Multi-Input Shape CNN-based Application Traffic Classification

Ui-Jun Baek, Jee-Tae Park, Jeong-Woo Choi
Seung-Woo Nam, Jae-Won Park, Gyeong-Min Yu, Myung-Sup Kim
Computer Information and Science
Korea University, Korea
{pb1069, pjj5846, choigoya97, nam131119, 2018270614, rudals2710, tmskim}@korea.ac.kr

Abstract

Due to the rapid advancement of the internet and online applications, network traffic classification has become a crucial topic in the field of network management. Recently, many traffic classification methods based on AI, particularly deep learning, have been proposed. However, these methods often focus solely on altering the hierarchical structure of deep learning models and do not consider the shape of the incoming traffic data. We propose a classification method that utilizes various input shapes that can be derived from fixed-length packet bytes. Our proposed approach achieves superior classification performance compared to previous research that only utilized two-dimensional square-like input shapes and one-dimensional linear input formats from publicly available datasets.[1]

Topics: traffic classification, traffic identification, convolutional neural network, DL-based classification.

Introduction and background

Traffic classification is a key technology in network traffic monitoring and analysis, involving grouping similar or related traffic and classifying it into predefined categories. This technique serves various crucial purposes. Firstly, it aims to troubleshoot network issues, such as locating faulty network devices, hardware/software misconfigurations, and points of packet loss within the network. Secondly, it ensures the overall acceptability of applications by managing quality of service (QoS), including bandwidth resources and cloud service usage. Thirdly, it plays a pivotal role in network security, enabling the distinction between normal and malicious traffic for security measurement and intrusion detection. Historically, widely used traffic classification methods include port-based classification and payload-based classification. However, these methods face limitations due to the application of dynamic ports and payload encryption. Although machine learning-based classification methods have been actively proposed, they fall short in analyzing the diverse traffic patterns in complex network environments. Recently, AI, especially deep learning-based classification methods, have gained attention. High-performance models based on Convolutional Neural Networks (CNN), which have shown remarkable performance in computer vision, have been applied in research.

Paper No.	Category	DL Method	Features	Shapes
[1] 2018	IDS	CNN	Header, Payload	Sqaure
[2] 2019	APP TC	CNN	Payload	Sqaure
[3] 2020	APP TC	CNN	Header, Payload	Sqaure
[4] 2021	APP TC	CNN, LSTM	Header	Linear
[5] 2022	APP TC	CNN	Header	Linear

Table 1: CNN-based TC studies

CNN is widely utilized for traffic classification in deep learning. It involves truncating packets to a fixed length before entering the learning model. These truncated packets are then transformed into 2D square-shaped or 1D linear-shaped vectors. Previous research has predominantly focused on reshaping raw packet bytes into 2D matrices resembling squares. From these matrices, spatial features are extracted for network traffic classification using CNN. These studies are consolidated in Table 1. Our investigation reveals the historical predominance of the two-dimensional square-like input shape. However, a recent shift has occurred towards the consistent adoption of 1-D linear input shapes. This shift is presumably informed by empirical studies indicating superior performance of linear input shape-based traffic classification models compared to their square-like counterparts [6]. Notably, the distinct characteristic of raw packet bytes is their lack of clear adjacency among data points, unlike conventional images where pixel relationships are evident. Unfortunately, most studies have not

This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-004)

adequately addressed the consideration of input shapes derived from raw packet bytes. In light of this gap, we propose an innovative traffic classification approach that capitalizes on various input shapes derived from fixed-length packet bytes. Our proposed method begins with the assumption that there may be alternative models that better represent packets than linear or rectangular shapes. Furthermore, the combination of various perspectives looking at the same input can be expected to have a similar effect to using ensemble techniques.Our method's efficacy is assessed using the publicly accessible ISCX VPN-non VPN 2016 dataset. Impressively, our approach outperforms prior research in classification performance. This holds true for both 2D square input shapes and 1D linear input shapes.

Goals, proposed method, novelty

We propose a method to extract distinctive features of packets through various input shapes that can be derived from fixed-length packet bytes, which are not limited to the aforementioned square-like or linear shapes. Assuming a fixed packet length of 784 bytes, there are a total of 15 possible shapes that can be extracted from it, as detailed in Table 2. The number of shapes is equal to the number of divisors of the packet length, and since the last shapes represent the same structure as the first shapes, it is excluded. Therefore, the final count of shapes is 14, which is one less than the number of divisors.

Shape.	Kernel Size	Shape	Kernel Size
(784, 1)	(6, 1)	(16, 49)	(2, 2)
(392, 2)	(6, 1)	(14, 56)	(2, 2)
(196 ,4)	(4, 1)	(8, 98)	(1, 4)
(98, 8)	(4, 1)	(4, 196)	(1, 4)
(56, 14)	(2, 2)	(2, 392)	(1, 6)
(49, 16)	(2, 2)	(1, 784)	(1, 6)
(28, 28)	(2, 2)		

Table 2: Shapes that can be derived from a packet with a length of 784 bytes.

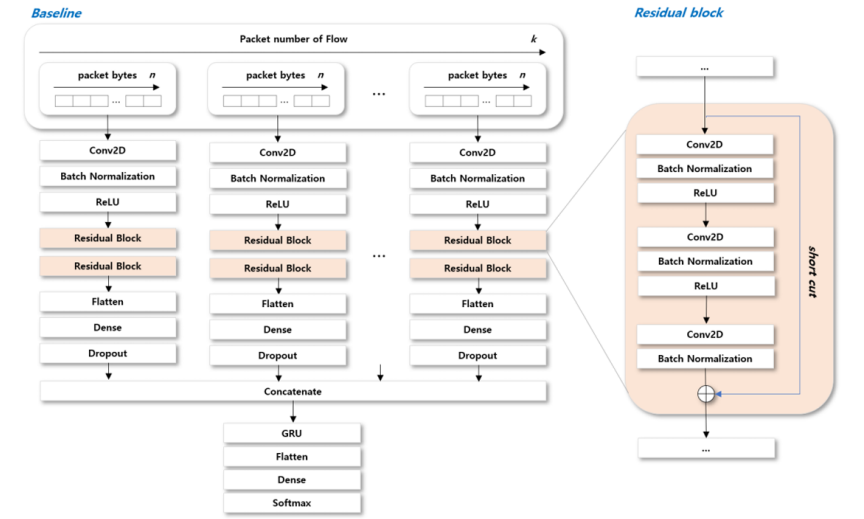

Fig.1: Baseline

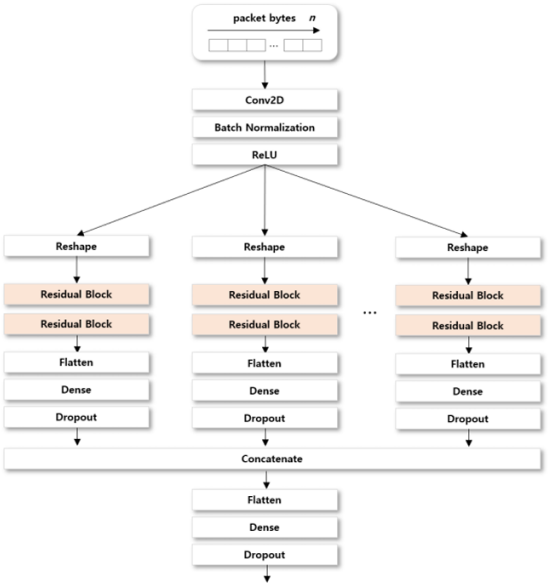

Fig.2: MISCNN scheme

The baseline model employed in this research comprises a blend of CNN and Gated Recurrent Unit (GRU), as depicted in Figure 1. Instead of aggregating them into a single input, utilizing each packet as an individual input ensures optimized performance. Consequently, the baseline processes multiple packets and subsequently consolidates them into a unified output at the GRU layer. To extract features from each packet-based input, a residual block is strategically positioned [7]. Each residual block is composed of three Convolution Layers, followed by the insertion of a Batch Normalization layer after every convolution operation, facilitating adjustment of output mean and variance. MISCNN is underpinned by a straightforward yet potent concept. Prior to entering the initial residual block, the input is partitioned into multiple shapes. The count of shapes for each input corresponds to the number of divisors in the packet vector size, denoted as 'n'. For instance, the frequently employed 7841 input configuration in earlier studies encompasses 15 distinct shapes (784*1, 392*2, 196*4, … , 1*784). At this juncture, the shape 1*784 is excluded due to its structural similarity to the shape 7*841 (though shapes like 2*392 and 392*2 differ noticeably). A overview of the MISCNN architecture is illustrated in Figure 2. These reshaping techniques find limited application in the realm of image recognition and computer vision, as they compromise the 2D spatial information of neighboring (vertical, horizontal) pixels inherent in typical images. On the contrary, since raw packets lack 2D spatial information, reshaping methods can be explored, enabling the training model to perceive the same input from diverse perspectives.

Experiments, results, analysis of the results

For evaluation, we employ the publicly accessible ISCX VPN-nonVPN 2016 dataset. This dataset, denoted as "ISCX VPN-nonVPN 2016," comprises raw pcap files featuring diverse applications, which we utilize to evaluate classification performance [6]. It encompasses human-generated traffic of varied types, coupled with information on the correlated applications. This data is garnered from both regular sessions and sessions encapsulated via VPN. This setup permits us to assign a three-view label (specifically, encapsulation, traffic type, and application) to any segmentation of raw network traffic, effectively forming a generic TC object. Each of these three-view labels corresponds to a distinct TC task that must be addressed. Each task is employed within our proposed model and subjected to comparative experiments. Packets extracted from the raw pcap files are amalgamated into flows based on the 5-tuple attributes (source IP, destination IP, source port, destination port, protocol) and reconstructed into bidirectional flows, taking directionality into account. Notably, nearly 60 percent of the aggregated two-way flows consist of only one UDP packet, leading to disruptions in the proper learning process. These instances have been filtered out, resulting in 27.8k bidirectional flows. Each flow sample contains the initial-k packets from the entire collection. When the number of packets in a flow falls below the pre-defined value k, an empty object filled with zeros is appended. The first-n bytes are extracted from the packets within each flow. If the packet size is less than the predetermined value n, the remaining vacant space within the packet object is zero-padded. The original shape of the packet data is k*1, which is then transformed to a p*q shape through the Reshape layer. Details of the hardware and software environment employed during the training process are outlined in Table 3. During model compilation, the Learning Rate was set to 25-e5. We employed Categorical Cross-entropy as the loss function and utilized the Adam Optimizer.

	List	Specification
Hardwar e	CPU	Intel(R) Xeon(R) CPU E5-2630 v4 @ 2.20GHz
	GPU	NVIDIA GP102 [TITAN Xp]
	RAM	120G
Software	Nvidia Driver	440.33.01
	CuDNN	cuDNN/7.6 for cuda 10.1
	Cuda	cuda/10.1
	Python	cuda/10.1
	Keras	keras 2.4.0

Table 3: Shapes that can be derived from a packet with a length of 784 bytes.

In order to assess the efficacy of the proposed approach, we conducted a comparative analysis between MISCNN, four prior studies, and baseline models across three distinct TC tasks. The resulting comparison outcomes are detailed in Table 4. Notably, MISCNN exhibited superior performance compared to its predecessors. Across the three tasks, it displayed notable improvements – a 4.83% enhancement in the Encapsulation task, a surpassing of 10% in the Category task, and a commendable 7% advancement in the Application task, widely regarded as the most challenging. Remarkably, the F-measure displayed substantial growth when contrasted with previous studies. Another salient point is that the Baseline's performance also exceeded that of earlier TC models. The Baseline, enriched with the skip-connection technique that has demonstrated excellence in the domain of image recognition and computer vision, exhibited superior performance in two metrics compared to previous TC models. This substantiates the utility of the skip-connection technique within the TC domain. Lastly, it is worth noting that the 1D CNN-based TC model more effectively extracts traffic characteristics compared to its 2D CNN-based counterpart, aligning with findings from prior investigations.

	Encapsulation (%)		Traffic Type (%)		Application (%)	
	Acc	F1	Acc	F1	Acc	F1
1D-CNN [6]	87.4	83.5	73.1	71.1	72.7	61.3
1D-CNN [8]	82.3	76.2	56.0	54.7	56.5	40.8
2D-CNN [8]	87.4	83.5	71.8	69.7	71.4	59.2
Distiller [9]	93.7	91.9	80.7	78.7	77.6	66.4
Baseline(1D)	97.4	95.2	91.2	91.3	81.0	78.7
Baseline(2D)	94.1	94.0	87.2	85.2	78.0	77.6
MISCNN	98.5	98.5	93.5	93.2	85	85.2
MISCNN (gain)	4.8	6.6	12	14.4	7.3	18.8

Table 4: comparison of MISCNN with previous studies

Conclusion

In this paper, we introduce a deep learning scheme aimed at comprehensively analyzing packets from multiple perspectives using diverse shapes derived from a single input. Our focus is on packet data, which lacks two-dimensional spatial information unlike typical images. We propose that reshaping packets into various forms offers a means to effectively observe and extract valuable features from raw packets. To validate our approach, we evaluate it using the publicly available ISCX VPN-nonVPN 2016 dataset, and we establish a baseline incorporating the skip-connection method. Subsequently, we implement the proposed method, termed MISCNN, atop the designed baseline, and compare its traffic classification (TC) performance against previous studies.

Our proposal contributes in three key ways. Firstly, in terms of overall comparison, MISCNN demonstrates a remarkable 14% accuracy improvement and an 18% enhancement in f-measure over state-of-the-art TC methods. This enhancement underscores that adopting diverse perspectives for packet observation mitigates overfitting in TC models and yields substantial performance gains. Secondly, performance comparisons based on experimental parameter variations reveal that employing multiple shapes, except for the Encapsulation task, contributes to accurate classification. Notably, considering that the Encapsulation task holds less significance and versatility in real-world networks compared to other tasks, integrating a broader array of shapes into the TC model can prove advantageous in common scenarios. Lastly, we perform comparative experiments within a novel dataset that addresses data imbalances, closely resembling real-world network conditions. This attests to the practical viability of the proposed MISCNN in actual network environments.

We also outline three areas for future exploration based on this study. The first pertains to reducing complexity. MISCNN, while accurate, can be excessively resource-intensive, limiting its practicality. Secondly, we suggest exploring the application of state-of-the-art techniques. Lastly, we plan to conduct research on XAI (Explainable AI) to discover improved justifications compared to existing methods.

References

[1] W. Wang et al., “HAST-IDS: Learning hierarchical spatial-temporal features using deep neural networks to improve intrusion detection,” IEEE access, vol. 6, pp. 1792–1806, 2017.
[2] H.-K. Lim, J.-B. Kim, J.-S. Heo, K. Kim, Y.-G. Hong, and Y.-H. Han, “Packet-based network traffic classification using deep learning,” in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC), 2019, pp. 046–051.
[3] J. Li and Z. Pan, “Network traffic classification based on deep learning,” KSII Transactions on Internet and Information Systems (TIIS), vol. 14, no. 11, pp. 4246–4267, 2020.
[4] X. Hu, C. Gu, and F. Wei, “CLD-Net: a network combining CNN and LSTM for internet encrypted traffic classification,” Security and Communication Networks, vol. 2021, 2021.
[5] S. Izadi, M. Ahmadi, and R. Nikbazm, “Network traffic classification using convolutional neural network and ant-lion optimization,” Computers and Electrical Engineering, vol. 101, p. 108024, 2022.
[6] W. Wang, M. Zhu, J. Wang, X. Zeng, and Z. Yang, “End-to-end encrypted traffic classification with one-dimensional convolution neural networks,” in 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), Jul. 2017, pp. 43–48. doi: 10.1109/ISI.2017.8004872.
[7] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.
[8] S. Rezaei and X. Liu, “Multitask Learning for Network Traffic Classification,” in 2020 29th International Conference on Computer Communications and Networks (ICCCN), Aug. 2020, pp. 1–9. doi: 10.1109/ICCCN49398.2020.9209652.
[9] G. Aceto, D. Ciuonzo, A. Montieri, and A. Pescapé, “DISTILLER: Encrypted traffic classification via multimodal multitask deep learning,” Journal of Network and Computer Applications, vol. 183, p. 102985, 2021.