One-dimensional CNN Model of Network Traffic Classification based on Transfer Learning

Lingyun Yang[1*], Yuning Dong[2], Zaijian Wang[1,3] and Feifei Gao[1]

[1] College of Physics and Electronic Information, Anhui Normal University
Wuhu, Anhui 241000 China
[e-mail: lingyun131@qq.com]
2 College of Communication and Information Engineering, Nanjing University of Posts and telecommunications
Nanjing, Jiangsu,210003, China
[e-mail: dongyn@njupt.edu.cn]
[3]The Key Laboratory of Grain Information Processing and Control at Henan University of Technology,
Ministry of Education, Zhengzhou, Henan, 450001, China.
[email: wangzaijian@ustc.edu, 28699540@qq.com]
[*]Corresponding author: Lingyun Yang

Abstract

There are some problems in network traffic classification (NTC), such as complicated statistical features and insufficient training samples, which may cause poor classification effect. A NTC architecture based on one-dimensional Convolutional Neural Network (CNN) and transfer learning is proposed to tackle these problems and improve the fine-grained classification performance. The key points of the proposed architecture include: (1) Model classification--by extracting normalized rate feature set from original data, plus existing statistical features to optimize the CNN NTC model. (2) To apply transfer learning in the classification to improve NTC performance. We collect two typical network flows data from Youku and YouTube, and verify the proposed method through extensive experiments. The results show that compared with existing methods, our method could improve the classification accuracy by around 3-5%for Youku, and by about 7 to 27% for YouTube.

1. Introduction

Network traffic classification (NTC) is the premise of many network behaviors, and how to classify the network flows accurately is one of the hot topics for years. Currently, the typical classification methods of network traffic classification are Machine Learning (ML) based algorithms. Research shows that ML algorithms have achieved good results in many classification domains [1-2]. Typical ML algorithms mostly consist of three main parts: original feature extraction, selection of optimal feature subset and classification algorithm [3-4]. However, the traditional machine learning algorithms are limited by expert experience, features, and classification methods, etc. Therefore, researchers try to use Deep Learning (DL) to improve the network traffic classification performance recently.

DL algorithms have been widely used in many fields [5-6]. Many researchers also tried to apply DL algorithms to NTC in recent years [7-8]. Authors [9-10] propose a new traffic classification method by inspecting packet head content and using deep learning algorithm, but if packet information of flows is protected or cannot be inspected, the classification performance will decline. [11] proposes an improved Stacked Auto Encoder (SAE) model for the multimedia traffic classification problems. [12] combines the convolutional neural network (CNN) and Swin Transformer to identify encrypted flows. [13] proposes a traffic classification method based on deep packet inspection (DPI) and CNN to improve satellite traffic classification performance.[14] proposes an improved Residual Convolutional Network for fine-grained traffic classification in Software Defined Networks (SDN). [15] utilizes Geometric Deep Learning and packet raw bytes metadata to classify encrypted flows. Most research of traffic classification on deep learning is still in primary stage and the research based on the transport layer in this paper (the packet information cannot be known) is relatively few.

Transfer learning (TL) is the use of existing knowledge and experience to learn new things, and the core is to find the similarities between existing knowledge and new things [16-17], the application of transfer learning is widely used in many fields but NTC is fewer. the authors [18] consider a new multi-task network traffic classification algorithm by combining the transfer learning algorithm and Deep Neural Network (DNN) model.[19] proposes a new network traffic classification algorithm based on the cross-silo model and transfer learning.[20] uses TL algorithm to realize network traffic classification with fewer samples.

Presently, the main problems in NTC research are as follows:

(1) The traditional NTC based on ML mainly contains selection of optimal feature subsets and classification algorithm, which mostly rely on the expert experience it's self-learning ability is largely relatively poor because of the relative independence. Therefore, this paper proposes a NTC architecture based on one-dimensional CNN to realize the classification who has better self-learning ability and strong robustness.

(2) Most research in the DL traffic classification realizes classification by learning the packet head information to obtain good classification results, however, with the enhancement of self-privacy protection, the effectiveness of the method may be largely declined. This paper proposes to realize the NTC based on 5-tuple group to avoid such problems.

(3) Research on NTC mostly focuses on coarse classification, that means, network traffic is simply divided into video, email, chat, etc. But the fine-grained classification is rarely. In the real network, the fine-grained classification is more needed to meet individual requirement and further optimization of network configuration.

(4) There are some other significance problems in traffic classification, which may impact the classification effect. For example, the traditional statistical features only show the global characteristics of the flows, which cannot reflect the characteristics of transmission process; More features are needed to help improve the fine-grained traffic classification results; The insufficient training samples may reduce the classification effect. Solving these problems, it is possible to effectively improve the classification performance.

For the problems caused by limited statistical features and insufficient training samples in fine-grained network traffic classification, a new NTC based on one-dimensional CNN model is proposed, and the TL is applied to further improve the performance of network traffic classification. The main innovations are as follows:

(1) CNN optimization. The network flows discussed in this paper based on the transport lay only contain 5-tuple (packet arrival time, packet size, source address, destination address and transmission protocol). The data is simple and long, and cannot reflect the characteristics of various flows, which should be pre-processed before classification. The traditional pre-processing is to extract various statistical features, such as variance of duration, mean of packet number etc. [21]. However, these features are not suitable for deep learning model. This paper proposes to calculate the standardized rate distribution probability, which reflects the transmission characteristic of network flow, and the network transmission standard of the flows indirectly. The features have better classification performance [22], and could help to improve the performance of network traffic classification.

Rate probability distribution features not only could be used as classification features, but also could help optimize CNN model as the weight step size. The training step size plays a crucial role in CNN model. If the step size is too high, over fitting is easily occur; when the step size is too small, the training duration and the difficulty of finding the optimal model will increase [23-24]. A dynamic step method based on probability distribution features is proposed in this paper. The major approach is the calculation of the difference value between the current and previous flow, and use this value as the next step size. If the difference value between the two network flows is small, the similarity is high, and the optimization step of the model is small; otherwise, the step size is large.

(2) Transfer Learning. This paper proposes to apply TL method to network traffic classification to resolve the insufficient training samples problem and further improve the performance of network traffic classification. Two key issues in TL are what to transfer and how to transfer. The method of TL in this paper mainly includes two points: firstly, realize the transfer learning of CNN classification model by transferring the key parameters of the model, the key parameters of model (deviation, kernel function value and weight size) determine the CNN classification performance. Secondly, when there are multiple sources domains, a new source domain should be established. The required source domain characteristics are roughly confirmed by the rate mean and variance of the probability distribution characteristics of flow. Experimental results show the reliability and effectiveness of the proposed method.

(3) Experiment. This paper collects different types of network flows to verify the effectiveness of the classification architecture. It mainly includes two aspects: firstly, the performance of CNN model, includes coarse classification and fine-grained classification, which reflects the improvement of classification accuracy and other parameters. The second is the performance verification of TL in traffic classification, which improves the classification performance.

The main structure of this paper is as follows: the first part is the introduction, which mainly introduces the research status of NTC and the main research of this paper; The second part is the improved one-dimensional CNN model for network traffic classification, which mainly includes the data pre-processing method and the step size improvement algorithm. The third is the TL algorithm, including the main parameters of TL and how to implement it; The fourth part is experimental verification, which verifies the effectiveness of proposed method; The last is summarized.

2. The 1-D convolutional neural networks

2.1 Data pre-processing

The one-dimensional CNN designed in this paper uses the rate probability distribution features to classify the flows. Two points are proposed, according to the characteristics of network traffic and one-dimensional CNN model. The first is data pre-processing, and the second is the adaption step algorithm of one-dimensional CNN. The data pre-processing process is shown in **Fig. 1**:
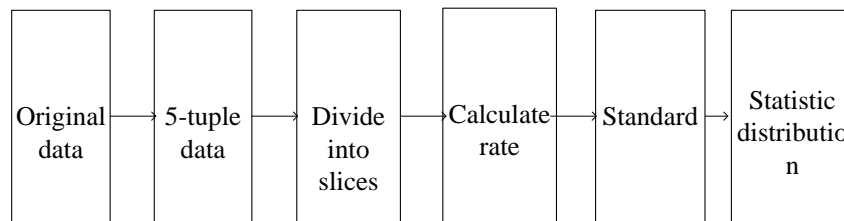
Fig. 1. The pre-processing of network flows data

The network traffic has some unique characteristics, which is collected by Wireshark, cannot be directly used for classification because of the following two reasons, firstly, raw data contains only the five-tuple group information, and cannot meet the classification requirements. Secondly, the number of original data values is large and repetitive, which cannot directly reflect the characteristics of network traffic. Therefore, it is necessary to pre-process the information before traffic classification. The pre-processing includes two contents: one is obtaining the five-tuple group, and the other is calculating the standard probability distribution. The detailed steps are as follows:

(1) Wireshark is a typical network software, which captures the required network data, but these data obtained by Wireshark contains kinds of redundancy contents. Therefore, the network traffic should be processed firstly, only retain five-tuple group (include the arrival time of the packet, the source address, the destination address, the transmission protocol, and the byte size of the packet). The advantage of this pre-processing is to simplify the data information, to better protect the user's privacy, and to better apply to encrypted network classification.

(2) Each encrypted network flows contains thousands of five-tuple groups. It is almost impossible to directly use such a large amount of data for one-dimensional CNN traffic classification, the cost is too expensive to do, that further pre-processing is needed. Different with most network traffic classification researchers, who mostly focus on the overall statistical features, this paper pay attention to the transmission process change characteristics of network flows. Therefore, the long network flow should be divided into smaller slices firstly, here, we

take one second as the basic unit.

(3) The transmission rate of each slice is calculated by formula (1).

$$V_i = \frac{\sum \text{packet size}}{\text{time}_{end} - \text{time}_0} \tag{1}$$

where, $\sum packet\ size$ represents the sum of all bytes in each slice, $time_{end} - time_0$ represents the duration from the first packet to the last one.

(4) Normalizing the rate value in [0, 1] by formula (2).

$$V_{NBR} = \frac{V_{BR} \cdot 8 \cdot 30}{1000 \cdot \min(30, FR)} \tag{2}$$

where, V_{NBR} represents the normalized rate value, V_{BR} represents the rate value, and FR represents the value of the video frame.

(5) Calculating the probability distribution values of all rates and make sure that the values are distributed in [0, 1].

In real network transmission, different transmission categories have distinctive definitions of the rate change. Rate characteristics reflect the essence of its network transmission, so it could better help to realize the fine-grained network traffic classification.

2.2 Adaptive step algorithm

The optimal training of CNN classification model is shown in **Fig. 2**. It takes the rate probability distribution features as the input data and continuously optimizes the classification model until it reaches the optimal state through convolution, pooling and output results. During the training, the step size is one of the key factors, a reasonable step value is useful to better realize the optimization of CNN model parameters and avoid over fitting.
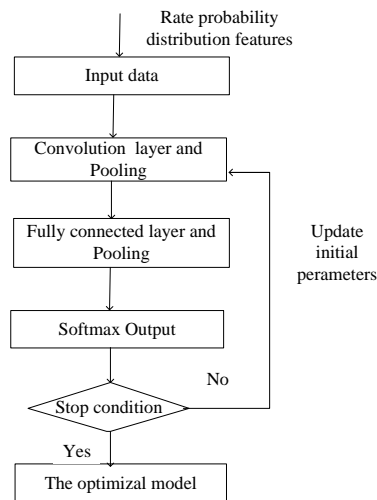
Fig. 2. The training of 1-D CNN classification model

Ihe gradient descent method in CNN is used to update the parameters. Assuming that $f(x)$ is a continuously updated parameter, it could be calculated as formula (3)

$$f(x_{i+1}) = f(x_i) - \delta \cdot \frac{\nabla f(x_i)}{\nabla x_i} \tag{3}$$

where, δ is the step size, which is also named learning rate etc., which reflects the speed of parameter update, the key parameter in the gradient descent method. δ value is between (0, 1), reasonable δ value could help to find the optimal value of parameters rapidly, to avoid problems such as too long iteration time or oscillation etc. $\frac{\nabla f(x_i)}{\nabla x_i}$ indicates the direction of gradient update.

This paper proposes to use the difference value between the current and the previous network flow rate distribution, and take its mean value as the step size, as in formula (4)

$$\delta = \sum \|y(i) - y(i-1)\|/(N-1) \tag{4}$$

where, $y(i)$ represents the probability distribution of the current network flow, $y(i-1)$ represents the distribution of the previous network flow, and N is the number of features contained in the probability distribution of the network flow.

The adaptive step's main characteristics are as follows:

(1) The probability distribution values are the known data, which has calculated in our first step of data pre-processing, and none special process is required here, so the calculation duration will not be increased;

(2) The probability distribution values are in (0, 1), and their difference value is also in (0, 1), and the step value changes following network flows.

(3) If the two network flows are highly similar, their probability distributions must be very similar too, the optimization step is relatively small. If the feature distribution similarity is low, a large step size is required.

The adaptive step size algorithm proposed in this paper is closely related to the current network flow characteristic, so it could realize the adaptive optimization of network traffic classification model effectively.

3. Transfer learning algorithm in classification

To solve the problem of insufficient samples in the traffic classification, this paper proposes to optimize the CNN classification model by the transfer learning one-dimensional CNN key parameters.

There are six key parameters for model optimization: two deviation variables (bias_1 represents the deviation after primary convolution and bias_2 represents the deviation after secondary convolution), two convolution variables (ker_1 represents the primary convolution kernel function and ker_1 represents the secondary convolution kernel function), the weighted value of secondary convolution w_{gh1} and output softmax function weighted value w_{gh2} in the one-dimensional CNN model. The relationship between them can be expressed by formula (5) and (6):

$$f_1(x+1) = \text{convolution}(f_1(x), \text{ker}\_1) + \text{bias}\_1 \tag{5}$$

$$f_2(x+1) = \text{convolution}(f_2(x), \text{ker}\_2, w_{gh1}) + \text{bias}\_2 \tag{6}$$

where, $f_1(x)$, $f_2(x)$ represents the update of the two states of the input data respectively, and $convolution(\ )$ represents the convolution between the two groups of data. Here, the update of each variable calculated by formula (3).

Based on many experiments in this paper, Tanh function is selected as the excitation equation, and the expression is

$$\text{Tanh}(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{7}$$

The output function adopts SoftMax function, which contains the second weight function w_{gh2}.

$$\text{fo}(x_k) = \frac{\exp\left(w_{gh2k}.x_k\right)}{\sum_i^{i=1...N} \exp\left(w_{gh2i}.x_i\right)} \tag{8}$$

where, Weight function w_{gh2} affects the output of the function directly, and also affects the loss function. Because the occurrence probability of the output function value is balanced, the weight function w_{gh2} initialization usually adopts to uniform distribution.

The loss function represents the difference between the actual value and the ideal value. Model optimization means that the loss function is minimum. Suppose that the ideal result of the classification is $a(k)$, and $fo(x_k)$ represents the actual value, then the calculation of the k-th loss function is

$$Loss\ (k\ ) = a(k) - fo(x_k) \tag{9}$$

Its cross-entropy loss function is

$$Loss = -\frac{exp\left(w_{gh2k}.x_k\right)}{\sum_i^{i=1...N} exp\left(w_{gh2i}.x_i\right)} \tag{10}$$

The smaller the loss function, the closer the corresponding classification model is to the optimal mode. The transfer learning of CNN model is aimed at the transfer learning of these six main model parameters in this paper.

The features contained in the source domain are represented as one-dimensional vectors $(v1_{si}\ ,\ v2_{si}, ... Vn_{si})$, where si represents the $i'th$ group of feature, and each feature group contains n features. The task of source domain is to build the optimal one-dimensional CNN model for network traffic classification by using the SD features.

The features contained in the target domain are represented as one-dimensional vectors $(v1_{ti}\ ,\ v2_{ti}, ... vm_{ti})$, where, ti represents the $i'th$ group of features in the target domain, and each feature group contains m features.

The relationship between source domain and target domain is described as follows:

(1) The space of source domain and two target domains are both composed of rate probability distribution features, and the pre-processing method of rate probability distribution in source domain and target domain is same.

(2) The source domain and target domain may have different spatial scopes.

(3) The task of both source domain and target domain is to realize the network traffic classification.

The main task of transfer learning between source domain and target domain contains two points:

(1) Transferring CNN classification model from source domain to target domain by transferring the key parameters of the model. The whole process of one-dimensional CNN model based on transfer learning is shown in **Fig. 3**.

(2) When there are multiple source domains, the similarity of network flows in the source domain and the target domain is calculated first, the appropriate source domain data is selected to help train the target domain model. The method of transfer learning classification method is shown in **Alg. 1**.
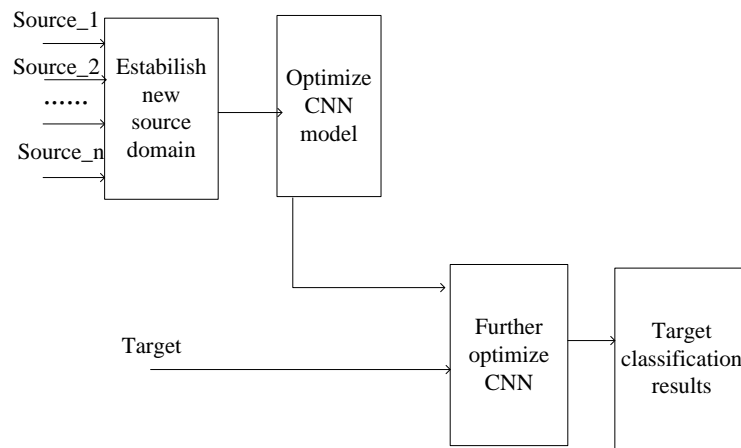
Fig. 3. The fine-grained classification method based on transfer learning

Alg. 1. The CNN classification model based on transfer learning

Input：**source domain, target domain**
Output：The CNN optimization parameters

（1）initialization step: M1=0; M2=0; M3=0;
    for i=1:P
      if $V_{meansi} \in (V_{mean1} \pm V_{dev1})$
       If $Me_{c1} < M_{p1}$
         $i \in \{NT_1\}$
         M1=M1+1
       else if $V_{meansi} \in (V_{mean2} \pm V_{dev2})$
           if $Me_{c2} < M_{p2}$
          $i \in \{NT_2\}$
          M2=M2+1
       else if $V_{meansi} \in (V_{mean3} \pm V_{dev3})$
           If $Me_{c3} < M_{p3}$
          $i \in \{NT_3\}$
          M3=M3+1
         end
       end
      $M = min(M_1, M_2, M_3)$
    $\{NT_1\}=[\{NT_1\}、\{NT_2\}、\{NT_3\}]$

---

(2) for i=1:N

    Transfer Learning and optimize parameters
   end
  $\mathbf{Loss_{max} = max([\ldots Loss_j \ldots])}$
Output: $\mathbf{bias_1,\ bias_2,\ ker_1, ker_2,\ w_{gh1},\ w_{gh2}}$

In **Alg. 1**, two-layer algorithm is used to achieve the selection of effective source data which is critical. The method is as follows:

(1) The selection of new source domain adopts a two-layer structure to improve the effectiveness of identifying network flows, and the network flows with higher similarity are selected to help optimize traffic classification model in the target domain.

(2) Pre-processing the labeled target domain data, calculating its average rate of network flow V_{meani} and the mean square deviation V_{devi}(here i=1,2,3). If the rate of source domain flow V_{meansi} is in the range (V_{meani}-V_{devi} , V_{meani}+V_{devi}) , The corresponding flow belongs to the target domain, Where, i indicates the different network flow type labels, our experiments classify the flows into three types (SD,FL, HD), so i=1,2,3, similarly hereinafter.

(3) Based on the above calculation, the relative error of probability distribution between the network flow in source domain and mean value of target domain is also calculated, as shown in formula (11),

$$Me_{ci} = (M_{psi} - M_{cdi})./M_{cdi} \tag{11}$$

Where, M_{cdi} means the average rate probability distribution of network flow in class $i'th$ target domain, M_{psi} represents the rate probability distribution of network flow in class $i'th$ source domain, Me_{ci} represents the relative probability distribution error of source domain network flow and target domain network flow (here i=1,2,3).

This method is effective and reliable for the following two reasons:

(1) Different types of network flows have different rate requirements due to specific transmission quality and content, so most network flow types could be distinguished by the mean and variance of the transmission rate. However, the network flows are easily affected by network environment, and the network flows collected are only a part of the long ones, which may not correctly reflect the transmission characteristics of the whole. Thus it is not enough to only consider the transmission rate factor, especially the network flows with similar rate, which is more likely to cause misjudgment.

(2) On the basis of above discrimination, this paper proposes the second discriminant step--probability distribution relative mean discriminant; The type of flow depends on relative difference mean of all network flows. if it is the same type of network flow, its difference must be lower than the relative average value of the flow. Next, paper will verify the method through experiments.

4. The experiments and analysis

The experiments consist of three parts, the first part is coarse classification, two sets of data are used to verify the CNN classification. The next two parts are fin-grained classification about Youku and YouTube flows, which divided the flows into three classes(standard definition video flows (Abbreviated as SD), Fluent video flows (Abbreviated as FL) and high definition flows (Abbreviated as HD)).Three sets are captured from different websites by ourselves, such as Youku, YouTube and Betta, Tencent and storm, These data was collected in the campus network of Nanjing University of Posts and telecommunications in September 2015 and January 2019 respectively, and the data transmission includes wired transmission and wireless transmission. The other group is from UNB university data set [25]. The CNN model and transfer learning method are both simulated by MATLAB, its version number is R 2016b.

The classification experiments randomly select 70% of data as the training samples and the other as the test samples, to verify the classification performance, repeat 10 to 20times, then average all the values, the times' number depends on the values are relatively stable or not. The network traffic classification criteria mainly include Accuracy, Recall, Precision and F-measure. The accuracy reflects the judgment result of the whole samples, and the recall, Precision and F-measures respectively reflect the classification effect of each type. The experiment is divided into three parts: (1) Coarse classification of network flows under one-dimensional CNN network model. (2) the one-dimensional CNN coarse classification model is applied to the fine-grained classification of Youku flows by transfer learning. (3) Based on the first two parts and transfer learning to establish an optimal model for YouTube traffic classification.

4.1 The coarse traffic classification based on one-dimensional CNN model

The Wireshark collects network flows from several typical websites. Five-tuple information were also extract (source address, destination address, data arrival time, packet size and transmission protocol), the data flow duration is 300s. The data is shown in **Table 1**.

Table 1. Data set 1

Name	Size (GB)	Date	Samples	Duration
Youku（non-real transmission）	259.09	2015.9	1080	300s
Betta（real transmission）	394.56	2019.9	600	300s
Tencent QQ (real transmission)	58.70	2019.6	300	300s
Storm（real transmission）	119.09	2019.3	300	300s

The CNN model is realized by MATLAB. The pre-processing is essential to traffic classification through CNN, because the raw data could not directly used as CNN original data, and the classification performance is not very good. Meanwhile, one-dimensional CNN is more suitable than two-dimensional CNN because of the less information in the flow. The traditional machine learning method takes SVM (support vector machine) as an example in this paper, which is implemented by Weka.

The pre-processing of CNN model proposed in this paper takes 1s as the basic unit, and calculates the flow's probability distribution features. During the classification experiments, we divide the data into 10 parts, randomly select seven parts as training data and three parts as test data, repeat for several times to find the average value and mean square deviation.

The two groups of coarse classified data are come from **Table 1**. The first group of data is 1000 groups of non-real-time data randomly selected from Youku website and 1000 groups of real-time data randomly selected from storm, Betta and Tencent QQ, the two compared method are the traditional ML algorithm to realize traffic classification [22] and a typical traffic classification method by deep learning [11].

Table 2. Coarse classification performance (mean ± standard deviation)

Flows		Precision	Recall	F-measure	Accuracy
[22]	non-real	0.992±0.001	0.897±0.002	0.945±0.001	0.948±0.002
	real	0.908±0.002	0.992±0.001	0.959±0.001	
[11]	non-real	0.994±0.002	0.918±0.005	0.953±0.003	0.960±0.004
	real	0.924±0.007	0.992±0.004	0.959±0.003	
This method	non-real	0.992±0.001	0.958±0.001	0.975±0.001	0.976±0.001
	real	0.959±0.002	0.993±0.001	0.976±0.001	

Table 2 shows that the classification performance of the one-dimensional CNN model is better than the traditional algorithm obviously, and the accuracy, precision, recall and F-measure values all have been improved. By one-dimensional CNN model, the overall accuracy of rough classification can reach 97.6%.

Compared with existing classification methods, one-dimensional CNN model could blur the subtle differences between data to a certain extent of these same type flows, but also retain the correlation, increase the robustness of data, thus improving the classification accuracy. In the data transmission, classification mostly be affected by the real-time network environment and reduce the accuracy of classification. However, the one-dimensional CNN model does not require the specific relationship between features, and increases the robustness of features by convolution and pooling in model training.

From **Table 2**, the performance of the one-dimensional CNN classification model proposed in this paper is suitable for network traffic classification. Good classification performance will also play a good foundation for the transfer learning of fine-grained classification models.

Additionally, this paper further verifies the effectiveness of one-dimensional CNN traffic classification through data set 2, which collected by University of New Brunswick (UNB). The data are shown in **Table 3**, it contains two data types: VPN and non-VPN.

Table 3. Data set 2

Flow type	Size	Samples
VPN	670.43M	99
non-VPN	944.82M	99

The data shown in **Table 3** collected VPN data and non-VPN data from Facebook, Hangouts, ICQ, Skype, Vimeo, AIM and other websites. The data processing method in this paper are as follows: (1) If the data duration of the sample is less than 300s, the sample participates in the classification as an independent flow. (2) If the duration of the sample flow is longer than 300s, the data flow is divided into several smaller slices according to the basic principle of 300s, and each slice participates in the classification as an independent one. (3) The number of samples of two types of network flows should be keep the same.

The data classification results are shown in **Table 4** the machine learning algorithm adopts ten-fold cross validation.

Table 4. Coarse traffic classification results of data set 2 (mean ± standard deviation)

Flows		Precision	Recall	F-measure	Accuracy
[22]	VPN	0.697±0.009	0.859±0.015	0.769±0.013	0.742±0.010
	Non-VPN	0.816±0.012	0.626±0.010	0.709±0.011	
[11]	VPN	0.789±0.010	0.949±0.013	0.869±0.012	0.848±0.009
	Non-VPN	0.945±0.009	0.747±0.010	0.842±0.009	
This method	VPN	0.879±0.010	0.949±0.009	0.914±0.010	0.909±0.009
	Non-VPN	0.966±0.009	0.868±0.010	0.907±0.009	

Table 4 shows the precise, recall, F-measure and accuracy values of VPN and non-VPN coarse classification of dataset 2. It shows that the methods proposed in this paper also could better realize the traffic classification, the accuracy improves from 74.2%, 84.8% to 90.9%. The main characteristics of dataset 2 are as follows: (1) The data sources are relatively broad, including conversation service flow, video service flow, email service flow, etc. (2) The length of data flow is different, which will have a certain impact on the classification result. (3) There are fewer samples than self-collected data set.

Both data sets verify the reliability of the coarse classification performance of one-dimensional CNN model.

4.2 The fine-grained classification of Youku

Youku is one of the representative websites in China, which collects three classes of network flows: standard definition video flows（Abbreviated as SD）, Fluent video flows（Abbreviated as FL） and high definition flows（Abbreviated as HD）. Youku data set is shown in **Table 5**, the duration of each flow is 600s, 180 samples are selected each classes .

Table 5. Data set 3

| Flow type	Size (GB)	Date	Samples	Duration
SD of Youku	42.8	2015.9	180	600s
FL of Youku	68.5	2015.9	180	600s
HD of Youku	145.5	2015.9	180	600s

The fine-grained classification of Youku flows takes the network flow classification model in previous section as the source domain model, realizes the model transfer to the target domain through the its key parameters, to improve the classification performance. **Table 6** shows the comparison of Youku traffic classification performance under three different methods.

Table 6.   Performance of fine-grained traffic classification of Youku (mean ± standard deviation)

Flows		Precision	Recall	F-measure	Accuracy
[22]	SD	0.853±0.015	0.906±0.007	0.879±0.011	0.900±0.010
	FL	0.886±0.009	0.867±0.011	0.876±0.010	
	HD	0.96±0.007	0.928±0.009	0.946±0.007	
[11]	SD	0.902±0.014	0.960±0.019	0.931±0.017	0.925±0.018
	FL	0.939±0.020	0.846±0.012	0.893±0.017	
	HD	0.937±0.015	0.969±0.025	0.953±0.019	
This method (non-transfer)	SD	0.899±0.010	0.944±0.011	0.922±0.010	0.938±0.010
	FL	0.937±0.011	0.889±0.008	0.913±0.008	
	HD	0.987±0.007	0.983±0.010	0.960±0.009	
This method (transfer)	SD	0.920±0.010	0.956±0.005	0.938±0.009	0.952±0.005
	FL	0.952±0.006	0.900±0.008	0.926±0.006	
	HD	0.987±0.005	0.998±0.001	0.970±0.003	

Table 6 shows that all the four methods of classification results of Youku network flows. The methods proposed in this paper (with transfer learning and non-transfer learning) have been improved classification results, compared with existing methods, the CNN model without transfer learning improve classification accuracy from 90.0%, 92.5% to 93.8%, and transfer learning helps further improve to 95.2%. The Precision, Recall and F-measure also improved with varying degrees. But the fluency data flow is in the middle of standard definition flow and HD flow, that it is easily affected by the network environment and misjudged in the actual network, that the classification results are relatively poor, and the advantage of this method is not obvious. However, the overall performance is still significantly improved.

The advantages of this method are mainly in two aspects: firstly, the rate probability distribution characteristics of network flow are studied and used to optimize the initialization value. Experiments show that model transfer learning plays an important role in the optimization. Secondly, to avoid over fitting, the step size through adaptive method is adopted, based on flow's probability distribution, which could help to improve the classification performance.
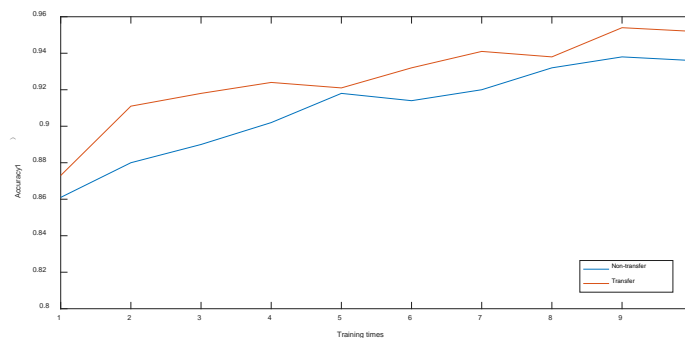
Fig. 4.   The relationship between accuracy and training times (Youku)

Fig. 4 shows the relationship between the accuracy results of traffic classification and the number of training times under transfer learning and non-transfer learning.

As in **Fig. 4**, transfer learning helps to improve the accuracy of fine-grained Youku model classification. The transfer learning of the model is to optimize the classification model of the target domain by using the existing classification model. The similarity and correlation between the source domain and the target domain and the features can help to realize the optimization of the target domain model quickly and improve the classification performance.

4.3 The fine-grained traffic classification of YouTube

YouTube is a video website owned by Google, is a representative video website provider in the world. To further verify the classification performance of transfer learning and one-dimensional CNN model.

Table 7. Data set 4

Flow type	Size (GB)	date	samples	duration
YouTube SD	29.6	2019.3	100	900s
YouTube FL	32.5	2019.3	100	900s
YouTube HD	44.3	2019.3	100	900s

As shown in **Table 7**, three classes of network flows from YouTube are collected from the campus network of Nanjing University of Posts and Telecommunications, each flow's duration is 900s, here, we select 100samples to realize the fine-grained classification.

Theoretically, the transmission characteristics of data flows from different websites are variance. Classifying YouTube flows is helpful to further verify the effectiveness of transfer learning and one-dimensional CNN model. **Table 8** shows the classification performance of YouTube under different method.

Table 8. Performance of traffic classification of YouTube (mean ± standard deviation)

Flows		Precision	Recall	F-measure	Accuracy
[22]	SD	0.697±0.012	0.767±0.011	0.730±0.011	0.683±0.011
	FL	0.529±0.012	0.600±0.008	0.563±0.010	
	HD	0.891±0.014	0.683±0.013	0.774±0.013	
[11]	SD	0.896±0.014	0.973±0.016	0.934±0.015	0.881±0.012
	FL	0.829±0.015	0.857±0.026	0.842±0.016	
	HD	0.924±0.010	0.813±0.015	0.869±0.012	
This method (non-transfer)	SD	0.987±0.008	0.844±0.010	0.916±0.008	0.929±0.009
	FL	0.853±0.007	0.967±0.008	0.910±0.007	
	HD	0.967±0.009	0.978±0.008	0.972±0.008	
This method (transfer)	SD	0.993±0.002	0.889±0.008	0.941±0.005	0.954±0.005
	FL	0.903±0.009	0.978±0.003	0.940±0.005	
	HD	0.973±0.003	0.994±0.003	0.983±0.002	

Table 8 shows that the method of the one-dimensional CNN model proposed in this paper has been greatly improved the values of precise, recall, F-measure, and accuracy. Compared with 68.3% in [22] and 88.1% in [11], The accuracy with non-transfer learning is 92.9%, transfer learning has increased to accuracy further to 95.4%, which verifies the effectiveness of transfer learning method and one-dimensional CNN model. it is found that Youku flow and YouTube flow have higher similarity, therefore, the transfer learning of Youku flow is helps the traffic classification model of YouTube. At the same time, **Fig. 5** shows the relationship between classification model accuracy of YouTube and model training times under transfer

learning and non-transfer learning.
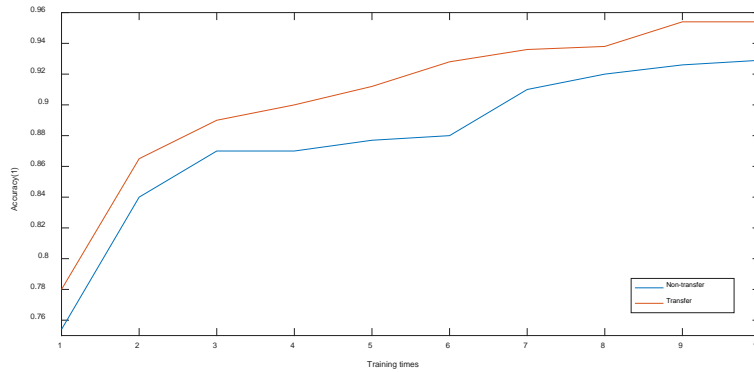
Fig. 5. The relationship between accuracy and training times (YouTube)

As in **Fig. 5**, the transfer learning helps to improve the classification accuracy in the fine-grained traffic classification of YouTube. The experiment shows that with the increase of training times, the accuracy of traffic classification increases gradually and tends to be stable finally.

4.4 Discussion

The effectiveness of the proposed traffic classification architecture based on one-dimensional CNN model and transfer learning is verified by three parts of typical network flows. The first part is the coarse classification, and two data sets are used in it. The first set flows are divided into real flows and non-real flows. The second set is collected by UNB, the flows are divided into VPN and non-VPN. Two experiments results verify that the CNN classification model proposed in this paper is helpful to improve the classification effects, compared with existing classification methods. The main reasons are as follows:

(1) The rate probability distribution features used in this paper could better reflect the characteristics of network transmission, compared with traditional features, thus the correlation between the probability distribution features is stronger, and could help to improve classification performance.

(2) The rate probability distribution features are used again to calculate the difference between the front and back features as the step size of CNN model. The experiments show that this method could improve the classification effect.

The second and third part are the fine-grained classification experiments of Youku and YouTube flows, both are divided the network flows into three classes (SD, FL and HD). These two-traffic classification are similar, so transfer learning could be used to improve the performance further. The experiments analyze the classification performance with transfer learning and without it. The results show that the CNN model with transfer learning could improve the classification performance. Mainly because:

(1) Based on existing model, the transfer learning in this paper transfers the model parameter, then the new data is added for next training to improve the classification performance.

(2) The transfer learning method compares the source domain data and the labeled data in the target domain. If the source data like target domain, they join it. The transfer learning

could help to solve the problem of insufficient training data and data imbalance, to improve the classification effect.

The one-dimensional CNN model and transfer learning formed the construction of classification model. Generally, the improved CNN model helps to improve the classification effect, and its performance is affected by the changes of rate transmission, the flow length etc. When the training samples are fewer or insufficient, transfer learning could help to improve the classification effect effectively. But if the labeled data itself could obtain a better classification model, its effect will not be obvious. The CNN model is the basis of transfer learning, if the CNN model could not be better optimized, the transfer learning parameters are not optimized too, then the classification performance will be worse, only transfer optimized parameters, the classification performance of the target domain could be improved.

5. Conclusion

This paper proposes a classification architecture for encrypted flows based on one-dimensional convolutional neural networks and transfer learning. In data pre-processing, discrete probability distribution features are used as the input data of one-dimensional CNN classification model, and the different value between the discrete probability distribution features of the current and previous flow is used as the optimized step size to improve the performance of classification model. Meanwhile, it is proposed to combine transfer learning algorithm with CNN model to further improve the classification effects.

The application research based on the deep learning models to realize network traffic classification is just the beginning, In the following research, authors will continue to focus on the optimized deep learning model in encrypted traffic classification continuously, and try to improve the fine-grained network traffic classification effect of more kinds of network flows. In the future, authors also mainly consider the classification of real-time network traffic to provide better basic guarantee for various network traffic services.

References

[1] R. Enisoglu and V. Rakocevic, "Low-Latency Internet Traffic Identification using Machine Learning with Trend-based Features," in *Proc. of 2023 International Wireless Communications and Mobile Computing (IWCMC)*, Marrakesh, Morocco, pp. 394-399, 2023. Article (CrossRef Link).

[2]   M. R. Choudhury, M. N, P. Acharjee and A. T. George, "Network Traffic Classification Using Supervised Learning Algorithms," in *Proc. of 2023 International Conference on Computer, Electrical & Communication Engineering (ICCECE)*, Kolkata, India, pp. 1-6, 2023. Article (CrossRef Link).

[3] M. S. Sheikh and Y. Peng, "Procedures, Criteria, and Machine Learning Techniques for Network Traffic Classification: A Survey," *IEEE Access*, vol. 10, pp. 61135-61158, 2022. Article (CrossRef Link).

[4] S. M. Rachmawati, D. -S. Kim and J. -M. Lee, "Machine Learning Algorithm in Network Traffic Classification," in *Proc. of 2021 International Conference on Information and Communication Technology Convergence (ICTC)*, Jeju Island, Korea, Republic of, pp. 1010-1013, 2021. Article (CrossRef Link).

[5] J. P. Briot, F. Pachet, "Music Generation by Deep Learning - Challenges and Directions," *Neural Computing and Applications*, vol.32, no.2, pp.981-993, 2020. Article (CrossRef Link)

[6]   D. Li, H. Hui, Y. Zhang, F. Tian, X. Yang, J. Liu, J. Tian, "Deep Learning for Virtual Histological Staining of Bright-Field Microscopic Images of Unlabeled Carotid Artery Tissue," *Molecular Imaging & Biology*, vol.22, no.5, pp.1301-1309, 2020. Article (CrossRef Link)

[7]   G. Aceto, D. Ciuonzo, A. Montieri and A. Pescapé, "Mobile Encrypted Traffic Classification Using Deep Learning: Experimental Evaluation, Lessons Learned, and Challenges," *IEEE Transactions on Network and Service Management*, vol.16, no.2, pp.445-458, 2019. Article (CrossRef Link)

[8]   P. Wang, X. Chen, F. Ye and Z. Sun, "A Survey of Techniques for Mobile Service Encrypted Traffic Classification Using Deep Learning," *IEEE Access*, vol.7, pp.54024-54033, 2019. Article (CrossRef Link)

[9]   M. Bahaa, A. Aboulmagd, K. Adel, H. Fawzy, N. Abdebaki, "nnDPI: A Novel Deep Packet Inspection Technique Using Word Embedding, Convolutional and Recurrent Neural Networks," in *Proc. of 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*, Giza, Egypt, pp.165-170, 2020. Article (CrossRef Link)

[10] S. Rezaei, X. Liu, "Multitask Learning for Network Traffic Classification," in *Proc. of 29th International Conference on Computer Communications and Networks (ICCCN)*, Honolulu, HI, USA, August, pp. 1-9, 2020. Article (CrossRef Link)

[11] A. Canovas, J. M. Jimenez, O. Romero, J. Lloret, "Multimedia Data Flow Traffic Classification Using Intelligent Models Based on Traffic Patterns," *IEEE Network*, vol.32, no.6, pp.100-107, 2018. Article (CrossRef Link)

[12]   Y. Wang, Y. Gao, X. Li and J. Yuan, "Encrypted Traffic Classification Model Based on SwinT-CNN," in *Proc. of 2023 4th International Conference on Computer Engineering and Application (ICCEA)*, Hangzhou, China, pp. 138-142, 2023. Article (CrossRef Link).

[13] X. Wan, X. Fu, J. Li, and J. Wang, "Research on Satellite Traffic Classification based on Deep packet recognition and convolution Neural Network," in *Proc. of 2023 8th International Conference on Computer and Communication Systems (ICCCS)*, Guangzhou, China, pp. 494-498, 2023. Article (CrossRef Link).

[14] C. Su, Y. Liu, and X. Xie, "Fine-grained Traffic Classification Based on Improved Residual Convolutional Network in Software Defined Networks," *IEEE Latin America Transactions*, vol. 21, no. 4, pp. 565-572, April 2023. Article (CrossRef Link).

[15] T. -L. Huoh, Y. Luo, P. Li, and T. Zhang, "Flow-Based Encrypted Network Traffic Classification with Graph Neural Networks," *IEEE Transactions on Network and Service Management*, vol. 20, no. 2, pp. 1224-1237, June 2023. Article (CrossRef Link).

[16] X. Liu, J. You, Y. Wu, T. Li, L. Li, Z. Zhang, J. Ge, "Attention-Based Bidirectional GRU Networks for Efficient HTTPS Traffic Classification," *Information Sciences*, vol.541, no.1, pp.297-315, 2020. Article (CrossRef Link)

[17] Zhang J, Li F, Wu H, F. Ye, "Autonomous Model Update Scheme for Deep Learning Based Network Traffic Classifiers," in *Proc. of 2019 IEEE Global Communications Conference (GLOBECOM)*, Waikoloa, Hi, USA, pp.1-6, 2020. Article (CrossRef Link)

[18] H. Sun, Y. Xiao, J. Wang, J. Wang, Q. Qi, J. Liao, X. Liu, "Common Knowledge Based and One-Shot Learning Enabled Multi-Task Traffic Classification," *IEEE Access*, vol. 7, pp. 39485-39495, 2019. Article (CrossRef Link)

[19] U.Majeed, S.S.Hassan, C.S.Hong, "Cross-Silo Model-Based Secure Federated Transfer Learning for Flow-Based Traffic Classification," in *Proc. of 35th International Conference on Information Networking (ICOIN)*, Jeju Island, Korea (South), pp.588-593, 2021. Article (CrossRef Link)

[20] Rezaei S, Liu X, "How to Achieve High Classification Accuracy with Just a Few Labels: A Semi-supervised Approach Using Sampled Packets," *ArXiv*, pp.1-15, 2020. Article (CrossRef Link)

[21] J. Kornycky, O. Abdul-Hameed, A. Kondoz and B. C. Barber, "Radio Frequency Traffic Classification Over WLAN," *IEEE/ACM Transactions on Networking*, 25(1), 56-68, 2017. Article (CrossRef Link).

[22] L.Yang, Y.Dong, W. Tian, Z. Wang, "The Study of New Features for Video Traffic Classification," *Multimedia Tools and Applications*, vol.78, no.12, pp. 15839-15859, 2019. Article (CrossRef Link)

[23] S. Indolia, A. K. Goswami, S.P. Mishra, P. Asopa, "Conceptual Understanding of Convolutional Neural Network- A Deep Learning Approach," *Procedia Computer Science*, vol.132, no.1, pp. 679-688, 2018. Article (CrossRef Link)

[24] Qin Z, Yu F, Liu C, et al, "How convolutional neural network see the world - A survey of convolutional neural network visualization methods," *American Institute of Mathematical Sciences*, vol.1, no.2, pp.148-180, 2018. Article (CrossRef Link)

[25] [Online]. Available: https://www.unb.ca/cic/datasets/vpn.html

Lingyun Yang received her Ph.D degree in signal processing System from Nanjing University of Posts and Telecommunications, Nanjing, China. She is currently a teacher with the College of Physics and Electronic Information, Anhui Normal University, Wuhu, China. Her current research interests focus on end-to-end QoE provisioning and multimedia streaming.

Yu-ning Dong received his B. E, M. E degrees from Nanjing University of Posts & Telecommunications (NUPT), Ph.D degree in Electrical Engineeringfrom Southeast University, and M.Phil degree in Computer Science from The Queen's University of Belfast (QUB). He is currently Professor with College of Communications and Information Engineering, NUPT. His research interests include multimedia communications, wireless networking, and Network Traffic Identification.

Zai-jian Wang received his BE degree from Anhui Polytechnic University in 2002, MSc degree (2005) from University of Science and Technology of China, and PhD degree (2015) from Nanjing University of Posts and Telecommunications. His current research focus on multimedia big data, end-to-end QoS provisioning and multimedia streaming. He is a professor at College of Physics and Electronic information, Anhui Normal University, Wuhu, China.

Feifei Gao received her master degree from Dalian University of Technology, her current research interests focus on signal processing and image processing, she is currently a senior engineer of the college of Physics and Electronic Information, Anhui Normal University, Wuhu, China.