Deep Leaning For Time Series

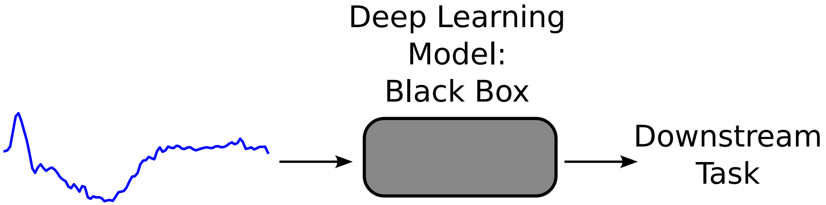
An Introduction to Machine Learning from Time Series

Ali Ismail-Fawaz & Germain Forestier

MSD, IRIMAS, Université de Haute-Alsace, Mulhouse France

European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases 2024

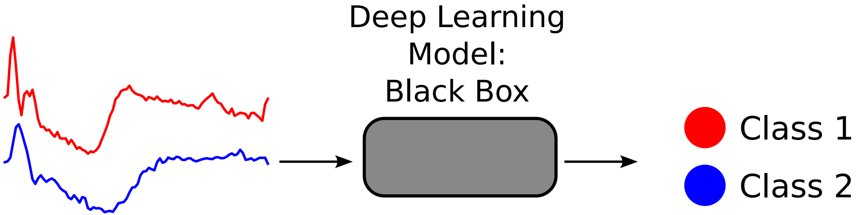
Deep Learning for Time Series Data



- Black Box: Mathematical function $f(\omega)$
- Goal: Find the optimal set of parameters ω to solve the downstream task
- $f(\omega)$ can be matrix transformations, convolution operations, recurrent operations etc.
- ω are found using a gradient based optimization algorithm

Deep Learning for Time Series Data

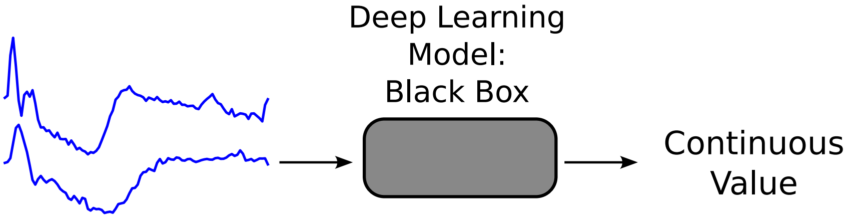
Classification:



- Black Box: Mathematical function $f(\omega)$
- Goal: Find the optimal set of parameters ω to solve the downstream task
- $f(\omega)$ can be matrix transformations, convolution operations, recurrent operations etc.
- ω are found using a gradient based optimization algorithm

Deep Learning for Time Series Data

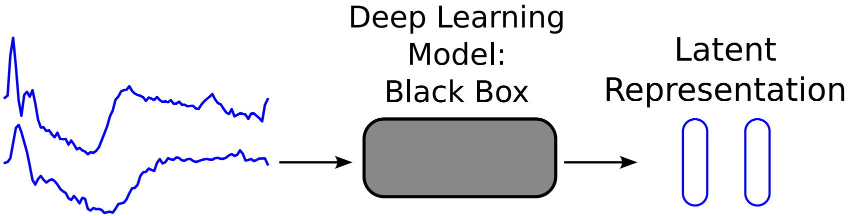
Regression:



- Black Box: Mathematical function $f(\omega)$
- Goal: Find the optimal set of parameters ω to solve the downstream task
- $f(\omega)$ can be matrix transformations, convolution operations, recurrent operations etc.
- ω are found using a gradient based optimization algorithm

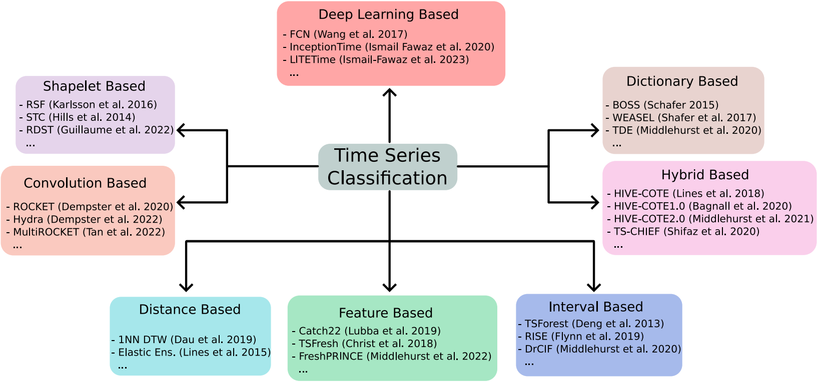
Deep Learning for Time Series Data

Representation Learning:



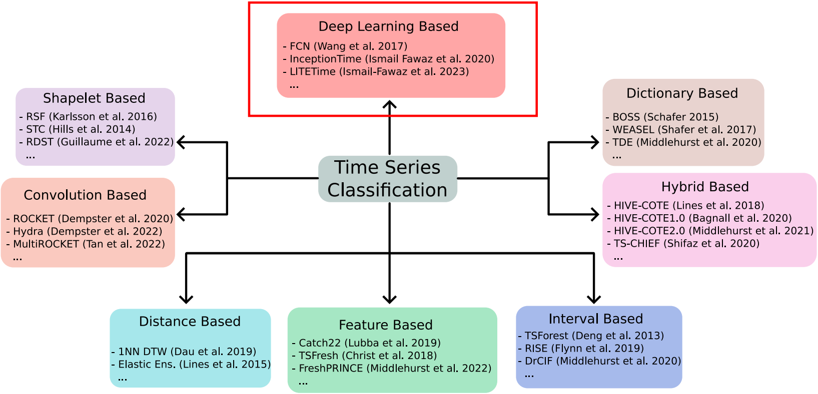
- Black Box: Mathematical function $f(\omega)$
- Goal: Find the optimal set of parameters ω to solve the downstream task
- $f(\omega)$ can be matrix transformations, convolution operations, recurrent operations etc.
- ω are found using a gradient based optimization algorithm

Time Series Classification: Taxonomy of Methods



Middlehurst, M., Schäfer, P., & Bagnall, A. (2024). Bake off redux: a review and experimental evaluation of recent time series classification algorithms. Data Mining and Knowledge Discovery, 1-74.

Time Series Classification: Taxonomy of Methods



Middlehurst, M., Schäfer, P., & Bagnall, A. (2024). Bake off redux: a review and experimental evaluation of recent time series classification algorithms. Data Mining and Knowledge Discovery, 1-74.

Why Deep Learning ?

- Around 2017, reviewers began to question the potential performance of deep learning for TSC while assessing papers on non-deep learning TSC methods.
- Deep learning has achieved great success in other data types, such as computer vision and natural language processing (NLP), so why not with time series?

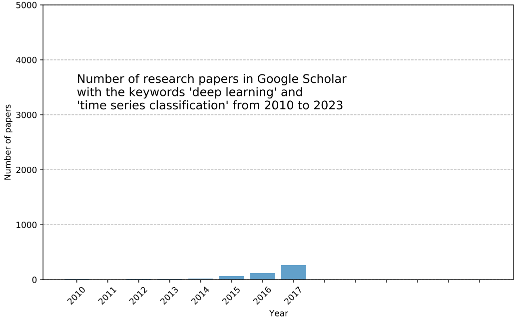


Figure: The number of research papers mentioning "deep learning" and "time series classification" increased rapidly in the last years:

Why Deep Learning ?

- Around 2017, reviewers began to question the potential performance of deep learning for TSC while assessing papers on non-deep learning TSC methods.
- Deep learning has achieved great success in other data types, such as computer vision and natural language processing (NLP), so why not with time series?

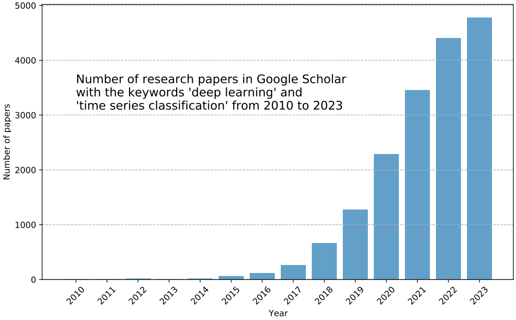
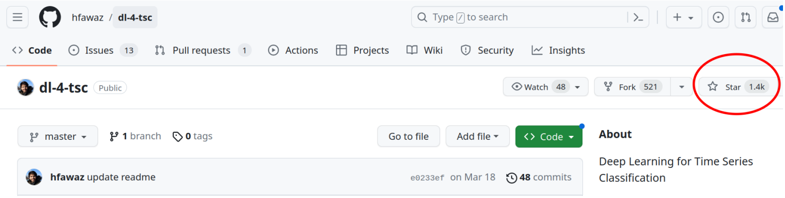


Figure: The number of research papers mentioning "deep learning" and "time series classification" increased rapidly in the last years.

DL4TSC - 2019

In **2019**, we presented a study of **Deep Learning for Time Series Classification** (cited more than 3.2K times (GoogleScholar)) [1].

- We selected models with enough details (or available code) to reproduce the model's architecture and benchmarked them on the UCR archive [2]
- We published the code on Github for reproducibility and got very positive feedback (>1.5K stars)

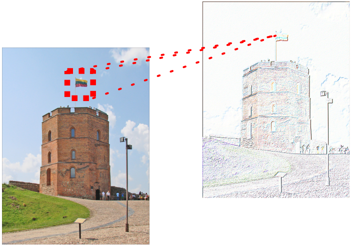


[1] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. (2019). Deep learning for time series classification: a review. Data mining and knowledge discovery, 33(4), 917-963.

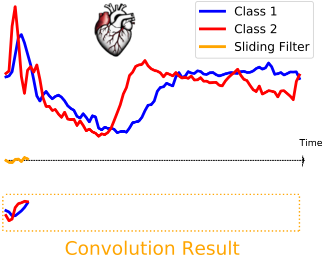
[2] Dau, H. A., Bagnall, A., Kamgar, K., Yeh, C. C. M., Zhu, Y., Gharghabi, S. & Keogh, E. (2019). The UCR time series archive. IEEE/CAA Journal of Automatica Sinica

DL4TSC - Some Architectures

Convolutions on Images vs Time Series



The result of a applying an edge detection convolution on an image

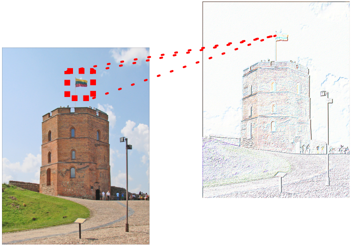


The result of applying a learned discriminative convolution on the ECG200 dataset

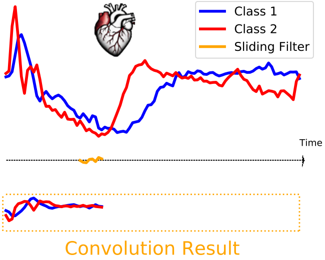
Image source: Gediminas Castle Tower of Vilnius, Jean-Pierre Dalbéra, flickr

DL4TSC - Some Architectures

Convolutions on Images vs Time Series



The result of a applying an edge detection convolution on an image

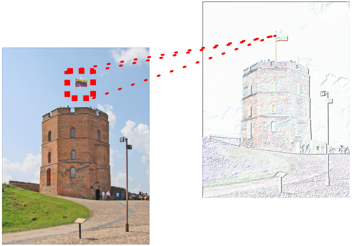


The result of applying a learned discriminative convolution on the ECG200 dataset

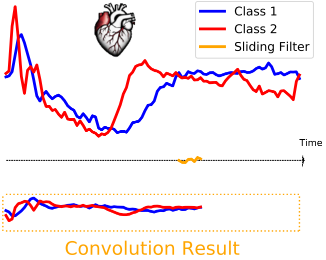
Image source: Gediminas Castle Tower of Vilnius, Jean-Pierre Dalbéra, flickr

DL4TSC - Some Architectures

Convolutions on Images vs Time Series



The result of a applying an edge detection convolution on an image

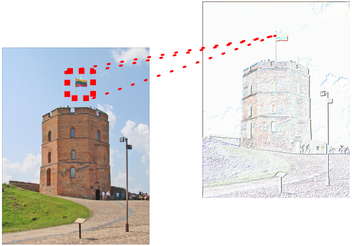


The result of applying a learned discriminative convolution on the ECG200 dataset

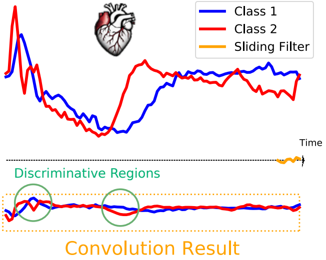
Image source: Gediminas Castle Tower of Vilnius, Jean-Pierre Dalbéra, flickr

DL4TSC - Some Architectures

Convolutions on Images vs Time Series



The result of a applying an edge detection convolution on an image

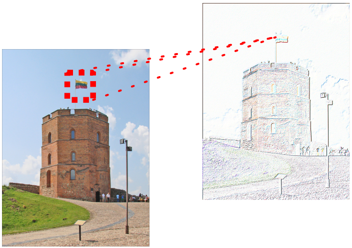


The result of applying a learned discriminative convolution on the ECG200 dataset

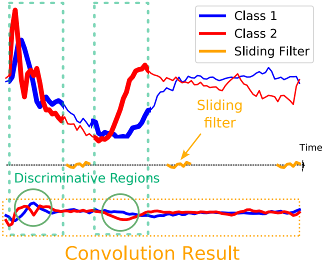
Image source: Gediminas Castle Tower of Vilnius, Jean-Pierre Dalbéra, flickr

DL4TSC - Some Architectures

Convolutions on Images vs Time Series



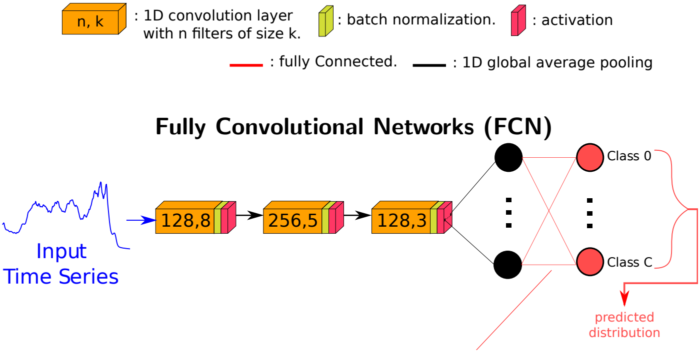
The result of a applying an edge detection convolution on an image



The result of applying a learned discriminative convolution on the ECG200 dataset

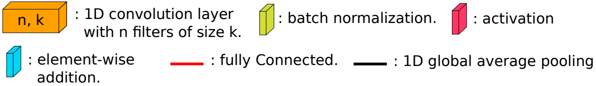
Image source: Gediminas Castle Tower of Vilnius, Jean-Pierre Dalbéra, flickr

DL4TSC - Some Architectures

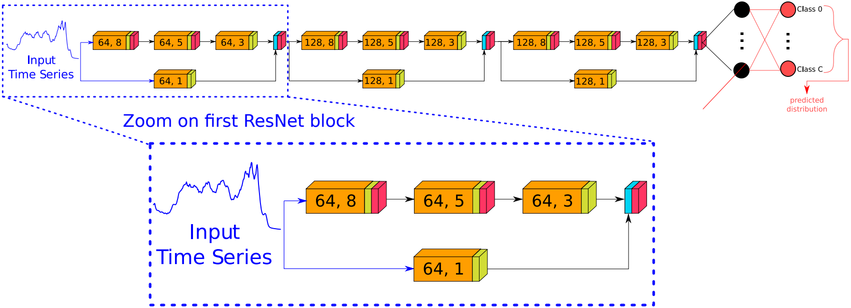


Wang, Z., Yan, W., & Oates, T. (2017, May). Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks (IJCNN) (pp. 1578-1585). IEEE.

DL4TSC - Some Architectures

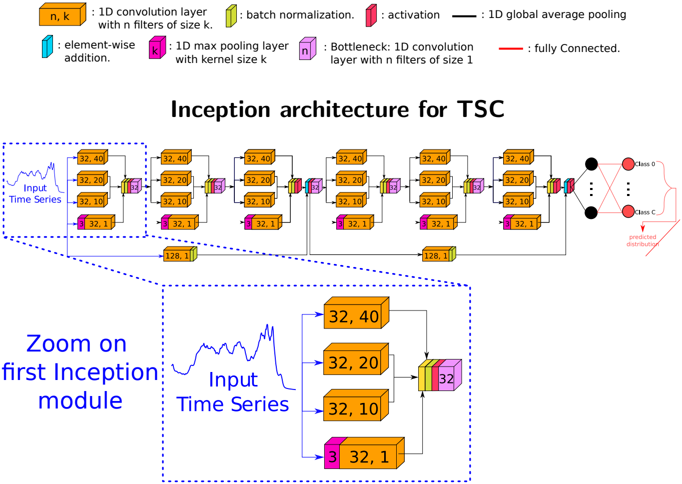


Residual Network (ResNet):



Wang, Z., Yan, W., & Oates, T. (2017, May). Time series classification from scratch with deep neural networks: A strong baseline. In 2017 International joint conference on neural networks (IJCNN) (pp. 1578-1585). IEEE.

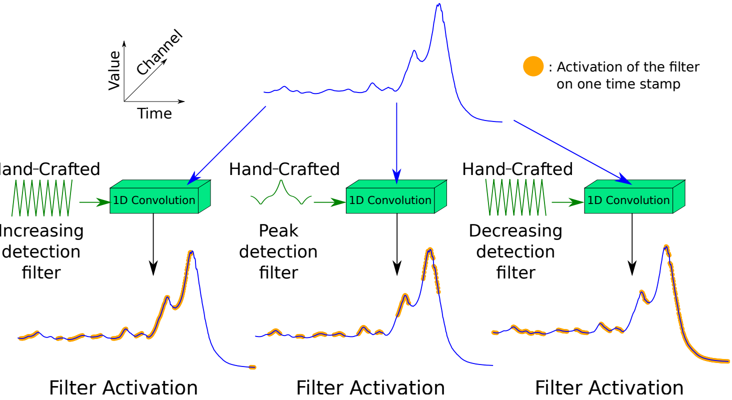
InceptionTime: Ensemble of Inception Models



Ismail Fawaz, Hassan, et al. "Inceptiontime: Finding alexnet for time series classification." Data Mining and Knowledge Discovery 34.6 (2020): 1936-1962.

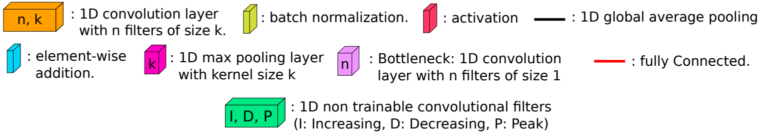
Hand-Crafted Filters

Bridging the Gap Between Random and Fully Learned Filters

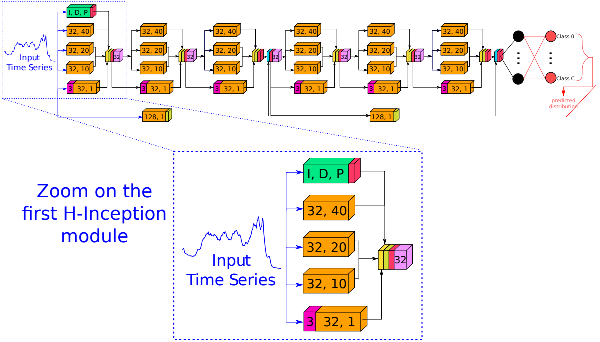


Ismail-Fawaz, A., Devanne, M., Weber, J., & Forestier, G. (2022). Deep learning for time series classification using new hand-crafted convolution filters. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 972-981)

H-InceptionTime: Ensemble of Hybrid Inception Models

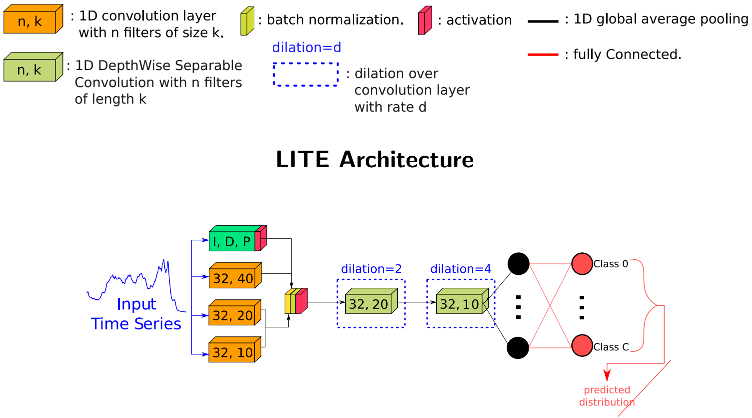


Hybrid Inception (H-Inception) Architecture



Ismail-Fawaz, A., Devanne, M., Weber, J., & Forestier, G. (2022). Deep learning for time series classification using new hand-crafted convolution filters. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 972-981)

LITETime: Ensemble of LITE Models

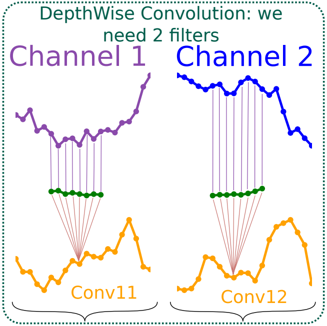


Ismail-Fawaz, A. et al. (2023). LITE: Light Inception with boosTing tEchniques for Time Series Classification. IEEE International Conference on Data Science and Advanced Analytics (DSAA)

Convolutions In LITETime

Standard vs Depthwise Separable Convolutions

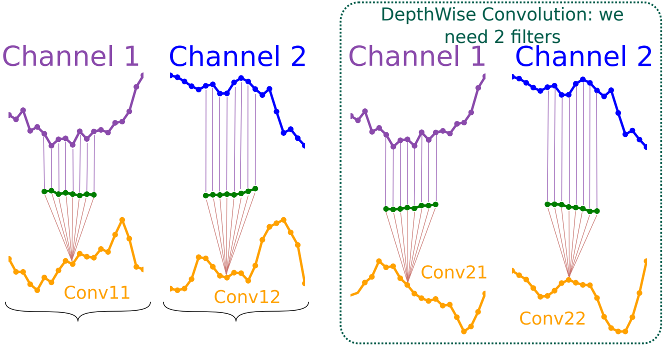
Standard Convolution = DepthWise Convolution + Summation



Convolutions In LITETime

Standard vs Depthwise Separable Convolutions

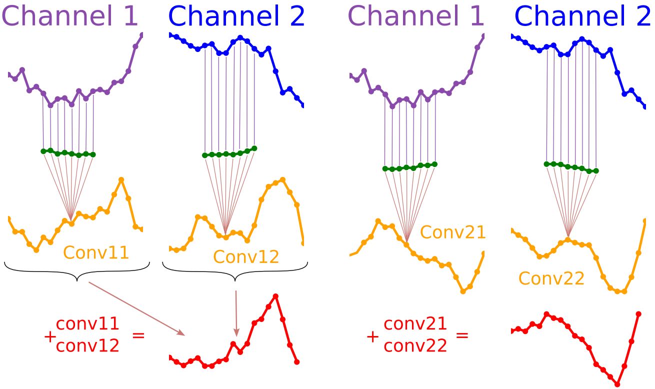
Standard Convolution = DepthWise Convolution + Summation



Convolutions In LITETime

Standard vs Depthwise Separable Convolutions

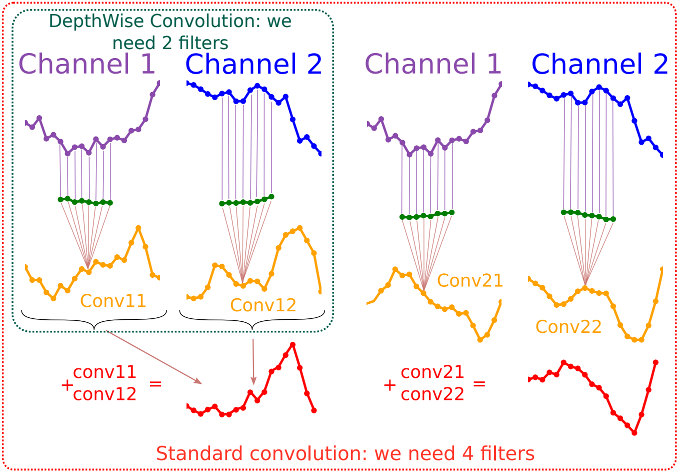
$$\text{Standard Convolution} = \text{DepthWise Convolution} + \text{Summation}$$



Convolutions In LITETime

Standard vs Depthwise Separable Convolutions

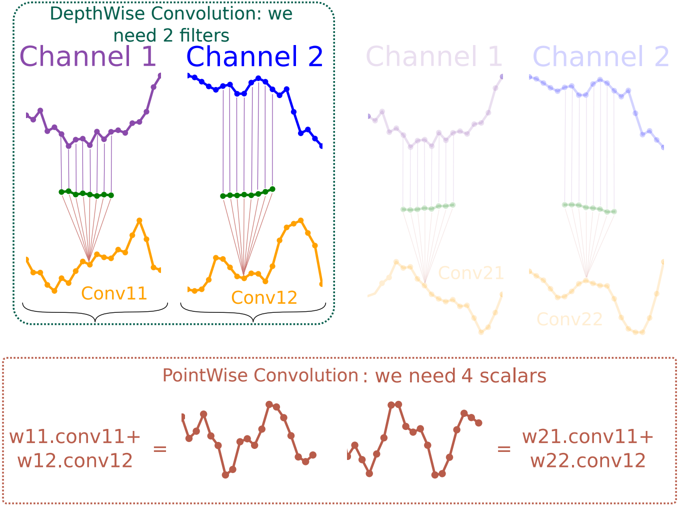
Standard Convolution = DepthWise Convolution + Summation



Convolutions In LITETime

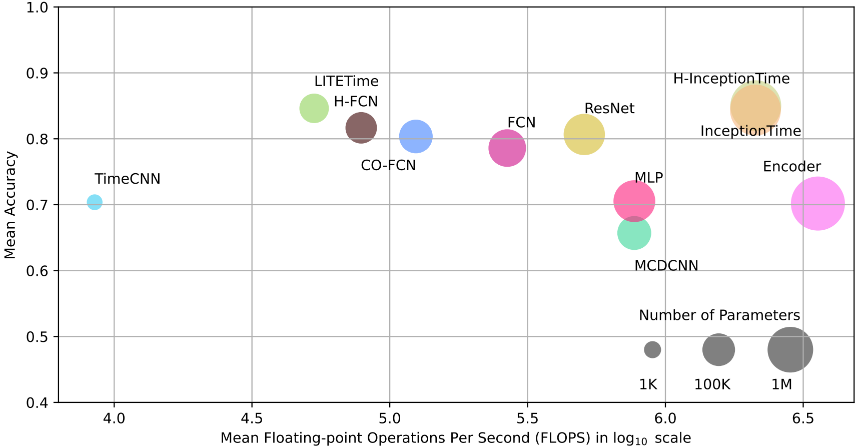
Standard vs Depthwise Separable Convolutions

DepthWise Separable Convolution = DepthWise + PointWise Convolution



Comparing all the architectures for univariate datasets

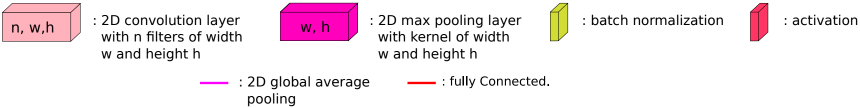
We created a dynamic website (updated regularly) to compare all these architectures in terms of performance and complexity:



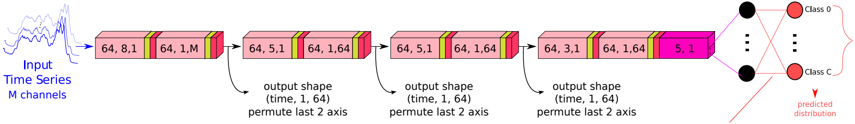
Try it out on : `https://msd-irimas.github.io/pages/dl4tsc/`

Deep Learning for Multivariate Time Series

CNN based architecture



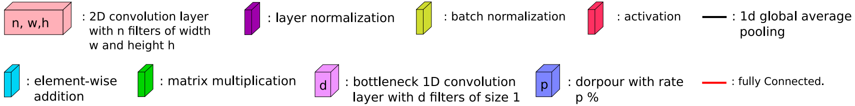
Disjoint-CNN Architecture



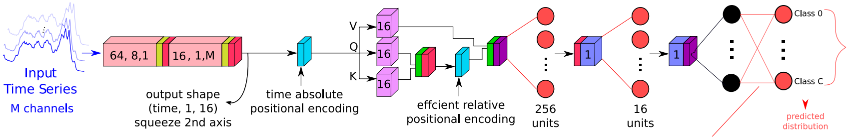
Foumani, Seyed Navid Mohammadi et al. "Disjoint-cnn for multivariate time series classification." International Conference on Data Mining Workshops (ICDMW) 2021.

Deep Learning for Multivariate Time Series

Architecture using Transfomer:



ConvTran Architecture

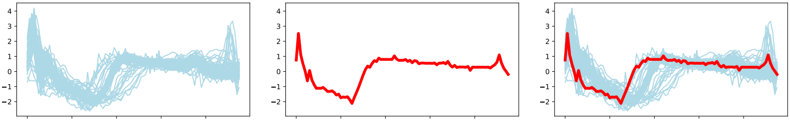


Foumani, Navid Mohammadi, et al. "Improving position encoding of transformers for multivariate time series classification." Data Mining and Knowledge Discovery 38.1 (2024): 22-48.

Regularization Techniques: Data Augmentation

How to create synthetic time series ?

- We averaged a set of time series and took the average as a new synthetic object
- We used weighted averages to generate multiple synthetic objects



Set of time series — Average time series — Augmented dataset

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition, 44(3), 678-693.

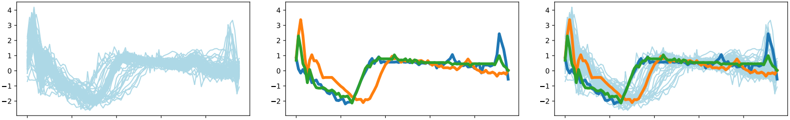
Forestier, G., et al. "Generating synthetic time series to augment sparse datasets." 2017 IEEE International Conference on Data Mining (ICDM). IEEE, 2017.

Ismail-Fawaz, A. et al. " ShapeDBA: Generating Effective Time Series Prototypes using ShapeDTW Barycenter Averaging." ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, 2023

Regularization Techniques: Data Augmentation

How to create synthetic time series ?

- We averaged a set of time series and took the average as a new synthetic object
- We used weighted averages to generate multiple synthetic objects



Set of time series | Weighted average time series | Augmented dataset

Petitjean, F., Ketterlin, A., & Gançarski, P. (2011). A global averaging method for dynamic time warping, with applications to clustering. Pattern Recognition, 44(3), 678-693.

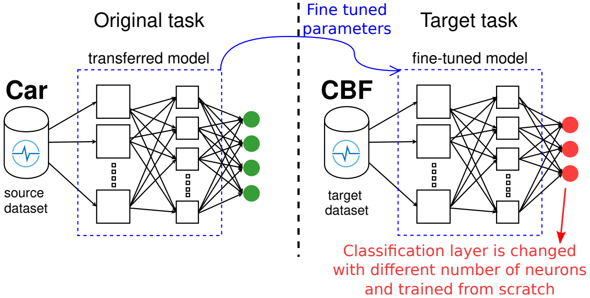
Forestier, G., et al. "Generating synthetic time series to augment sparse datasets." 2017 IEEE International Conference on Data Mining (ICDM). IEEE, 2017.

Ismail-Fawaz, A. et al. " ShapeDBA: Generating Effective Time Series Prototypes using ShapeDTW Barycenter Averaging." ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, 2023

Regularization Techniques: Transfer Learning

Transfer Learning:

1. Train a base network on a source dataset
2. Transfer the learned features (the network's weights) to a second network and adapt the last layer (class-dependent)
3. Re-train or fine-tune the transferred network on a target dataset

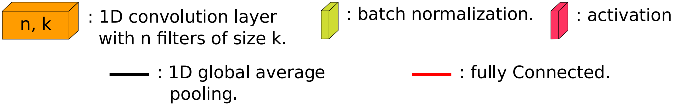


Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., & Muller, P. A. Transfer learning for time series classification. IEEE International Conference on Big Data 2018.

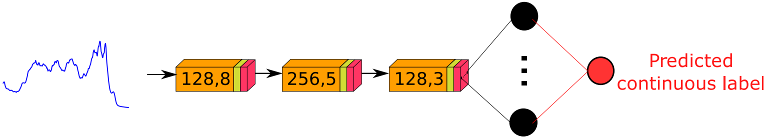
Ismail-Fawaz, A. et al. " Finding foundation models for time series classification with a pretext task." PAKDD International Workshop on Temporal Analytics 2024.

Deep Learning for Time Series Extrinsic Regression

All deep learners are simply a "black box" adaptable to any downstream task.



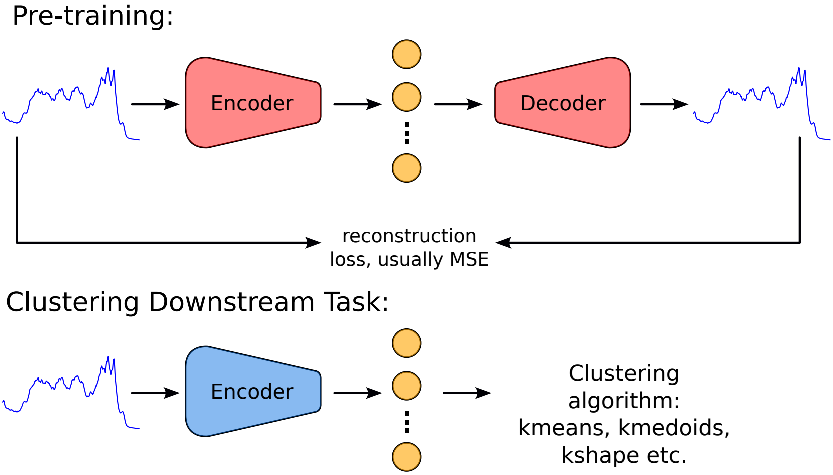
The FCN Architecture Adapted for Extrinsic Regression



Ismail-Fawaz, A. et al. "Weighted Average of Human Motion Sequences for Improving Rehabilitation Assessment" AALTD ECML/PKDD 2024.

Deep Learning for Time Series Clustering

Auto-Encoder based method:

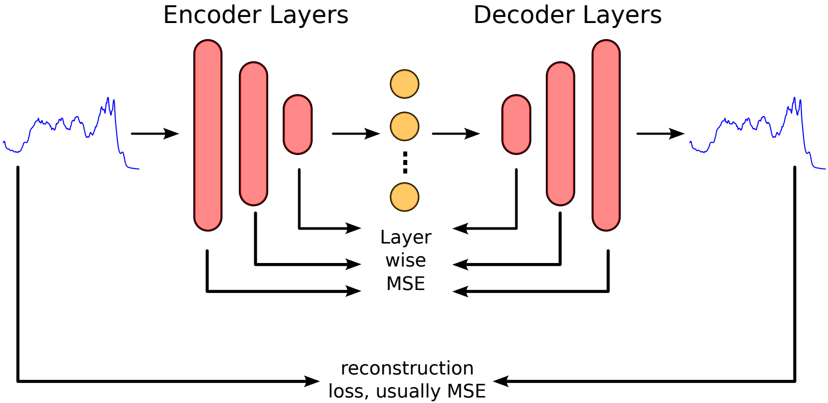


Lafabregue, Baptiste, et al. "End-to-end deep representation learning for time series clustering: a comparative study." Data Mining and Knowledge Discovery 2022.

Deep Learning for Time Series Clustering

Auto-Encoder based method with layer wise reconstruction (only for symmetrical Auto-Encoders):

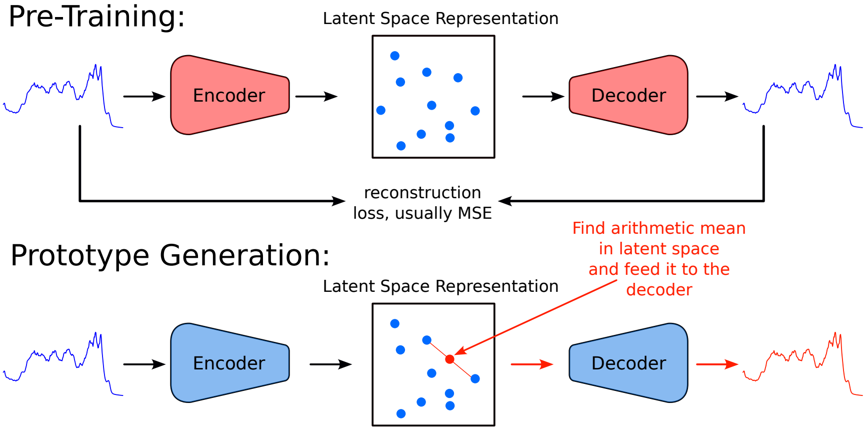
Pre-training:



Lafabregue, Baptiste, et al. "End-to-end deep representation learning for time series clustering: a comparative study." Data Mining and Knowledge Discovery 2022.

Deep Learning for Time Series Prototyping Generation

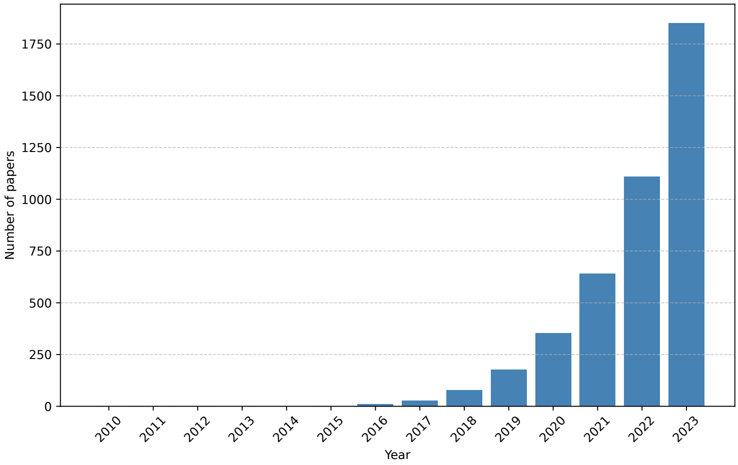
Auto-Encoder based approach with latent space mean selection:



Terefe, Tsegamlak, et al. "Time series averaging using multi-tasking autoencoder." IEEE 32nd International Conference on Tools with Artificial Intelligence (ICTAI) 2020.

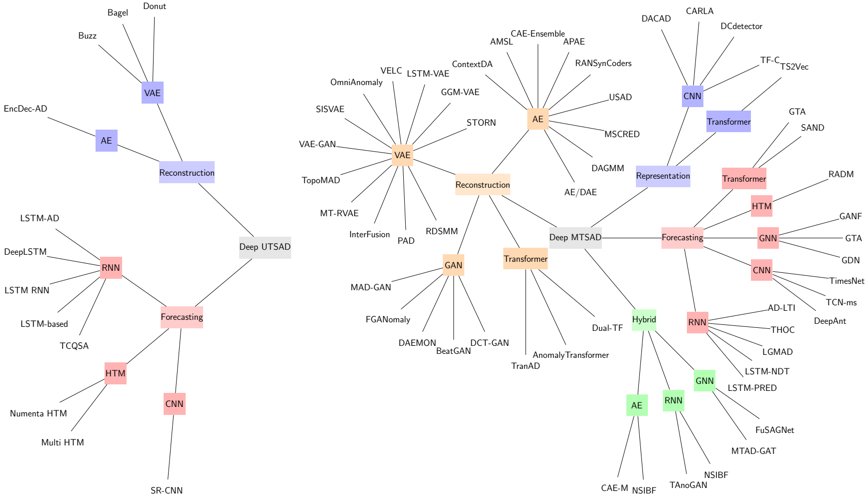
Terefe, Tsegamlak, et al. "Estimating time series averages from latent space of multi-tasking neural networks." Knowledge and Information Systems 2023.

Deep Learning for Time Series Anomaly Detection



Deep Learning for Time Series Anomaly Detection

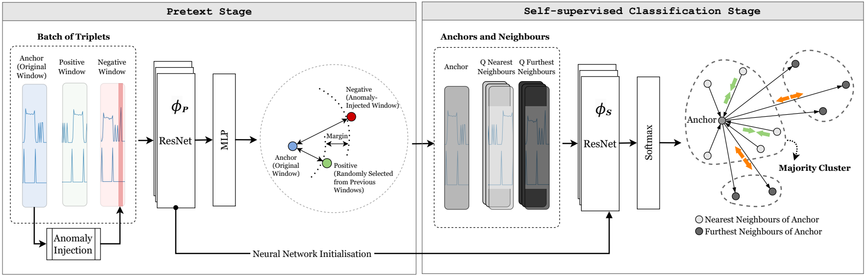
Deep Learning models for univariate/multivariate TSAD:



Zahra Z. Darban et al. "Deep Learning for Time Series Anomaly Detection: A Survey." ACM Computing Surveys 2024.

Deep Learning for Time Series Anomaly Detection

Introducing pair selection approaches for contrastive learning using anomaly injection techniques, and nearest and furthest neighbors - CARLA 2024



Zahra Z. Darban et al. "CARLA: Self-supervised Contrastive Representation Learning for Time Series Anomaly Detection." Pattern Recognition 2024.