Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC PapersOnLine 56-2 (2023) 7716–7721

FRAN-X: An improved diagnostic transfer learning approach with application to ball bearings fault diagnosis

L. Pitturelli, M. Mazzoleni, L. Rillosi, F. Previdi *

* *Department of Management, Information and Production engineering University of Bergamo, via Marconi 5, 24044 Dalmine (BG), Italy (e-mail: leandro.pitturelli@unibg.it).*

Abstract:
Data-driven diagnostic methods are attractive from an industrial and practical perspective due to their limited amount of required prior knowledge about the process or component under monitoring. However, these methods usually require a large amount of healthy and possibly faulty labeled data. Often, gathering and manually labeling a vast dataset is not feasible in real scenarios. Transfer learning has emerged as an answer to the labeling problem, exploiting the idea that the diagnostic knowledge could be reused across multiple different, but related, machines and operating conditions. In this work, we introduce several improvements to the Feature Representation and Alignment Network (FRAN) architecture described in (Chen et al., 2020) devised with the diagnostic transfer learning purpose. Our approach, named FRAN-X, presents improved transfer and diagnostics performance between identical machines in different operating conditions, and it is computationally lighter than its original counterpart. The FRAN-X approach is evaluated on the CWRU-bearing dataset and on experimental data collected from a Computerized Numerical Control (CNC) workcenter machine.

Keywords: Transfer learning; Fault diagnosis

1. INTRODUCTION

Data-driven diagnostic approaches include the use of machine learning algorithms for fault diagnosis (Gao et al., 2015). While signal-based and model-based diagnostic approaches require some prior information about the machine components under diagnosis, a knowledge-based method is expected to automatically detect and recognize the health states of the components using its available measurements (Mazzoleni et al., 2021). Common assumptions when using supervised machine learning are that: (i) sufficient healthy and faulty labeled data are available to train the models; (ii) the train and test data distributions are the same. However, in practical engineering scenarios, both assumptions might be violated, primarily due to the facts that machines run in healthy conditions, and performing experiments with faults requires time and machine availability. It follows that: (i) lots of healthy data are collected in operative, running scenarios; (ii) if available at all, faulty data are most commonly collected in ad-hoc, laboratory settings. This leads to situations where training (laboratory) and test (actual machine usage) data are collected under *different experimental settings*. If not properly tackled, this *domain shift* problem makes models developed on the first setting unusable on the second one (Valceschini et al., 2021).

A first solution for the employment of a data-drive diagnostic method in these situations would be to use anomaly detection algorithms that can be designed directly with healthy data from running operating conditions (Hendrickx et al., 2020; Mazzoleni et al., 2022). Alternatively, *transfer learning* methods, powered by deep neural networks architectures, can be leveraged to *reutilize the diagnostic knowledge* extracted from one problem domain to another, e.g. when dealing with similar machines that work under different operating conditions of speed, load, vibration transmission path, humidity, and so on (Yang et al., 2019). Practical use of transfer learning approaches consist first in the collection of a labeled dataset (*source domain dataset*) under laboratory or controlled experimental conditions. The model built with this source dataset is then employed to process the unlabeled (or poorly labeled) dataset collected under actual experimental conditions (*target domain dataset*). The envisaged result is an automatic features extraction method from raw signals, less dependent on the prior knowledge of the diagnosticians, able to reduce signal processing times required to grasp the measurements properties in healthy and faulty conditions (Lei et al., 2016).

Of the various approaches to transfer learning (Lei et al., 2020), feature-based ones learn a feature mapping to map the source and target data into a common features space, where distribution discrepancy of the features is measured by a distance metric and minimized. Then, a classifier, trained with the source-domain labeled samples, is employed to work on the target domain data based on features with a similar distribution (Yang et al., 2019), see Fig. 1. In the context of cross-domain fault diagnosis of rotating components, the Feature Representation Alignment Network (FRAN) has been proposed in (Chen et al., 2020)

10.1016/j.ifacol.2023.10.1175

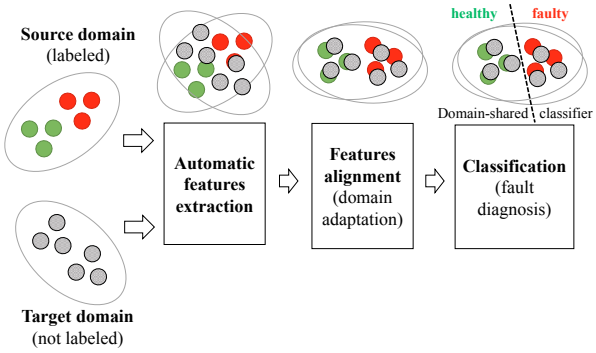

Fig. 1. Steps of feature-based transfer learning approaches.

as a neural-network-based algorithm comprising automatic features extraction, domain adaptation and classification, comparing favorably with previous transfer approaches as (Lu et al., 2017) on the public Case Western Reserve University (CWRU) dataset related to bearing faults data (Case Western Reserve University, 2012). Rolling elements bearings are the most common part of a mechanical system in engineering and one of the main contributors to the machine failures. The monitoring of defective bearings has been extensively studied, with a standard approach for their diagnosis based on envelope analysis of vibration data (Randall and Antoni, 2011). Envelope analysis is a powerful method for components rotating at constant speed, and it can be applied also to other mechanical components that are not, but resemble, bearings (Valceschini et al., 2022). However, the fault symptoms are related to the magnitude of specific frequencies that depend on the rotation speed. If another component rotates at a different velocity, the rationale of the diagnostic algorithm has to be redesigned. Thus, transfer learning approaches are envisaged in this setting. In this work, we propose an improvement over the FRAN model, called the FRAN-X model. The improvements consist in: (i) the introduction of recent ResNeXt blocks and skip connections in place of standard convolutional layers in the features extraction component (encoder); (ii) reduced dimension of the automatically extracted features vectors; (iii) increasing of the receptive field of the encoder, with consequent balancing of the number of parameters between the encoder and the classifier. The improvements lead to *increased transfer capability* and *computational speed* due to lowered memory requirements. The improved approach is evaluated on the CWRU bearing dataset and on an experimental dataset collected by the authors from a Computerized Numerical Control (CNC) machine, consisting in experiments with an inner-race faulty bearing (Mazzoleni et al., 2020).

The remainder of this paper is organized as follows. Section 2 introduces transfer learning problems in fault diagnosis and the proposed FRAN-X architecture. Section 3 describes the data source used for evaluation of the diagnostic algorithms. Section 4 reports the experimental results, while Section 5 is devoted to concluding remarks.

2. TRANSFER LEARNING FOR FAULT DIAGNOSIS OF ROTATING COMPONENTS

For transfer learning in fault diagnosis, the diagnostic knowledge is expected to be reused from one or multiple diagnosis tasks (source domain) to other related but different ones (target domain). The source domain $\mathcal{D}^s = \{(\boldsymbol{x}_i^s, y_i^s)\}_{i=1}^{N^s}$ is composed by n labeled measurements where $\boldsymbol{x}_i^s \in \mathcal{X}^s \subseteq \mathbb{R}^n$ is the i-th measurements vector and y_i its corresponding diagnostic label (e.g. `healthy`, `fault 1`, `fault 2`, etc.). The source dataset $\mathcal{D}^s$ is sampled from the source space $\mathcal{X}^s \times \mathcal{Y}^s$, with $\mathcal{Y}^s$ the set of source labels. The source input set $\mathcal{X}^s$ is endowed with a probability distribution function $p(\mathcal{X}^s)$. The target domain $\mathcal{D}^t = \{\boldsymbol{x}_j^t\}_{j=1}^{M}$ consists of m unlabeled measurements $\boldsymbol{x}_i^t \in \mathcal{X}^t \subseteq \mathbb{R}^m$. The target dataset $\mathcal{D}^t$ is sampled from the target space $\mathcal{X}^t \times \mathcal{Y}^t$, with $\mathcal{Y}^t$ the set of target labels. The target input space $\mathcal{X}^t$ has probability distribution function $p(\mathcal{X}^t)$. Due to different operating conditions, the source and target domain are different so that $p(\mathcal{X}^s) \neq p(\mathcal{X}^t)$. For transfer learning in fault diagnosis, it is assumed that $\mathcal{Y}^t \subseteq \mathcal{Y}^s$ (Lei et al., 2020). The aim is to learn a model able to perform well on the target domain, using $\mathcal{D}^s$ and $\mathcal{D}^t$ as training information.

2.1 FRAN: Feature Representation Alignment Network

The FRAN architecture consists of two sub-models. The first sub-model performs an unsupervised *domain adaptation* task, with the aim to train a feature extractor function $g(\cdot)$ (*encoder*) able to produce features from raw signals for both the source and target domains. The second sub-model is the classifier $f(\cdot)$ needed to perform fault classification.

The encoder $g(\cdot)$ produces the set of source features $\mathcal{Z}^s = g(\mathcal{X}^s)$ and the set of target features $\mathcal{Z}^t = g(\mathcal{X}^t)$. The encoder has two objectives both towards features alignment between source and target domains: (i) a maximization of the mutual information $-\mathcal{L}_M(\mathcal{Z}^s, \mathcal{Z}^t)$ between the target features $\mathcal{Z}^t$ and the entire features set $\mathcal{Z}^s \cup \mathcal{Z}^t$; (ii) a minimization of the Maximum Mean Discrepancy (MMD) (Gretton et al., 2012) cost $\mathcal{L}_D(\mathcal{Z}^s)$ applied to features computed from $g(\cdot)$. The first objective has the aim to generate informative features domains; the second objective has the aim of minimizing the discrepancy of the source and target features distributions. The classifier $f(\cdot)$ has the aim of minimizing the classification error $\mathcal{L}_C(\mathcal{X}^s, \mathcal{Y}^s)$ on source data (for which labels are available). The models $g(\cdot)$ and $f(\cdot)$ are trained simultaneously by minimizing the cost

$$\mathcal{L}(\mathcal{Z}^s, \mathcal{Z}^t, \mathcal{Y}^s) = \alpha \cdot \mathcal{L}_M(\mathcal{Z}^s, \mathcal{Z}^t) + \beta \cdot \mathcal{L}_D(\mathcal{Z}^s, \mathcal{Z}^t) + \\ + \mathcal{L}_C(\mathcal{X}^s, \mathcal{Y}^s), \quad (1)$$

where $\alpha, \beta \in \mathbb{R}_{\geq 0}$ are weighting coefficients. The diagnosis stage takes the target domain data as input. These data are then processed by the encoder $g(\cdot)$ to produce features which will be classified by $f(\cdot)$ to obtain the fault labels estimate.

The encoder $g(\cdot)$ consists of two 1-D convolutional layers, each followed by a batch normalization, and a Sigmoid activation function followed by average-pooling layers. The kernel size for both 1-D convolutional layers is 4, and the output channels are 32 and 64, respectively. The classifier $f(\cdot)$ consists of two fully connected layers with ReLU activation, with 1000 and 3 output channels, respectively [1].

[1] These and other configuration parameters can be devised from the code made available by the authors at

2.2 FRAN-X: an improved architecture for transfer learning

The proposed transfer learning architecture, named FRAN-X, maintains the the original structure and cost function of FRAN but it presents a more streamlined architecture with modifications to improve the model transfer capacity, diagnostic accuracy on target data and computational requirements. The differences are as follows, see Fig. 2:

(1) the classic convolutional layers of the encoder $g(\cdot)$ are replaced by a series of recently introduced ConvNeXt blocks (Liu et al., 2022);
(2) ResNet skip connections (He et al., 2016) are introduced to avoid the vanishing gradient problem, thus allowing the stacking of more layers;
(3) the sum operation of the residual part and the skip connections uses decrease-increase-decrease approach in the number of ConvNeXt filters;
(4) apart from the last layer of $f(\cdot)$ that computes the final classification score, the Softmax and ReLu activation functions in FRAN are replaced by the recently introduced Gaussian Error Linear Unit (GELU), removing consequently the dropout layer in $f(\cdot)$. GELU units (Hendrycks and Gimpel, 2016) are mostly used in transformers like BERT (Devlin et al., 2019);
(5) the 1-D kernel dimension of the first encoder layer is increased from 4 to 11. As suggested in (Peng et al., 2019), wide convolution kernels are better to extract low frequency features, while narrow convolution kernels are better at extracting high-frequency features. Using a larger kernel results in a better performance when dealing with high-frequency noise;
(6) in the new architecture, the number of parameters for the encoder $g(\cdot)$ is increased from 8.6 k to 36.7 k, while the number of parameters of the classifier $f(\cdot)$ has been significantly reduced from 19.2 M to 26.3 k. This change leads to a massive reduction in the overall number of parameters.

3. DATASETS DESCRIPTION AND PROCESSING

3.1 CWRU bearing dataset

The CWRU bearing dataset is an open-source dataset of vibration signals widely used to evaluate fault diagnosis algorithms. The bearing experiments were carried out on a 2 hp electric motor, and the vibration data were collected using accelerometers deployed on both the drive-end (DE) and fan-end (FE) of the motor housing. An electrical discharge machining was used to inject three types of faults on the ball bearings supporting the motor shaft: (i) inner race fault, (ii) outer race fault and (ii) ball fault, with fault range sizes from 0.007 to 0.028 inches. Each fault class and size combination is recorded under different loads from to 0 hp to 3 hp while the motor RPMs are kept roughly constant at 1750 RPM. The sampling frequencies of drive-end fault data is 12 kHz and 48 kHz, and the sampling frequency of the normal baseline data and fan-end fault data are both 12 kHz.

https://github.com/JiahongChen/FRAN. Last access: 8 Nov. 2022. The dimension of the last layer of $f(\cdot)$ depends on the number of classes of the considered problem (which is 3 in the CWRU dataset case).

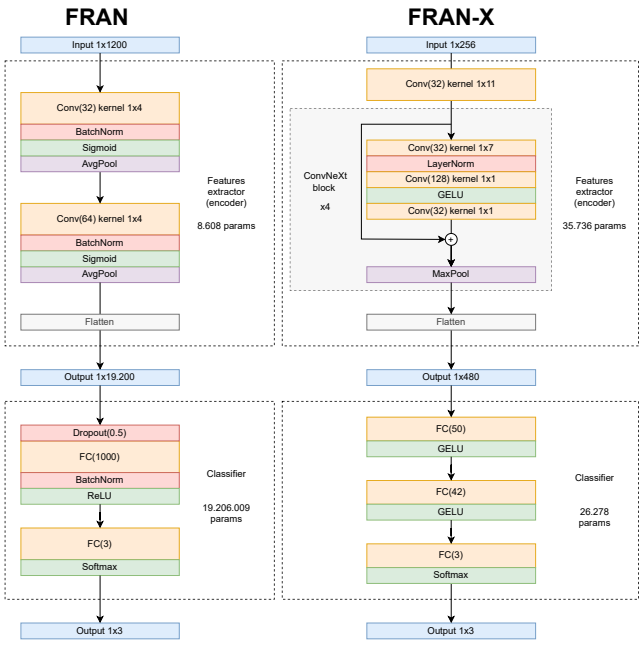

Fig. 2. Comparison between the architecture of the proposed FRAN-X model and the original FRAN model.

We divided the CWRU dataset in two sub-datasets, sharing the same sampling rate of 12 kHz:

- vibration signals in the CWRUA dataset are collected only from the drive-end with motor load at 0 hp;
- vibration signals in the CWRUB dataset are collected with motor load of 2 hp.

For both CWRUA and CWRUB data, only the class `Normal`, `Inner` race fault, and `Outer` race fault are considered. These datasets are denoted as `CWRUA3` and `CWRUB3`. Here, all fault sizes are kept together. We then considered a further subdivision based on sensor type, e.g. fan-end (FE) and drive-end (DE), and fault size (007, 014, 021 inches), while there isn't any division in motor horsepower. The resulting datasets are denoted as `DExxx` or `FExxx`, with `xxx` indicating the fault size.

3.2 CNC workcenter bearing dataset

Here the focus is on the upper bearing of the vertical shaft (Y direction) of a 5-axis CNC machine (Mazzoleni et al., 2020), see Fig. 3. The vertical shaft is supported by two threaded rods. Each rod has a set of four ball bearings which allow the rod to spin. We injected a severe pitting on one bearing, mounting an accelerometer on the bearing cage to collect its vibrations at a sample frequency of 12.8 kHz, during ascent and descent movements at different constant speeds. The load is kept constant to 0 N.

The workcenter dataset contains recordings for two classes: `Normal` and `Inner race` fault. Domain transfer is evaluated between different speeds considering the same CNC machine, denoted as `CNC50RPM` and `CNC1500RPM`.

3.3 Data preparation

The preprocessing scheme that takes the raw vibration data from the CWRU or CNC datasets and prepares

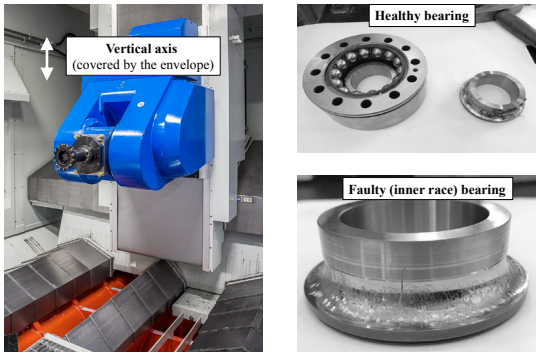

Fig. 3. CNC workcenter with healthy and faulty bearing.

these data for training and testing is as follows. Each raw signal is split into many fixed-size subsamples, using a rolling window approach. The window size and overlap are set respectively to 256 samples with 80% of overlap. The window size differs from FRAN, where 1200 samples were considered. Next, a min-max normalization is applied to compress the signal's amplitude to the range $[-1, 1]$. Before splitting both source and target into the training and test sets, a random shuffle of all the data is applied. An example of CWRU data is shown in Fig. 4.

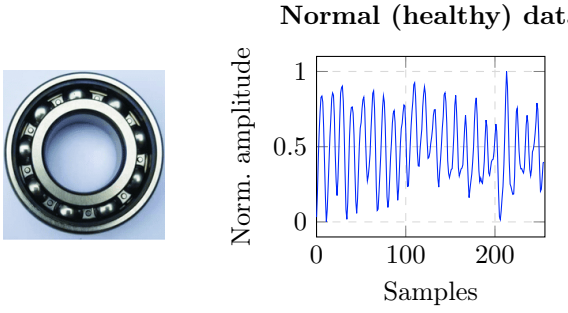

Fig. 4. CWRU healthy class component and data.

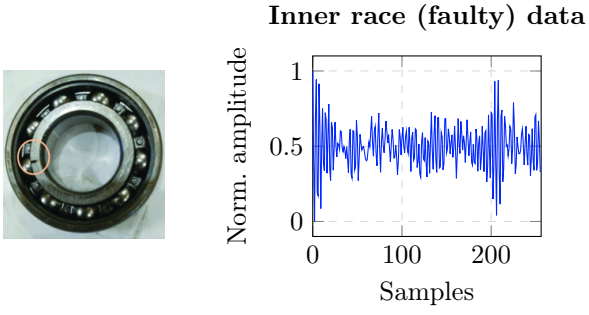

Fig. 5. CWRU inner race fault class component and data.

4. EXPERIMENTAL EVALUATION

4.1 Transfer learning on CWRU dataset

The CNC dataset is used to evaluate transfer capabilities over *different bearing loads*. The FRAN-X hyperparameters are tuned by grid search using the training sets of source and target data, considering the single transfer scenario from `CWRUA3` (source) to `CWRUB3` (target) and evaluating the classification accuracy on the target dataset. The best hyperparameters found are $\alpha = 1$ and $\beta = 0$. Interestingly, for this transfer scenario the introduced motifications make not necessary the MMD cost $\mathcal{L}_D(\cdot, \cdot)$ in (1), suggesting that good separable and similar features are generated by $g(\cdot)$ by considering only the mutual information cost $\mathcal{L}_M(\cdot, \cdot)$ in (1).

Transfer scenario 1: `CWRUA3` → `CWRUB3`. First, we compared the transfer learning capabilities of the FRAN-X model considering the `CWRUA3` (source) and `CWRUB3` (target) scenario, while comparing with FRAN in the same setting. Both datasets contain the same number of samples per class. When transfer learning is disabled (OFF) the hyperparameters α, β in (1) are set to zero, so that only the classification cost is minimized during model training. Fig. 6 shows the faults classification accuracy of FRAN-X on the test target set while training for 60 epochs. FRAN-X achieves higher accuracy on the target domain when transfer learning is enabled (ON), reaching an accuracy of 98.3%. The FRAN model reports an accuracy of 91.3% after 100 epochs with transfer ON. The FRAN-X model is able to achieve better accuracy than FRAN (on this problem) also when transfer for FRAN-X is disabled.

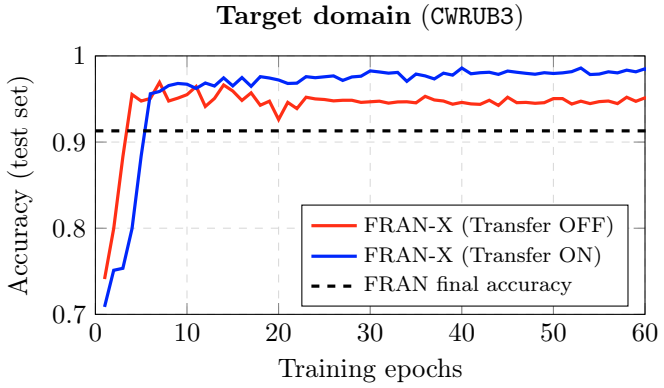

Fig. 6. FRAN-X transfer capabilities on the `CWRUA3`(source) → `CWRUB3`(target) scenario.

Transfer scenario 2: `CWRUB3` → `CWRUA3`. The same configuration as the previous one applies, but with source and target domains swapped. The FRAN-X model achieves an accuracy of 96.6% after 60 epochs, while the FRAN model reports an accuracy of 91.7% after 100 epochs.

Transfer scenario 3: sensors swap. In this third experiment, the source and target domains differ by the sensor used to collect the data. The source domain is composed of data collected using the drive-end (DE) sensor, while the target domain is composed of data collected using the fan-end (FE) sensor. The fault size is fixed at 0.021 inches. The results show an accuracy of 98.4% and 99.8% on the `DE021` → `FE021` and `FE021` → `DE021` configurations, respectively. Compared to the FRAN, this results in a 10.2% improvement in accuracy in the first configuration and a slight decrease of -0.2% in the second configuration.

Transfer scenario 4: sensors and fault sizes swap. In this experiment, both the sensor and the fault size are different between the source and target domains. The first data batch `FE007` ↔ `DE014` is performed with data from 0.007 inches faults collected using the FE sensor and data collected using the DE sensor with a fault size of 0.014 inches. The model achieves an accuracy of 99.9% on the `FE007` → `DE014` scenario, and 99.3% on the `DE014` → `FE007` case. Compared to the FRAN, this results in a 11.5% improvement in accuracy in the first scenario and a 8.2% improvement in the second scenario. The second data batch `DE007` ↔ `FE014` is performed with a similar configuration as the previous one. Interestingly, while the `DE007` → `FE014` configuration achieves an accuracy of 93.2%, which represents a +3.7% improvement over FRAN, the `FE014` → `DE007` accuracy drops during training, and the early stopping callback interrupts the training after a few epochs. This may be because the `FE014` → `DE007` configuration is more challenging than the `DE007` → `FE014` configuration, and the model is not able to generalize well to this target domain. Further investigation is needed to fully understand the reason for this behavior.

Table 1 summarizes the fault classification results on the CWRU dataset.

Source	Target	FRAN	FRAN-X	Change
CWRUA3	CWRUB3	91.3%	98.3%	**+7.0%**
CWRUB3	CWRUA3	91.7%	96.6%	**+4.9%**
DE021	FE021	88.2%	98.4%	**+10.2%**
FE021	DE021	100.0%	99.8%	-0.2%
FE007	DE014	88.4%	99.9%	**+11.5%**
DE014	FE007	91.1%	99.3%	**+8.2%**
DE007	FE014	89.5%	93.2%	**+3.7%**
FE014	DE007	84.5%	49.2%	-35.3%

Table 1. Comparison between FRAN and FRAN-X accuracy on multiple CWRU configurations.

4.2 Transfer learning on CNC dataset

The CNC dataset is used to evaluate transfer capabilities over different *bearing rotation speeds*. Here, two classes are present: the `Normal` (healthy) class and the `Inner race` (faulty) one. The FRAN-X model is trained from scratch using the same set hyperparameters as before.

Transfer scenario 1: `CNC50RPM` → `CNC1500RPM`. When training with transfer disabled (OFF), the model reaches an accuracy of 74.7% on the test target domain. When domain transfer is enabled (ON), the model can achieve an accuracy of 95.0% on the test target domain. The evolution of the accuracies is shown in Fig. 7.

Transfer scenario 2: `CNC1500RPM` → `CNC50RPM`. As shown in Fig. 8, without domain adaptation the model is not able to reach an adequate performance, reaching 50.0% accuracy on the test target set, which is equivalent to perform a random classification guessing. With the help of domain transfer, the FRAN-X model reaches an accuracy of 98.3% on the test target set.

Table 2 summarizes the fault classification results on the CNC dataset.

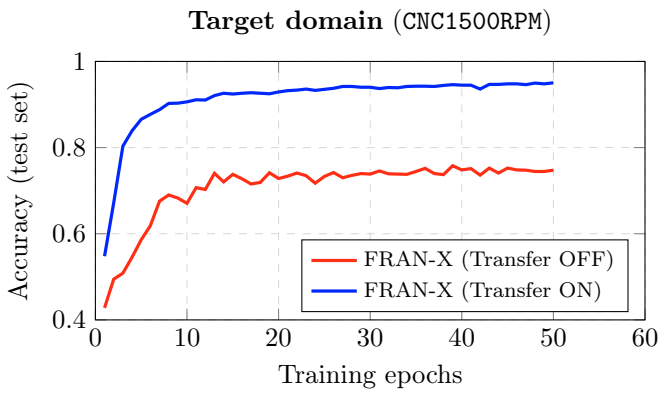

Fig. 7. FRAN-X transfer capabilities on the `CNC50RPM`(source) → `CNC1500RPM`(target) scenario.

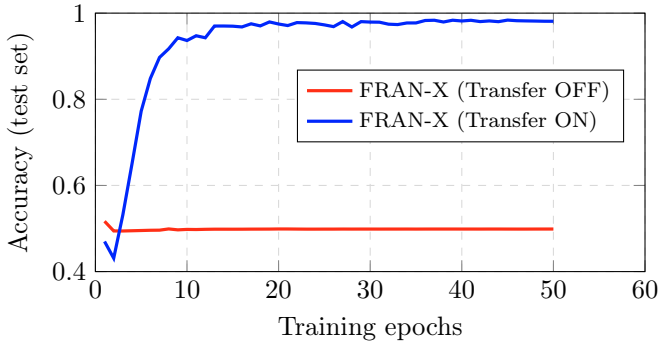

Fig. 8. FRAN-X transfer capabilities on the `CNC1500RPM`(source) → `CNC50RPM`(target) scenario.

4.3 Computational aspects

The FRAN-X has been developed using the PyTorch Lightning framework on three hardware configurations: (i) `Intel i7-6700HQ 2.60GHz CPU` (ii) `Nvidia Tesla T4 GPU`; (iii) `Nvidia GTX960M GPU`. Compared with FRAN, the FRAN-X model has a significantly lower number of parameters (see Fig. 2), which makes it possible to train and test FRAN-X on both high-end and low-end hardware (such as a laptop). Table 3 reports inference times of FRAN and FRAN-X on the considered hardware platforms. The FRAN-X model performs considerably better on CPU. The FRAN model requires a 7GB Graphics Processor Unit to run (Chen et al., 2020). The smaller footprint of FRAN-X removes this requirement, allowing the model to run on smaller GPUs such as the `Nvidia GTX960M GPU`, where the FRAN model could not be run.

5. CONCLUSION

Transfer learning is a fascinating approach for fault diagnosis in complex industrial environments, requiring very few prior process knowledge from the diagnostician. Experimental results on two different datasets showed its potential. However, several challenges are still open. First, we should try to understand if the features extracted by the encoder have physical meaning (e.g. frequencies magnitudes in case of rotating component and vibration data).

Source	Target	Transfer OFF	Transfer ON	Change [%]
CNC50RPM	CNC1500RPM	74.7%	95.0%	**+20.3%**
CNC1500RPM	CNC50RPM	49.8%	98.3%	**+48.5%**

Table 2. Summary of FRAN-X accuracy on the CNC dataset, with transfer learning ON and OFF.

Hw	FRAN	FRAN-X
Intel i7-6700HQ 2.60GHz CPU	34.3 ms	**0.6 ms**
Nvidia Tesla T4 GPU	0.63 ms	**0.43 ms**
Nvidia GTX960M GPU	-	**0.62 ms**

Table 3. Inference time for FRAN and FRAN-X on different hardware platforms.

In this work, the approach was limited to the Transfer to Identical Machines (TIM) scenario Lei et al. (2020). The Transfer to Different Machine (TDM), where a model trained on one machine is used to diagnose a different one, is much more involved and requires further research.

REFERENCES

Case Western Reserve University (2012). Bearing data center seeded fault data. URL https://engineering.case.edu/bearingdatacenter.

Chen, J., Wang, J., Zhu, J., Lee, T.H., and de Silva, C.W. (2020). Unsupervised cross-domain fault diagnosis using feature representation alignment networks for rotating machinery. *IEEE/ASME Transactions on Mechatronics*, 26(5), 2770–2781. doi:10.1109/TMECH.2020.3046277.

Devlin, J., Chang, M.W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, 4171–4186. Association for Computational Linguistics, Minneapolis, Minnesota. doi:10.18653/v1/N19-1423.

Gao, Z., Cecati, C., and Ding, S.X. (2015). A survey of fault diagnosis and fault-tolerant techniques part ii: Fault diagnosis with knowledge-based and hybrid/active approaches. *IEEE Transactions on Industrial Electronics*, 62(6), 3768–3774. doi:10.1109/TIE.2015.2419013.

Gretton, A., Borgwardt, K.M., Rasch, M.J., Schölkopf, B., and Smola, A. (2012). A kernel two-sample test. *The Journal of Machine Learning Research*, 13(1), 723–773.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778. doi:10.1109/CVPR.2016.90.

Hendrickx, K., Meert, W., Mollet, Y., Gyselinck, J., Cornelis, B., Gryllias, K., and Davis, J. (2020). A general anomaly detection framework for fleet-based condition monitoring of machines. *Mechanical Systems and Signal Processing*, 139, 106585.

Hendrycks, D. and Gimpel, K. (2016). Gaussian error linear units (gelus). doi:10.48550/ARXIV.1606.08415.

Lei, Y., Jia, F., Lin, J., Xing, S., and Ding, S.X. (2016). An intelligent fault diagnosis method using unsupervised feature learning towards mechanical big data. *IEEE Transactions on Industrial Electronics*, 63(5), 3137–3147. doi:10.1109/TIE.2016.2519325.

Lei, Y., Yang, B., Jiang, X., Jia, F., Li, N., and Nandi, A.K. (2020). Applications of machine learning to machine fault diagnosis: A review and roadmap. *Mechanical Systems and Signal Processing*, 138, 106587. doi:10.1016/j.ymssp.2019.106587.

Liu, Z., Mao, H., Wu, C.Y., Feichtenhofer, C., Darrell, T., and Xie, S. (2022). A convnet for the 2020s. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11976–11986. doi:10.1109/CVPR52688.2022.01167.

Lu, W., Liang, B., Cheng, Y., Meng, D., Yang, J., and Zhang, T. (2017). Deep model based domain adaptation for fault diagnosis. *IEEE Transactions on Industrial Electronics*, 64(3), 2296–2305. doi:10.1109/TIE.2016.2627020.

Mazzoleni, M., Scandella, M., Maurelli, L., and Previdi, F. (2020). Mechatronics applications of condition monitoring using a statistical change detection method. *IFAC-PapersOnLine*, 53(2), 92–97. doi:10.1016/j.ifacol.2020.12.100. 21st IFAC World Congress.

Mazzoleni, M., Rito, G.D., and Previdi, F. (2021). *Electro-Mechanical Actuators for the More Electric Aircraft*. Springer International Publishing. doi:10.1007/978-3-030-61799-8.

Mazzoleni, M., Sarda, K., Acernese, A., Russo, L., Manfredi, L., Glielmo, L., and Del Vecchio, C. (2022). A fuzzy logic-based approach for fault diagnosis and condition monitoring of industry 4.0 manufacturing processes. *Engineering Applications of Artificial Intelligence*, 115, 105317. doi:10.1016/j.engappai.2022.105317.

Peng, D., Liu, Z., Wang, H., Qin, Y., and Jia, L. (2019). A novel deeper one-dimensional cnn with residual learning for fault diagnosis of wheelset bearings in high-speed trains. *IEEE Access*, 7, 10278–10293. doi:10.1109/ACCESS.2018.2888842.

Randall, R.B. and Antoni, J. (2011). Rolling element bearing diagnostics - a tutorial. *Mechanical Systems and Signal Processing*, 25(2), 485–520. doi:10.1016/j.ymssp.2010.07.017.

Valceschini, N., Mazzoleni, M., Pitturelli, L., and Previdi, F. (2022). Experimental fault detection of input gripping pliers in bottling plants. *IFAC-PapersOnLine*, 55(6), 778–783. doi:10.1016/j.ifacol.2022.07.221. 11th IFAC Symposium on Fault Detection, Supervision and Safety for Technical Processes (SAFEPROCESS).

Valceschini, N., Mazzoleni, M., and Previdi, F. (2021). Inertial load classification of low-cost electro-mechanical systems under dataset shift with fast end of line testing. *Engineering Applications of Artificial Intelligence*, 105, 104446. doi:10.1016/j.engappai.2021.104446.

Yang, B., Lei, Y., Jia, F., and Xing, S. (2019). An intelligent fault diagnosis approach based on transfer learning from laboratory bearings to locomotive bearings. *Mechanical Systems and Signal Processing*, 122, 692–706. doi:10.1016/j.ymssp.2018.12.051.