BULLETIN OF THE POLISH ACADEMY OF SCIENCES
TECHNICAL SCIENCES, Vol. 71(4), 2023, Article number: e145676
DOI: 10.24425/bpasts.2023.145676

Anonymous traffic classification based on three-dimensional Markov images and deep learning

Xin TANG[1,2], **Huanzhou LI**[1,2]*, **Jian ZHANG**[1,2], **Zhangguo TANG**[1,2], **Han WANG**[1,2], and **Cheng CAI**[1,2]

[1] School of Physics and Electronic Engineering, Sichuan Normal University, Chengdu 610101, Sichuan, China
[2] Institute of Network and Communication Technology, Sichuan Normal University, Chengdu 610101, Sichuan, China

Abstract. Illegal elements use the characteristics of an anonymous network hidden service mechanism to build a dark network and conduct various illegal activities, which brings a serious challenge to network security. The existing anonymous traffic classification methods suffer from cumbersome feature selection and difficult feature information extraction, resulting in low accuracy of classification. To solve this problem, a classification method based on three-dimensional Markov images and output self-attention convolutional neural network is proposed. This method first divides and cleans anonymous traffic data packets according to sessions, then converts the cleaned traffic data into three-dimensional Markov images according to the transition probability matrix of bytes, and finally inputs the images to the output self-attention convolution neural network to train the model and perform classification. The experimental results show that the classification accuracy and F1-score of the proposed method for Tor, I2P, Freenet, and ZeroNet can exceed 98.5%, and the average classification accuracy and F1-score for 8 kinds of user behaviors of each type of anonymous traffic can reach 93.7%. The proposed method significantly improves the classification effect of anonymous traffic compared with the existing methods.

Key words: anonymous network; traffic classification; three-dimensional Markov images; output self-attention; deep learning.

1. INTRODUCTION

The anonymous network is a technology and system that provides users with communication privacy protection, and its goal is to hide the network relationship between senders, receivers, and messages. It usually uses message relaying, traffic obfuscation, and data encryption to hide important information in packets from both sides of the communication. Multi-hop reverse proxies as well as resource-sharing storage are often used to mask the real address of the service provider and ensure that anonymous services are untraceable and unlocatable [1]. According to the network structure, anonymous networks can be divided into P2P anonymous networks and non-P2P anonymous networks. P2P anonymous networks usually design special nodes to maintain network status and node information, which weakens the concept of server and client. Compared with non-P2P anonymous networks, P2P anonymous networks realize the decentralization of the anonymous network.

With the development of anonymity network technology, anonymity networks such as Tor, I2P, Freenet, and ZeroNet are widely used and have a huge number of users. Tor is based on the second generation of onion routing technology. Users communicate anonymously through Tor, which can effectively protect privacy [2, 3]. I2P is a P2P anonymous network using one-way tunnel technology. The applications created on it can realize anonymous communication and have strong scalability, self-organization, and recovery capabilities [4]. Freenet is a completely self-organized P2P anonymous network and is one of the mainstream anonymous networks used in the early days. Compared with other anonymous networks, Freenet is faster and more stable [5]. ZeroNet is a new type of anonymity network. It does not provide anonymity protection by default, but users can communicate anonymously through the built-in Tor feature. This makes its users more secretive and it is increasingly difficult to discover communicating nodes and site and node relationships [6].

At present, the main application scenario of an anonymous network is Internet communication. Users can use anonymous technology to communicate anonymously on the Internet without worrying about privacy leakage. However, the anonymity of the anonymous network is also used by a large number of criminals. They use the anonymity of the anonymous network to build darknets and engage in various network criminal activities, such as drug trafficking, arms smuggling, network fraud, and human trafficking, which seriously threaten the network security of Internet users. The abuse of anonymous networks brings great challenges to network security, and the detection of traffic among them can effectively identify and curb these illegal activities. Tor, I2P, Freenet, and ZeroNet are the four mainstream dark network forms nowadays, so it is of great practical value to study the traffic of these types of anonymous networks.

To improve the classification effect of anonymous traffic, this paper proposes an anonymous traffic classification method based on three-dimensional Markov images and output self-attention convolutional neural network (OSACNN). The cleaned traffic session data are first converted into three-

*e-mail: lihuanzhou@sicnu.edu.cn

Manuscript submitted 2022-10-14, revised 2023-02-25, initially accepted for publication 2023-04-01, published in August 2023.

dimensional Markov images, and then the useful features are automatically extracted from the images for classification by OSACNN. The main contributions are as follows:

- A three-dimensional Markov image conversion method based on a transfer probability matrix is proposed, which can generate anonymous traffic images of fixed size, effectively extract traffic information and reduce information redundancy.
- The self-attention mechanism is introduced into the output space of the convolutional neural network, and an anonymous traffic image classification model (OSACNN) is proposed, which improves the classification accuracy.
- An anonymous traffic classification method combining three-dimensional Markov images and OSACNN is proposed, which has a better classification effect than other methods.

2. RELATED WORK

Network traffic identification and classification technology is a very important part of network defense and network anomaly detection. In recent years, many scholars have focused on identifying and classifying anonymous network traffic. This paper summarizes related research from the perspective of anonymous traffic classification objects.

Papers [7, 8] studied the identification of Tor traffic; [9] analyzed the NTCP communication protocol of I2P, and realized the detection and identification of I2P network traffic; [10] used a machine learning approach to classify Freenet traffic. However, these studies are limited to simple 2-class identification of normal traffic and anonymous traffic and do not go deep into the level of user behavior.

The paper [11] proposed a Tor traffic identification and multi-level classification framework based on network traffic characteristics, which realized the identification of mobile anonymous traffic, traffic types of anonymous traffic, and user behavior. [12] proposed an encrypted Tor traffic detection and classification method based on a deep neural network (DNN). This method achieves 99.89% accuracy in identifying Tor traffic on the ISCXTor2016 dataset, and 95.6% accuracy in classifying user behavior of Tor traffic. [13] considered the characteristics of the payloads of Tor and non-Tor data packets from 8 different user behaviors and extracted relevant features to train the machine learning model. [14] designed a new deep learning model AttCorr for the classification of Tor, which extracted the original features of the ingress and egress streams of the Tor network, and then generated samples and input them to AttCorr for training. These studies explore user behavioral classifications of traffic but only analyze a single type of anonymous traffic.

[15] used improved convolutional long short memory (CNN-LSTM) and convolutional gradient recursive unit (CNN-GRU) to perform experiments on the classification of Tor and VPN traffic, but the article only focused on these two types of traffic, and the classification accuracy of user behavior was only 89%. [16] used the CIC-Darknet2020 dataset to evaluate the classification effects of SVM, RF, CNN, and AC-GAN. The RF model achieved the best results, but it only focused on Tor and VPN traffic. [17, 18] studied the classification of three kinds of darknet traffic (Tor, I2P, and JonDonym), but the classification only covers a small range of user behavior, and the accuracy rate is only 66.76%. [19] captured the real traffic of Tor, I2P, Freenet, and ZeroNet, and publishes the dataset. They extracted 26 time-based anonymous traffic features and trained a hierarchical classifier composed of six local classifiers. Their method achieves 99.42% accuracy in identifying anonymous traffic, 96.9% accuracy in classifying four types of anonymous traffic, and 91.6% accuracy in classifying user behavior.

The research on the above anonymous traffic classification is summarized as follows:

1. Most of the research focuses on Tor network traffic, and there is less research work on other mainstream anonymous traffic such as I2P, Freenet, and ZeroNet.
2. The accuracy rate of simple two-classification identification of anonymous traffic is more than 99%, but the accuracy of multi-classification of anonymous traffic and user behavior classification still needs to be improved.
3. Current traditional machine learning-based methods rely heavily on feature selection and require a lot of work to filter and extract various features of the raw traffic with unsatisfactory accuracy. Most methods based on deep learning directly convert feature data or bytes in traffic packets into images. The disadvantage of converting feature data into images is the same as that of traditional machine learning methods. The disadvantage of converting traffic bytes into images is that the information extraction of traffic data is difficult. As a result, the accuracy of classification based on deep learning methods is also unsatisfactory.

In this paper, we investigate the mainstream anonymous traffic multi-classification and anonymous traffic user behavior classification and propose an anonymous traffic classification method based on three-dimensional Markov images and output self-attentive convolutional neural network. The proposed method can effectively extract information from anonymous traffic while avoiding the tedious feature selection process, which significantly improves the effectiveness of classification.

3. MATERIALS AND METHODS

3.1. Markov image

In anonymous traffic classification, traditional machine learning methods have certain limitations, and methods based on anonymous traffic images and deep learning become a new way because they avoid the complex feature selection process. Most of the current classification methods based on anonymous traffic images need to unify the length of the data and directly convert the bytes of the data packet into pixel data in the image. These methods do not consider the information loss caused by the interception process of traffic data and the information redundancy caused by direct conversion, and cannot effectively extract information.

A Markov chain is a stochastic process in the state space that undergoes a transition from one state to another. This process requires a “memoryless” property: the probability that a state

will undergo a transition at a given moment depends only on the state before it. [20] first proposed the Markov image for the classification of malware. Markov image is a fixed-size pixel matrix that reflects data statistical characteristics and can better extract information from data. The conversion process is shown below.

The data file to be converted is regarded as a byte stream, and the byte distribution of similar data is very similar. The byte stream can be represented as a random process as follows:

$$B_i, \quad i \in \{0,1,\ldots,N-1\}, \tag{1}$$

where B_i is the value of the i-th byte, and N is the length of the data file, $B_i \in \{0,1,\ldots,255\}$. Suppose that for two adjacent bytes X_i and X_{i+1}, the probability of X_{i+1} is only related to X_i, then the random variable B_i is a Markov chain:

$$P(X_{i+1}|X_0,X_1,\ldots,X_i) = P(X_{i+1}|X_i). \tag{2}$$

The byte transfer probability matrix is as follows:

$$P(X_{i+1}|X_i) = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,255} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,255} \\ \vdots & \vdots & p_{x_i,x_{i+1}} & \vdots \\ p_{255,0} & p_{255,1} & \cdots & p_{255,255} \end{bmatrix}, \tag{3}$$

where $p_{x_i,x_{i+1}}$ indicates the transfer probability. x_i and x_{i+1} are actual values of X_i and X_{i+1}, x_i and $x_{i+1} \in \{0,1,\ldots,255\}$.

The transfer probability is calculated as follows:

$$p_{x_i,x_{i+1}} = \frac{P(x_i,x_{i+1})}{P(x_i)} = \frac{f(x_i,x_{i+1})}{\sum_{j=0}^{255} f(x_i,j)}. \tag{4}$$

In the formula, $f(x_i,x_{i+1})$ means that x_i is followed by the frequency of x_{i+1}, $p_{x_i,x_{i+1}} \in (0,1]$.

To map the range of values from 0–255 of the grayscale map, a fixed-size Markov image is obtained by enlarging the value of the transfer probability matrix by a factor of 255 and converting it to an image. At a higher level, Markov images are equivalent to graphing the state transfer probability matrix. Compared with other images, the Markov image overcomes their shortcomings and has the following characteristics:

1. It does not need to unify the length of the data and can retain complete data information.
2. It is a statistical image, representing the distribution characteristics of data. Compared with direct conversion, it avoids a lot of information redundancy.

Traffic data itself is a one-dimensional byte stream, which is more consistent with the concept of Markov image. Therefore, the Markov image is more suitable for traffic data classification. [21] proposed a method to convert encrypted traffic into Markov images for classification and achieved good results. However, the traffic byte data is not a Markov chain in the strict sense, and the Markov image only considers the relationship between adjacent bytes, which will cause the loss of effective information.

3.2. The LeNet-5 model

The LeNet-5 model was first proposed in [22], which is a classical convolutional neural network. The network consists of a total of 7 layers, which include convolutional layers C1, C3, and C5, pooling layers S2, and S4, a fully connected layer F6, and an output layer: Output. The output is a Gaussian connected layer that uses the softmax function to classify the output image. Figure 1 shows the structure of the LeNet-5 model.

The three-dimensional Markov image is a statistical image, and the data in the image is relatively scattered. A small number of convolution layers can extract features from the image very well, while the deep convolution network is difficult to converge. Therefore, the OSACNN designed in this study refers to the network structure of the LeNet-5 model. Compared with other deep models, it has fewer convolutional layers and is suitable for the classification of three-dimensional Markov images.

4. THE PROPOSED METHOD

This section describes the classification method proposed in this paper, and its framework is shown in Fig. 2. It includes two parts: the transformation of three-dimensional Markov images and the classification of anonymous traffic images.

Part 1: Convert anonymous traffic data to a three-dimensional Markov image. Experimental studies show that sessions are a better form of traffic representation [23]. "Session segmentation" is the segmentation of raw traffic data by session. First, the input packets are segmented according to the session. Second, to avoid the influence of IP addresses and MAC addresses in session packets on the model learning of traffic characteristics, this paper uses randomly generated data to cover the original address information. Finally, according to the transition proba-

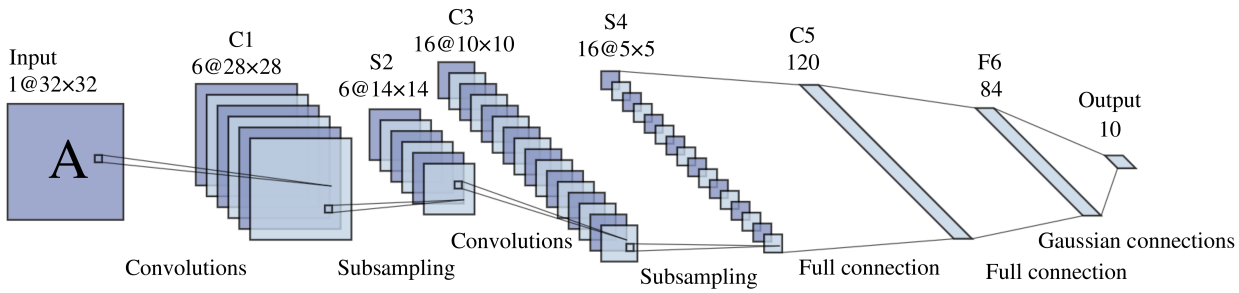

Fig. 1. The LeNet-5 model structure

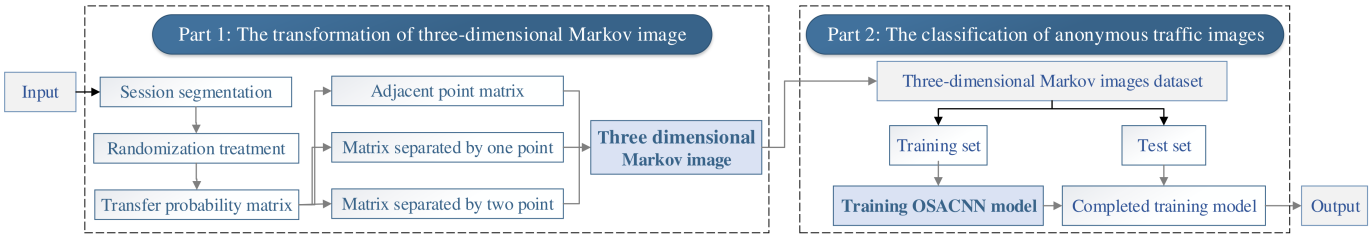

Fig. 2. Framework of the proposed method

bility matrix between data bytes, the session data packet is converted into a three-dimensional Markov image to form a three-dimensional Markov images dataset.

Part 2: Building OSACNN to classify anonymous traffic images. The converted three-dimensional Markov images dataset is divided into a training set and a test set. Use the samples in the training set to train the model of OSACNN. The test set is used to evaluate the classification effect of the proposed method.

4.1. Conversion of three-dimensional Markov image

In this paper, we propose a three-dimensional Markov image transformation method, which considers the connection between adjacent traffic bytes and interval traffic bytes to solve the shortcomings of traditional Markov images in traffic classification. The traffic characteristic information that should exist between interval bytes is explained from the perspective of a multi-order Markov chain, and the transfer probability of adjacent bytes is introduced as the information weight between interval bytes. Since the correlation information between interval traffic bytes decreases as the interval distance increases, this paper selects the interval bytes with close distance and fills the traffic information of adjacent points, the traffic information of interval 1 byte, and the traffic information of interval 2 bytes into RGB three channels to form a three-dimensional Markov image. The three-dimensional Markov image contains more traffic information than the traditional Markov picture of a single channel. Figure 3 is a schematic diagram of traffic byte information padding, X_i on behalf of the traffic bytes, x_i means the actual value of X_i, and N represents the length of the session traffic data.

Assuming that for two adjacent traffic bytes X_i and X_{i+1}, the probability of X_{i+1} is only related to X_i. Then the traffic bytes are a Markov chain. Noting the transfer probability matrix of adjacent bytes as P_1, then $P_1 = P(X_{i+1}|X_i)$, it has the following values:

$$P_1 = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,255} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,255} \\ \vdots & \vdots & p_{x_i,x_{i+1}} & \vdots \\ p_{255,0} & p_{255,1} & \cdots & p_{255,255} \end{bmatrix}, \tag{5}$$

where $i \in (0, N-2]$. The information data of adjacent traffic bytes is noted as D_1, which has the following value:

$$D_1 = 255 \times P_1 . \tag{6}$$

Assuming that for traffic bytes X_i, X_{i+1} and X_{i+2}, the probability of X_{i+2} is only related to X_i and X_{i+1}, then the traffic bytes are a second-order Markov chain:

$$P(X_{i+2}|X_0, X_1, ..., X_{i+1}) = P(X_{i+2}|X_i, X_{i+1}). \tag{7}$$

If only the relationship between X_{i+2} and X_i, X_{i+2} and X_{i+1} is considered, then $P(X_{i+2}|X_i, X_{i+1})$ can be approximated by

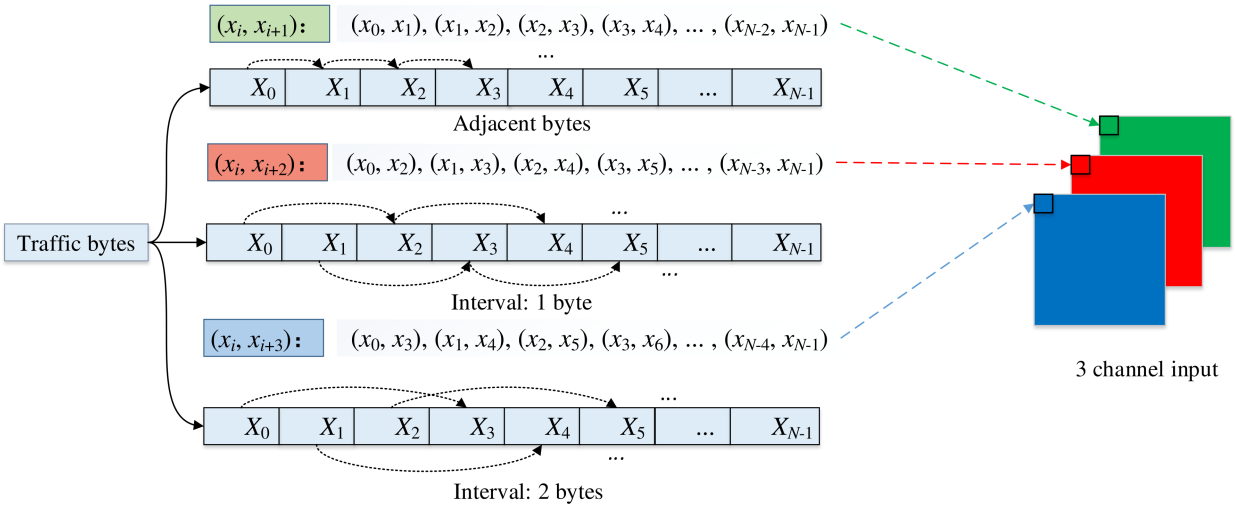

Fig. 3. Traffic byte information filling schematic

a combination of the basic transfer probability matrices as follows:

$$P(X_{i+2}|X_i,X_{i+1}) \approx \{P(X_{i+2}|X_i),P(X_{i+2}|X_{i+1})\}, \tag{8}$$

where $P(X_{i+2}|X_{i+1})$ is the transfer probability matrix of adjacent bytes (P_1); $P(X_{i+2}|X_i)$ is the transfer probability matrix of interval 1 byte, which obviously represents the traffic information of interval 1 byte. The transfer probability matrix with an interval of 1 byte is written as P_2, then $P_2 = P(X_{i+2}|X_i)$, which has the following values:

$$P_2 = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,255} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,255} \\ \vdots & \vdots & p_{x_i,x_{i+2}} & \vdots \\ p_{255,0} & p_{255,1} & \cdots & p_{255,255} \end{bmatrix}, \tag{9}$$

where $i \in (0,N-3]$. Since there is less valid information between interval traffic bytes than between adjacent traffic bytes, the traffic byte data converge more closely to a Markov chain. Therefore, the transfer probability of adjacent traffic bytes is introduced as the weight of the interval traffic byte information. The values of the transfer probability matrix P_2' after the weight transformation are shown below:

$$P_2' = P_2 \cdot P_1. \tag{10}$$

The traffic information data with a 1-byte interval is recorded as D_2, and its value is:

$$D_2 = 255 \times P_2'. \tag{11}$$

Similarly, assuming that for traffic bytes X_i, X_{i+1}, X_{i+2}, and X_{i+3}, the probability of X_{i+3} is only related to X_i, X_{i+1}, and X_{i+2}, the traffic byte is a third-order Markov chain:

$$P(X_{i+3}|X_0,X_1,\ldots,X_{i+2}) = P(X_{i+3}|X_i,X_{i+1},X_{i+2}). \tag{12}$$

If only the relationship between X_{i+3} and X_i, X_{i+3} and X_{i+1}, X_{i+3} and X_{i+2} is considered, then $P(X_{i+3}|X_i,X_{i+1},X_{i+2})$ can likewise be approximated by a combination of the basic transfer probability matrices as follows:

$$P(X_{i+3}|X_i,X_{i+1},X_{i+2}) \approx \{P(X_{i+3}|X_i),P(X_{i+3}|X_{i+1}),P(X_{i+3}|X_{i+2})\}, \tag{13}$$

where $P(X_{i+3}|X_{i+2})$ is the transfer probability matrix of adjacent bytes (P_1); $P(X_{i+3}|X_{i+1})$ is the transfer probability matrix of interval 1 byte (P_2); $P(X_{i+3}|X_i)$ is the transfer probability matrix of interval 2 bytes, which can be on behalf of the traffic information of interval 2 bytes. Noting the transfer probability matrix with an interval of 2 bytes as P_3, then $P_3 = P(X_{i+3}|X_i)$, which has the following values:

$$P_3 = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,255} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,255} \\ \vdots & \vdots & p_{x_i,x_{i+3}} & \vdots \\ p_{255,0} & p_{255,1} & \cdots & p_{255,255} \end{bmatrix}, \tag{14}$$

where $i \in (0,N-4]$. The transfer probabilities of adjacent traffic bytes are introduced as weights, and the values of the transfer probability matrix P_3' after the weight transformation are shown below:

$$P_3' = P_3 \cdot P_1. \tag{15}$$

The traffic information data with 2 bytes interval is recorded as D_3, and its value is:

$$D_3 = 255 \times P_3'. \tag{16}$$

The data of D_2, D_1, and D_3 are filled into the R, G, and B channels of the image respectively, which completes the conversion of the three-dimensional Markov image.

To facilitate human eye distinction, the inverse transformation of the base color of the image is performed by equation (17), taking the opposing events of the transfer probability matrix:

$$P^{**} = 1 - P^*, \tag{17}$$

where P^* represents the untransformed transfer probability matrix and P^{**} means the value after performing the base color inversion transformation, and P^{**} is used as the fill data of the image to realize the base color inversion operation. Figure 4 shows the comparison of a Markov image and a three-dimensional Markov image for the same session traffic, where Figs. 4a and 4b are the Markov image and its inversion, respectively; and Figs. 4c and 4d is the three-dimensional Markov image and its inversion, respectively.

As seen in Fig. 4, the three-dimensional Markov image contains more traffic information in its RGB three channels while retaining the data distribution characteristics.

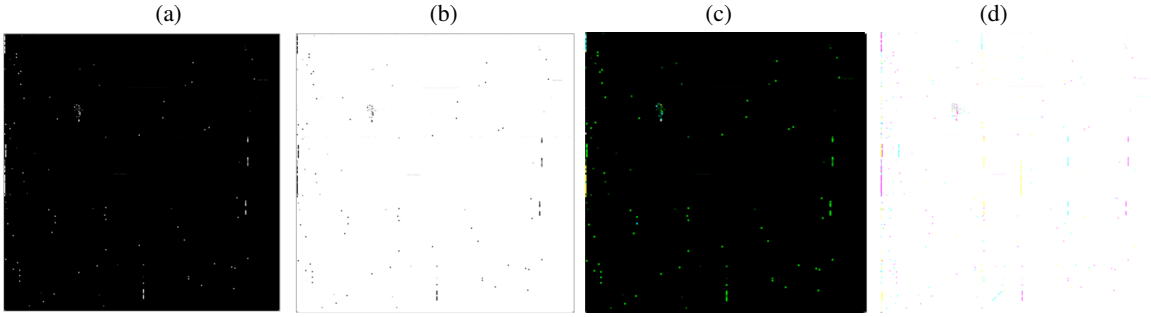

Fig. 4. Comparison of Markov image and three-dimensional Markov image

4.2. Anonymous traffic classification

Attention mechanisms allow neural networks to focus on critical information and reduce the attention of other information, thus improving the efficiency and accuracy of task processing. The self-attention mechanism is a variation of the attention mechanism that reduces the reliance on external information and is better at capturing the internal relevance of data or features [24, 25]. To improve the classification effect of anonymous traffic, this paper refers to the convolutional structure of the LeNet-5 model, introduces the idea of a self-attentive mechanism to the output space of a convolutional neural network, and constructs an output self-attentive convolutional neural network model (OSACNN). OSACNN assigns small weights to output values with small values and then performs the calculation of loss values, to reduce the neural network learning of useless information in output values with small values. Focusing attention on a more reasonable range of intervals to improve the performance of classification. Compared with traditional CNN, OSACNNN can improve the classification effect of neural network models on unbalanced datasets to a certain extent. The structure of the OSACNN network is shown in Fig. 5, which consists of a 10-layer network, including the Convolution layer, the Max-Pool layer, the Flatten layer, the Dense layer, the Dropout layer, and the Attention layer.

Layer 1 is the Convolution layer with the activation function ReLU. 80 convolution kernels are used in this layer, each with a size of 5×5. Each convolution kernel is convolved with the original input image, and 80 feature maps are obtained, with a size of 126×126.

Layer 2 is the Max-Pool layer, which uses maximum pooling. The size of the feature map is 63×63 after pooling, and the number of feature maps is kept constant.

Layer 3 is the Convolution layer with the activation function ReLU. 130 convolution kernels are used in this layer, and the size of each convolution kernel is 4×4. 130 feature maps are obtained after convolution, and the size of the feature maps is 30×30.

Layer 4 is the Max-Pool layer, which uses maximum pooling. The size of the feature map after pooling is 15×15, and the number of feature maps is 130.

Layer 5 is the Flatten layer, which acts as a transition from the convolutional layer to the fully connected layer, and one-dimensionalizes the multidimensional input. It unfolds the data coming from layer 4 into one-dimensional data, and the size of the unfolded data is 1×29250.

Layers 6 and 7 are Dense layers, which take the features extracted earlier and vary them nonlinearly to extract the association between features.

Layer 8 is the Dropout layer, which is used to prevent the model from overfitting and improve its robustness of the model.

Layer 9 is the Dense layer, which maps the data to the output dimension.

Layer 10 is the Output self-attention layer, which calculates the weights based on the data passed from layer 8, then multiplies the weights with the original data passed, and finally outputs the predicted labels for the anonymous traffic categories.

The calculation steps of the output self-attention are shown in Fig. 6.

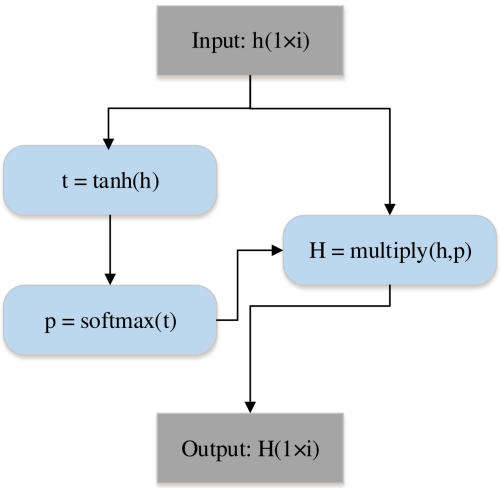

Fig. 6. Output self-attention calculation steps

The function in Fig. 6 is calculated as follows:

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}}, \tag{18}$$

$$\text{softmax}(z_i) = \frac{e(z_i)}{\sum_j e(z_j)}, \tag{19}$$

$$\text{multiply}(a, b) = a \cdot b. \tag{20}$$

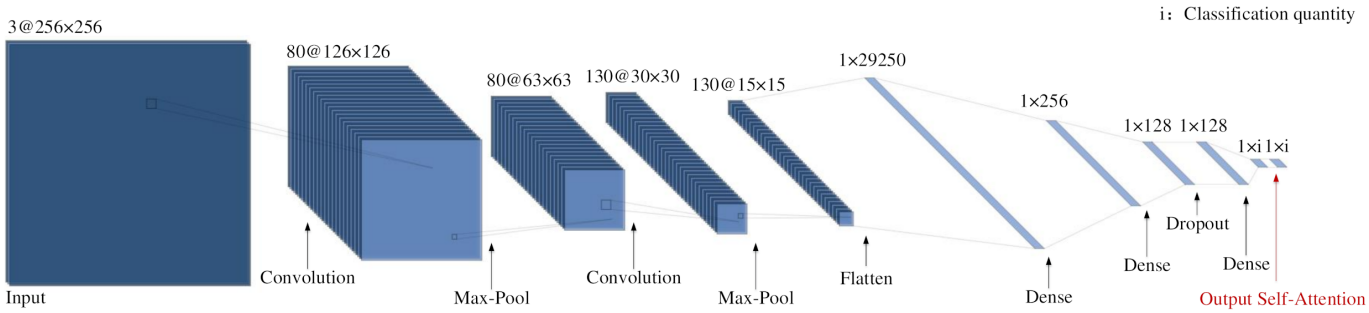

Fig. 5. OSACNN model structure diagram

In equation (20), a and b are the arrays that satisfy the inner product condition. The computational effect of the output self-attention is verified with the following input function:

$$h(x) = x^3 \quad (x > 0). \tag{21}$$

Six values of x are arbitrarily selected for calculation, and the values of each key process are shown in Table 1, where x is the selected coordinate value, $h(x)$ is the original input value, $p(x)$ is the calculated weight value, and $H(x)$ is the output value after the weight change.

Table 1
Values for key processes

x	0	0.4	0.8	1.2	1.6	2
$h(x)$	0	0.064	0.512	1.728	4.096	8
$p(x)$	0.085	0.091	0.137	0.219	0.232	0.233
$H(x)$	0	0.005	0.070	0.378	0.954	1.864

Figure 7 shows a line graph of the effect of performing the output self-attention calculation.

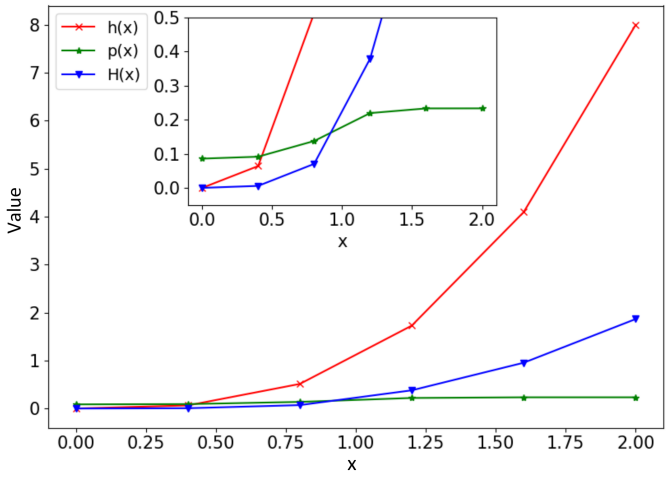

Fig. 7. Output self-attention calculation steps

As can be seen in Fig. 7, the proposed output self-attention in this paper assigns weights of similar size to larger output values, and for smaller output values in the interval, the smaller the value is, the smaller the weight is, reducing the overall impact of smaller output values on the model training.

In the OSACNN training phase, the Adam optimizer is used to learn the parameters of the model and cross-entropy loss is used to train the network. The initial values of the convolution kernel are generated randomly. The parameter values are continuously updated in real time until the model is fitted.

5. EXPERIMENTAL RESULTS AND ANALYSIS

The experimental environment of this experiment is shown in Table 2.

Table 2
Experimental environment

Parameters	Numerical value
Operating System	Windows 10
CPU	AMD Ryzen 7 4800H
GPU	GTX 1650
RAM	16G
Hard Disk	512G SSD
Python	3.8
Tensorflow	2.3

5.1. Dataset and evaluation indicators

5.1.1. Dataset

There are few publicly available anonymous traffic datasets, and this paper has collected relevant information through references and online websites, as shown in Table 3.

Table 3
Anonymous traffic dataset information

Source	Dataset	Content	Release date/year
CIC	ISCXTor2016	Tor	2016
CIC	CIC-Darknet2020	Tor, VPN	2020
Hu, *et al.*	Darknet-dataset-2020	Tor, I2P, ZeroNet, Freenet	2020

Most of the publicly available datasets are from the Canadian Institute for Cybersecurity (CIC), which released two traffic datasets in 2016, ISCXTor2016 and ISCXVPN2016, containing different types of user behavior traffic. In 2020 CIC merged these two datasets to form the CIC-Darknet2020 dataset. Many studies were conducted on these datasets, but they have two obvious problems. The first is that the dataset contains only two types of traffic, Tor and VPN, and lacks data on other mainstream anonymous traffic; the second is that the data were captured a relatively long time ago and cannot be fully applied to the current classification situation in the context of evolving anonymization network technologies.

The experiments in this paper use the Darknet-dataset-2020 dataset collected by Hu *et al.* [19] in 2020. This dataset contains eight types of user behavior traffic (browsing, chat, email, audio streaming, video streaming, file transfer, P2P, VoIP) from four mainstream anonymous traffic categories, Tor, I2P, ZeroNet, and Freenet, with a total of 10.7 GB size data. After the session slicing of this dataset, the detailed data distribution is shown in Table 4.

Table 4
Darknet-dataset-2020 dataset distribution

Type	Browsing	Chat	Email	FileT	P2P	Audio	Video	VoIP	Total
Freenet	390	226	358	178	–	–	348	–	1500
I2p	477	364	747	1137	981	–	–	–	3706
Tor	365	118	134	967	674	352	253	385	3248
ZeroNet	6997	1026	372	1057	559	392	698	–	11101

5.1.2. Evaluation indicators

The experiments visualize the classification results by confusion matrix and evaluate the performance of the classification using accuracy, precision, recall, and F1-score. Table 5 shows the definition of the confusion matrix. The TP, FP, FN, and TN represent True Positive, False Positive, False Negative, and True Negative, respectively.

Table 5
Confusion matrix

Real Results	Predicted results	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

The accuracy represents the proportion of correctly classified samples to the total number of samples and is calculated as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN}. \tag{22}$$

The precision represents the proportion of samples correctly classified in this class to the number of all samples assigned to that class and is calculated as follows:

$$\text{Precision} = \frac{TP}{TP+FP}. \tag{23}$$

The recall represents the proportion of samples that were correctly classified to all samples that should have been correctly classified and is calculated as follows:

$$\text{Recall} = \frac{TP}{TP+FN}. \tag{24}$$

Relying on a single metric does not provide a more comprehensive assessment of the performance of the classifier, so the performance is evaluated using the F1-score, a combined metric of accuracy and completeness, which is calculated as follows:

$$\text{F1-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{25}$$

5.2. Experimental results

1. Comparison experiments of different feature images

To investigate the effectiveness of three-dimensional Markov images in classifying anonymous traffic, the results of grayscale images, traditional Markov images, and three-dimensional Markov images in classifying anonymous traffic are compared. The experimental dataset and the data pre-processing are kept in the same way. Since the length of the traffic data is inconsistent after session slicing, converting grayscale images requires uniform processing of the length of the session data to facilitate training and classification. The connection data and part of the content data in the first part of the traffic session generally best reflect the intrinsic characteristics of the traffic, so the first 256 bytes of data of each session are intercepted, and those less than 256 bytes are complemented with 0 at the end, and the intercepted data are converted into a grayscale image of 16×16 size. The model used in the experiments for the traditional Markov images and the three-dimensional Markov images (Td-Markov images) is OSACNN, and the model used for the grayscale images is modified on OSACNN due to the image size by removing a Dense layer from the model. Table 6 shows the results of the experimental comparison under the conversion of anonymous traffic to different feature images, and the numbers in parentheses represent the number of classifications, which are 4, 5, 5, 8, and 7 classifications.

As seen from the data in Table 6, grayscale images cannot extract complete information from traffic data with widely varying lengths, leading to their poor results in anonymous traffic classification, while traditional Markov images avoid the complexity of pre-processing and retain most of the useful content in the traffic, so the classification is better. The three-dimensional Markov images extract more useful traffic information while retaining the advantages of traditional Markov images, so its classification is the best.

To reflect the computational cost introduced by the proposed image conversion method, it is illustrated in two dimensions: image storage space and image conversion time. Table 7 compares the detailed cost data of conventional Markov image conversion with the three-dimensional Markov image conversion.

When faced with large traffic session data, both images consume more computational resources. From the data in Table 7, it is clear that the generation of the three Markov images requires more time and cost. On top of storage space, since both images are stored in JPG format, due to the compression fea-

Table 6
Experimental comparison of different feature images

Feature image	Grayscale images		Markov images		**Td-Markov images**	
	Accuracy	F1-score	Accuracy	F1-score	**Accuracy**	**F1-score**
Anonymous traffic (4)	85.92%	85.13%	97.17%	97.17%	**98.56%**	**98.55%**
Freenet behavior (5)	58.00%	57.90%	88.33%	88.34%	**92.33%**	**92.25%**
I2P behavior (5)	58.78%	56.78%	91.08%	91.11%	**94.05%**	**94.05%**
Tor behavior (8)	65.46%	64.52%	90.93%	90.97%	**91.09%**	**90.93%**
ZeroNet behavior (7)	63.06%	48.77%	95.09%	94.99%	**97.34%**	**97.32%**

Table 7
Image conversion cost comparison

Feature image	Average storage space	Total storage space	Average conversion time	Total conversion time
Markov images	12.75 KB	243.46 MB	0.51 S	9955 S
Td-Markov images	**12.07 KB**	**230.62 MB**	**1.88 S**	**36879 S**

ture of this format, it makes the storage space of three Markov images rather smaller.

2. Adding output self-attention comparison experiments

To demonstrate the effect of adding the self-attention mechanism on the output space, experimental validation is performed. Table 8 shows the comparison of classification effects before and after adding output self-attention.

Table 8
Comparison of the classification effect of adding output self-attention

Whether to add?	No		**Yes**	
	Accuracy	F1-score	**Accuracy**	**F1-score**
Anonymous traffic (4)	98.53%	98.52%	**98.56%**	**98.55%**
Freenet behavior (5)	91.00%	90.97%	**92.33%**	**92.25%**
I2P behavior (5)	93.37%	93.39%	**94.05%**	**94.05%**
Tor behavior (8)	91.56%	91.52%	**91.09%**	**90.93%**
ZeroNet behavior (7)	97.16%	97.12%	**97.34%**	**97.32%**

From the data in Table 8, it can be seen that the classification effect of anonymous traffic is improved after adding the self-attention mechanism to the output space. Tor traffic has a smaller amount of data, but the largest number of classifications. With the attention focused on the larger output values, the model cannot learn enough about the features with smaller output values. Without the support of data volume, the output self-attention causes the classification effect of Tor to decrease instead.

3. Comparison experiment of different anonymous traffic classification methods

A total of five sets of experiments are conducted by the proposed method for four types of anonymous traffic multi-classification of Tor, I2P, ZeroNet, and Freenet, and the classification of user behavior for each type of anonymous traffic. The confusion matrices of the proposed method in the five sets of classification experiments are shown in Fig. 8, where Fig. 8a shows the confusion matrix of the four anonymous traffic multi-classifications; Figs. 8b, 8c, 8d, and 8e show the confusion matrices of the Freenet traffic user behavior classification, I2P traffic user behavior classification, ZeroNet traffic user behavior classification, and Tor traffic user behavior classification, respectively. The classification performance of the methods in this paper is shown in Table 9, which contains the accuracy, precision, recall, and F1-score.

Table 9
Classification performance of the proposed method

	Accuracy	Precision	Recall	F1-score
Anonymous traffic (4)	98.56%	98.58%	98.56%	98.55%
Freenet behavior (5)	92.33%	92.32%	92.33%	92.25%
I2P behavior (5)	94.05%	94.07%	94.05%	94.05%
ZeroNet behavior (7)	97.34%	97.32%	97.34%	97.32%
Tor behavior (8)	91.09%	91.12%	91.09%	90.93%

From the data in Fig. 8 and Table 9, it can be seen that the proposed method can achieve 98.56% accuracy in multi-classifying four types of anonymous traffic, Tor, I2P, ZeroNet, and Freenet, and the average accuracy in classifying the user behavior of each type of anonymous traffic can reach 93.70%. Since deep learning relies heavily on the amount of data, the data of Freenet traffic and Tor traffic in the dataset used in this paper is small, resulting in lower classification results for these two types of traffic than the other two.

To verify the effectiveness of the proposed classification method (Tpm), this paper is compared with other anonymous traffic classification methods. Table 10 shows the comparison

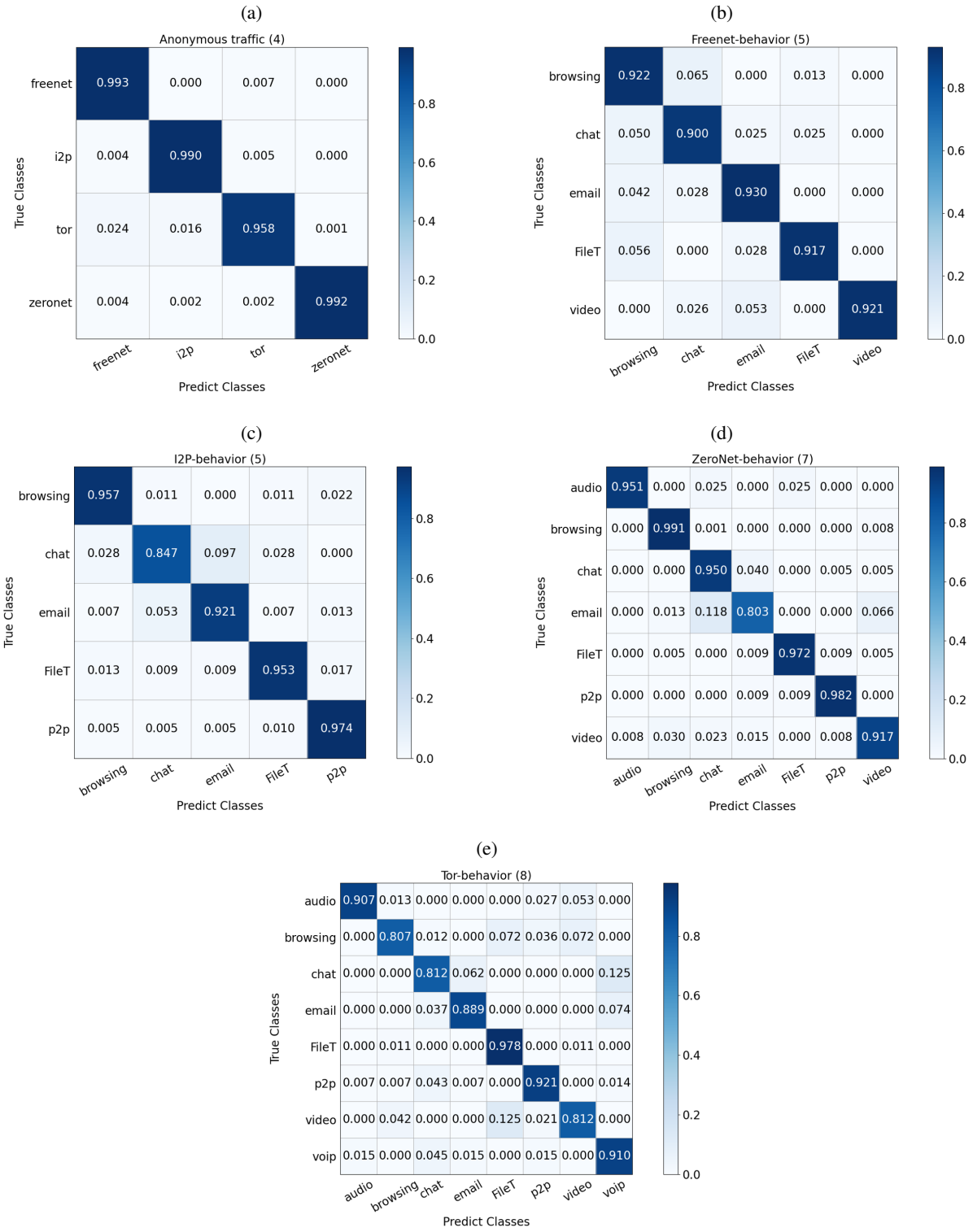

Fig. 8. Confusion matrix for 5 sets of classification experiments

results of different anonymous traffic classification methods, where the result data of other classification methods are obtained from the literature [19].

As can be seen from the data in Table 10, Freenet traffic and Tor traffic are not well classified due to fewer session data, but overall, the proposed method outperforms all other methods in classification. Compared with RF, XGBoost, and LightGBM algorithms, which perform better in anonymous traffic classification, the proposed method in this paper has a greater improvement in accuracy and F1-score, and can effectively distinguish different anonymous network traffic and the user behavior of anonymous network traffic.

Table 10
Comparison of the effects of different classification methods

Methods	Anonymous traffic (4)		Freenet behavior (5)		I2P behavior (5)		Tor behavior (8)		ZeroNet behavior (7)	
	accuracy	F1-score	accuracy	F1-score	accuracy	F1-score	accuracy	F1-score	accuracy	F1-score
LR	62.80%	62.89%	41.79%	36.42%	33.91%	29.20%	52.26%	51.89%	51.42%	48.11%
DT	96.16%	96.16%	96.66%	96.66%	90.51%	90.52%	83.79%	83.82%	90.63%	90.65%
RF	96.55%	96.55%	96.66%	96.67%	90.76%	90.76%	84.64%	84.65%	91.23%	91.18%
XGBoost	96.85%	96.85%	96.75%	96.74%	91.29%	91.28%	85.35%	85.38%	92.98%	92.98%
GBDT	95.23%	95.23%	96.85%	96.85%	90.84%	90.82%	84.29%	84.27%	86.67%	86.16%
LightGBM	96.33%	96.33%	96.46%	96.46%	90.27%	90.25%	85.24%	85.28%	93.17%	93.18%
LSTM	87.18%	87.17%	–	–	–	–	73.04%	73.48%	66.31%	63.76%
Tpm	**98.56%**	**98.55%**	**92.33%**	**92.25%**	**94.05%**	**94.05%**	**91.09%**	**90.93%**	**97.34%**	**97.32%**

6. CONCLUSIONS

Anonymous networks pose a significant risk to people's cyberspace security, so it is necessary to identify and classify anonymous traffic among them. In this paper, we propose an anonymous traffic classification method based on three-dimensional Markov images and output self-attention convolutional neural networks, which converts anonymous traffic into three-dimensional Markov images and classifies them by OSACNN. Compared with other conventional images, three-dimensional Markov images can retain useful information in traffic data to a greater extent. OSACNN incorporates self-attention in the output space, which reduces the learning of useless information by the neural network in the case of small output values and is more conducive to the learning of feature information by the model. Compared with other classification methods, the proposed method can significantly improve the effectiveness of anonymous traffic classification. The three-dimensional Markov images used in this paper require more computational resources. How to simplify the method of computation and find a better way to assign weights in the image transformation will be the focus of the next research. Few data on Freenet traffic and Tor traffic in the dataset used in this paper lead to low accuracy in classifying these two types of traffic. The next step is planned to supplement the corresponding anonymous traffic data to further improve the classification effect and generalization ability of the method.

ACKNOWLEDGEMENTS

The authors express their acknowledgment of the anonymous review.

REFERENCES

[1] L. Junzhou, Y. Ming, L. Zhen, W. Wenjia, and G. Xiaodan, "Anonymous Communication and Darknet: A Survey," *J. Comput. Res. Dev.*, vol. 56, p. 103, 2019, doi: 10.7544/issn1000-1239.2019.20180769.

[2] R. Dingledine, N. Mathewson, and P. Syverson, "Tor: The Second-Generation Onion Router," *J. Frankl. Inst.*, 2004, doi: 10.1016/0016-0032(45)90142-6.

[3] M.G. Reed, P.F. Syverson, and D.M. Goldschlag, "Anonymous connections and onion routing," *EEE J. Sel. Areas Commun.*, vol. 16, no. 4, pp. 482–494, 1998, doi: 10.1109/49.668972.

[4] F. Astolfi, J. Kroese, and J. Van Oorschot, "I2p-the invisible internet project," Leiden University Web Technology Report, 2015.

[5] I. Clarke, O. Sandberg, B. Wiley, and T.W. Hong, "Freenet: A distributed anonymous information storage and retrieval system," in *Designing privacy enhancing technologies*, 2001: Springer, pp. 46–66, doi: 10.1007/3-540-44702-4_4.

[6] "Open, free and uncensorable websites, using Bitcoin cryptography and BitTorrent network," ZeroNet, 2022. [Online]. Available: https://zeronet.io/zh.

[7] R. Jansen, M. Juarez, R. Galvez, T. Elahi, and C. Diaz, "Inside Job: Applying Traffic Analysis to Measure Tor from Within," in *NDSS*, 2018, doi: 10.14722/ndss.2018.23279.

[8] W. Juan, C. Shimin, Z. Jun, H. Bin, and S. Lei, "Identification of Tor Anonymous Network Traffic Based on Machine Learning," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, 2021, pp. 150–153, doi: 10.1109/ICCWAMTIP53232.2021.9674056.

[9] H. Yin and Y. He, "I2P anonymous traffic detection and identification," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2019, pp. 157–162, doi: 10.1109/ICACCS.2019.8728517.

[10] S. Lee, S.-h. Shin, and B.-h. Roh, "Classification of Freenet Traffic Flow Based on Machine," *J. Commun.*, vol. 13, no. 11, pp. 654–660, 2018, doi: 10.12720/jcm.13.11.654-660.

[11] L. Wang, H. Mei, and V.S. Sheng, "Multilevel identification and classification analysis of tor on mobile and pc platforms," *IEEE Trans. Ind. Inf.*, vol. 17, no. 2, pp. 1079–1088, 2020, doi: 10.1109/TII.2020.2988870.

[12] D. Sarkar, P. Vinod, and S.Y. Yerima, "Detection of Tor traffic using deep learning," in *2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA)*, 2020, pp. 1–8, doi: 10.1109/AICCSA50499.2020.9316533.

[13] P. Choorod and G. Weir, "Tor traffic classification based on encrypted payload characteristics," in *2021 National Computing Colleges Conference (NCCC)*, 2021, pp. 1–6, doi: 10.1109/NCCC49330.2021.9428874.

[14] J. Li, C. Gu, X. Zhang, X. Chen, and W. Liu, "Attcorr: A novel deep learning model for flow correlation attacks on tor," in *2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE)*, 2021, pp. 427–430, doi: 10.1109/ICCECE51280.2021.9342534.

[15] M.B. Sarwar, M.K. Hanif, R. Talib, M. Younas, and M.U. Sarwar, "DarkDetect: darknet traffic detection and categorization using modified convolution-long short-term memory," *IEEE Access.*, vol. 9, pp. 113705–113713, 2021, doi: 10.1109/ACCESS.2021.3105000.

[16] N. Rust-Nguyen, S. Sharma, and M. Stamp, "Darknet traffic classification and adversarial attacks using machine learning," *Comput. Secur.*, vol. 127, 2023, doi: 10.1016/j.cose.2023.103098.

[17] A. Montieri, D. Ciuonzo, G. Aceto, and A. Pescape, "Anonymity services tor, i2p, jondonym: classifying in the dark (web)," *IEEE Trans. Dependable Secur. Comput.*, vol. 17, no. 3, pp. 662–675, 2018, doi: 10.1109/TDSC.2018.2804394.

[18] A. Montieri, D. Ciuonzo, G. Bovenzi, V. Persico, and A. Pescapé, "A dive into the dark web: Hierarchical traffic classification of anonymity tools," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1043–1054, 2019, doi: 10.1109/TNSE.2019.2901994.

[19] Y. Hu, F. Zou, L. Li, and P. Yi, "Traffic classification of user behaviors in tor, i2p, zeronet, freenet," in *2020 IEEE 19th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, 2020, pp. 418–424, doi: 10.1109/TrustCom50675.2020.00064.

[20] B. Yuan, J. Wang, D. Liu, W. Guo, P. Wu, and X. Bao, "Byte-level malware classification based on markov images and deep learning," *Comput. Secur.*, vol. 92, p. 101740, 2020, doi: 10.1016/j.cose.2020.101740.

[21] Z. Tang, J. Wang, B. Yuan, H. Li, J. Zhang, and H. Wang, "Markov-GAN: Markov image enhancement method for malicious encrypted traffic classification," *IET Inf. Secur.*, 2022, doi: 10.1049/ise2.12071.

[22] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi: 10.1109/5.726791.

[23] W. Wang, M. Zhu, X. Zeng, X. Ye, and Y. Sheng, "Malware traffic classification using convolutional neural network for representation learning," in *2017 International conference on information networking (ICOIN)*, 2017, pp. 712–717, doi: 10.1109/ICOIN.2017.7899588.

[24] D. Bahdanau, K. Cho, and Y. Bengio, "Neural Machine Translation by Jointly Learning to Align and Translate," in *6th International Conference on Learning Representations (ICLR 2018)*, 2014, doi: 10.48550/arXiv.1409.0473.

[25] A. Vaswani *et al.*, "Attention is all you need," in *Proc. 31st Conference on Neural Information Processing Systems (NIPS 2017)*, 2017, doi: 10.48550/arXiv.1706.03762.