Reducing Complexity of Deep Learning for Time Series Classification Using New Hand-Crafted Convolution Filters

Ali Ismail-Fawaz[1], Maxime Devanne[1], Stefano Berretti[2], Jonathan Weber[1], and Germain Forestier[1,3]

[1] IRIMAS, Université de Haute-Alsace, Mulhouse France
{ali-el-hadi.ismail-fawaz, maxime.devanne, jonathan.weber, germain.forestier}@uha.fr
[2] MICC, University of Florence, Folrence Italy
stefano.berretti@unifi.it
[3] DSAI, Monash University, Melbourne Australia

Abstract. Deep learning for Time Series Classification (TSC) has become one relevant subject in the literature for this task. It is used for wide applications in multiple domains ranging from medical data, action recognition and robotics. In the last decade, Convolutional Neural Networks (CNNs) have shown to be the best base architecture to use when dealing with deep learning for TSC ever since the release of the UCR archive, the largest repository for TSC datasets. The UCR archive includes a variety of 128 datasets of univariate time series data, where the task is to correctly classify the samples to their corresponding annotation. Deep learning models face two main challenges. The first one is represented by overfitting and the consequent incapacity of generalizing to new unseen samples. With CNN based architectures, this is commonly due to the fact that the learned filters tend to detect specific patterns in the training set instead of generic ones. The second challenge is complexity wise, which limits its usability in real world scenarios such as embedded systems. In this work, we propose to address these two challenges with one solution: hand-crafting some generic non-learned convolutional filters to detect generic patterns. These hand-crafted filters can replace the usability of the first layer in the CNN model, resulting in a significant reduction in the number of parameters. The proposed architecture is evaluated on 128 datasets of the UCR archive and the results reveal a significant improvement in performance compared to other approaches as well as the reduction in terms of complexity.

Keywords: Time Series Classification, Deep Learning, Time Series, Hand-Crafted Filters

1 Introduction

Time Series is a type of sequential data that is almost in every domain these days. Many tasks can be applied over this kind of data, including averaging [1], data augmentation [2], clustering [3], regression [4], classification [5, 6]. Time Series Classification (TSC) is extensively investigated in the literature such as the task of surface recognition by robots [7]. For instance, TSC can be used for evaluating surgical performance, [8] human motion action recognition, [9, 10] surface type detection from robots movements [7] etc. The availability of the UCR archive [11] made it possible to test multiple machine learning tools to be benchmarked over the 128 datasets available. Ismail Fawaz et al. [12] presented a comprehensive review with exhaustive experiments to compare deep learning models for the task of TSC. This review concluded that Convolutional Neural Networks (CNNs) based architectures are more suitable for the task of TSC on the UCR archive. One of the winning CNN based architecture is the Fully Convolutional Network (FCN) [13], made of three convolutional layers in cascade followed by Batch Normalization and ReLU activation function. Moreover, the authors in [14] proposed InceptionTime, adapted from Inception-v3 for image classification. InceptionTime is currently the state-of-the-art for the task of TSC on the UCR archive. Some other work addressed deep learning for TSC using self-supervised approach [15] and knowledge distillation [16]. Even though deep learners present a success for the task of TSC, nevertheless, they do suffer from two main problems. First, deep learners tend to overfit on the training examples, which leads, in the case of CNNs, to learning filters to detect specific unique patterns in the training examples. Second, deep learners have large complexity, this limits their deployment into real world machines such as embedded systems.

In this work, we address these two problems and propose new hand-crafted convolution filters for time series data. We define three different hand-crafted filters to detect generic patterns that are independent from the data: increasing trend, decreasing trend and peaks. These filters, given that they are generic and non-learned, can help overcome the overfitting problem when replacing the first convolutional layer in the network. By combining the proposed hand-crafted convolution filters with the FCN [13] architecture, we proposed the Custom Only FCN (CO-FCN), which replaces the first learned layer in FCN by the hand-crafted filters.

We show that the CO-FCN reduces the number of parameters of FCN by almost 47%. Evaluation on the 128 datasets of the UCR archive show that CO-FCN can outperform FCN with a statistical significance in difference of performance. This is the higlight of our contributions in this work:

- Proposing new hand-crafted convolution filters;
- Adapting an existing CNN network for Time Series Classification to use the hand-crafted filters to reduce its number of parameters and increase its performance;
- Extensive experiments on 128 datasets of the UCR archive to evaluate the performance of the hand-crafted filters and the proposed architecture.

2 Proposed Method

2.1 Definitions

First, we list some definitions to facilitate the understanding of the rest of this work:

Univariate Time Series: Let $\mathbf{x}$ be a univariate time series of length L, a sequence of data points equally separated in time.

Univariate Time Series Dataset: A dataset $\mathcal{D} = \{(\mathbf{x}_0, y_0), ..., (\mathbf{x}_N, y_N)\}$ is a set of N pairs of univariate time series of length L and a label y associated to it.

One Dimensional Convolution: An operation using a filter $\mathbf{w}$ of length k on a time series $\mathbf{x}$ to obtain $\mathbf{s} = \mathbf{x} * \mathbf{w}$ as follows:

$$\forall t \in [0; L-1] \qquad \mathbf{s}[t] = \sum_{i=0}^{k-1} \mathbf{x}[t+i].\mathbf{w}[i] \tag{1}$$

A detailed version of convolution on time series can be seen in Figure 1.

Activation of a Filter: When the convolution operation results in a positive response, the filter is considered as activated.

Increasing Trend: A sub-sequence of a time series $\mathbf{x}$, where the values are strictly increasing in time.

Decreasing Trend: A sub-sequence of a time series $\mathbf{x}$, where the values are strictly decreasing in time.

Stationary Trend: A sub-sequence of a time series $\mathbf{x}$, where the values vary of a small difference ϵ.

Peak: A sub-sequence of a time series $\mathbf{x}$, where the values changed with a large variation increasingly and then decreasingly.

In what follows, we detail the hand-crafted convolution filters.

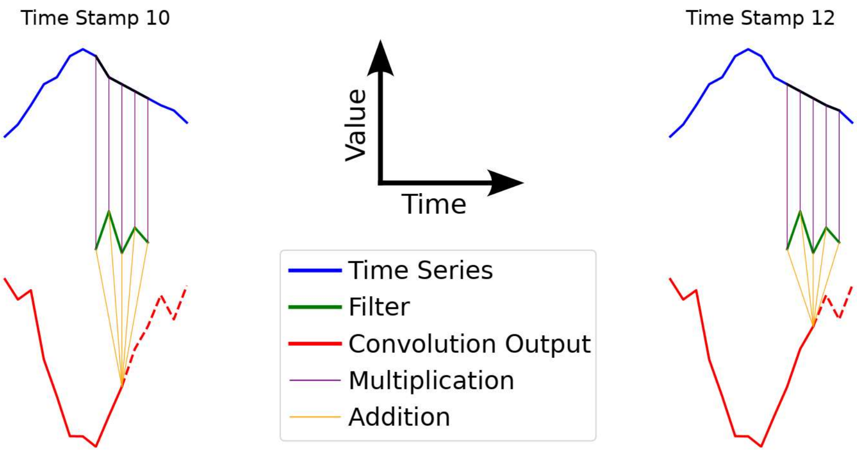

Fig. 1: One dimensional convolution filter being convolved with a one dimensional time series.

2.2 Hand-Crafted Convolutional Filters

A summary of the three hand-crafted filters proposed in this work is given in Figure 2. The information gathered from the gradient of the time series can give rise to the increasing and decreasing trends. For this reason, both increasing and decreasing trends detection filters are simply an oscillation between -1 and 1. Motivated by the Sobel filters proposed on image contour detection [17], we adapt to a one-dimensional case the detection of peaks by mimicking the inverse of the second order derivative of the Gaussian.

2.3 FCN Adaptation: Custom Only FCN (CO-FCN)

By replacing the first convolutional layer in the original FCN architecture and replacing it by the hand-crafted filters, we obtain the Custom Only FCN (CO-FCN) presented in Figure 3. The FCN model presents $264,704$ trainable parameters, where as CO-FCN has only $122,496$ parameters to train.

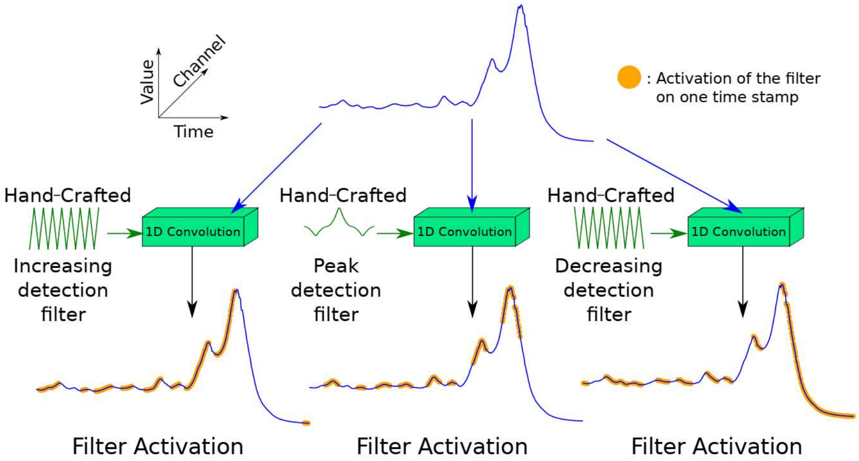

Fig. 2: The three hand-crafted convolutional filters applied on the Meat dataset of the UCR archive.

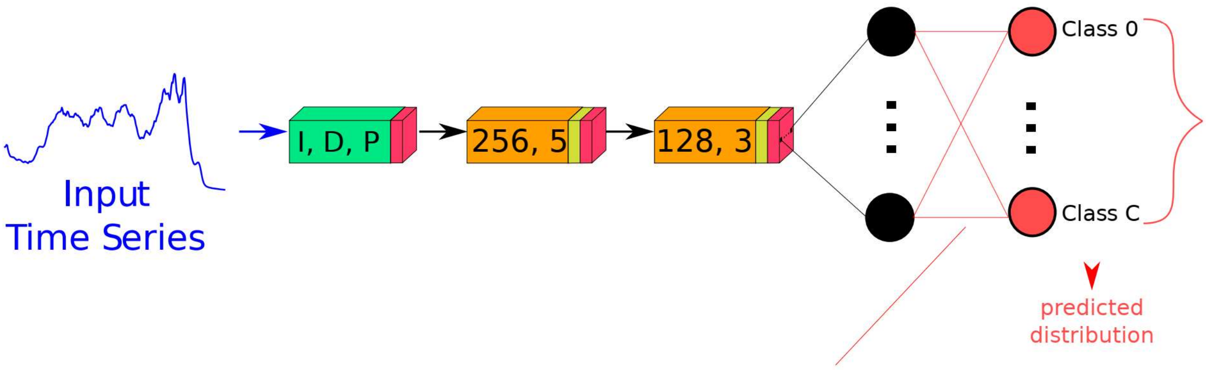

Fig. 3: The Custom Only FCN (CO-FCN) architecture. The hand-crafted filters are used in the first layer (in green) followed by a ReLU activation (in magenta). The second and third layers are made of convolution blocks (in orange: (n_filters, kernel_size)) followed by batch normalization (in oily) and ReLU activation. The last layers are composed of a 1D global average pooling (in black) and a linear classification layer (in red).

3 Results and Discussion

3.1 Experimental Setup

To have a fair comparison between FCN and CO-FCN, we used the same number of epochs, the same optimizer and its initial parameters and the same batch size. We also used a learning rate decay that monitors the training loss function. Five different initialization are done and the performance presented in this work is the average over all of them. The best model on the training loss during training is saved and used for the evaluation phase.

We use the set of 128 datasets of the UCR archive [11], where each dataset is split into train and test sets. The evaluation is done over the test set using the classification accuracy. All of the datasets are z-normalized before training in order to have a zero mean and unit variance.

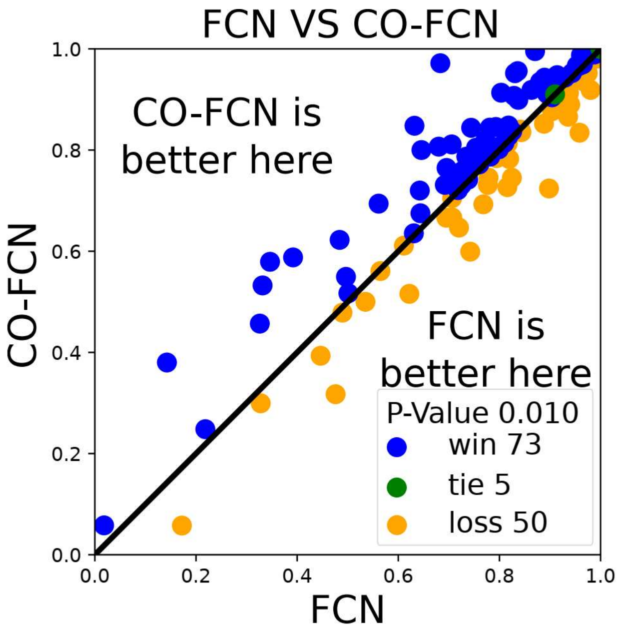

Fig. 4: One-vs-One scatter plot between FCN and CO-FCN. Each point represents one dataset, on the x-axis the accuracy of the FCN is presented and the one of CO-FCN on the y-axis.

3.2 CO-FCN vs FCN

To compare the performance of those two models, we evaluated their performance using the accuracy metric on the test set of all the datasets of the UCR archive. In Figure 4, the One-vs-One scatter plot between FCN and CO-FCN is presented. We counted the number of times each model wins and the number of ties, the difference in performance helps to produce the p-value (see legend in Figure 4) produced by the Wilcoxon Signed Rank test [18]. This p-value represents a % of statistical significance for the difference in performance between two comparates. This %, the p-value, if less than a given threshold (usually set to 5%) means that the difference in performance is statistically significant. If that last condition is not true then no conclusion can be made on the significance of difference in performance. This comparison concludes that CO-FCN outperforms FCN with a difference in performance that is statistical significant.

3.3 Comparing with State-of-the-Art

To compare the performance of CO-FCN to other deep learning approaches, we present the Multi Comparison Matrix (MCM) [19] in Figure 5. The MCM presents a pairwise comparison and a multi-classifier comparison at the same time. Each cell presents the Win/Tie/Loss count between two classifiers and the difference in average accuracy and the p-value using the Wilcoxon test. The MCM in Figure 5 also orders the competitors following their average performance on the accuracy metric over all datasets used. CO-FCN, even though coming third on the average performance, it can be seen that it beats significantly FCN and is not statistically significant than ResNet, which is 4.1 larger than CO-FCN in terms of number of parameters.

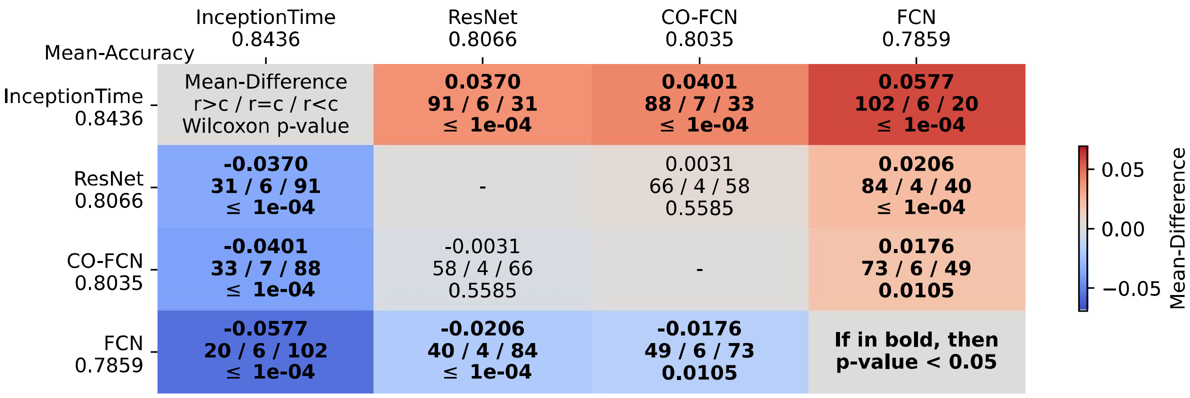

Fig. 5: A Multi Comparison Matrix (MCM) benchmarking the state of the art models including the proposed CO-FCN architecture.

4 Conclusions

We presented in this work new hand-crafted convolution filters that are non-learned and generic to any time series data. These filters are able to replace the first convolutional layer of the Fully Convolutional Network proposed for TSC. This replacement constructs the CO-FCN with almost half the number of parameters of FCN. The proposed model is not only less complex than the FCN but also outperforms it on the majority of the UCR archive datasets. This new approach would help embed deep learners for TSC tasks into small embedded systems such as robotics systems.

5 Acknowledgements

This work was supported by the ANR DELEGATION project (grant ANR-21-CE23-0014) of the French Agence Nationale de la Recherche. The authors would like to acknowledge the High Performance Computing Center of the University of Strasbourg for supporting this work by providing scientific sup- port and access to computing resources. Part of the computing resources were funded by the Equipex Equip@Meso project (Programme Investissements d'Avenir) and the CPER Alsacalcul/Big Data. The authors would also like to thank the creators and providers of the UCR Archive.

References

1. Ismail-Fawaz, A., Ismail Fawaz, H., Petitjean, F., Devanne, M., Weber, J., Berretti, S., Webb, G.I., Forestier, G.: Shapedba: Generating effective time series prototypes using shapedtw barycenter averaging. In: ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data. (2023)
2. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Data augmentation using synthetic data for time series classification with deep residual networks. In: ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data. (2018)
3. Holder, C., Middlehurst, M., Bagnall, A.: A review and evaluation of elastic distance functions for time series clustering. Knowledge and Information Systems (2023) 1–45
4. Guijo-Rubio, D., Middlehurst, M., Arcencio, G., Silva, D.F., Bagnall, A.: Unsupervised feature based algorithms for time series extrinsic regression. arXiv preprint arXiv:2305.01429 (2023)

5. Middlehurst, M., Schäfer, P., Bagnall, A.: Bake off redux: a review and experimental evaluation of recent time series classification algorithms. arXiv preprint arXiv:2304.13029 (2023)
6. Ismail-Fawaz, A., Maxime, D., Stefano, B., Jonathan, W., Germain, F.: Lite: Light inception with boosting techniques for time series classification. In: International Conference on Data Science and Advanced Analytics (DSAA). (2023)
7. Vail, D., Veloso, M.: Learning from accelerometer data on a legged robot. IFAC Proceedings Volumes **37**(8) (2004) 822–827
8. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Evaluating surgical skills from kinematic data using convolutional neural networks. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part IV 11, Springer (2018) 214–221
9. Devanne, M., Wannous, H., Berretti, S., Pala, P., Daoudi, M., Del Bimbo, A.: 3-d human action recognition by shape analysis of motion trajectories on riemannian manifold. IEEE transactions on cybernetics **45**(7) (2014) 1340–1352
10. Devanne, M., Wannous, H., Berretti, S., Pala, P., Daoudi, M., Del Bimbo, A.: Space-time pose representation for 3d human action recognition. In: New Trends in Image Analysis and Processing–ICIAP 2013: ICIAP 2013 International Workshops, Naples, Italy, September 9-13, 2013. Proceedings 17, Springer (2013) 456–464
11. Dau, H.A., Bagnall, A., Kamgar, K., Yeh, C.C.M., Zhu, Y., Gharghabi, S., Ratanamahatana, C.A., Keogh, E.: The ucr time series archive. IEEE/CAA Journal of Automatica Sinica **6**(6) (2019) 1293–1305
12. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data mining and knowledge discovery **33**(4) (2019) 917–963
13. Wang, Z., Yan, W., Oates, T.: Time series classification from scratch with deep neural networks: A strong baseline. In: 2017 International joint conference on neural networks (IJCNN), IEEE (2017) 1578–1585
14. Ismail Fawaz, H., Lucas, B., Forestier, G., Pelletier, C., Schmidt, D.F., Weber, J., Webb, G.I., Idoumghar, L., Muller, P.A., Petitjean, F.: Inceptiontime: Finding alexnet for time series classification. Data Mining and Knowledge Discovery **34**(6) (2020) 1936–1962
15. Ismail-Fawaz, A., Devanne, M., Weber, J., Forestier, G.: Enhancing time series classification with self-supervised learning. In: 15th International Conference on Agents and Artificial Intelligence: ICAART 2023, INSTICC (2023)
16. Ay, E., Devanne, M., Weber, J., Forestier, G.: A study of knowledge distillation in fully convolutional network for time series classification. In: 2022 International Joint Conference on Neural Networks (IJCNN), IEEE (2022) 1–8
17. Bogdan, V., Bonchiş, C., Orhei, C.: Custom extended sobel filters. arXiv preprint arXiv:1910.00138 (2019)
18. Wilcoxon, F.: Individual comparisons by ranking methods. In: Breakthroughs in statistics. Springer (1992) 196–202
19. Ismail-Fawaz, A., Dempster, A., Tan, C.W., Herrmann, M., Miller, L., Schmidt, D.F., Berretti, S., Weber, J., Devanne, M., Forestier, G., et al.: An approach to multiple comparison benchmark evaluations that is stable under manipulation of the comparate set. arXiv preprint arXiv:2305.11921 (2023)