

The Application of Deep Learning for Network Traffic Classification

Jingran Yang *

School of Beijing University of Posts and Telecommunications, Beijing, China

* Corresponding author email: hhhhh2o@bupt.edu.cn

Abstract. The classification, detection, and analysis of routine network traffic has been a hot topic for businesses and research institutions due to the proliferation of Internet of Things devices and the explosive development of networks. Traditional methods for categorizing network traffic primarily employ common machine learning algorithms e.g., decision trees and plain Bayes algorithms, but as deep learning technology advances, more and more traffic classifications are being successfully applied. This study examines existing deep learning-based network traffic classification techniques and focuses on the categorization of computer network traffic. Firstly, the research background of the topic is introduced, and then the traffic classification based on deep learning is mainly described, which includes traffic classification based on Stacked Autoencoder, traffic classification based on Convolutional Neural Network and traffic classification based on Recurrent Neural Networks. Following investigation, this paper comes to the conclusion that Long Short-Term Memory and Convolutional Neural Network models are the best deep learning models for traffic classification, with three-dimensional Convolutional Neural Network outperforming the others.

Keywords: Network Traffic Classification; Deep Learning; SAE; CNN; LSTM.

1. Introduction

The proliferation of IoT devices, 5G deployment, and enterprise migration to the cloud have all contributed to the exponential expansion of networks in recent years, which has resulted in a wide variety of different types of network traffic for information transfer and communication. Network management and security are severely challenged by the variety of network traffic. Thus, it is becoming more crucial than ever to recognize and categorize network traffic, including encrypted communication, detecting malicious traffic or network attacks, and enhancing the effectiveness of network operations. They are a requirement for network operation, management, and security.

Techniques for classifying traffic have changed dramatically over time. The four kinds of conventional traffic classification methods can be broadly classified as port-based, deep packet inspection (DPI), statistical-based, and behavior-based model [1]. Port-based traffic classification use transport layer protocol UDP/TCP port numbers for classification. This technique was particularly successful in the early phases of classifying network traffic [2]. However, the classification accuracy of this method has significantly fallen to 50%-70% due to the continual increase in the number of applications and the advent of port number disguising techniques [3]. Deep packet inspection learns about the application layer from the packet payload, which is then scanned and matched against the feature library to identify the traffic type. It makes up for the drawback of being influenced by port changes [4]. However, it is unable to recognize a significant number of location features to classify encrypted communications due to the limited feature base. When integrated with machine learning, statistical-based traffic classification employs the same five-tuple of network flows as the statistical features of packets as input [5], but the packets' multiple redundant pieces of information significantly reduce the classification accuracy. Behavior-based traffic classification methods are also based on traditional machine learning, which analyzes behavioral features and provides a rule-based representation of host communication behavior by observing communication patterns between hosts. The disadvantage of it is that the behavioral model library has a significant impact on the classification structure. Both statistical-based and behavior-based traffic classifications use classic machine learning [6], which lowers the need for labor and simplifies computation while increasing accuracy.

However, manual feature set building is still necessary and has a significant impact on accuracy. Therefore, finding high-quality feature sets has become the focus of such research directions.

Deep learning can be trained to choose features autonomously; thus, it doesn't need a specific expert to research and choose them. It is an extremely desirable approach for traffic classification because of this feature. Deep learning can learn extremely complicated patterns since it has a larger learning capacity than traditional Machine Learning (ML) techniques [7, 8]. Deep learning as an end-to-end approach can learn the non-linear relationship between the original input and the corresponding output by combining these two features, therefore there is no need to break the problem down into smaller sub-problems of feature selection and classification. The Stacked Self-Encoders (SAE) traffic classification method, the Convolutional Neural Network (CNN)-based traffic classification algorithm, and the RNN-based traffic classification algorithm are some of the more well-known deep learning techniques used in the field of network traffic. Besides, raw computer network traffic data is generally not available for direct input into deep learning models because they need a specific data format. Therefore, raw traffic data must be converted into a format that is compatible with deep learning models.

Based on the current progress, this paper systematically summarizes and compares deep learning-based traffic classification methods such as stacked SAEs, One-dimensional Convolutional Neural Networks (1D-CNNs), Two-dimensional Convolutional Neural Networks (2D-CNNs), Three-dimensional Convolutional Neural Networks (3D-CNNs) and Long and Short-term Memory Networks (LSTMs) in the vein of the development of deep learning in the field of network traffic, with a focus on how the transformation of raw traffic, in a way is compatible with deep learning models, can be carried out.

2. Method

2.1 Datasets

Multiple sources, such as actual traffic or publicly accessible statistics, provided the network traffic data used in the studies examined in this paper. ISCX VPN-nonVPN, USTC-TF2016, DARPA1998, ISCX2012, QUIC Dataset+Unlabeled Waikato Dataset, Ariel Dataset, KDDCup-99, UNSW-NB15, CTU- Mixed, and CIC-IDS-2017 are a few of the publicly accessible datasets.

The two most frequently used datasets in this paper's statistics are ISCX VPN-nonVPN (8 articles) and USTC-TFC2016 (3 articles). A number of authors used unpublished datasets that they captured and processed themselves to generate.

Also, in the large variety of datasets, only ISCX VPN-non-VPN is encrypted traffic, the traffic in the other datasets is unencrypted.

2.2 Models

2.2.1 SAE

The SAE model is a deep neural network model consisting of a multilayer AE model, where the output of the first layer of the self-encoder is used as input to the next layer of the self-encoder, and the last layer is a classifier.

Wang was the innovator of deep learning techniques used to end-to-end network traffic classification and made the initial proposal for the use of deep learning in network traffic classification [9]. Since then, a lot of studies have drawn research ideas from it and suggested various methodologies based on his article.

In [10], the authors trained and compared the 1D-CNN and SAE models on the ISCX VPN-nonVPN dataset. Both models exhibited high accuracy, while the CNN model slightly outperformed the SAE model in terms of F1 scores.

2.2.2 1D-CNN

Researchers easily associate traffic with 1D-CNN because its byte data is linearly organized. By analogizing network traffic (bytes, frames, packets, and flows) to letters, words, phrases, and articles, 1D-CNN considers the bytes in traffic as characters in natural language and employs text classification algorithms in NLP to achieve traffic classification.

Due to the concept of [11], Chen et al. produced a vector with a size of 784 bytes as input to a 1D-CNN [12]. To categorize the traffic, the authors used three datasets: ISCX VPNnonVPN, ISCX-IDS-2012, and ISCX - IDS -2017.

In order to divide up network traffic, Wangwei et al. suggested using 1D-CNN [13], with the data converted as in [14]. The ISCX vpn-nonvpn dataset is processed using CNNs in this method. The authors compared the proposed 1D-CNN method with the 2D-CNN method and showed that the 2D method achieved worse results than the 1D method.

2.2.3 2D-CNN

The basic method based on 2D-CNN is to convert network traffic's byte data into byte-level greyscale images, turning the problem of traffic classification into a problem of picture classification.

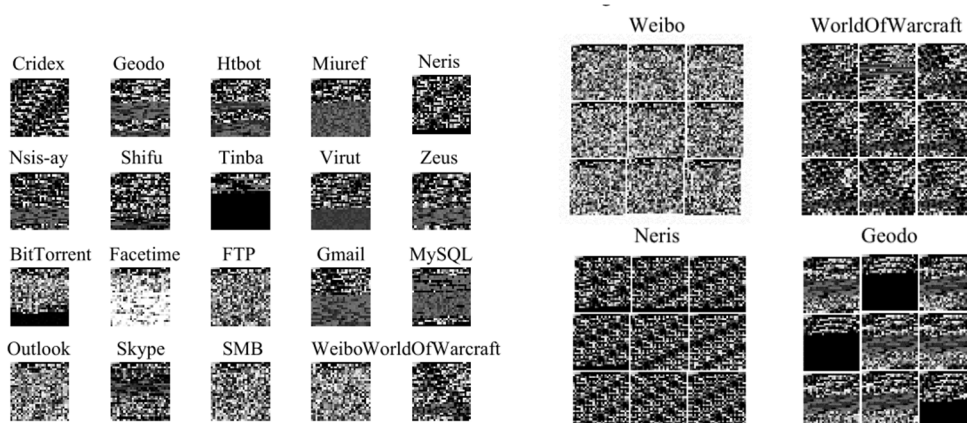


Fig 1. Grayscale image generated after 2D-CNN pre-processing of data

With the help of the transformed greyscale images, it can be observed that while the transformed graph textures from different flows are very different, those from the same flow are essentially the same (Figure 1). It serves as a foundation for additional 2D-CNN classification visualizations.

Wei et al. [14], was the first to apply the concept of [9]. This traffic classification used CNN algorithms. They also compared the data input using only L7 layers with that using all layers and obtained the result that the precision, recall, and f1 scores were higher using all layers than those using only L7 layers.

The paper by Wang et al. successfully applied the transformation concept of by training a CNN network with the TensorFlow framework [2, 15]. The CNN received a 28×28 -byte input matrix and produced an average F1 score of 98.5%.

To perform the conversion of traffic using thermal encoding, Wang et al. devised the HAST-IDS method [16]. Thermal coding is used to turn each byte in a vector into its unique vector, which is then concatenated to create a matrix. The 256×600 byte images produced the best classification results for the ISCX-IDS-2012 dataset, whereas the 256×800 byte images performed better for the DARPA 1998 dataset.

2.2.4 3D-CNN

Researchers have looked into the 3D-CNN approach to better utilize the time-series characteristics of network traffic data that one-dimensional and two-dimensional neural networks do not have. 3D-CNN converts the packet header data of the traffic into a two-dimensional image and adds a third dimension: temporal features, forming three-dimensional data as input, similar to video processing,

where the grey-scale image of the traffic data is the video. Each frame, with multiple frames arranged in chronological order, becomes a video. Data pre-processing requires the order of packets in each stream to be processed as a third dimension.

Analytical comparisons of the 1D, 2D, 3D, and MLP traffic classification methods were made by Zhang et al. [17]. The initial vectors for a 1D-CNN were 1456 bytes in size; in 2D, they were compressed into a 39×39 -byte matrix; and in 3D, they were compressed into a $22 \times 22 \times 3$ -byte RGB color cube. Network traffic information was gathered from the authors' own recorded dataset as well as the ISCX VPN-nonVPN dataset. It was demonstrated that the MLP classifier performed worse than all CNN-based classifiers. However, because the MLP-based classifiers are lighter and support higher bandwidths in comparison to the others, they are more computationally efficient than the others. With better classification accuracy and computational efficiency, the 3D-CNN-based classifier performs best overall.

2.2.5 LSTM

CNN's network as a whole lacks "memory" because the output signal is solely related to the input signal and not the order in which the input signal was received. In order to classify the whole network traffic by combining the state of the front and back frames, we need to introduce LSTM. The LSTM has a stronger capacity for learning temporal aspects than the CNN, since it incorporates forgetting gates to regulate whether the output of earlier moments contributes to the output of that moment.

Zou et al. used LSTM to implement a recurrent neural network and compared the results to 1D-CNN [18], finding an average accuracy improvement of 5%. It achieves the highest values in terms of accuracy and recall for most types of flows, and the extracted flow time series features perform well in cryptographic flow classification.

Experimental data that is used to classify traffic also produces results that are practically useful when using the LSTM, but the network structure is more complex than that of SAE and CNN. There are more parameters to learn, and the computational complexity is higher.

2.3 Evaluation Metrics

The following metrics are commonly used in the field of network traffic classification.

TP: Positive case for a correct prediction (actual positive, predicted positive)

FP: Positive case for incorrect prediction (actual negative, predicted positive)

TN: negative case for a correct prediction (actual negative, predicted negative)

FN: negative case for incorrect prediction (actual positive, predicted negative)

Precision (P) : Fraction of predicted positive cases that are indeed positive, as a proportion of all predicted positive cases

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

Recall (R): Proportion of samples in the category correctly predicted by the model, the fraction predicted to be positive and indeed positive, as a proportion of all categories that are indeed positive

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F1 score (F1): Combined precision and recall metrics, simple weighted average of precision and recall

$$F1 - score = \frac{2TP}{FP+2TP+FN} \quad (3)$$

When dealing with data sets with imbalances, studies should focus on F1 (or precision and recall) rather than accuracy.

3. Result and Discussion

The percentage of each method's dataset, precision, recall, and F1-score are provided in the Table 1 along with examples of typical articles for each method.

The Table 1 shows that the SAE approach performs the least well of all the methods. With regard to the CNN methods, the 1D method outperforms the 2D method, and the 3D method outperforms

neither of the other two evaluation metrics listed. However, when speed and bandwidth are taken into account [17], 3D performs better overall than both 2D and 1D. Although it is more sophisticated than 3 D, the LSTM approach performs comparable to the 3D-CNN method.

Table 1. Comparison of methods

Method	Dataset	Precision(%)	Recall(%)	F1-score(%)
SAE[10]	ISCX VPN-nonVPN	44	57	30
1D-CNN[12]	ISCX VPN-nonVPN	98.65	98.64	98.64
2D-CNN[15]	UNSW-NB15	98.65	98.40	98.52
3D-CNN[17]	ISCX VPN-nonVPN	96.40	96.33	96.40
LSTM[18]	ISCX VPN-nonVPN	95.2	97.4	96.1

4. Conclusion

In this work, numerous methods for classifying network traffic using deep learning were examined, such as SAE, CNN, LSTM, as well as their underlying concepts and effectiveness. The most efficient deep learning models at the moment are the LSTM and CNN models, with 3D-CNN outperforming the rest in terms of performance. While deep learning research in web traffic classification has advanced and made significant strides, it also faces a number of issues and difficulties. In light of the current situation, where it is challenging to collect labelled data, it is necessary to further explore semi-supervised and unsupervised models to give them the ability to categorize network flows with little to no labeling.

References

- [1] Biersack E, Callegari C, Matijasevic M. Data traffic monitoring and analysis. Lecture Notes in Computer Science, 2013, 5(23): 12561-12570.
- [2] Carela-Espanol V, Barlet-Ros P, Solé-Simó M, et al. K-dimensional trees for continuous traffic classification. International Workshop on Traffic Monitoring and Analysis. Springer, Berlin, Heidelberg, 2010: 141-154.
- [3] Moore A W, Papagiannaki K. Toward the accurate identification of network applications. International workshop on passive and active network measurement. Springer, Berlin, Heidelberg, 2005: 41-54.
- [4] Li G, et al. Network traffic classification method based on DPI and Machine learning (in Chinese). Journal of Guilin University of Electronic Technology, 2012, 32(2): 140-144.
- [5] Zhao S, Zhang Y, Chang P. Network traffic classification using tri-training based on statistical flow characteristics. 2017 IEEE Trustcom/BigDataSE/ICCESS. IEEE, 2017: 323-330.
- [6] Karagiannis T, Papagiannaki K, Faloutsos M. BLINC: multilevel traffic classification in the dark. Proceedings of the 2005 conference on Applications, technologies, architectures, and protocols for computer communications. 2005: 229-240.
- [7] Song, Y et al. Deep learning-based automatic segmentation of images in cardiac radiography: A promising challenge. Computer Methods and Programs in Biomedicine, 2022, 106821.
- [8] Yu, Q. Pose-guided matching based on deep learning for assessing quality of action on rehabilitation training. Biomedical Signal Processing and Control, 72, 2022, 103323.
- [9] Wang Z. The applications of deep learning on traffic identification. BlackHat USA, 2015, 24(11): 1-10.
- [10] Lotfollahi M et al. Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Computing, 2020, 24(3): 1999-2012.
- [11] Wang W, Sheng Y, Wang J, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE access, 2017, 6: 1792-1806.
- [12] He Y, Li W. Image-based encrypted traffic classification with convolution neural networks. 2020 IEEE Fifth International Conference on Data Science in Cyberspace (DSC). IEEE, 2020: 271-278.

- [13] Wang W et al. End-to-end encrypted traffic classification with one-dimensional convolution neural networks. 2017 IEEE international conference on intelligence and security informatics (ISI). IEEE, 2017: 43-48.
- [14] Wang W, Zhu M, Zeng X, et al. Malware traffic classification using convolutional neural network for representation learning. 2017 International conference on information networking (ICOIN). IEEE, 2017: 712-717.
- [15] Wang Y, An J, Huang W. Using CNN-based representation learning method for malicious traffic identification. 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS). IEEE, 2018: 400-404.
- [16] Wang W, Sheng Y, Wang J, et al. HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection. IEEE access, 2017, 6: 1792-1806.
- [17] Zhang J, Li F, Ye F, et al. Autonomous unknown-application filtering and labeling for dl-based traffic classifier update. IEEE INFOCOM 2020-IEEE Conference on Computer Communications. IEEE, 2020: 397-405.
- [18] Zou Z, Ge J, Zheng H, et al. Encrypted traffic classification with a convolutional long short-term memory neural network. 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS). IEEE, 2018: 329-334.