervised methods to train models to classify network flows with a small number of labels even without labels.

(4) The security of network traffic classification needs to be further improved. Rahim Taheri et al.[53] present an architecture for learning ipped data which rects our main focus in the malware detection system. However, these security defense researches are all aimed at machine learning algorithm. Whether the traffic classification based on deep learning will suffer from these security threats, and what defense measures should be taken, is still a worthy research direction.

(5) Optimization of the algorithm. When using deep learning algorithm for network traffic classification, we mainly consider the realization of an end-to-end classification method by using deep learning, which does not need to extract features manually, but does not consider the further optimization of the algorithm. In order to better the performance of the classifier, in addition to the classification performance of the algorithm, we should also consider the further optimization of the algorithm when designing the algorithm.

(6) The security of deep learning. For deep learning, although the parameter model is known, but the algorithm implementation process is inexplicable to people, it is a black box, so the security of deep learning can not be proved, and the security of network traffic classification based on deep learning also needs to be further considered.

(7) Lightweight implementation of the algorithm. Using deep learning algorithm to classify network traffic needs a lot of data, which consumes a lot of computing power and time. In the real network, considering the network delay and congestion control, we need to lighten the existing deep learning algorithm to meet the needs of some specific network scenarios.

6. Conclusion

With the increasing complexity of network traffic, traffic classification becomes more and more important. We analyzed the research background and progress of network traffic classification. Then, we summarize and compare traffic classification based on deep learning such as stack autoencoder, one-dimensional convolution neural network, two-dimensional convolution neural network, three-dimensional convolution neural network and long short-term memory network from their respective principles, preprocessing methods, using models, technical implementation and classification results. Moreover, the traffic classification based on deep learning is compared with methods based on port number, deep packets detection and machine learning. Finally, the future research direction and trend of network traffic classification based on deep learning are prospected.

With the continuous progress of deep learning, network traffic classification is full of opportunities and challenges, and there is still much room for improvement. It is hoped that we can provide a more comprehensive understanding of network traffic classification based on deep learning and reference for network planning, network management and network security, and expect that traffic classification can make further progress.

Acknowledgements

This research was partially supported by the National Key Research and Development Plan Key Projects of China under Grant No. 2017YFB0802800 and the National Natural Science Foundation of China Grant No. 61473149.

The authors would like to thank the Associate Editor and the referees for their help and valuable comments.

References

[1] China Internet Network Information Center, "Internet usage," *Statistical Report on the Development of Internet in China*, pp. 17-18, February 2019.

[2] Aceto G, Ciuonzo D, Montieri A, "Mobile encrypted traffic classification using deep learning," in *Proc. of 2018 Network Traffic Measurement and Analysis Conference (TMA)*, pp. 1-8, 2018. Article (CrossRef Link).

[3] Wang W, Zhu M, Zeng X, "Malware traffic classification using convolutional neural network for representation learning," in *Proc. of 2017 International Conference on Information Networking (ICOIN)*, pp. 712-717, 2017.

[4] Ducange P, Mannarà G, Marcelloni F, "A novel approach for internet traffic classification based on multi-objective evolutionary fuzzy classifiers," in *Proc. of 2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), IEEE*, pp. 1-6, 2017. Article (CrossRef Link).

[5] Shone N, Ngoc T N, Phai V D, "A deep learning approach to network intrusion detection," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 2, no.1, pp. 41-50, 2018. Article (CrossRef Link).

[6] Sun G, Liang L, Chen T, Xiao F, Fang L, "Network traffic classification based on transfer learning," *Computers & electrical engineering*, vol. 69, pp. 920-927, 2018. Article (CrossRef Link).

[7] Moore A W, Zuev D, "Discriminators for use in Flow-based classification," *Technical Report IRC-TR-04-028, Intel Research, Cambridge*, 2005. Article (CrossRef Link).

[8] Zhang J, Chen X, Xiang Y, "Robust network traffic classification," *IEEE/ACM Transactions on Networking (TON)*, vol. 23, no. 4, pp. 1257-1270, 2015. Article (CrossRef Link).

[9] He G, Yang M, Luo J, "A novel application classification attack against Tor," *Concurrency and Computation: Practice and Experience*, vol. 27, no.18, pp. 5640–5661, 2015. Article (CrossRef Link).

[10] McGaughey D, Semeniuk D, Smith R, "A systematic approach of feature selection for encrypted network traffic classification," in *Proc. of the 2018 Annual IEEE International Systems Conference, Piscataway, NJ: IEEE*, pp. 1- 8, 2018. Article (CrossRef Link).

[11] Sarkar A, Dasgupta S, Naskar S K, "Says Who? Deep Learning Models for Joint Speech Recognition, Segmentation and Diarization," in *Proc. of 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary,* pp. 5229-5233, 2018. Article (CrossRef Link).

[12] Lopez-Martin M, Carro B, Sanchez-Esguevillas A, "Network Traffic Classifier with Convolutional and Recurrent Neural Networks for Internet of Things," *IEEE Access*, vol. 5, pp. 18042-18050, 2017. Article (CrossRef Link).

[13] Wang W, "HAST-IDS: Learning Hierarchical Spatial-Temporal Features Using Deep Neural Networks to Improve Intrusion Detection," *IEEE Access*, vol. 6, pp. 1792-1806, 2018. Article (CrossRef Link).

[14] Wu K, Chen Z, Li W, "A Novel Intrusion Detection Model for a Massive Network Using Convolutional Neural Networks," *IEEE Access*, vol. 6, pp. 50850-50859, 2018. Article (CrossRef Link).

[15] Adie H T R, Pradana I A, "Parallel Computing Accelerated Image Inpainting using GPU CUDA, Theano and Tensorflow," in *Proc. of 2018 10th International Conference on Information Technology and Electrical Engineering (ICITEE), Kuta*, pp. 621-625, 2018. Article (CrossRef Link).

[16] Perera P, Tian Y C, Fidge C, "A comparison of supervised machine learning algorithms for classification of communications network traffic," in *Proc. of International Conference on Neural Information Processing. Springer, Cham*, pp. 445-454, 2017. Article (CrossRef Link).

[17] Zhang L , Cui Y, "Application of Machine Learning in Cyberspace Security Research," *Chinese Journal of Computers*, vol. 41, no 9, pp. 1943-1975, 2018. Article (CrossRef Link).

[18] Ge Y F, Peng F , Feng X X, "Homology Analysis of Malicious Code Based on Dynamic BP Neural Network," *Journal of Chinese Computer Systems*, vol. 37, no. 11,pp. 2527-2531, 2016. Article (CrossRef Link).

[19] HE H, "A network traffic classification method using support vector machine with feature weighted-degree," *Journal of Digital Information Management*, vol. 15, no. 5, pp. 76-83, 2017.

[20] Moore A W, Zuev D, "Internet traffic classification using Bayesian analysis techniques," *ACM SIGMETRICS Performance Evaluation Review*, vol. 33, no. 1, 2005. Article (CrossRef Link).

[21] KONG L, HUANG G, WU K, "Identification of abnormal network traffic using support vector machine," in *Proc. of the 2017 18th International Conference on Parallel and Distributed Computing, Applications and Technologies, Piscataway, NJ: IEEE*, pp. 288-292, 2017. Article (CrossRef Link).

[22] Wang R, Feng D G, "Feature extraction and detection method of malicious code based on semantics," *Journal of Software*, vol. 23, no. 2, pp. 378-393, 2012. Article (CrossRef Link).

[23] Vu L, Bui C T, Nguyen Q U, "A deep learning based method for handling imbalanced problem in network traffic classification," in *Proc. of the Eighth International Symposium on Information and Communication Technology, ACM*, pp. 333-339, 2017. Article (CrossRef Link).

[24] Rao Z, Niu W, Zhang X S, "Tor anonymous traffic identification based on gravitational clustering," *Peer-to-Peer Networking and Applications*, vol. 11, pp. 592–601, 2018. Article (CrossRef Link).

[25] Velan P, Čermák M, Čeleda P, "A survey of methods for encrypted traffic classification and analysis," *International Journal of Network Management*, vol. 25, no.5, pp. 355-374, 2015. Article (CrossRef Link).

[26] Deng Z, Qian G, Chen Z, "Identifying Tor Anonymous Traffic Based on Gravitational Clustering Analysis," in *Proc. of International Conference on Intelligent Human-Machine Systems and Cybernetics, IEEE*, pp. 79-83, 2017. Article (CrossRef Link).

[27] Song L, Chang L, "Application of mutation particle swarm optimization BP neural network in malicious code detection," *Journal of Intelligent Systems*, vol. 8, no. 6, pp. 558-563, 2013. Article (CrossRef Link).

[28] Carela-Español V, Barlet-Ros P, Mula-Valls O, "An autonomic traffic classification system for network operation and management," *Journal of Network and Systems Management*, vol. 23, no. 3, pp. 401-419, 2015. Article (CrossRef Link).

[29] Liu J, Zheng C, Guo L, "Understanding the Network Traffic Constraints for Deep Packet Inspection by Passive Measurement," in *Proc. of 2018 3rd International Conference on Information Systems Engineering (ICISE), IEEE*, pp. 26-32, 2018. Article (CrossRef Link).

[30] Ho T L, Cho S J, Oh S R, "Parallel multiple pattern matching schemes based on cuckoo filter for deep packet inspection on graphics processing units," *IET Information Security*, vol. 12, no. 4, pp. 381-388, 2018. Article (CrossRef Link).

[31] Araújo I M, Natalino C, Santana Á L, “Accelerating VNF-based Deep Packet Inspection with the use of GPUs,” in *Proc. of 2018 20th International Conference on Transparent Optical Networks (ICTON), IEEE*, pp. 1-4, 2018. Article (CrossRef Link).

[32] Sivaprasad A, Ghawalkar N, Hodge S, “Machine learning based traffic classification using statistical analysis,” *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 3, pp. 187-191, 2018.

[33] Krizhevsky A, Sutskever I, Hinton G E, “ImageNet classification with deep convolutional neural networks,” in *Proc. of International Conference on Neural Information Processing Systems, Curran Associates Inc.*, pp. 1097-1105, 2012. Article (CrossRef Link).

[34] Deshpande P, Sharma S C, Peddoju S K, “HIDS: A host based intrusion detection system for cloud computing environment,” *International Journal of System Assurance Engineering and Management*, vol. 9, no. 3, pp. 567-576, 2018. Article (CrossRef Link).

[35] LeCun Y, Bengio Y, Hinton G, “Deep learning,” *Nature*, no. 521, pp. 436-444, 2015. Article (CrossRef Link).

[36] Tongaonkar A, Torres R, Iliofotou M, “Towards self-adaptive network traffic classification,” *Computer Communications*, vol. 56, pp. 35-46, 2015. Article (CrossRef Link).

[37] Amaral P, Dinis J, Pinto P, “Machine learning in software defined networks: Data collection and traffic classification,” in *Proc. of 2016 IEEE 24th International Conference on Network Protocols (ICNP), IEEE*, pp. 1-5, 2016. Article (CrossRef Link).

[38] Ertam F, Avcı E, “A new approach for internet traffic classification: GA-WK-ELM,” *Measurement*, vol. 95, pp.135-142, 2017. Article (CrossRef Link).

[39] Zhu Z, Liang D, Zhang S, “Traffic-sign detection and classification in the wild,” in *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2110-2118, 2016. Article (CrossRef Link).

[40] Aceto G, Ciuonzo D, Montieri A, “Traffic classification of mobile apps through multi-classification,” in *Proc. of GLOBECOM 2017-2017 IEEE Global Communications Conference, IEEE*, pp. 1-6, 2017. Article (CrossRef Link).

[41] Dong Y, Zhao J, Jin J, “Novel feature selection and classification of Internet video traffic based on a hierarchical scheme,” *Computer Networks*, vol. 119, pp. 102-111, 2017. Article (CrossRef Link).

[42] Avci O, Abdeljaber O, Kiranyaz S, “Structural damage detection in real time: implementation of 1D convolutional neural networks for SHM applications,” *Structural Health Monitoring & Damage Detection, Springer, Cham*, vol. 7, pp. 49-54, 2017. Article (CrossRef Link).

[43] Ran J, Kong X C, “A Self-adaptive Traffic Classification System with Unknown Flow Detection,” in *Proc. of 3rd IEEE International Conference on Computer and Communications, IEEE*, pp. 1215-1220, 2017. Article (CrossRef Link).

[44] Ran J, Chen Y, Li S, “Three-dimensional Convolutional Neural Network based Traffic Classification for Wireless Communications,” in *Proc. of 2018 IEEE Global Conference on Signal and Information Processing (Global SIP), Anaheim, CA, USA*, pp. 624-627, 2018. Article (CrossRef Link).

[45] Stevanovic M, Pedersen J M, “An analysis of network traffic classification for botnet detection,” in *Proc. of 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment (CyberSA), IEEE*, pp. 1-8, 2015. Article (CrossRef Link).

[46] Gómez S E, Martínez B C, Sánchez-Esguevillas A J, “Ensemble network traffic classification: Algorithm comparison and novel ensemble scheme proposal,” *Computer Networks*, vol. 127, pp. 68-80, 2017. Article (CrossRef Link).

[47] Höchst J, Baumgärtner L, Hollick M, “Unsupervised traffic flow classification using a neural autoencoder,” in *Proc. of 2017 IEEE 42nd Conference on Local Computer Networks (LCN), IEEE*, pp. 523-526, 2017. Article (CrossRef Link).

[48] Karim F, Majumdar S, Darabi H, “Multivariate lstm-fcns for time series classification,” *Neural Networks*, vol. 116, pp. 237-245, 2019. Article (CrossRef Link).

[49] Kim T Y, Cho S B, “Web traffic anomaly detection using C-LSTM neural networks,” *Expert Systems with Applications*, vol. 106, pp. 66-76, 2018. Article (CrossRef Link).

[50] Zou Z, Ge J, Zheng H, "Encrypted Traffic Classification with a Convolutional Long Short-Term Memory Neural Network," in *Proc. of 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), IEEE*, pp. 329-334, 2018. Article (CrossRef Link).

[51] Hinton G E, Salakhutdinov R R, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504-507, 2006. Article (CrossRef Link).

[52] Shao H, Tang L, Dong L, "A Research of Network Applications Classification Based on Deep Learning," in *Proc. of International Conference on Machine Learning and Intelligent Communications. Springer, Cham*, pp. 13-21, 2018. Article (CrossRef Link).

[53] Taheri R, Javidan R, Shojafar M, "On Defending Against Label Flipping Attacks on Malware Detection Systems," *Neural Computing and Applications*, vol. 32, pp. 14781–14800, 2020. Article (CrossRef Link).

Junwei Li received his M.S. degree in computer technology from Huazhong University of Science and Technology, Wuhan, China , in 2009. He is currently a Ph.D. student at Army Engineering University. He is a lecturer in Xinxiang University. His research interests include deep learning, traffic classification and cyberspace security.

Zhisong Pan received the Diploma degree in computer science and technology from the PLA Information Engineering University, Zhengzhou, China, in 1996 and the Ph.D. degree in computer science and technology from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2003. He is currently a professor at Army Engineering University, a doctoral supervisor and a senior member of CCF. His research interests focus on machine learning, pattern recognition and cyberspace security.