Generative Adversarial Networks used for Latent Space Optimization: A Comparative Study for Partial Discharge Analysis

Ryad Zemouri[1], Mélanie Lévesque[2], Olivier Kokoko[3], and Claude Hudon[4]

[1,2,3,4] *Institut de Recherche d'Hydro-Québec (IREQ), Varennes, QC, J3X1S1 Canada*
zemouri.ryad@hydroquebec.com
levesque.melanie2@hydroquebec.com
kokoko.olivier2@hydroquebec.com
hudon.claude@hydroquebec.com

[1] *Cedric-Lab, CNAM, HESAM université, 292, rue Saint-Martin, 750141 Paris cedex 03, France*
ryad.zemouri@cnam.fr

ABSTRACT

Hydrogenerators are complex equipment with many components where more than 100 failure mechanisms can be active. During normal operation, the high voltage stator, one of the main components of hydrogenerators, is always subjected to Partial Discharge (PD) activity. Multiple sources of PD activity can be active simultaneously. Global PD signals which include all active PD sources are obtained from periodic measurements made on hydrogenerators while they are in operation. PD measurements are an effective diagnostic tool for evaluating the integrity of the stator winding, similar to signals coming from an electrocardiogram for the health status of a human. Quantifying PD activity is still a challenge in the industry since the recognition of the type of PD is not trivial and still requires expert judgement. Since the degradation rate of all active PD sources is different, automatic classification of PD source is thus essential to monitor their evolution. With that goal in mind, an extensive effort was initiated in 2019 to automatically recognize individual PD sources from 2D Partial Discharge Analyzer (PDA) files using Deep Learning (DL) techniques. In this context, this paper presents the use of a Generative Adversarial Network (GAN) in combination with A Variational Autoencoder (VAE) for increasing the representativeness of each PD sources in the VAE latent space.

1. INTRODUCTION

PD activity is caused by local concentrations of electrical stress and occurs within voids in insulation or around insulating system exposed to high voltage. During normal operation of high voltage hydrogenerators, PD activity from multiple sources is always present in the stator insulation. Each of these PD sources has its own insulation degradation rate as well as its risk of failure. At Hydro-Québec, PD measurements on hydrogenerators have been performed over the past 30 years using the Partial Discharge Analyzer (PDA) instrument, a two-dimensional (2D) pulse height analyzer. More than 33 000 periodic measurement files have been recorded using this instrument. The PDA instrument is used as a first level diagnostic tool. In addition, since the early 2000s, when PDA measurements indicate an intense discharge activity or a sudden increase, Phase Resolved Partial Discharge (PRPD) measurements are made. This second level diagnostic tool which gives a three-dimensional (3D) representation of the PD activity, is used by experts to recognize the different types of PD sources that may occur in the stator insulation. More than 6000 PRPD measurement files have been recorded using the PRPD instrument. Typical pattern giving by these two PD instruments are shown in figure 1. Each measurement file obtained using the PDA or the PRPD instrument is available in a home-built database called MIDA (Methodology for Integrated Diagnostic of Generator). Measurement files in the MIDA database are not publicly available due to the sensitivity of these data for Hydro-Québec.

Suitable recognition of active PD sources is essential to improve prognostic model of hydrogenerators and to reduce the risk of in-service failure. However, automatic recognition of PD sources is not straightforward and cannot be based on ground rules alone, experts are still required. An analogy would be the diagnosis of the heart using an electrocardiogram where different characteristic signatures can be recognized by medical specialist and associated with health issues.

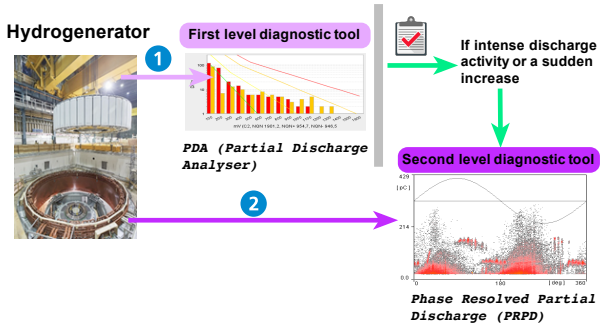

Figure 1. A two-step diagnosis tools for hydrogenerator. A first level of analysis performed by the PDA instrument and a second level is made using the PRPD instrument.

A PDA measurement file has the form of a histogram as it can be seen in Figure 1. The discharge rate (PD/s) is plotted against the amplitude in mV for the positive and negative discharge pulses, respectively Pd_i^+ (in red) and Pd_i^- (in yellow), for each of the 16 amplitude channel i of the horizontal axis. Each PD signal is then represented by a 2D matrix:

$$\mathbf{Pd} = \begin{bmatrix} Pd_1^+, ..., Pd_i^+, ..., Pd_{16}^+ \\ Pd_1^-, ..., Pd_i^-, ..., Pd_{16}^- \end{bmatrix} \quad (1)$$

Here, specific features based on the ratio of positive to negative pulses have been extracted from the PDA database for only the following three PD sources: symmetric PD patterns, negative asymmetry PD patterns and positive asymmetry PD patterns. These three PD sources represent 70% of the entire PDA database which yields approximately 23 000 measurement files. Analysis of those files suggests an imbalance in the distribution of each PD sources where symmetric PD patterns account for about 65% of these files while negative asymmetry and positive asymmetry PD patterns represent 15% and 20% respectively. In this paper, a comparative study to optimize the latent space based on the PDA files coming from these three PD sources is presented. The combination of two DL techniques is used: The Variational Autoencoder (VAE) and the Generative Adversarial Network (GAN).

In industrial applications, two major difficulties can be encountered: unbalanced and unlabelled data. On one side, often more operating data are available for healthy modes than for other unhealthy modes. On the other side, most data collected are not labelled. GANs are therefore an interesting alternative to address these two problems by generating artificial data to compensate for data imbalance and minority oversampling (Mullick, Datta, & Das, 2019) (Pan, Chen, Xie, Chang, & Zhou, 2020) (Zou, Li, & Xu, 2020) (Zhou, Yang, Fujita, Chen, & Wen, 2020) (Gao, Deng, & Yue, 2020). In the PHM domain, GANs can be used for two purposes:

- Diagnosis and fault detection (Dai, Wang, Huang, Shi, & Zhu, 2020) (Ducoffe, Haloui, & Gupta, 2019) (Han, Liu, Yang, & Jiang, 2019) (Liu et al., 2018) (Shao, Wang, & Yan, 2019) (Wang, Huang, Hu, & Yang, 2018) (Wang, Wang, & Wang, 2018) (Zheng & Gupta, 2020) (Zhang et al., 2020) (Pan et al., 2020) (Zou et al., 2020) (Zhou et al., 2020) (Mao, Liu, Ding, & Li, 2019) (Farajzadeh-Zanjani, Hallaji, Razavi-Far, Saif, & Parvania, 2021),
- Prognostic (Khan, Prosvirin, & Kim, 2018) (Huang, Tang, VanZwieten, Liu, & Xiao, 2019) (Que, Xiong, & Xu, 2019) (Doulamis et al., 2020) (Li, Zhang, Ma, Luo, & Li, 2020) (Bao, Miao, Wang, Yang, & Zhang, 2020).

The paper is organized as follows. The proposed approach for PD analysis is first introduced in section 2. Thereafter, both generative models (VAE and GAN) are briefly described in section 3. Experimental results and data analysis are presented in section 4 while the last section of the paper summarizes the conclusion and future works.

2. The proposed approach for PD analysis

The general scheme of the proposed approach is illustrated in Figure 2. A Variational Autoencoder (VAE) is used for dimension reduction and projection into a 2D latent space to analyze the training data. Analyzing the data is an essential step before the classification phase. Projection into a 2D latent space allows to see if clusters of features are emerging, thus facilitating the interpretation of the results obtained by the classifier. Data space in the latent space of the VAE is restructured and reorganized in a continuous way. This characteristic is due to the regularization layer z which ensures continuity around a given point on one hand, (each new input data point is represented by a cluster of points uniformly distributed around the mean), and on the other hand, the Kullback-Leibler term of the loss function provides a Gaussian distribution of all the latent space around the origin point. The problem is how to optimize this latent space and obtain the best distribution for each PD source in order to maximize the chances of a good classification? This is an open question as it is not easy to quantify the quality of a low dimensional feature space. What is certain is that the optimization of the VAE learning is directly related to the quality of the learning database, i.e. a large size and a perfectly balanced database, which is not the case with the three PD sources from the PDA database. The Figure 3 illustrates this causal relationship of the VAE latent space performance.

The objective of this study is to compare the quality of the latent space obtained from the expert rules with a latent space obtained directly from the input signal in an *End-to-End* approach. The first method concerns an original unsupervised DL method for PD source recognition. Instead of using labeled PD measurement files for a supervised learning process, rules developed by PD experts were used to create a feature vector from recognizable PD patterns (purple path in Figure 2). Indeed, labelling enough PD measurement files for

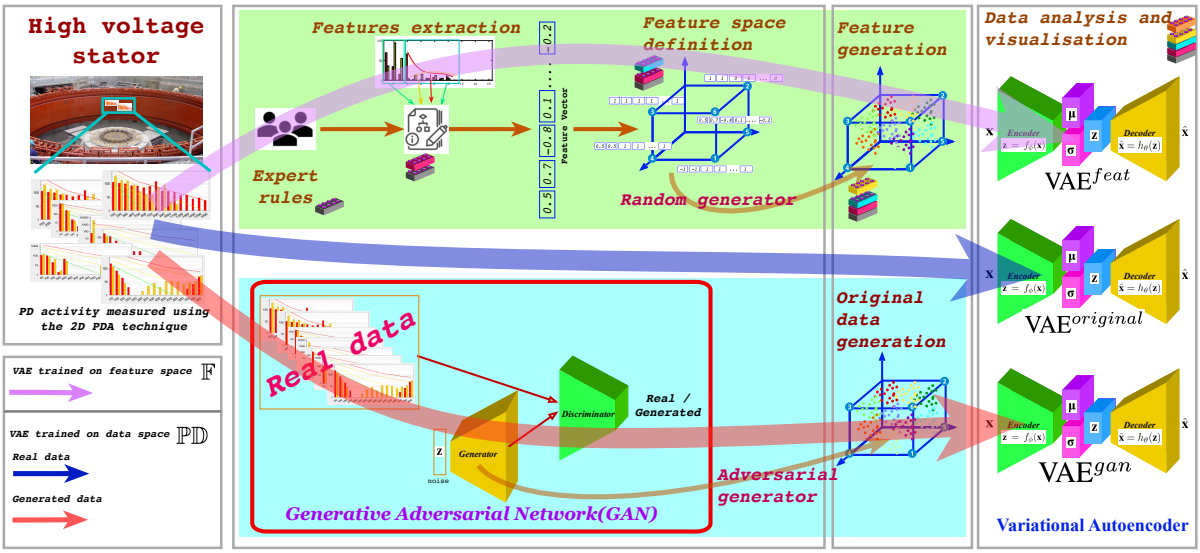

Figure 2. The DL framework of the proposed method for the PD analysis.

a supervised approach is very time-consuming and therefore cannot be implemented. This is a common problem in the industry where more and more operational data is available, but in most of the time, is not labeled by experts. In the proposed approach, expert knowledge is injected into a characteristic feature vector. The resulting feature space is defined by the feature vector and delimited by extreme data points. A random data generator is then used to artificially increase the VAE learning base by generating data points within the feature definition space (referred as VAE^{feat} in Figure 2). The whole feature definition space is thus covered while taking care to create a balanced learning base to optimize the VAE learning process. In the second method, the VAE is trained directly on the original PDs signals without any feature extraction. To do so, two different ways were tested: one VAE built from real PD data (referred as VAE^{original} in Figure 2: the blue path) and a second VAE obtained from synthetic data generated by a set of Generative Adversarial Networks (GANs) (referred as VAE^{gan} in Figure 2: the red path). The use of GANs artificially increased the number of PD measurement files in the database to optimize the learning of the VAE.

3. Theoretical Background

3.1. The variational autoencoders

Autoencoder (AE) represents one of the first generative models trained to recreate or reproduce the input vector $\mathbf{x}$. The AE is composed by two main structures: an encoder and a decoder, which are multilayered neural networks (NNs) parameterized by ϕ and θ, respectively. The first part encodes the input data $\mathbf{x}$ into a latent representation $\mathbf{z}$ by the encoder function $\mathbf{z} = f_\phi(\mathbf{x})$, whereas the second NN decodes this latent

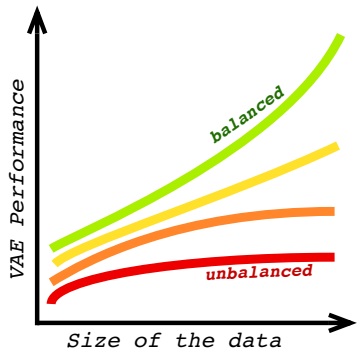

Figure 3. Causal relationship between the VAE latent space performance and the quality of the learning data.

representation onto $\hat{\mathbf{x}} = h_\theta(\mathbf{z})$ which is an approximation or a reconstruction of the original data. In an AE, an equal number of units are used in the input/output layers while fewer units are used in the latent space. The variational form of the AE becomes a popular generative model by combining the Bayesian inference and the efficiency of the NNs to obtain a nonlinear low-dimensional latent space use in industrial applications such as fault detection (Proteau, Zemouri, Tahan, & Thomas, 2020), (Huang, Chen, & Huang, 2019), (Lee, Kwak, Tsui, & Kim, 2019), (Martin, Droguett, Meruane, & das Chagas Moura, 2018). The Bayesian inference is obtained by an additional layer used for sampling the latent vector $\mathbf{z}$ with a prior specified distribution $p(\mathbf{z})$, usually assumed to be a standard Gaussian $\mathcal{N}(0, \mathbf{I})$, where $\mathbf{I}$ is the identity matrix. Each element z_i of the latent layer is obtained as follows: $z_i = \mu_i + \sigma_i.\epsilon$ where μ_i and σ_i are the i^{th} components of the mean and standard deviation vectors, ϵ is a random variable

following a standard Normal distribution ($\epsilon \sim \mathcal{N}(0,1)$). Unlike the AE, which generates the latent vector $\mathbf{z}$, the VAE generates vector of means μ_i and standard deviations σ_i. This allows to have more continuity in the latent space than the original AE. The VAE loss function has two terms $\mathcal{L} = \mathcal{L}_{rec} + \mathcal{L}_{kl}$ The first term $\mathcal{L}_{rec} = -\mathbb{E}_{q_\phi(\mathbf{z}|\mathbf{x})}[log\, p_\theta(\mathbf{x} \mid \mathbf{z})]$ is the reconstruction loss function, which allows to minimize the difference between the input and output instances. Both the negative expected log-likelihood (e.g., the cross-entropy function) and the mean squared error (MSE) can be used. When the sigmoid function is used in the output layer, the derivatives of MSE and cross-entropy can have similar forms. The second term $\mathcal{L}_{kl} = \mathbb{D}_{kl}(q_\phi(\mathbf{z} \mid \mathbf{x}) \parallel p(\mathbf{z}))$ corresponds to the Kullback–Leibler ($\mathbb{D}_{kl}$) divergence loss term that forces the generation of a latent vector with the specified Normal distribution (Kingma, 2017). The $\mathbb{D}_{kl}$ divergence is a theoretical measure of proximity between two densities $q(x)$ and $p(x)$. It is asymmetric ($\mathbb{D}_{kl}(q \parallel p) \neq \mathbb{D}_{kl}(p \parallel q)$) and nonnegative. It is minimized when $q(x) = p(x)$. Thus, the $\mathbb{D}_{kl}$ divergence term measures how close the conditional distribution density $q_\phi(\mathbf{z} \mid \mathbf{x})$ of the encoded latent vectors is from the desired Normal distribution $p(\mathbf{z})$. The value of $\mathbb{D}_{kl}$ is zero when two probability distributions are the same, which forces the encoder of VAE $q_\phi(\mathbf{z} \mid \mathbf{x})$ to learn the latent variables that follow a multivariate Normal distribution over a k-dimensional latent space. One of the major advantages of the Variational form compared to the classic version of the AE is achieved using the Kullback-Leibler divergence loss term ($\mathbb{D}_{kl}$), which generates a more continuous and easier to interpolate latent space. Instead of encoding an input as a single point, a Normal distribution is associated with encoding each input instance. This continuity of latent space allows the decoder not only to be able to reproduce an input vector, but also to generate new data from the latent space.

3.2. The generative adversarial networks

The reconstruction loss function $\mathcal{L}_{rec}$ of VAE is not efficient for data generation compared to the adversarial learning technics (Goodfellow et al., 2014). The main idea of GANs is to create an additional NN, called a discriminator $Dis(x)$, that will learn to distinguish between the real data and the data generated by the generator $Gen(z)$. The learning process then consists in successively training the generator to generate new data and the discriminator to dissociate between real and generated data. The learning process converges when the generator reaches the point of luring the discriminator. The discriminator $Dis(x)$ is optimized by maximizing the probability of distinguishing between real and generated data while the generator $Gen(z)$ is trained simultaneously to minimize $log(1 - Dis(Gen(z)))$. Thus, the goal of the whole adversarial training can be summarized as a two-player min-max game with the value function $V(Gen, Dis)$ (Goodfellow et al., 2014):

$$\min_{Gen} \max_{Dis} V(Gen, Dis) = \mathbb{E}_x[\Phi_x] + \mathbb{E}_z[\Psi_z] \qquad (2)$$

where $\Phi_x = log(Dis(x))$ and $\Psi_z = log(1 - Dis(Gen(z)))$. These techniques are much more efficient than the reconstruction loss function $\mathcal{L}_{rec}$ for generating data from a data space, for example, the latent space of VAE. Unlike the latent space of the VAE where the data are structured according to the Kullback-Leibler divergence loss term, the data space generated by GANs is structured in a random way, making it less efficient in managing the data.

4. Experimental results and data analysis

4.1. VAE trained on the feature space

As shown in Figure 2, the first VAE is built from a feature space definition obtained from extraction of expert rules that have been integrated in a function f_{ext}. For further details on f_{ext}, see our previous work (Zemouri et al., 2020). The feature vector $\mathbf{F}$ is obtained by transforming the data space $\mathbb{PD}$ into a feature space $\mathbb{F}$: $\mathbb{PD} \rightarrow \mathbb{F}$, $\mathbf{Pd} \rightarrow \mathbf{F} = f_{ext}(\mathbf{Pd})$, where $\mathbf{F} \in [-1, +1]^{16}$. In order to optimize the VAE learning, the size of the learning database was increased by generating additional points $\{\mathbf{F}_j\}$ within the feature space. The algorithm 1 gives the details of the whole procedure which allows to obtain a sufficiently large and well-balanced learning database in order to consider all possible combinations. The parameter α is used to define the boundary between the PD sources when generating each random variable ε. In this case, $\alpha = 0,5$. A total of 465k points have been generated ($Nb_{max} = 5000$) for learning the VAE^{feat}. The Figure 4 gives the projection of the PD measurement database in a 2D latent space $\mathbb{Z}$ obtained by the encoder Enc^{feat} with $\mathbf{Z} = Enc^{feat}(\mathbf{F})$. It is important to note that this latent space respects a certain consistency in the distribution of data during encoding (see the visual landmarks in Figure 4).

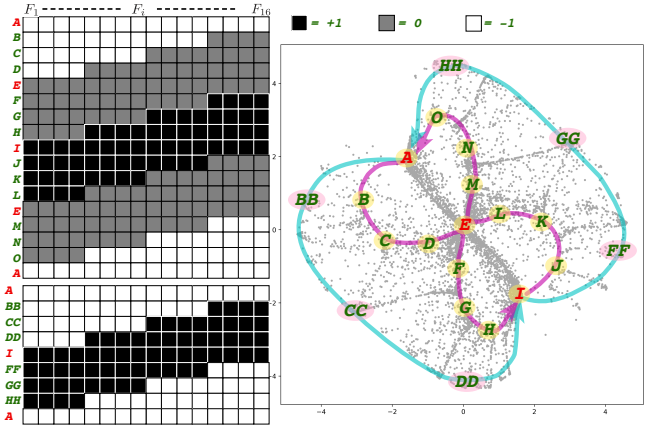

Figure 4. The projection of the PD measurement database in a 2D latent space $\mathbb{Z}$ with the corresponding visual landmarks of the feature vector $\mathbf{F}$.

Algorithm 1: Feature space generation algorithm

Data:
$\varepsilon = Rand(a_1, a_2)$: return a random number $\varepsilon \in [a_1, a_2]$
Nb_{max}: size of the generated data
$\hat{\mathbf{F}}_j(i)$: is the i^{th} feature F_i of the j^{th} generated vector $\hat{\mathbf{F}}_j$
Result: Generate a training dataSet $\{\hat{\mathbf{F}}_j\}$ for the VAE
$j = 0$
while $j < Nb_{max}$ **do**
  $\hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-1, -\alpha)|_{i=1 \textbf{ to } 16}$;
  $\hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-\alpha, +\alpha)|_{i=1 \textbf{ to } 16}$;
  $\hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(+\alpha, +1)|_{i=1 \textbf{ to } 16}$;
  for $k \leftarrow 1$ **to** 15 **do**
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-1, -\alpha)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(-\alpha, +\alpha)|_{i=k+1 \textbf{ to } 16} \end{cases}$
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-1, -\alpha)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(+\alpha, 1)|_{i=k+1 \textbf{ to } 16} \end{cases}$
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-\alpha, +\alpha)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(-1, -\alpha)|_{i=k+1 \textbf{ to } 16} \end{cases}$
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(-\alpha, +\alpha)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(+\alpha, 1)|_{i=k+1 \textbf{ to } 16} \end{cases}$
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(+\alpha, +1)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(-1, -\alpha)|_{i=k+1 \textbf{ to } 16} \end{cases}$
    $\begin{cases} \hat{\mathbf{F}}_{j\leftarrow j+1}(i) = Rand(+\alpha, +1)|_{i=1 \textbf{ to } k} \\ \hat{\mathbf{F}}_j(i) = Rand(-\alpha, +\alpha)|_{i=k+1 \textbf{ to } 16} \end{cases}$

4.2. VAE trained on the data space

In addition to the VAE^{feat}, two other VAEs have been built directly from the PD data space $\mathbb{PD}$. The first one, VAE^{original} is trained on the whole real PD database $\{\mathbf{Pd}_j\}$, while the second one, VAE^{gan} is obtained from a training performed exclusively on data generated by GANs. The algorithm 2 gives the procedure adopted for learning the GANs to generate an artificial PD database $\{\hat{\mathbf{Pd}}_j\}$. Instead of creating a single generator $Gen(z)$ derived from an adversarial learning performed on the whole PD database $\{\mathbf{Pd}_j\}$, several generators $Gen_k(z)$ obtained from subsets have been created. These subsets are extracted from the latent space $\mathbb{Z}$ obtained by the previous model VAE^{feat}. Figure 5 gives the distribution of the areas $\mathcal{A}_k$ which forms these subsets $\{\mathbf{Pd}_j\}_k$. A total of 35 areas were selected to cover the entire latent space. The selection of these areas is crucial as it helps to compensate for the real dataset imbalance. The objective is to cover the entire latent space by considering small areas in order to have a maximum chance of capturing minority PD patterns. Once the learning of the GANs is performed, the obtained k generators $Gen_k(z)$ were exploited to produce several artificial PD datasets $\{\hat{\mathbf{Pd}}_j\}_k$ used to learn the VAE^{gan}. As shown in the table 1, three dataset configurations were tested where the size of the data generated by the generators $Gen_k(z)$ were varied. As an illustration, the Figure 6 shows some examples of PD patterns generated by generators $Gen_k(z)$.

Algorithm 2: Training of the GAN

Data: $\mathcal{A}_k$: selected areas from the feature latent space
Result: $\{Gen_k\}$: set of k generators
for $k \leftarrow 1$ **to** Nb_{Area} **do**
  Step 1: Select the k^{th} area $\mathcal{A}_k$ from the feature latent space (Figure 5) ;
  Step 2: From $\mathcal{A}_k$ extract $\{\mathbf{Z}_j^{feat}\}_k$ where $\mathbf{Z}_j^{feat} \in \mathcal{A}_k$;
  Step 3: From $\{\mathbf{Z}_j^{feat}\}_k$ extract $\{\mathbf{Pd}_j\}_k$ where $\mathbf{Z}_j^{feat} = Enc(f_{ext}(\mathbf{Pd}_j))$;
  Step 4: Train the k^{th} generator Gen_k on $\{\mathbf{Pd}_j\}_k$;

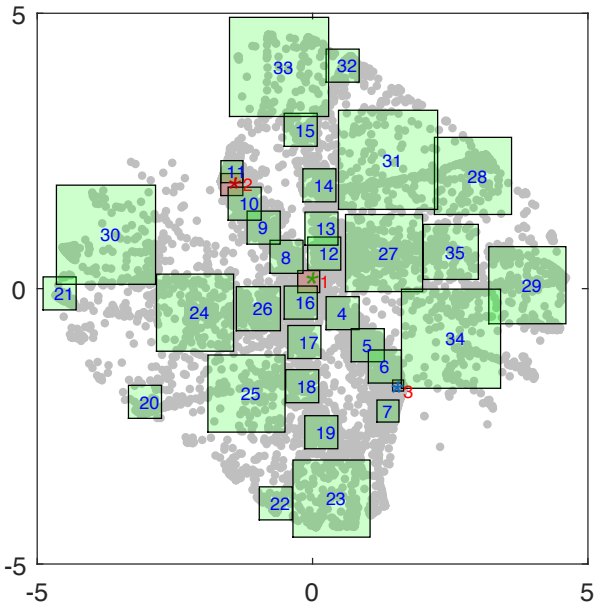

Figure 5. The distribution of the areas $\mathcal{A}_k$ over the 2D latent space $\mathbb{Z}$.

Table 1. Size of the generated data used to learn the VAE^{gan}.

	Size of the generated subset by each generator Gen_k	Total generated data
VAE^{gan1}	1 000	35 000
VAE^{gan2}	10 000	350 000
VAE^{gan3}	100 000	3 500 000

4.3. Neural network structure

The structure of the NNs used is as follows. For the VAE^{feat}, a network with fully connected (FC) layers has been used. The encoder has 5 hidden layers composed successively by 512, 256, 128, 64 and 32 neurons with a hyperbolic tangent activation function. For the VAE^{original} and VAE^{gan} networks, a mixed of convolutional layers (CL) and FC layers

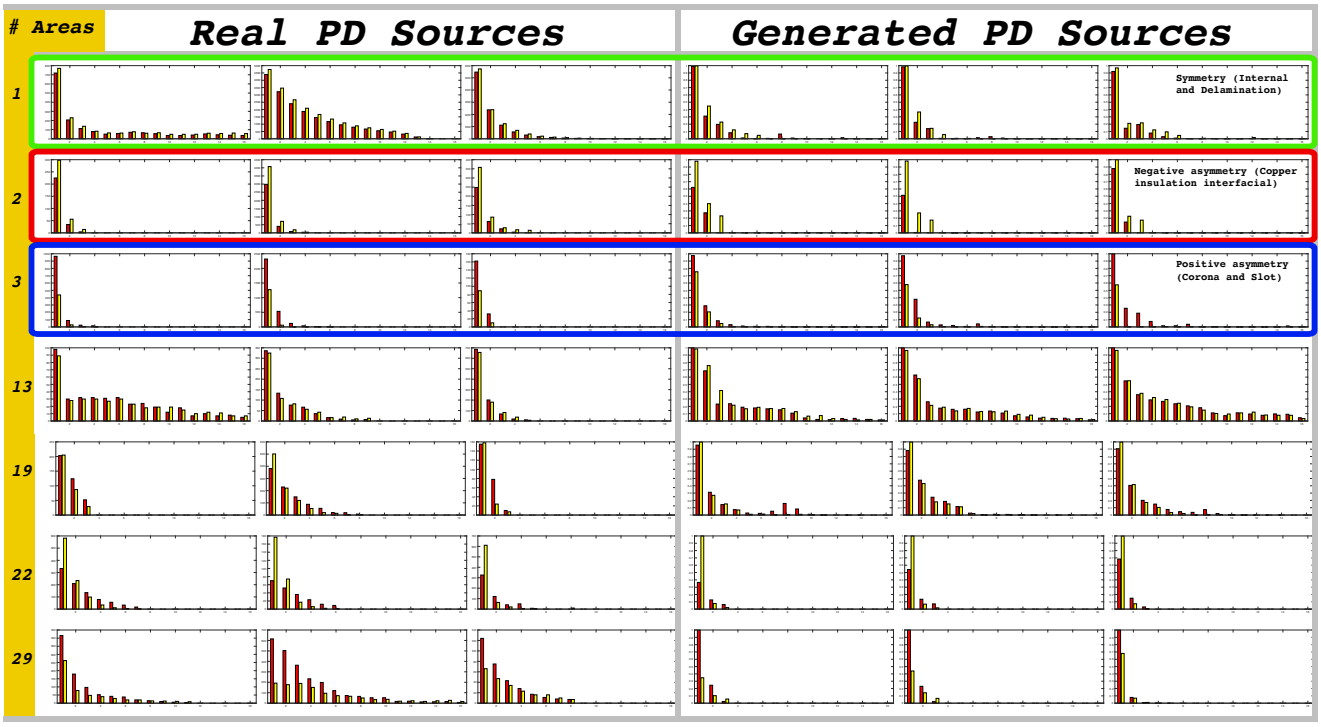

Figure 6. Example of PD generation with the generators $Gen_k(z)$.

was chosen. A 3 CL was constructed in the encoder network with 2×2 kernel and respectively 128, 64 and 32 filters. A residual block is added after each CL and all the residual blocks contain two 2×2 kernel CL with the same number of filters (Zemouri, 2020). Lastly, three FC layers with respectively 1024, 512 and 256 neurons are added. Each of the CL and the FC layers are followed by a LeakyReLU activation layer. The decoders have exactly the same structure as the encoders but inverted.

4.4. Visual data analysis

The three PD sources are recognizable from the features latent space: three areas $\mathcal{A}_k$ with $k = 1, 2, 3$ in red in the Figure 5 and described in table 2. Area 1 corresponds to symmetric PD patterns which is distinguished by an equal distribution between positive and negative PDs, area 2 and 3 correspond respectively to negative and positive asymmetry PD patterns, characterized respectively by a clear superiority of negative / positive PD pulses. These three types of PD patterns are characterized by the feature vectors *E*, *A*, *I* in Figure 4.

Table 2. The three main PD sources.

#	PD source	Feature description
1	Symmetric	$\sum_{i=1}^{16} Pd_i^+ \approx \sum_{i=1}^{16} Pd_i^-$
2	Negative asymmetry	$\sum_{i=1}^{16} Pd_i^+ << \sum_{i=1}^{16} Pd_i^-$
3	Positive asymmetry	$\sum_{i=1}^{16} Pd_i^+ >> \sum_{i=1}^{16} Pd_i^-$

Figure 8 gives the correspondence between the different latent spaces: the corresponding original PDs in the data space $\mathbb{PD}$ were extracted from the areas $\{\mathcal{A}_k\}$ of the Figure 5 and projected into the other latent spaces. For example, the green cluster (symmetric PD patterns) corresponds to the areas $\{\mathcal{A}_k\}$ of the latent space VAE^{feat} with $k = 1, 4, 8$. The original PD data of these three areas $\{\mathcal{A}_k\}$ were identified and then projected into the latent spaces of the VAE^{original} and the three VAE^{gan}. As it can be seen in Figure 8 three main clusters of data points are formed. These clusters correspond respectively to the three main PD sources mentioned above, namely symmetric, negative asymmetry and positive asymmetry. For all of the VAE models, the green cluster representing the PD source of symmetric signals, is positioned between the red and blue clusters representing respectively the negative and positive asymmetry PD sources. By analyzing the original PD sources projected on the different latent space, moving away from the center of the green cluster in the direction of the blue or red one, the more important the asymmetry is. This result proves that the models trained on the data space have well captured the asymmetric relationship that exists between the different PD patterns. The second characteristic that the VAE models captured was the density in each channel of the PD pattern matrix. This characteristic does not appear in the latent space of the VAE^{feat} model because it has not been integrated into the function f_{ext}. It should also be noted that VAE^{gan} models trained exclusively on data generated artificially by the GANs have captured the two characteristics as

well as the VAE^{original} model trained on the real data. This is an important result that validates the quality of the artificial data generation by the GAN. In addition to the three main PD sources, hybrid PD patterns could also be present. These particular PD sources are characterized by the presence of several sources at the same time, for example the presence of positive and negative asymmetries in the same PD pattern. An attempt to integrate this feature into the function f_{ext} was made, but this aspect has yet to be further developed. However, the VAE^{gan} and VAE^{original} models failed to capture this characteristic, as shown in Figure 8. This is due to the low presence of these PD patterns in the overall available data. GANs have not been able to capture these PD patterns as they are certainly mixed with other patterns within the various selected areas $\mathcal{A}_k$. The model failed to disentangle these two characteristics in the latent space. An interesting prospect to explore would be to exploit the recent work on the unsupervised learning of disentangled representations (Locatello et al., 2018) (Eastwood & Williams, 2018) (Chen, Li, Grosse, & Duvenaud, 2018) (Kim & Mnih, 2019) (Duan et al., 2019) (Hristov, Angelov, Burke, Lascarides, & Ramamoorthy, 2019).

The equation 3 gives the Pearson correlation coefficient r_p between the original data space $\mathbb{PD}$ and the target space $\mathbb{Z}$ computed for each of the $\mathcal{A}_k$ areas presented in Figure 5. The Pearson correlation coefficient was used here to quantify the quality of latent spaces. $\overline{\mathbf{Pd}}$ and $\overline{\mathbf{Z}}$ represent respectively the means of each cluster k in respectively the original data space $\mathbb{PD}$ and the target space $\mathbb{Z}$. The Euclidean norm $\|.\|$ is used for the distance calculation. A cross-correlation coefficient is also given between the different $\mathcal{A}_k$ areas. Figure 7 gives all the obtained correlation results: correlation within each area $\mathcal{A}_k$ and cross-correlations between two areas. the VAE^{feat} model has the lowest correlation score r_p. This is certainly due to the expert rules, used for feature extraction, which need to be further optimized.

$$r_p = \frac{\sum_{i=1}^{N} \|\mathbf{Pd}_i - \overline{\mathbf{Pd}}\|.\|\mathbf{Z}_i - \overline{\mathbf{Z}}\|}{\sqrt{\sum_{i=1}^{N} \|\mathbf{Pd}_i - \overline{\mathbf{Pd}}\|^2}.\sqrt{\sum_{i=1}^{N} \|\mathbf{Z}_i - \overline{\mathbf{Z}}\|^2}} \tag{3}$$

5. Conclusion and future works

In this paper, the latent space properties of the Variational Autoencoder for the Partial Discharge analysis have been exploited. The quality of the low dimensional latent space obtained from the expert rules was compared with a latent space obtained directly from the input signal. For this purpose, the Generative Adversarial Networks were used to artificially enhance the learning database of PDs. An important result is that the latent space of the VAEs trained exclusively on

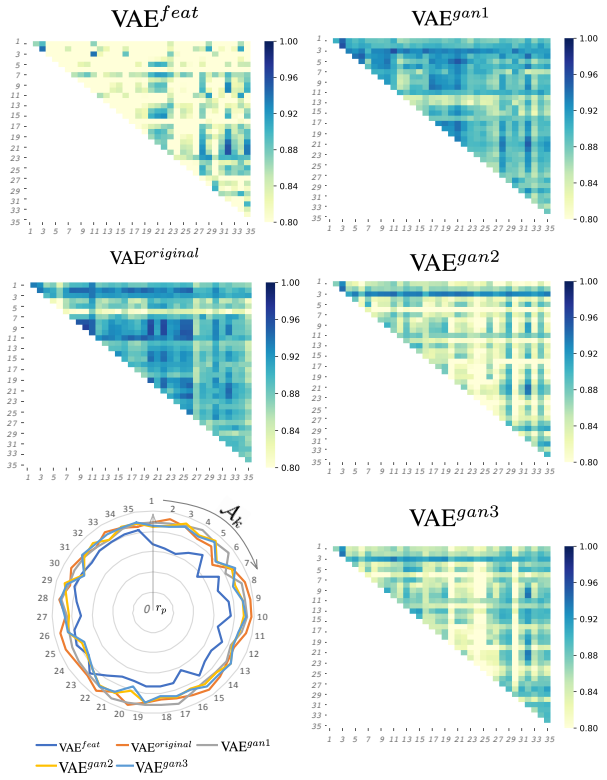

Figure 7. All the obtained correlation results: correlation within each area $\mathcal{A}_k$ and cross-correlations between two areas given by the heatmaps.

generated data has captured the same characteristics as the VAE trained on real data. This first level validation allows us to approve the quality of the data artificially generated by the GANs. In future works, focus will be made on isolating underrepresented PD patterns such as hybrid PD sources in order to increase their significance by the GANs. Furthermore, the final objective of this study is to automatically recognize each individual PD source. A reference PD pattern set will be developed and labeled by experts. The encoder part would then be kept to extract characteristic features of each PD sources, used as input for a classifier. As in other feature representation learning papers (Franceschi, Dieuleveut, & Jaggi, 2019), a classifier will be trained on top of the different feature spaces to showcase the superiority of certain spaces over the others. Afterwards, several degradation phenomena leading to failures will be analyzed in the latent space. The objective would then be to obtain a particular degradation pattern in the form of a trajectory on the latent space that will characterize each degradation phenomenon.

References

Bao, W., Miao, X., Wang, H., Yang, G., & Zhang, H. (2020). Remaining useful life assessment of slewing bearing

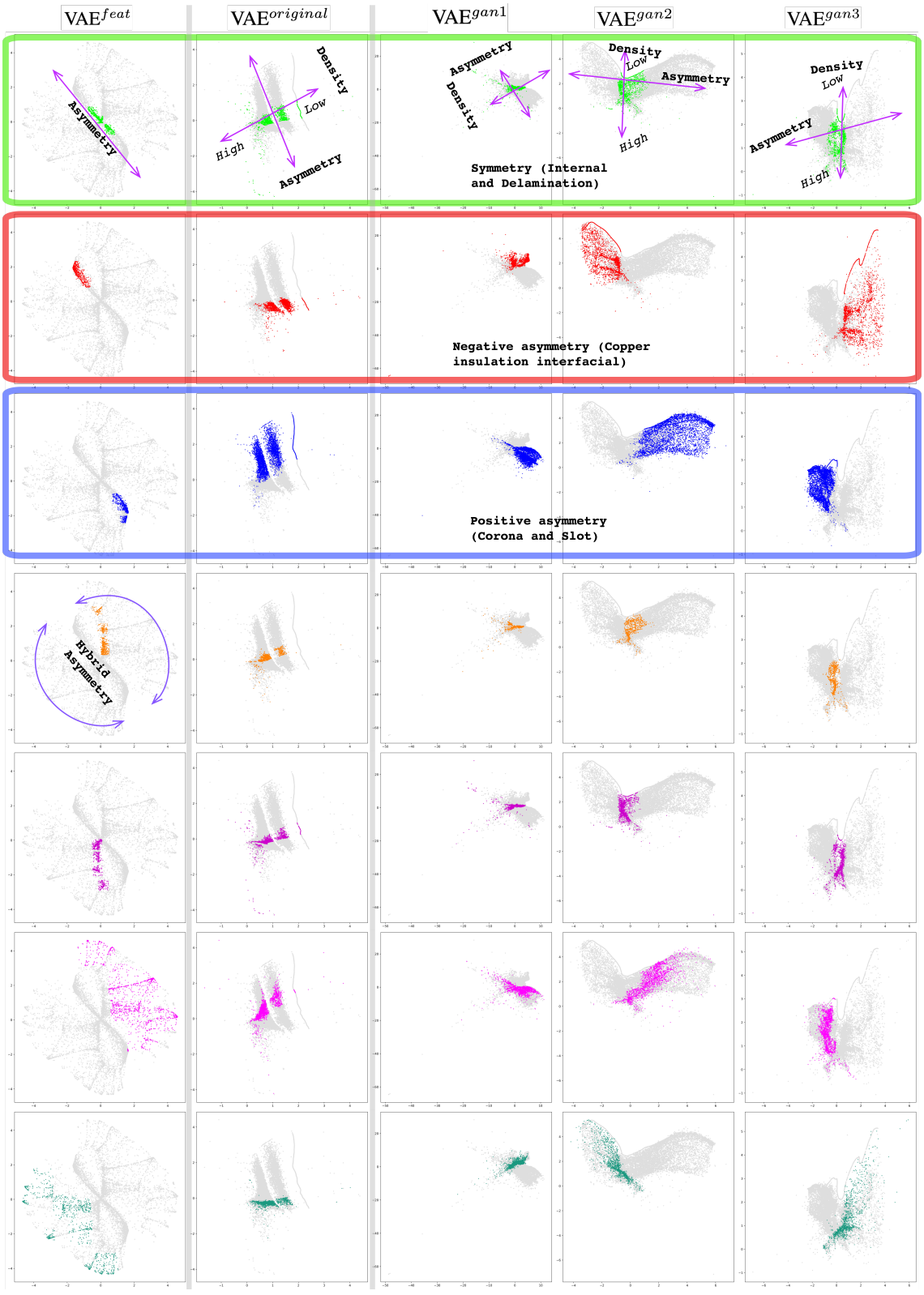

Figure 8. The correspondence between the different latent spaces. Figure 5 and 6 can be used to better understand the relationship between each area.

based on spatial-temporal sequence. *IEEE Access*, *8*, 9739-9750. doi: 10.1109/ACCESS.2020.2965285

Chen, T. Q., Li, X., Grosse, R. B., & Duvenaud, D. (2018). Isolating sources of disentanglement in variational autoencoders. *CoRR*, *abs/1802.04942*.

Dai, J., Wang, J., Huang, W., Shi, J., & Zhu, Z. (2020, Oct). Machinery health monitoring based on unsupervised feature learning via generative adversarial networks. *IEEE/ASME Transactions on Mechatronics*, *25*(5), 2252-2263. doi: 10.1109/TMECH.2020.3012179

Doulamis, A. D., Hou, G., Xu, S., Zhou, N., Yang, L., & Fu, Q. (2020). Remaining useful life estimation using deep convolutional generative adversarial networks based on an autoencoder scheme. *Computational Intelligence and Neuroscience*, *2020*, 9601389. doi: 10.1155/2020/9601389

Duan, S., Watters, N., Matthey, L., Burgess, C. P., Lerchner, A., & Higgins, I. (2019). A heuristic for unsupervised model selection for variational disentangled representation learning. *CoRR*, *abs/1905.12614*.

Ducoffe, M., Haloui, I., & Gupta, J. S. (2019). Anomaly detection on times series with wasserstein gan applied to phm. *International Journal of Prognostics and Health Management: Special Issue on Deep Learning and Emerging Analytics*, *10*(25), 12.

Eastwood, C., & Williams, C. K. I. (2018). A framework for the quantitative evaluation of disentangled representations. In *International conference on learning representations.*

Farajzadeh-Zanjani, M., Hallaji, E., Razavi-Far, R., Saif, M., & Parvania, M. (2021). Adversarial semi-supervised learning for diagnosing faults and attacks in power grids. *IEEE Transactions on Smart Grid*, 1-1. doi: 10.1109/TSG.2021.3061395

Franceschi, J., Dieuleveut, A., & Jaggi, M. (2019). Unsupervised scalable representation learning for multivariate time series. *CoRR*, *abs/1901.10738*.

Gao, X., Deng, F., & Yue, X. (2020). Data augmentation in fault diagnosis based on the wasserstein generative adversarial network with gradient penalty. *Neurocomputing*, *396*, 487 - 494. doi: https://doi.org/10.1016/j.neucom.2018.10.109

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., … Bengio, Y. (2014). Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems 27* (pp. 2672–2680). Curran Associates, Inc.

Han, T., Liu, C., Yang, W., & Jiang, D. (2019). A novel adversarial learning framework in deep convolutional neural network for intelligent diagnosis of mechanical faults. *Knowledge-Based Systems*, *165*, 474 - 487. doi: https://doi.org/10.1016/j.knosys.2018.12.019

Hristov, Y., Angelov, D., Burke, M., Lascarides, A., & Ramamoorthy, S. (2019). Disentangled relational representations for explaining and learning from demonstration. *CoRR*, *abs/1907.13627*.

Huang, Y., Chen, C., & Huang, C. (2019). Motor fault detection and feature extraction using rnn-based variational autoencoder. *IEEE Access*, *7*, 139086-139096. doi: 10.1109/ACCESS.2019.2940769

Huang, Y., Tang, Y., VanZwieten, J., Liu, J., & Xiao, X. (2019, May). An adversarial learning approach for machine prognostic health management. In *2019 international conference on high performance big data and intelligent systems (hpbd is)* (p. 163-168). doi: 10.1109/HPBDIS.2019.8735480

Khan, S. A., Prosvirin, A. E., & Kim, J. (2018, Feb). Towards bearing health prognosis using generative adversarial networks: Modeling bearing degradation. In *2018 international conference on advancements in computational sciences (icacs)* (p. 1-6). doi: 10.1109/ICACS.2018.8333495

Kim, H., & Mnih, A. (2019). Disentangling by factorising. *arXiv*.

Kingma, D. (2017). *Variational inference & deep learning: A new synthesis* (Doctoral dissertation, Faculty of Science (FNWI), Informatics Institute (IVI), University of Amsterdam).

Lee, S., Kwak, M., Tsui, K.-L., & Kim, S. B. (2019). Process monitoring using variational autoencoder for high-dimensional nonlinear processes. *Engineering Applications of Artificial Intelligence*, *83*, 13 - 27. doi: https://doi.org/10.1016/j.engappai.2019.04.013

Li, X., Zhang, W., Ma, H., Luo, Z., & Li, X. (2020). Data alignments in machinery remaining useful life prediction using deep adversarial neural networks. *Knowledge-Based Systems*, *197*, 105843. doi: https://doi.org/10.1016/j.knosys.2020.105843

Liu, H., Zhou, J., Xu, Y., Zheng, Y., Peng, X., & Jiang, W. (2018). Unsupervised fault diagnosis of rolling bearings using a deep neural network based on generative adversarial networks. *Neurocomputing*, *315*, 412 - 424. doi: https://doi.org/10.1016/j.neucom.2018.07.034

Locatello, F., Bauer, S., Lucic, M., Gelly, S., Schölkopf, B., & Bachem, O. (2018). Challenging common assumptions in the unsupervised learning of disentangled representations. *CoRR*, *abs/1811.12359*.

Mao, W., Liu, Y., Ding, L., & Li, Y. (2019). Imbalanced fault diagnosis of rolling bearing based on generative adversarial network: A comparative study. *IEEE Access*, *7*, 9515-9530. doi: 10.1109/ACCESS.2018.2890693

Martin, G. S., Droguett, E. L., Meruane, V., & das Chagas Moura, M. (2018). Deep variational autoencoders: A promising tool for dimensionality reduction and ball bearing elements fault diagnosis. *Structural Health Monitoring*, *0*(0), 1475921718788299. doi: 10.1177/1475921718788299

Mullick, S. S., Datta, S., & Das, S. (2019). Generative adversarial minority oversampling. *CoRR*, *abs/1903.09730*.

Pan, T., Chen, J., Xie, J., Chang, Y., & Zhou, Z. (2020). Intelligent fault identification for industrial automation system via multi-scale convolutional generative adversarial network with partially labeled samples. *ISA Transactions*, *101*, 379 - 389. doi: https://doi.org/10.1016/j.isatra.2020.01.014

Proteau, A., Zemouri, R., Tahan, A., & Thomas, M. (2020). Dimension reduction and 2d-visualization for early change of state detection in a machining process with a variational autoencoder approach. *The International Journal of Advanced Manufacturing Technology*, *111*(11), 3597–3611.

Que, Z. J., Xiong, Y., & Xu, Z. G. (2019, Dec). A semi-supervised approach for steam turbine health prognostics based on gan and pf. In *2019 ieee international conference on industrial engineering and engineering management (ieem)* (p. 1476-1480). doi: 10.1109/IEEM44572.2019.8978717

Shao, S., Wang, P., & Yan, R. (2019). Generative adversarial networks for data augmentation in machine fault diagnosis. *Computers in Industry*, *106*, 85 - 93. doi: https://doi.org/10.1016/j.compind.2019.01.001

Wang, X., Huang, H., Hu, Y., & Yang, Y. (2018, Sep.). Partial discharge pattern recognition with data augmentation based on generative adversarial networks. In *International conference on condition monitoring and diagnosis, cmd* (p. 1-4). doi: 10.1109/CMD.2018.8535718

Wang, Z., Wang, J., & Wang, Y. (2018). An intelligent diagnosis scheme based on generative adversarial learning deep neural networks and its application to planetary gearbox fault pattern recognition. *Neurocomputing*, *310*, 213 - 222. doi: https://doi.org/10.1016/j.neucom.2018.05.024

Zemouri, R. (2020). Semi-supervised adversarial variational autoencoder. *Machine Learning and Knowledge Extraction (MAKE)*, *2*(3), 361-378. doi: https://doi.org/10.3390/make2030020

Zemouri, R., Lévesque, M., Amyot, N., Hudon, C., Kokoko, O., & Tahan, S. A. (2020). Deep convolutional variational autoencoder as a 2d-visualization tool for partial discharge source classification in hydrogenerators. *IEEE Access*, *8*, 5438-5454. doi: 10.1109/ACCESS.2019.2962775

Zhang, W., Li, X., Jia, X.-D., Ma, H., Luo, Z., & Li, X. (2020). Machinery fault diagnosis with imbalanced data using deep generative adversarial networks. *Measurement*, *152*, 107377. doi: https://doi.org/10.1016/j.measurement.2019.107377

Zheng, S., & Gupta, C. (2020, May). Discriminant generative adversarial networks with its application to equipment health classification. In *Icassp 2020 - 2020 ieee international conference on acoustics, speech and signal processing (icassp)* (p. 3067-3071). doi: 10.1109/ICASSP40776.2020.9053475

Zhou, F., Yang, S., Fujita, H., Chen, D., & Wen, C. (2020). Deep learning fault diagnosis method based on global optimization gan for unbalanced data. *Knowledge-Based Systems*, *187*, 104837. doi: https://doi.org/10.1016/j.knosys.2019.07.008

Zou, L., Li, Y., & Xu, F. (2020). An adversarial denoising convolutional neural network for fault diagnosis of rotating machinery under noisy environment and limited sample size case. *Neurocomputing*, *407*, 105 - 120. doi: https://doi.org/10.1016/j.neucom.2020.04.074