Opening the Deep Pandora Box: Explainable Traffic Classification

Cedric Beliard, Alessandro Finamore, Dario Rossi
Huawei Technologies Co. Ltd – first.last@huawei.com

***Abstract*—Fostered by the tremendous success in the image recognition field, recently there has been a strong push for the adoption of Convolutional Neural Networks (CNN) in networks, especially at the edge, assisted by low-power hardware equipment (known as "tensor processing units") for the acceleration of CNN-related computations. The availability of such hardware has re-ignited the interest for traffic classification approaches that are based on Deep Learning. However, unlike tree-based approaches that are easy to interpret, CNNs are in essence represented by a large number of weights, whose interpretation is particularly obscure for the human operators. Since human operators will need to deal, troubleshoot, and maintain these automatically learned models, that will replace the more easily human-readable heuristic rules of DPI classification engine, there is a clear need to open the "deep pandora box", and make it easily accessible for network domain experts. In this demonstration, we shed light in the inference process of a commercial-grade classification engine dealing with hundreds of classes, enriching the classification workflow with tools to enable better understanding of the inner mechanics of both the traffic and the models.**

I. Introduction

Traffic classification is definitively not a new subject, with seminal work [1] dating back over a decade ago. However, whereas the first wave of traffic classification research, which is well covered in [2], essentially focused on extracting features for classifying a relatively small (no more than 15 applications) and static application population, the network landscape has changed in the meanwhile. Particularly, the tremendous push toward encryption in the Post-Snowden era make traffic classification a prominent and unavoidable tool for correct network operation, which prompted a new breed of research, surveyed in [3]. This re-ignited interest for the approaches in the industry, that is now looking with greater interest at actively deploying statistical classification approaches introduced in the last decade [4], and that so far mostly remained an academic exercise, as recently surveyed in [5]. One reason of the limited success of academic-style systems is that, wheres statistical classification techniques have gained success in academia, commercial DPI tools are able to handle *hundreds* to *thousands* of application signatures, whereas academic classification has been limited to a few *tens* of classes, at best. *In this demo, we leverage a unique dataset comprising 15 million labeled flows across 350 classes, that is thus commercial-grade (for which we cannot make the demo interface available on the public Internet).*

Another reason is of non-technical nature, and it intrinsically due to lack of interpretability of machine-learning models (eg. SVM and CNN), especially when compared to rule-based DPI that human operators are accustomed to. Whereas tree-based machine learning models (e.g. C45) are relatively easy to interpret, satisfactory accuracy can be achieved only by ensemble methods (e.g., Random Forest, XGBoost, CatBoost). Hence, there is a tradeoff between interpretability and complexity. Beside their excellent discriminant power, CNNs are appealing especially when coupled with dedicated "tensor processing units" (e.g., Huawei Ascend[6], Google's Coral[7]) which can accelerate AI workflows [8]. This is something particularly sensitive, especially at the network edge. For instance, these dedicated engines operates on few tens of Watts, reducing significantly the operational cost of CNN with respect to, e.g., classic GPU deployments (that consumes several hundreds of Watts). Although less pressing than for other fields (medical, juridical, etc.)[9] the lack of interpretability of CNN can be seen as possibly blocking their commercial success in traditional telcos business, where the use of CNN is not as widespread in the image field yet. *For these reasons, in this demo we focus on opening the "deep pandora box" and illustrating the steps of our CNN-based commercial-grade classification engine – which assists telco domain experts interaction in the transition from DPI to CNN.*

The above are overtly recognized as major blocking points for deployment of behavioral classification techniques as recently pointed out in [5], which we both address in this work.

II. CNN for traffic classification

Traffic classification is far from being a green field, with seminal work dating back to over a decade ago work [1], [10]. In this work we leverage the well know "early traffic classification" techniques [4] dating back to 2007, that are based on the size and direction of a few packets of a flow, and that our own previous work demonstrated to perform *several million classification per seconds* on COTS systems in 2012 [11]. A novel ingredient that was previously missing on the table is the current emergence of hardware assisted tensor processing units [7], [6], which bring CNN based models in an operationally efficient point.

A. Demo at a glance

As represented in top left of Fig.1, the input of the CNN is a time-series of packet size and direction[4], which is well fit for 1D-CNN models. The design of (one of) our CNN model(s) is sketched in the bottom of Fig.1: our CNN design stacks a series of convolutional (3x3 filters) with ReLU activation and max-pooling layers. The first layers of CNN have a

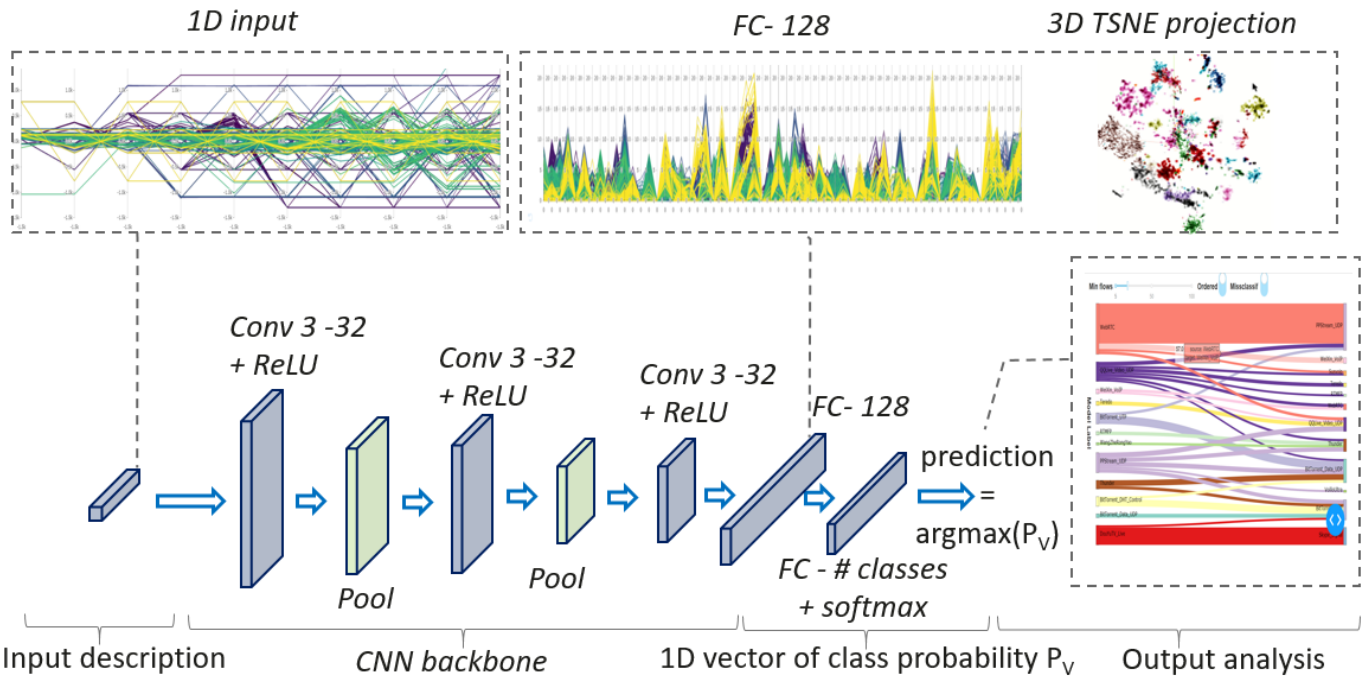

Figure 1: Synoptic of the demo, showcasing a few of the different 3Djs illustrations that are available in the dashboard. From left to right, top to bottom, the boxed illustrations extracted from the dashboard represents respectively: (i) 1D input representation in the original space, (ii) F128 feature extractor with 3D TSNE projection of one of the intermediate layers of the CNN backbone, (iii) arc diagram representation of the confusion matrix contrasting ground-truth input vs CNN output labels. All visualization are interactive in the demo.

readily physical interpretation, as they construct classic signal processing properties related to the time-series (e.g., batch of stacked filters). The subsequent layers effectively perform non-linear transformation of these information, projecting the original 10-valued input into a higher 128-dimensional space, which is shown in the top right of Fig.1, along with a three-dimensional TSNE projection: the picture shows well that applications of different classes are well separable (already in a 3D space), so that the softmax output of the latest full convolutional layer has very high accuracy. Additionally, the dashboard allows to interactively explore the classification output to (i) compare it with the ground truth label, which is instrumental to understand at a fine-grain level the accuracy performance, as the confusion matrix (depicted as an arc diagram in the bottom right of Fig.1) and (ii) project backward into the CNN layers what is the most salient "features" that ultimately led to the classification results, to help human operators interacting with the products familiarize with understanding of the classification process.

B. Demo workflow

The demo dashboard, developed in `Voila` (a dashboarding framework based on Jupyter notebook), features interactive `plotly` charts rendered with `D3js`, allow users to understand the different stages of the classification process. For instance, applications can be selectively toggled in input/CNN layers/output shown in Fig.1. Similarly, the arc diagram shown in the picture allows to show the full elements of the confusion matrix (possibly sorting them by larger amount, so that most important applications appear on top) or to only show the elements that are *not* on the diagonal (to let the misclassifications standout, and backtrack the origin of the misclassification in the layers, up to possibly similarity in the input space) – which helps operators putting order back in the opened pandora box.

References

[1] M. Roughan et al., "Class-of-service mapping for qos: a statistical signature-based approach to ip traffic classification," in *ACM IMC*, 2004.

[2] T. T. Nguyen and G. J. Armitage, "A survey of techniques for internet traffic classification using machine learning." *IEEE Communications Surveys and Tutorials*, vol. 10, no. 1-4, pp. 56–76, 2008.

[3] P. Velan et al. , "A survey of methods for encrypted traffic classification and analysis," *International Journal of Network Management*, vol. 25, no. 5, pp. 355–374, 2015.

[4] M. Crotti et al, "Traffic classification through simple statistical fingerprinting," *ACM SIGCOMM CCR*, vol. 37, no. 1, pp. 5–16, 2007.

[5] F. Pacheco et al., "Towards the deployment of machine learning solutions in network traffic classification: A systematic survey," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2018.

[6] https://www.hisilicon.com/en/Media-Center/News/Key-Information-About-the-Huawei-Ascend310.

[7] https://coral.ai/.

[8] J. L. Hennessy and D. A. Patterson, "A new golden age for computer architecture," *Commun. ACM*, p. 48–60, Jan. 2019.

[9] https://www.aiforhumanity.fr/en/.

[10] D. Bonfiglio et al, "Revealing skype traffic: When randomness plays with you," *ACM SIGCOMM*.

[11] P.M. Santiago del Rio et al., "Wire-speed statistical classification of network traffic on commodity hardware," in *ACM IMC*, 2012.