

# Attention gated encoder-decoder for ultrasonic signal denoising

Nabil Jai Mansouri<sup>1,2</sup>, Ghizlane Khaissidi<sup>2</sup>, Gilles Despaux<sup>1</sup>, Mostafa Mrabti<sup>2</sup>, Emmanuel Le Clézio<sup>1</sup>

<sup>1</sup>Laboratory Institute of Electronics and Systems (IES), University of Montpellier,  
French National Centre for Scientific Research (CNRS), Montpellier, France

<sup>2</sup>Laboratory of Computer and Inter-disciplinary Physics (LIP), Ecole Normale Supérieure (ENS),  
Sidi Mohamed Ben Abdellah University, Fez, Morocco

---

## Article Info

### Article history:

Received Jun 24, 2022

Revised Jan 18, 2023

Accepted Jan 30, 2023

---

### Keywords:

Attention

Deep learning

Noise reduction

Ultrasonic signal

---

## ABSTRACT

Ultrasound imaging is one of the most widely used non-destructive testing methods. The transducer emits pulses that travel through the imaged samples and are reflected by echo-forming impedance. The resulting ultrasonic signals usually contain noise. Most of the traditional noise reduction algorithms require high skills and prior knowledge of noise distribution, which has a crucial impact on their performances. As a result, these methods generally yield a loss of information, significantly influencing the final data and deeply limiting both sensitivity and resolution of imaging devices in medical and industrial applications. In the present study, a denoising method based on an attention-gated convolutional autoencoder is proposed to fill this gap. To evaluate its performance, the suggested protocol is compared to widely used methods such as butterworth filtering (BF), discrete wavelet transforms (DWT), principal component analysis (PCA), and convolutional autoencoder (CAE) methods. Results proved that better denoising can be achieved especially when the original signal-to-noise ratio (SNR) is very low and the sound waves' traces are distorted by noise. Moreover, the initial SNR was improved by up to 30 dB and the resulting Pearson correlation coefficient was maintained over 99% even for ultrasonic signals with poor initial SNR.

*This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.*



---

## Corresponding Author:

Nabil Jai Mansouri

Institute of Electronics and Systems, University of Montpellier

Montpellier, France

Email: nabil.jai-mansouri@umontpellier.fr, nabil.jaimansouri@usmba.ac.ma

---

## 1. INTRODUCTION

One of the most used non-destructive control methods is ultrasound imaging. It is applied for medical purposes, as it allows the acquisition of images of internal organs, that help diagnose pain causes [1], cancers [2], and fetal assessments [3]. Its application extends as well to the industrial domain where it was deployed in the fault diagnosis of rolling element bearings [4] or for non-destructive testing of nuclear reactors [5]. The main idea of this method aims to emit ultrasounds, using a transducer, that will penetrate materials and be reflected on the different layers of the imaged sample. The reflected power or the corresponding time of flight (ToF) of the ultrasonic signals will be used to quantify the grayscale of each pixel in the final image. The ToF can be estimated using different methods such as Threshold, Akaike information criterion method, and Cross-correlation [6]. Unfortunately, these methods' performance can be heavily degraded due to poor signal-to-noise ratio (SNR), hence the necessity of an efficient noise reduction. Because of the noise randomness, noise reduction becomes a challenging task requiring high skills, knowledge of the signal properties, and advanced denoising algorithms. At present moment, the methods used for signal denoising can be classified into two clusters: classical signal processing methods based on mathematical models, and learning methods. The performances of these methods depend on the complexity of the noise to

be suppressed. Traditional frequency filtering methods are recommended when the noise does not share the same frequency properties as the noise-free signal. These methods permit the selection of the informative frequency range of the signal. But in certain cases, the noise spectrum overlaps with the spectrum of the clean signal, and the classical filtering methods' results are no longer in perfect adequation. For this reason, wavelet filtering methods are recommended. These methods reduce noise on signal following three steps. First, the signal is transformed to the wavelet domain using Mallat's algorithm [7], resulting in a set of approximation and detail coefficients. These latter are then thresholded as they refer to high-frequency terms (noise), contrary to the approximation coefficients that identify the relevant low-frequency information. Last, a reconstruction of the signal is performed leveraging the thresholded detail and the approximation coefficients. From the widely used wavelet-based filtering methods, we cite the discrete wavelet transform (DWT), the stationary wavelet transform (SWT), and the wavelet packets (WP) [8]. Matz *et al.* [9] compared these methods and proved the WP's greatness. The main limitation of wavelet-based denoising methods is their requirement of an adequate choice of the basic function, threshold method as well as a precise estimation of the threshold value. Therefore, learning methods were proposed to overtake the necessity of these signal processing skills and to perform a greater noise reduction.

The learning methods are categorized into two sub-fields, the supervised ones that learn during the training to map from an input sample to a corresponding output, and the unsupervised algorithms that aim to blindly extract hidden features yielding sufficient details to perform the clustering, dimensionality reduction, or denoising. The Machine Learning algorithms as principal components analysis (PCA), independent component analysis (ICA), and singular value decomposition (SVD), were combined by researchers with traditional signal processing methods for better noise reduction [10]. Lately, deep learning (DL) has gained researchers' attention in a wide variety of fields, in parallel with the improvement of computing powers [11]. In signal and image processing, several researchers used DL algorithms for denoising applications [12]–[15]. Precisely, Gao *et al.* [16] designed a reversible mapping algorithm between a two-dimensional visual image and a one-dimensional ultrasonic signal. They built an autoencoder able to extract complex features, and perform the denoising of the ultrasonic signal. Their method showed more adaptability and robustness than PCA, SVD, and wavelet algorithms. Xu *et al.* [17] removed grain noise by clustering the correlative signals using the K-means algorithm. They trained autoencoders on the different configurations, and leveraged the trained models to perform the noise reduction. Contrariwise, Antczak [18] proposed deep recurrent neural networks (RNN) to denoise Electrocardiographic signals. The model was trained on simulated data, and fine-tuned on real data. The results showed better performances when compared to undecimated wavelet transform and bandpass filtering. In addition to that, the results proved that pretraining on synthetic data before fine-tuning on real ones improved the DL model performances.

RNNs such as gated recurrent unit (GRU) or long short-term memory (LSTM) were proposed to deal with sequential data [19]. This type of neural networks was employed for several tasks as time-series' forecasting [20], text classification [21], speech translation [22], and many more natural language processing (NLP) applications. The main challenge in training RNN is the vanishing gradient problem, which tends to penalize the network performances in extracting relevant information from data. Alongside, convolutional neural networks (CNN) have known promising improvements like the ResNet architecture [23] that tackles vanishing gradient phenomena through skip connections and ended up outperforming RNNs. Thus, CNN offered higher performances and attracted researchers for sequential problem modeling. Song *et al.* [24] proposed an unsupervised multispectral denoising method applied to satellite imagery using a wavelet sub-band cycle-consistent adversarial network. Results proved a particularity of preserving high-frequency information, representing edges in the case of satellite images. In the meanwhile, it removed successfully the noise patterns. Sharma and Pramanik [25] proposed a U-net-based DL model to reduce noise and enhance the resolution in acoustic resolution photoacoustic microscopy. The model performance has been validated on in vivo rat vasculature imaging. Furthermore, Dang *et al.* [26] designed a dual-path-transformer-based full-band and sub-band fusion network for speech enhancement purposes. The method is based on an encoder of a Transformer model. This submodel is formed by a positional encoding layer, multi-head attention, and fully connected layers. Their method outmatched the state-of-the-art methods on the voice cloning toolkit (VCTK), diverse environments multichannel acoustic noise database (DEMAND), and deep noise suppression (DNS) benchmarking datasets. However, the drawback of the Transformer based architectures is the need for very large datasets limiting their application to several domains.

In this work, an attention u-shaped convolutional autoencoder (Att-CAE) for ultrasonic signal denoising is proposed. The neural network was trained on a wide variety of synthetic ultrasonic signals offering the possibility to learn a mapping from noisy signals to completely noise-free ones. Moreover, it enables highlighting the proposed method's performance in the most critical conditions. The proposed model was compared to DWT, butterworth filtering (BF), PCA, and convolutional autoencoders (CAE) methods.

2. METHODOLOGY

2.1. Denoising attention convolutional autoencoder

To perform the denoising of the ultrasonic signals, an attention U-shaped convolutional autoencoder was proposed to learn the signal features, suppress the noise, and provide the corresponding denoised signal. Firstly, the autoencoder is an unsupervised neural network that learns through an encoding-decoding process. It compresses the input data into a reduced space, called a latent-space representation, which is then leveraged to reconstruct the input along with the decoding operation. Vanilla autoencoders are artificial neural networks (ANN). Due to the weak feature extraction performed by the ANN, and thanks to CNN's ability to learn the spatial information, a transition to CAE was proposed, where the ANNs of the vanilla autoencoder are replaced by CNNs. The output of each convolutional layer is expressed in (1).

$$a_i = \sigma(F_i * a_{i-1} + b_i) \tag{1}$$

Where  $a_{i-1}$  is the activation of the previous layer,  $F_i$  and  $b_i$  are the filters weights and biases from the previous layer to the current one and  $\times$  denotes the convolution [27]. The main limitation of the convolutional autoencoders is that the model could potentially learn to simply copy its inputs, and poorly perform on other signal distributions. On that ground, denoising autoencoders were proposed to force the encoder to learn complex features of its inputs in a noisy environment and end-up in providing the decoder with sufficient useful information for optimal reconstruction along with noise suppression.

In the proposed model, the rectified linear unit activation function (ReLU) has been used, except for the last layer where it was replaced by the hyperbolic tangent activation function (tanh). The *tanh* was required to reconstruct the dynamic range of the denoised signal. The ReLU activation function is employed as it prevents the vanishing gradient problem thanks to its derivative, allowing the autoencoder to learn faster, perform better, and generalize well. Another benefit of using the ReLU activation function is pushing the latent-space representation units to zero, enabling an indirect control of the average number of zeros in the latent space, resulting in the representation sparsity [28]. Inspired by the U-Net architecture [29], the model was designed very carefully in a manner to have a symmetrically shaped encoder and decoder, allowing the cross-connection between their same-sized layers [30]. These connections are concatenations of the outputs of the same shaped layers from the encoder and the decoder. They ensure the reusability of features lost along the encoding process in the decoding operation, allowing the final computations to be aware of the small and basic details of the input signal. These cross-connections permit the gradient injection in the top layers as well, helping them decide in which direction the weights should be moved to minimize the cost function. Thus, the top layers, performing the basic feature extractions, surpass the vanishing gradient phenomena and well adjust their filters' parameters. Furthermore, it has been confirmed that the introduction of skip connections affects the loss landscape offering the possibility to minimize highly non-convex loss functions [31]. The main limitation of these connections is that basic features extracted at the encoder's top layers are processed equally to deeper features from the decoder. To tackle this issue, an attention mechanism was proposed for features weighting before concatenation, discriminating the relevant information to the task learned by the DL model [32]–[34]. These mechanisms filter the neurons' activations, whether during the forward or the backward pass. During the backpropagation, the gradients emerging from the noisy background are shrunk so that the model parameters in the top layers get updated based on pertinent information to the denoising task. The used attention mechanism architecture is shown in Figure 1. The use of the attention mechanism at the level of the skip connections helps improve the DL model noise reduction using shallower models. Considering overfitting phenomena, dropout was used as a regularization method [35].

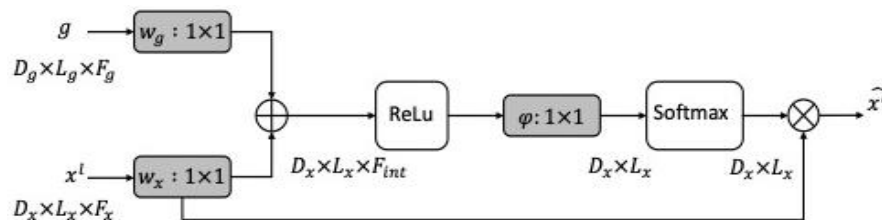


Figure 1. Architecture of the employed additive attention mechanism. Each decoder layer's output is scaled using coefficients learned from the same decoder layer's activations, representing the input features ( $x^l$ ), and the gating signals ( $g$ ) originating from the same-sized encoder's layers [32]

Another important factor in the Att-CAE training is the optimization algorithm. In this study, the results of the most popular gradient-based algorithms were compared: Firstly, the stochastic-gradient descent algorithm (SGD) [36], as an accelerated schemes method, and then two common adaptative methods called adaptative momentum estimation (Adam) [37], and AdaBelief [38]. The Adam optimization algorithm was selected to optimize the model parameters. The adopted algorithm showed better denoising results and converged slightly faster than Adabelief. This optimization has been performed by maximizing the peak signal-to-noise ratio (PSNR) as a loss function. The latter has been chosen as it prevents the DL model from over-smoothing the dynamic range of the signal since the denoised signals should conserve the echoes considered as the most relevant segments of the ultrasonic signal. The architecture of the DL model is shown in Figure 2.

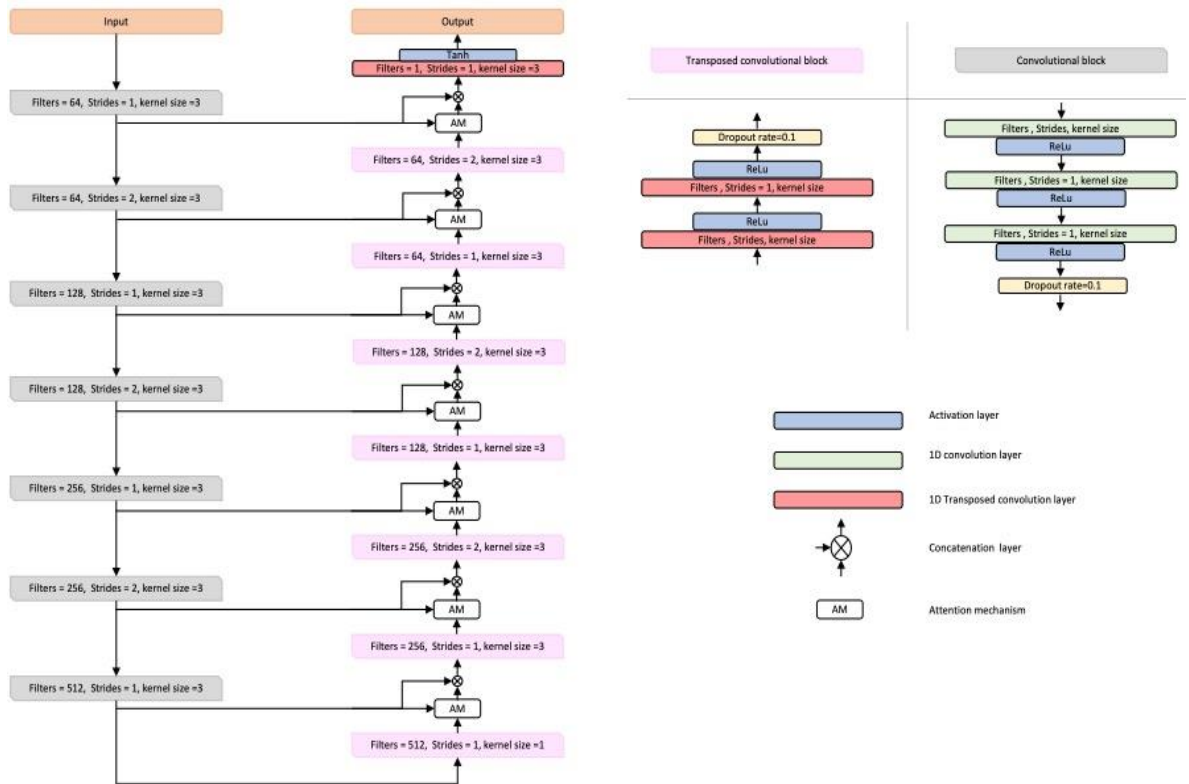


Figure 2. The proposed Att-CAE architecture. The U-shaped model compresses the noisy signals along with the encoding branch in silver, and then decompressed in the decoding violet branch. The skip connections concatenate weighted features learned by the encoder’s blocks to the same-size decoder blocks’ outputs. The weighting is performed by the Attention mechanism presented in Figure 1

**2.2. Data generation**

Deep neural networks (DNN) require a very large amount of data to learn. As it was proven that pretraining on synthetic data before fine-tuning on experimental one’s results in better performances [39], two datasets were developed to train and validate the proposed denoising DL model. As most of the present ultrasonic imaging devices digitalize the signals through an analog to digital converter (ADC) [40], these two sets are composed of 100,000 ultrasonic noise-free signals of 1,000 data points Figure 3(a). They contain a 3 cycles pulse with a 10 MHz center frequency transducer and its first echo. The original pulse acting as a reference signal is positioned at 0.25 ms and shifted by an  $\epsilon$  ranging between  $-\Delta t/2$  and  $+\Delta t/2$ , where  $\Delta t$  is the signal temporal sampling step. The  $\epsilon$  is made to represent a non-synchronization between the excitation trigger and the signal sampling at the level of the signal generation. Echoes are then arbitrarily shifted between  $200\Delta t + \epsilon$  and  $500\Delta t + \epsilon$ . In this simulation,  $\Delta t$  is equal to 5 ns, corresponding to a 0.2 GHz sampling frequency, and the propagation times range between 1  $\mu$ s and 2.5  $\mu$ s.

For the first training database, gaussian white noise (GWN) was added at 10 dB SNR. Considering the second training database, GWN noise was added to the clean signals, but at different intensities ranging from 10 dB to 50 dB Figure 3(b), in a manner to have a uniform SNR distribution. The distributions of the

databases are explained by the fact that during our experimental study, it has been observed that training the model at first on the 10 dB SNR database and then on the second database results in better noise reduction on poor SNR signals. Providing the model at first with low SNR signals to denoise forces the learning of features.

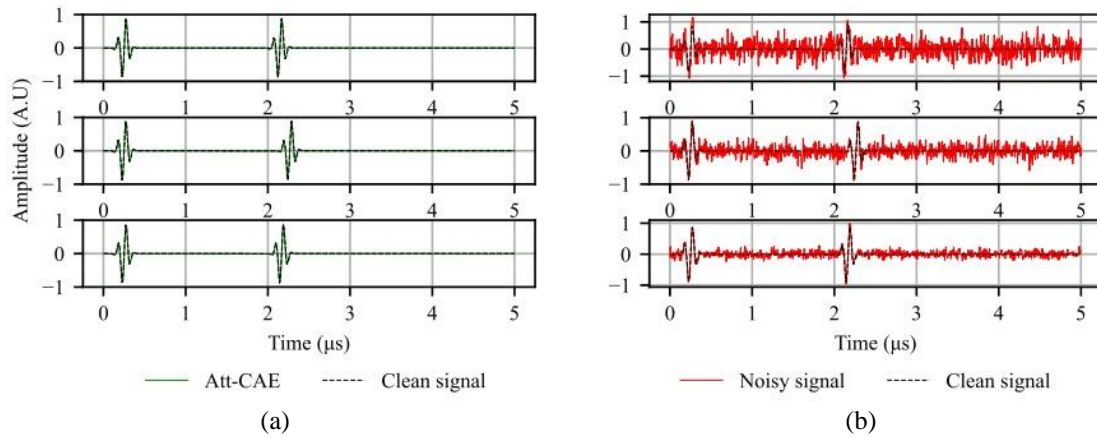


Figure 3. Comparison between, (a) denoised signals in green and (b) highly noisy signals in red at 10 dB, 15 dB and 20 dB SNR for (1), (2), and (3) respectively, and the initial clean signals in black

In more complex conditions. Afterward, training on the second database permits the generalization to other noise intensities. The noise was added as expressed in (2).

$$y_i = x_i + n_i \tag{2}$$

Where  $x_i$  represents the clean signal,  $n$  is a GWN and  $y_i$  the resulting noisy signal from the addition of the noise to the clean signal. The noise amplitude added to the clean signals is calculated in (3).

$$A_{noise} = A_{signal} \cdot 10^{\frac{-SNR_{dB}}{20}} \tag{3}$$

### 3. RESULTS AND DISCUSSION

#### 3.1. Noise reduction

In order to validate our Att-CAE, we created 9 databases similar to the initial ones as test sets containing 100 noise-free signals  $x_i$  along with their corresponding noisy versions  $y_i$ . These noisy signals' SNR range between 10 dB and 50 dB with a step of 5, denoted from database 1 to 9 respectively. These databases' noisy signals were denoised by means of the Att-CAE, CAE, Machine Learning methods such as PCA, and classical signal processing methods such as BF and DWT. We then analyzed the denoising performance of each algorithm comparing the SNR enhancement and the Pearson correlation coefficient (P'r) by (4) and (5).

$$SNR = 10 \log_{10} \left( \frac{P_s}{\sigma_{noise}^2} \right) \tag{4}$$

Where  $P_s$  refers to the clean signal power, and  $\sigma_{noise}$  donates the noise standard deviation.

$$P'r = \frac{\sum_{i=0}^n ((x_i - \bar{x})(y_i - \bar{y}))}{\sqrt{\sum_{i=0}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^n (y_i - \bar{y})^2}} \tag{5}$$

Among them,  $\bar{x}$  and  $\bar{y}$  are the mean values of  $(x_i)$  and  $(y_i)$  respectively, and  $n$  is the length of the signal. SNR donates the ratio between the power of the ultrasonic signal and the corrupting noise power. The higher the SNR value, the better the noise reduction. The P'r donates the linear relationship between the clean signal and the denoised one. The coefficient ranges from -1 to 1, for a negative correlation and a positive correlation, respectively. The closer the value to 1, the greater the correlation between the clean signal and the de-noised one.

To evaluate the method's effectiveness, we tuned the parameters of the wavelet algorithm, the results of which will be compared to the Att-CAE results. The wavelet algorithm decomposes the signals on a basis of functions yielding coefficients corresponding to the local signal similarity to these basis elements [41]. These coefficients will then be filtered using the Bayes Shrink algorithm where each wavelet subband is thresholded and reconstructed to provide the denoised signal [42]. The choice of the basis function is made in a manner to maximize the SNR results on each evaluation database. It is also conventional to choose a mother wavelet that correlates to the signal [43]. For the BF method, we designed a low pass filter that cuts off all the frequencies superior to 25 MHz, and the filter order was set to 5. PCA used a Bayesian model to automatically choose the components that should be retained. To allow a precise comparison between our Att-CAE model and existing DL denoising protocols, CAEs were trained on the same data distribution and include the same methodology and hyperparameters as our model.

To compare the proposed method results to the other methods, we denoised the evaluation databases' samples using our Att-CAE, and compared our results with those from CAE, PCA, DWT, and the BF. The SNR and P'r means of each database's denoised signals were also calculated. Figure 4 presents the comparison of the SNR distributions before and after denoising. It shows that the method proposed in this paper demonstrates a  $\approx 30$  dB improvement on signals with SNR ranging from 10 dB to 20 dB. The second-best method is CAE performing a 2 dB weaker SNR enhancement. Then PCA performs a 12 dB increase slightly better than DWT and BF methods with SNR improvements limited to  $\approx 10$  dB and  $\approx 6$  dB respectively. For the signals with initial SNR ranging between 25 dB and 40 dB, the proposed method raised the SNR values by more than 25 dB outmatching the CAE by more than 4 dB. The wavelet method and BF were limited to 7 dB increases, moderately surpassed by the PCA method which achieved a 9 dB SNR improvement. For SNR values ranging from 45 dB to 50 dB corresponding to low noise levels, our denoising method performed  $\approx 15$  dB enhancement, outperforming all the other methods limited to an 8 dB upgrade. Furthermore, the proposed method achieved better SNR enhancements than the method proposed by Gao *et al.* [16] and Xu *et al.* [17] limited to 6 dB and 18 dB SNR increases, respectively. Moreover, Sun and Lu [44] improved a wavelet threshold processing function for noise reduction on ultrasonic signals. Their work enabled the detection of defect echoes lost by the traditional thresholding functions. However, their SNR improvement was limited to 6 dB compared to the method proposed in this paper.

Thereafter, the P'r is analyzed in Figure 5 as an accurate comparison metric linked to the covariance computed between the denoised sample and its corresponding noise-free signal. For SNR higher than 30 dB, all the methods have similar P'r coefficients. However, for higher noise powers, P'r coefficients of the PCA, DWT, and BF methods undergo dramatic decreases. Meanwhile, the Att-CAE proposed in this paper alongside a slightly lower CAE method resulted in considerably higher values ( $\approx 1$ ). This metric then confirms that the Att-CAE compared to other methods considerably reduces the noise, even for very poor SNR. In brief, whatever the SNR and in particular in very noisy situations, the Att-CAE outperforms traditional and Machine Learning methods. This yields better noise reductions that will lead to an optimization of signal position identification essential for imaging techniques.

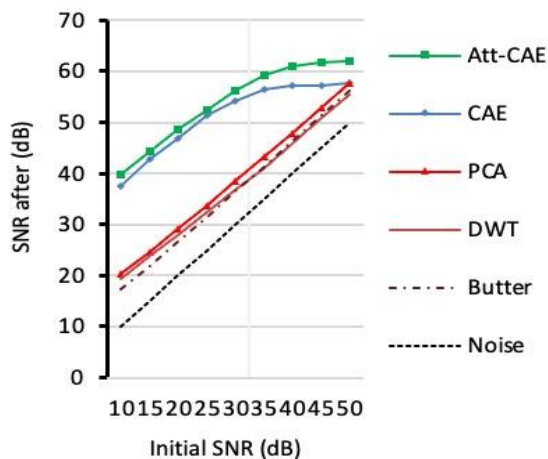


Figure 4. SNR distributions' means before and after denoising by mean of the compared methods

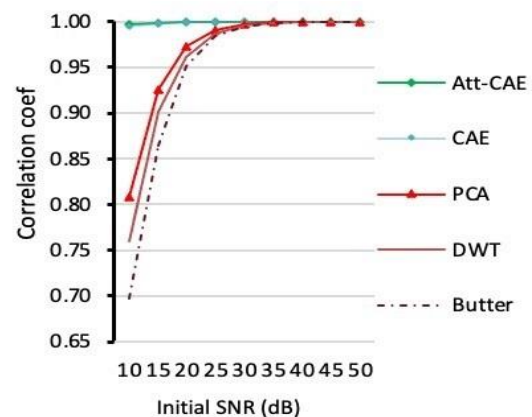


Figure 5. Correlation coefficients' means of each evaluation database after noise reduction using the proposed method, BF, DWT, PCA, and the CAE



### 3.2. Method's reusability and carbon footprint

The method proposed in this paper requires minor reparameterization and no pre-knowledge of the signals' features or high skills, which prevents exhaustive parameters' optimization computations, contrary to other traditional methods where often several characteristics of the signal are essential to the denoising process as well as an optimization of the parameters for each configuration. Moreover, the distributed-computing feature promotes the proposed method compared to the other ones. Lately, huge DL models demand more and more computation power and energy, which raises concerns about the environmental effects of these algorithms. For this purpose, we proposed an Att-CAE that could be recycled using transfer learning approaches [45]. These techniques permit retrieval of similar performances on different data distributions for correlative applications preventing full retraining from scratch and resulting in a reduction of the carbon footprint of the method for a greener algorithm. The model proposed in this paper was trained on NVIDIA's Tesla K80 GPU, and its training carbon footprint was estimated to be less than  $5.04 \text{ kgCO}_2\text{eq}$  [46].

## 4. CONCLUSION

In this paper, an Att-CAE is proposed for ultrasonic signal denoising. The DL model learns a compressed representation of the noisy ultrasonic signals providing the most relevant features of the latter. This representation is then leveraged to reconstruct the denoised signals. Attention gates were employed at the level of the skip connections to filter the features shared between the encoder and the decoder layers. The method proposed in this paper was compared to BF, DWT, PCA, and CAE methods, where the experiments collated the signal-to-noise ratios and Pearson correlation coefficient results. The DL model showed considerable improvements in the SNR values up to 30 dB outmatching the compared methods' SNR enhancement. Considering the correlation coefficients, the proposed method alongside the CAE resulted in very high values ( $\approx 1$ ) on signals with different noise intensities, which proves the efficient noise suppression, while the PCA, DWT, and BF methods achieved similar values only when the SNRs of the noisy signals were higher than 35 dB. The proposed deep learning model effectively denoises signals at different noise levels and recovers the signal waveform even when the signal is heavily corrupted by the noise. Such an efficient algorithm will lead to an improvement in the ultrasonic imaging process, enhancing the resolution of medical, industrial, and other applications based on this technology. Future work will then focus on the application of the proposed DL method to the estimation of the times of flight of ultrasonic signals. The quality of the Att-CAE pulse detection should then allow an enhancement of the axial resolution of imaging devices. With a similar objective, this method will also be extended to higher frequencies where noise becomes a dominant problem due to ultrasound attenuation.

## REFERENCES




- [1] J. Knez, A. Day, and D. Jurkovic, "Ultrasound imaging in the management of bleeding and pain in early pregnancy," *Best Practice & Research Clinical Obstetrics & Gynaecology*, vol. 28, no. 5, pp. 621–636, 2014, doi: 10.1016/j.bpobgyn.2014.04.003.
- [2] R. H. Perera *et al.*, "Real time ultrasound molecular imaging of prostate cancer with PSMA-targeted nanobubbles," *Nanomedicine: Nanotechnology, Biology and Medicine*, vol. 28, p. 102213, 2020, doi: 10.1016/j.nano.2020.102213.
- [3] H. Werner *et al.*, "An interactive experiment combining ultrasound, magnetic resonance imaging, and force feedback technology to physically feel the fetus during pregnancy," *European Journal of Radiology*, vol. 110, pp. 128–129, 2019, doi: 10.1016/j.ejrad.2018.11.020.
- [4] A. Rai and S. H. Upadhyay, "A review on signal processing techniques utilized in the fault diagnosis of rolling element bearings," *Tribology International*, vol. 96, pp. 289–306, 2016, doi: 10.1016/j.triboint.2015.12.037.
- [5] D. Laux, D. Baron, G. Despaux, A. I. Kellerbauer, and M. Kinoshita, "Determination of high burn-up nuclear fuel elastic properties with acoustic microscopy," *Journal of Nuclear Materials*, vol. 420, no. 1–3, pp. 94–100, 2012, doi: 10.1016/j.jnucmat.2011.09.010.
- [6] L. Espinosa, J. Bacca, F. Prieto, P. Lasaygues, and L. Brancheriau, "Accuracy on the time-of-flight estimation for ultrasonic waves applied to non-destructive evaluation of standing trees: A comparative experimental study," *Acta Acustica united with Acustica*, vol. 104, no. 3, pp. 429–439, 2018, doi: 10.3813/AAA.919186.
- [7] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *Fundamental Papers in Wavelet Theory*. Princeton University Press, pp. 494–513, 2009, doi: 10.1515/9781400827268.494.
- [8] C. Beale, C. Niezrecki, and M. Inalpolat, "An adaptive wavelet packet denoising algorithm for enhanced active acoustic damage detection from wind turbine blades," *Mechanical Systems and Signal Processing*, vol. 142, p. 106754, 2020, doi: 10.1016/j.ymsp.2020.106754.
- [9] V. Matz, R. Smid, S. Starman, and M. Kreidl, "Signal-to-noise ratio enhancement based on wavelet filtering in ultrasonic testing," *Ultrasonics*, vol. 49, no. 8, pp. 752–759, 2009, doi: 10.1016/j.ultras.2009.05.010.
- [10] Z. Talebhighi, F. Bazzazi, and A. Sadr, "Design and simulation of ultrasonic denoising algorithm using wavelet transform and ICA," in *2010 The 2nd International Conference on Computer and Automation Engineering (ICCAE)*, 2010, pp. 739–743, doi: 10.1109/ICCAE.2010.5451260.
- [11] Y. Zhang, W. Li, Z. Li, and T. Ning, "Dual attention per-pixel filter network for spatially varying image deblurring," *Digital Signal Processing*, vol. 113, p. 103008, 2021, doi: 10.1016/j.dsp.2021.103008.
- [12] H. Wang, Z. Liu, D. Peng, and Z. Cheng, "Attention-guided joint learning CNN with noise robustness for bearing fault diagnosis and vibration signal denoising," *ISA Transactions*, vol. 128, pp. 470–484, 2022, doi: 10.1016/j.isatra.2021.11.028.
- [13] R. Tibi, P. Hammond, R. Brogan, C. J. Young, and K. Koper, "Deep learning denoising applied to regional distance seismic data in Utah," *Bulletin of the Seismological Society of America*, vol. 111, no. 2, pp. 775–790, 2021, doi: 10.1785/0120200292.

- [14] V. Dalal and S. Bhairannawar, "Efficient de-noising technique for electroencephalogram signal processing," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 11, no. 2, p. 603, 2022, doi: 10.11591/ijai.v11.i2.pp603-612.
- [15] A. Rasti-Meymandi and A. Ghaffari, "A deep learning-based framework For ECG signal denoising based on stacked cardiac cycle tensor," *Biomedical Signal Processing and Control*, vol. 71, p. 103275, 2022, doi: 10.1016/j.bspc.2021.103275.
- [16] F. Gao, B. Li, L. Chen, X. Wei, Z. Shang, and C. He, "Ultrasonic signal denoising based on autoencoder," *Review of Scientific Instruments*, vol. 91, no. 4, 2020, doi: 10.1063/1.5136269.
- [17] W. Xu, X. Li, J. Zhang, Z. Xue, and J. Cao, "Ultrasonic signal enhancement for coarse grain materials by machine learning analysis," *Ultrasonics*, vol. 117, p. 106550, 2021, doi: 10.1016/j.ultras.2021.106550.
- [18] K. Antczak, "Deep recurrent neural networks for ECG signal denoising," pp. 1–8, 2018.
- [19] K. Cho, B. van Merriënboer, D. Bahdanau, and Y. Bengio, "On the properties of neural machine translation: encoder–decoder approaches," in *Proceedings of SSSST-8, Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation*, 2014, pp. 103–111, doi: 10.3115/v1/W14-4012.
- [20] H. Hewamalage, C. Bergmeir, and K. Bandara, "Recurrent neural networks for time series forecasting: Current status and future directions," *International Journal of Forecasting*, vol. 37, no. 1, pp. 388–427, 2021, doi: 10.1016/j.ijforecast.2020.06.008.
- [21] P. Liu, X. Qiu, and X. Huang, "Recurrent neural network for text classification with multi-task learning," *IJCAI International Joint Conference on Artificial Intelligence*, pp. 2873–2879, 2016.
- [22] D. Liu, M. Du, X. Li, Y. Hu, and L. Dai, "The USTC-NELSLIP systems for simultaneous speech translation task at IWSLT 2021," in *Proceedings of the 18th International Conference on Spoken Language Translation (IWSLT 2021)*, 2021, pp. 30–38, doi: 10.18653/v1/2021.iwslt-1.2.
- [23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [24] J. Song, J.-H. Jeong, D.-S. Park, H.-H. Kim, D.-C. Seo, and J. C. Ye, "Unsupervised denoising for satellite imagery using wavelet subband CycleGAN," pp. 1–11, 2020.
- [25] A. Sharma and M. Pramanik, "Convolutional neural network for resolution enhancement and noise reduction in acoustic resolution photoacoustic microscopy," *Biomedical Optics Express*, vol. 11, no. 12, p. 6826, 2020, doi: 10.1364/BOE.411257.
- [26] F. Dang, H. Chen, and P. Zhang, "DPT-FSNet: Dual-path transformer based full-band and sub-band fusion network for speech enhancement," in *ICASSP 2022 - 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 6857–6861, doi: 10.1109/ICASSP43922.2022.9746171.
- [27] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, "Object recognition with gradient-based learning," pp. 319–345, 1999, doi: 10.1007/3-540-46805-6\_19.
- [28] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Journal of Machine Learning Research*, 2011, pp. 315–323.
- [29] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNET++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020, doi: 10.1109/TMI.2019.2959609.
- [30] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Advances in Neural Information Processing Systems*, 2017, pp. 6389–6399.
- [31] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10008, pp. 179–187, 2016, doi: 10.1007/978-3-319-46976-8\_19.
- [32] O. Oktay *et al.*, "Attention U-Net: Learning where to look for the pancreas," 2018.
- [33] J. Schlemper *et al.*, "Attention gated networks: Learning to leverage salient regions in medical images," *Medical Image Analysis*, vol. 53, pp. 197–207, 2019, doi: 10.1016/j.media.2019.01.012.
- [34] A. Vaswani *et al.*, "Attention is all you need," *Advances in Neural Information Processing Systems*, pp. 5999–6009, 2017.
- [35] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [36] H. Robbins and S. Monro, "A stochastic approximation method," *The Annals of Mathematical Statistics*, vol. 22, no. 3, pp. 400–407, 1951, doi: 10.1214/aoms/1177729586.
- [37] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations*, 2014, pp. 1–15.
- [38] J. Zhuang *et al.*, "AdaBelief optimizer: Adapting stepsizes by the belief in observed gradients," 2020.
- [39] Z. Yan, Z. Zhang, and S. Liu, "Improving performance of seismic fault detection by fine-tuning the convolutional neural network pre-trained with synthetic samples," *Energies*, vol. 14, no. 12, p. 3650, 2021, doi: 10.3390/en14123650.
- [40] R. J. M. da Fonseca, L. Ferdj-Allah, G. Despau, A. Boudour, L. Robert, and J. Attal, "Scanning acoustic microscopy? Recent applications in materials science," *Advanced Materials*, vol. 5, no. 7–8, pp. 508–519, 1993, doi: 10.1002/adma.19930050703.
- [41] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Transactions on Image Processing*, vol. 9, no. 9, pp. 1532–1546, 2000, doi: 10.1109/83.862633.
- [42] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994, doi: 10.1093/biomet/81.3.425.
- [43] R. Cohen, "Signal denoising using wavelets project report," 2012.
- [44] Z. Sun and J. Lu, "An ultrasonic signal denoising method for EMU wheel trackside fault diagnosis system based on improved threshold function," *IEEE Access*, vol. 9, pp. 96244–96256, 2021, doi: 10.1109/ACCESS.2021.3093482.
- [45] N. Tajbakhsh *et al.*, "Convolutional neural networks for medical image analysis: Full training or fine tuning?," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016, doi: 10.1109/TMI.2016.2535302.
- [46] A. Lacoste, A. Luccioni, V. Schmidt, and T. Dandres, "Quantifying the carbon emissions of machine learning," 2019.






## BIOGRAPHIES OF AUTHORS






**Nabil Jai Mansouri**    received his engineering degree from the National School of Applied Sciences of Fez, Morocco. He is pursuing his Ph.D. degree at the Sidi Mohamed Ben Abdellah University and the University of Montpellier. His Ph.D. thesis focuses on Deep Learning methods for ultrasonic signal processing. He can be contacted at emails: [nabil.jai-mansouri@umontpellier.fr](mailto:nabil.jai-mansouri@umontpellier.fr), [nabil.jaimansouri@usmba.ac.ma](mailto:nabil.jaimansouri@usmba.ac.ma).






**Ghizlane Khaissidi**   , National Ph.D. holder in 2009 of the Sidi Mohamed Ben Abdellah University in Fez in Image Processing and Computer science. Currently a Professor at the National School of Applied Sciences (ENSA), University USMBA Fez (Morocco), and member of the Lab of computing and interdisciplinary physics (L.I.P.I). Her research activities concern image processing and its applications in medicine, heritage preservation (indexing of old manuscripts), societal dimension of applications (applications for the blind and visually impaired), and handwriting analysis for the detection of neurodegenerative pathologies. Machine learning, Deep learning, and data analysis. She can be contacted at email: [ghizlane.khaissidi@usmba.ac.ma](mailto:ghizlane.khaissidi@usmba.ac.ma).






**Gilles Despaux**    graduated from a School of Engineers in Robotics in 1990. He received his Ph. D. in "Electrical Engineering" in 1993 from Montpellier University, France, where he was Assistant Professor until 2006 and then Professor. He is currently head of the acoustic team of the Institute of Electronics and Systems and is the Director of the master's degree in Electrical Engineering. His research area concerns material characterization and imaging by Acoustic Microscopy which he first studied at Stanford Univ. during his Master's Degree Training period in 1990. He is a member of the French Society of Acoustics and an IEEE Member. He can be contacted at email: [gilles.despaux@umontpellier.fr](mailto:gilles.despaux@umontpellier.fr).



**Mostafa Mrabti**    obtained a Ph.D. degree from the USMBA University in 1996, Fes Morocco. He is a Professor at the National School of Applied Sciences (ENSA), University USMBA Fez (Morocco), and a member of the LIPI laboratory. He is the author of many publications. His research interests are automatic control, signal processing, and information. He can be contacted at email: [mostafa.mrabti@usmba.ac.ma](mailto:mostafa.mrabti@usmba.ac.ma).



**Emmanuel Le Clézio**    graduated in mathematics in 1995 and electronics in 1997 from the Univ. of Rennes I, France, and received the Master of Acoustics from the Univ. of Le Mans, France, in 1998. He received his Ph.D. degree in mechanics (acoustics) from the Univ. of Bordeaux 1, France, in 2001. For 10 years he was an Assistant Professor of Physics and Telecommunications at the University of Tours (IUT-Blois) and since 2011 Professor of Electronics at the University of Montpellier. His research area concerns complex material characterization by acoustic methods. He is a member of the French Society of Acoustics and an IEEE Member. He can be contacted at email: [emmanuel.le-clezio@umontpellier.fr](mailto:emmanuel.le-clezio@umontpellier.fr).