smart cities

MDPI

Article

Denoising Method for Pipeline Leakage Voiceprint in Utility Tunnel Using an Enhanced StarGAN

Qi-Wen Tian [1], Yu-Fei Chen [1], Shi-Wan Zhang [1], Hui-Qing Lan [1,2,*] and Jie Gao [1]

[1] Key Laboratory of Vehicle Advanced Manufacturing, Measuring and Control Technology (Ministry of Education), Beijing Jiaotong University, Beijing 100044, China; 23111415@bjtu.edu.cn (Q.-W.T.); cyf@bjtu.edu.cn (Y.-F.C.); 23222023@bjtu.edu.cn (S.-W.Z.); 25110492@bjtu.edu.cn (J.G.)

[2] School of Mechanical, Electronic and Control Engineering, Tangshan Research Institute of Beijing Jiaotong University, Tangshan 063000, China

* Correspondence: hqlan@bjtu.edu.cn

Highlights

What are the main findings?

- Enhanced StarGAN denoises multiple utility tunnel noise types using one unified leakage-voiceprint model.
- The proposed method achieved better overall denoising performance than representative deep learning and signal-decomposition-based methods, with improved SNR, PSNR, and SSR while preserving leakage-related voiceprint features.

What are the implications of the main findings?

- It improves real-time acoustic preprocessing for robust pipeline leakage monitoring in smart utility tunnels.
- Its lightweight plug-and-play design supports integration with denoising and recognition networks.

Abstract

Leakage of water pipelines in urban underground utility tunnels poses a major threat to tunnel safety and operation; therefore, voiceprint recognition is adopted for real-time pipeline monitoring. Although utility tunnels are less affected by outdoor interference, multiple internal noises still degrade voiceprint recognition accuracy. To address this problem, this study proposes an enhanced StarGAN-based denoising method using a single network to handle multiple noise types. Unlike the original StarGAN-VC2 developed for voice conversion, the proposed model is specifically redesigned for leakage voiceprint denoising by integrating MFCC-based representation, a lightweight bottleneck, channel attention, residual feature preservation, and U-Net-style reconstruction. Experimental and engineering application results show that the denoised signals achieve improvements of 3–7 dB in SNR, 3–4 dB in PSNR, and 3–4 in SSR. The model also demonstrates strong generalization capability and plug-and-play applicability, enabling integration with conventional denoising and voiceprint recognition networks. These results indicate that the proposed method can effectively suppress diverse utility tunnel noises while preserving leakage-related voiceprint features.

Keywords: utility tunnels; denoising; StarGAN; voiceprint recognition

Check for updates

Academic Editor: Isam Shahrour

Received: 30 May 2026
Revised: 3 July 2026
Accepted: 7 July 2026
Published: 9 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Urban water supply pipelines installed in underground utility tunnels are critical components of modern urban infrastructure and ensure reliable water delivery to residents [1–3]. Ensuring the safe and reliable operation of such pipelines is therefore essential. Leakage represents one of the most common and serious failure modes in water supply systems, while current detection practices still rely mainly on video inspection and periodic manual patrols [4,5]. Although widely used, conventional inspection approaches cannot provide continuous, real-time condition monitoring. Delayed leakage detection may lead to water loss, economic losses, and degradation of service quality. Therefore, an effective real-time leakage detection method is needed to improve the resilience, safety, and sustainability of urban water supply systems.

Compared with open environments, confined utility tunnels provide a relatively stable acoustic field with limited external interference, creating favorable conditions for acoustic-based monitoring and voiceprint recognition [6–8]. Voiceprint recognition for utility tunnel monitoring generally requires a representative acoustic database, task-specific recognition models, and acoustic sensors or microphones for real-time signal acquisition and analysis. Such a monitoring strategy enables efficient and cost-effective leakage surveillance in confined tunnel spaces. Reliable voiceprint recognition, however, depends on acoustic signals with clear and discriminative features that can characterize fault states.

Although confined utility tunnels provide a comparatively stable acoustic field, the internal acoustic scene remains complex. Major interference sources include electrical discharge and arcing associated with electrical faults [9], ventilation system noise [10], gas pipeline leakage noise [11], personnel inspection activities, pump operation noise, and electromagnetic interference generated by communication equipment [3]. Water supply pipeline leakage produces distinctive voiceprint features, but the leakage sound is accompanied by secondary acoustic components, including reverberation, impact noise generated by high-pressure water jets striking tunnel walls, and water-impact noise produced when leakage flow hits the tunnel floor. The primary leakage sound and secondary acoustic components coexist with other tunnel noise sources and often overlap in the time-frequency domain. Such overlap can mask or distort leakage-related voiceprint features, reduce the separability of leakage signals, and ultimately limit the recognition accuracy of voiceprint-based leakage detection.

Existing denoising methods can generally be divided into traditional signal processing techniques and deep learning-based approaches [12–15]. Traditional signal processing is widely used in engineering because little or no training data is required. Typical methods include time-domain filtering, frequency-domain filtering, spectral subtraction, Wiener filtering, and adaptive signal decomposition. Time-domain filtering, such as mean filtering [16,17], suppresses high-frequency noise by averaging samples within a sliding window, but excessive smoothing may weaken leakage-related acoustic details. For water-supply pipeline leak detection, Guo et al. [18] proposed EMD-based adaptive noise cancellation, in which IMF components weakly correlated with leakage were removed using cross-correlation information from signals collected on both sides of a suspected leak. The method increased the SNR of simulated leakage signals by up to 16 dB and produced clearer time-delay peaks in practical vibro-acoustic pipeline signals. Frequency-domain filtering, including bandpass filtering [19,20], suppresses background noise by retaining selected frequency bands, but fixed bands are often insufficient for non-stationary leakage signals. Li et al. [21] therefore developed a VMD-based adaptive denoising method for water-supply pipeline leakage localization. By defining a signal clarity index, leakage-related modes were selected adaptively under unknown background noise, improving localization performance in simulations and experiments. Spectral subtraction [22,23]

estimates the noise spectrum and subtracts it from the short-time power spectrum of noisy signals; however, musical noise and feature distortion may occur when noise overlaps with leakage voiceprint components. For low-SNR leakage localization, Li et al. [24] proposed a CEEMDAN-ICA noise reduction method combined with cross-spectral analysis. The method used cross-spectral analysis to determine leakage-related frequency bands, CEEMDAN to decompose leakage vibration signals, and FastICA to separate leakage source components from reconstructed signals and virtual noise channels, thereby improving SNR, time-delay estimation, and localization accuracy. Wiener filtering [25,26] estimates the power spectra of noise and target signals and designs an optimal filter according to the minimum mean square error criterion, but the performance of Wiener filtering depends on stable noise and accurate spectral estimation. Butarbutar [27] evaluated adaptive Wiener filtering for recognition systems under different environments. Overall, traditional denoising methods are easy to implement and can improve leakage signal quality, but preserving discriminative leakage voiceprint features in complex utility tunnels remains challenging for reliable multi-noise monitoring.

In contrast to traditional signal processing methods, deep learning-based denoising methods have attracted considerable research interest because they can learn nonlinear mappings from noisy signals to clean representations [28–30]. For example, Peng et al. [31] proposed a hybrid denoising framework that combined deep learning with signal processing and achieved better performance than conventional wavelet-based methods. Wu et al. [32] developed an end-to-end self-supervised voice denoising model trained only on noisy audio samples, while Alamdari et al. [33] introduced a deep learning framework for real-world audio denoising without requiring clean speech references. Both studies reduced the dependence on clean target signals while maintaining competitive denoising performance. Kim et al. [34] used deep denoising autoencoders to remove industrial noise from audio data for anomaly detection, and Yang et al. [35] designed a fully convolutional network with dense and residual connections to enhance weak signal representations under complex noise. In addition, Su et al. applied GAN-based models, including HiFi-GAN [36] and HiFi-GAN-2 [37], to audio denoising, showing that adversarial learning can suppress noise while preserving important voiceprint features. Overall, deep learning-based methods show strong potential for complex acoustic enhancement, but large training datasets, high computational costs, and limited generalization under unseen tunnel noise conditions remain major challenges for practical utility tunnel monitoring.

Existing denoising methods perform effectively under relatively simple noise conditions but still face major challenges in complex mixed-noise environments. Traditional signal processing methods usually depend on prior noise estimation, empirical parameter selection, or explicit noise separation, which may remove leakage-related voiceprint details together with interference components and reduce the reliability of subsequent fault identification. Deep learning-based methods can learn nonlinear feature representations from noisy acoustic signals, but model performance depends strongly on the noise conditions covered by training data, and retraining or fine-tuning may be required for new noise categories. In utility tunnels, multiple internal noise sources with diverse acoustic characteristics may overlap with leakage sounds in the time-frequency domain, making robust extraction of leakage-related voiceprint features difficult. Practical monitoring also requires rapid denoising, low-latency inference, and acceptable computational cost. StarGAN [38] is a multi-domain generative adversarial network that can learn mappings among multiple domains within a unified framework. StarGAN-VC [39] and StarGAN-VC2 [40] have demonstrated strong multi-domain acoustic transformation capability in voice conversion tasks. Therefore, StarGAN is suitable for handling multiple utility tunnel noise categories using a single denoising model. However, the original StarGAN architecture remains

complex and computationally demanding for engineering leakage voiceprint denoising. Task-specific optimization is needed to improve training stability, denoising performance, feature preservation, and deployment efficiency.

To adapt StarGAN-VC2 to utility tunnel acoustic denoising, this study proposes an enhanced StarGAN framework for pipeline leakage voiceprint signals, as shown in Figure 1. Different from the original StarGAN-VC2 designed for speaker-domain voice conversion, the proposed model reformulates multi-domain generation as a multi-noise denoising task that maps leakage voiceprints contaminated by different tunnel noises to the clean leakage domain. MFCC features [41] are used as the input representation to capture compact time-frequency characteristics of leakage voiceprints. To improve denoising capability and computational efficiency, the original nine-layer ResNet bottleneck [42] is simplified to five layers, and SE-based channel attention [43] is introduced to emphasize informative leakage-related feature channels. In addition, residual skip connections and U-Net-style encoder–decoder connections [44] are employed to preserve shallow acoustic details and improve MFCC feature reconstruction during denoising. The acoustic feature dimension is further reduced from 512 to 128, considering the relatively simple and periodic characteristics of engineering leakage signals. These modifications reduce the training time from approximately 40 h to 7.5 h while maintaining generative quality.

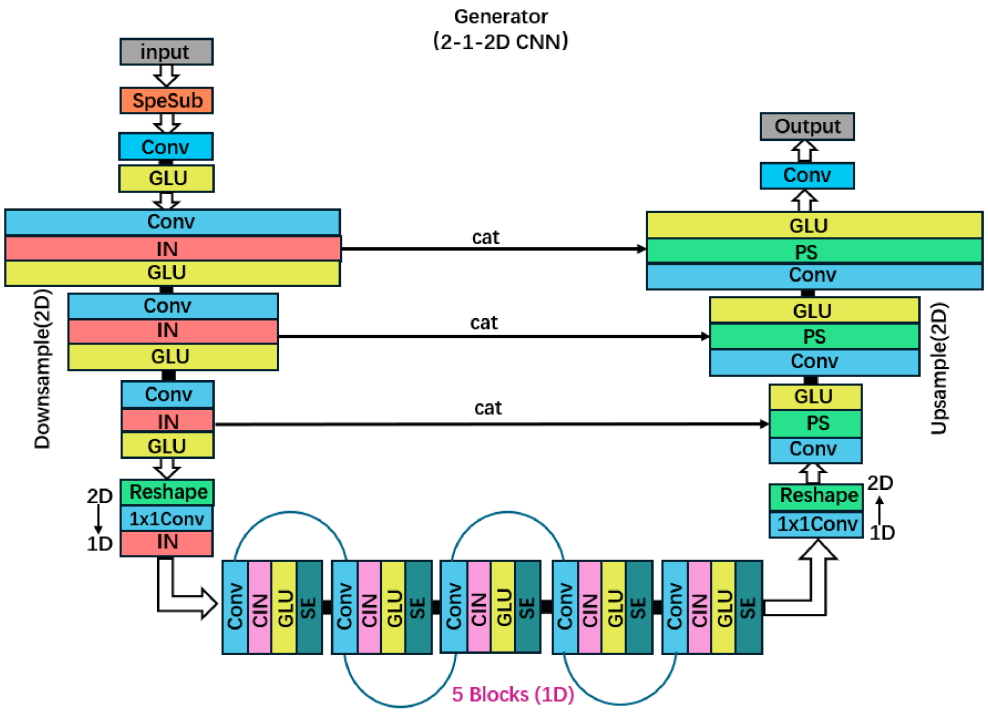


Figure 1. Generator Model Architecture Diagram.

The main contributions of this study are therefore threefold. First, an enhanced StarGAN-based multi-noise denoising model is developed for pipeline leakage voiceprints in utility tunnels. Second, a lightweight and feature-preserving generator is designed to balance noise suppression, leakage feature preservation, and training efficiency. Third, a comprehensive experimental framework is established, including multi-noise dataset construction, quantitative evaluation, generalization analysis, integration with spectral subtraction, downstream voiceprint recognition validation, and field-like pilot tests. The overall workflow of this study is illustrated in Figure 2.

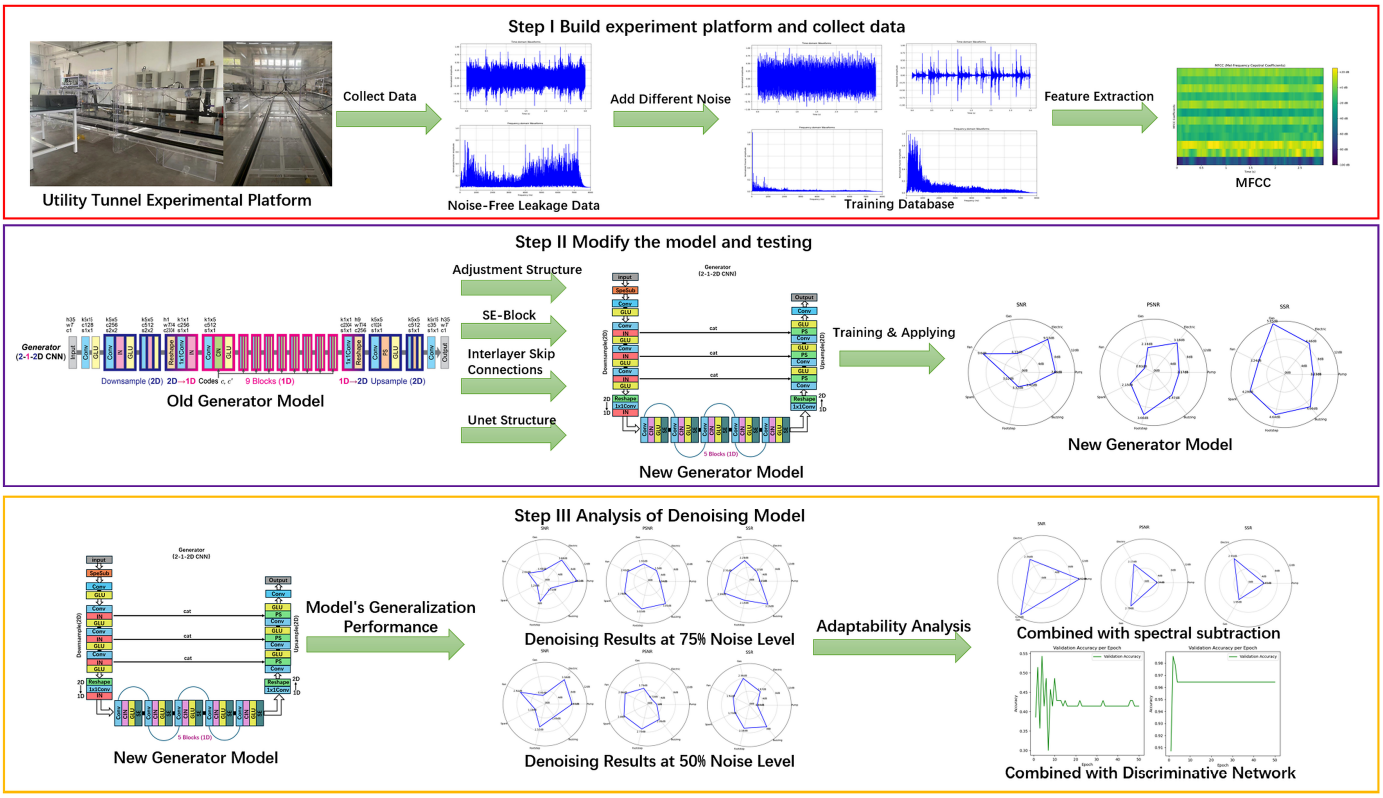


Figure 2. Article Structure Diagram.

2. Methods

2.1. StarGAN-Vc2

StarGAN is a generative adversarial network (GAN) architecture designed to learn mappings among multiple target domains within a unified model. Unlike conventional domain-specific GANs that require separate models for different transformation tasks, StarGAN performs multi-domain transformation through a shared generator and a domain-conditioned discriminator. The generator produces samples according to target-domain labels, whereas the discriminator evaluates generated samples and guides domain-consistent generation. To jointly optimize adversarial realism, domain consistency, and content preservation, the StarGAN framework employs multiple loss terms. The main loss functions used in the StarGAN framework are described as follows:

Confronting losses, for generator G:

$$\mathcal{L}_{\text{adv}}(G, D) = -E_{x,c_y}\left[\log D\left(G\left(x, c_y\right), c_y\right)\right] \tag{1}$$

For discriminator D:

$$\mathcal{L}_D = -E_{y,c_y}\left[\log D\left(y, c_y\right)\right] - E_{x,c_x}\left[\log\left(1 - D\left(G\left(x, c_y\right), c_y\right)\right)\right] \tag{2}$$

where x represents the source data, y denotes the true data from the target domain, and c_y and c_x are the labels for the target and source domains, respectively.

Refactoring loss

$$\mathcal{L}_{\text{rec}}(G) = E_{x,c_x}\left[\|G(G(x, c_x), c_x) - x\|_1\right] \tag{3}$$

Self-reconstruction loss

$$\mathcal{L}_{\text{self}-\text{rec}}(G) = E_{y,c_y}\left[\left\|G\left(G\left(y,c_y\right),c_y\right) - y\right\|_1\right] \tag{4}$$

Conditional loss

$$\mathcal{L}_{\text{cls}}(G) = -\mathrm{E}_{x,c_y}\left[\log D\left(G\left(x,c_y\right),c_y\right)\right] \tag{5}$$

Here, $\mathcal{L}_{\text{self}-\text{rec}}$ is the weight coefficient for the self-reconstruction loss.
Total loss, for generator G:

$$\mathcal{L}_G = \lambda_{\text{adv}}\mathcal{L}_{\text{adv}}(G,D) + \lambda_{\text{rec}}\mathcal{L}_{\text{rec}}(G) + \lambda_{\text{cls}}\mathcal{L}_{\text{cls}}(G) + \lambda_{\text{self}-\text{rec}}\mathcal{L}_{\text{self}-\text{rec}}(G) \tag{6}$$

Since the discriminator is optimized using the adversarial loss defined in Equation (2), the duplicate discriminator loss expression is omitted.

2.2. Proposed Enhanced StarGAN for Leakage Voiceprint Denoising

Although StarGAN-VC2 provides an effective multi-domain generation framework, its original architecture was mainly designed for voice conversion and cannot be directly optimized for pipeline leakage voiceprint denoising in utility tunnels. Therefore, the proposed enhanced StarGAN was redesigned for this task. Its main improvements are summarized as follows.

Task-specific denoising formulation:

The original StarGAN-VC2 is reformulated from speaker-domain voice conversion into a multi-domain denoising model, which maps leakage voiceprints contaminated by different tunnel noises to the clean leakage domain.

MFCC-based representation:

MFCC feature maps are used as the network input to provide compact time-frequency representations while preserving leakage-related acoustic characteristics during denoising.

Lightweight bottleneck:

The original nine-layer ResNet bottleneck is reduced to five ResNet-based layers, which decreases redundant feature transformation and computational cost while retaining sufficient reconstruction capability.

Attention and residual preservation:

SE-based channel attention and residual skip connections are introduced to emphasize informative MFCC channels, suppress noise-dominated responses, and improve feature propagation.

U-Net-style reconstruction:

Encoder–decoder skip connections are employed to preserve shallow time-frequency details and improve the reconstruction quality of denoised leakage voiceprints.

These modifications enable the proposed model to better balance multi-noise suppression, leakage-related feature preservation, and computational efficiency for practical utility tunnel acoustic monitoring.

2.3. Denoising Result Discrimination Method

2.3.1. SNR and I-SNR

Signal-to-Noise Ratio (SNR) is used to measure the ratio between the effective components of a signal and the noise components. It is typically expressed in decibels (dB);

a higher SNR value indicates better signal quality. For a signal x and its noise n, the calculation formula for SNR is as follows:

$$SNR = 10\log_{10}\left(\frac{P_s}{P_n}\right) \tag{7}$$

where P_s is the power of the signal, typically calculated as the mean of the square of the signal,

$$P_s = \frac{1}{n}\sum_{i=1}^{n} x_i^2 \tag{8}$$

and P_n is the power of the noise, typically calculated as the mean of the square of the noise,

$$P_n = \frac{1}{n}\sum_{i=1}^{n} n_i^2 \tag{9}$$

The improved Signal-to-Noise Ratio (i-SNR) is an indicator used to evaluate the effectiveness of denoising. It quantifies the efficacy of the denoising algorithm by comparing the signal-to-noise ratio of the denoised signal with that of the original signal. Its calculation formula is as follows:

$$\text{i-SNR} = \text{SNR}_{\text{post}} - \text{SNR}_{\text{pre}} \tag{10}$$

2.3.2. PSNR

Peak signal-to-noise ratio (PSNR) is a widely used metric for evaluating reconstruction quality after denoising, compression, or signal enhancement. In this study, PSNR was used to assess the similarity between the denoised MFCC feature map and the corresponding clean leakage MFCC feature map. PSNR is expressed in decibels (dB), and a larger PSNR value indicates better denoising performance and lower reconstruction error. The PSNR calculation is based on mean squared error (MSE), which quantifies the pixel-level difference between the clean MFCC feature map and the denoised MFCC feature map. The calculation procedure is defined as follows:

Mean Squared Error:

$$MSE(I,K) = \frac{1}{N \cdot M}\sum_{i=1}^{N}\sum_{j=1}^{M}(I(i,j) - K(i,j))^2 \tag{11}$$

where $I(i,j)$ and $K(i,j)$ represent the values of the original image and the denoised image at the pixel position (i,j) respectively. N and M are the dimensions of the image (number of rows and columns). The calculated MSE is the mean of the squared differences of all pixel values in the image.

The formula for calculating $PSNR$ through MSE is:

$$PSNR = 10 \cdot \log_{10}\left(\frac{I_{max}^2}{MSE}\right) \tag{12}$$

where I_{max} is the maximum pixel value in the image.

2.3.3. SSR (Signal-to-Signal Ratio)

SSR is used to quantify the ratio between signals. It is commonly utilized to assess the relationship between the denoised signal and the original signal. Since SSR is defined as a direct signal-energy ratio in this study, it is treated as a dimensionless metric.

$$SSR = \frac{\left\| I_{\text{orig}} \right\|^2}{\left\| I_{\text{denoised}} - I_{\text{orig}} \right\|^2} \tag{13}$$

where I_{orig} is the original image, $I_{denoised}$ is the denoised image, and $\| \cdot \|^2$ represents the sum of the squares of the signal.

2.4. Data Collection Experimental Platform

The structure of the utility tunnel experimental platform is shown in Figure 3. The platform was constructed using 10 mm-thick acrylic plates and was divided into three sections, each measuring 2.1 m in length, resulting in a total length of 6.3 m. The tunnel had an overall width of 1.02 m and a height of 1.1 m. A DN50 water supply pipeline and a DN50 gas supply pipeline were installed inside the tunnel. Leakage conditions were simulated by drilling holes in the pipelines. Three leak aperture sizes, namely 4, 6, and 8 mm, were used in the experiments. For each aperture size, three internal pressure levels of 0.40, 0.45, and 0.50 MPa were applied. Sampling positions were arranged at distances of 0.3, 0.5, and 0.8 m from the leak aperture. The experimental configuration increased the number of samples while maintaining data diversity and controlling excessive variability, thereby reducing the risk of overfitting caused by insufficient or highly homogeneous training data. The sampling rate was set to 48 kHz, and the data acquisition conditions are summarized in Table 1.



Figure 3. Experiment Platform for Data Collection.

Table 1. Method for Collecting Voiceprints of Pipeline Leaks in the Utility Tunnel.

Type	Location from Entrance (m)	Leakage Size (mm)	Pipeline Pressure (MPa)
Water Leakage	0.3/0.5/0.8	4/6/8	0.4/0.045/0.5

After data acquisition was completed, the leakage audio recordings were segmented into 3 s audio clips. Different noise types were then added to the collected leakage audio signals, including pump noise, electric noise, gas leakage noise, fan noise, electrical spark noise, footstep noise, and buzzing noise. The time-domain and frequency-domain characteristics of these noise sources are shown in Figures 4 and 5. For each noise category, ten independent audio clips longer than 3 s were selected. Random temporal positions within the selected clips were used for noise addition to increase the diversity of the generated dataset. The noise signals were directly superimposed onto the leakage signals without additional gain adjustment or attenuation. As a result, eight categories of data were generated, including seven categories of noisy leakage voiceprint data and one category of clean leakage voiceprint data without additional noise.

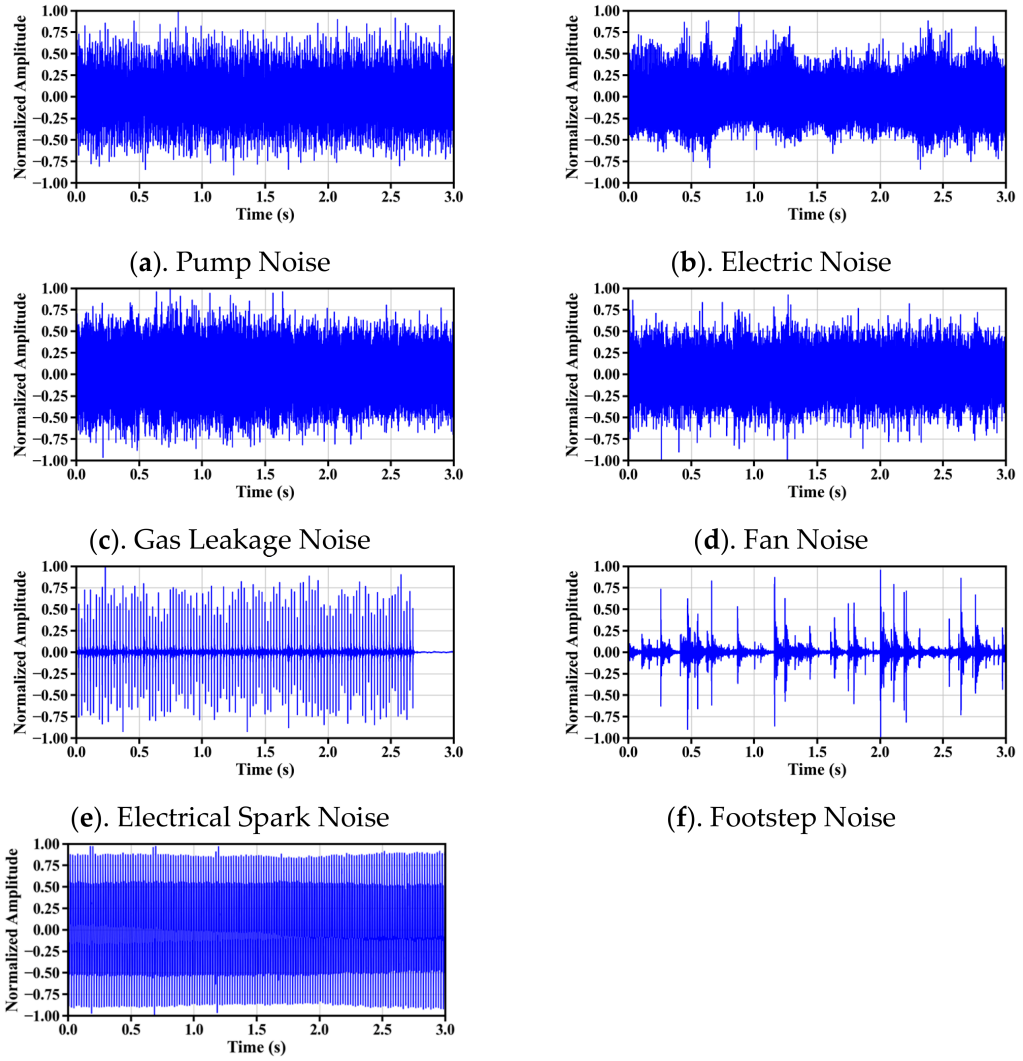


Figure 4. Time Domain Graphs of Different Noises.

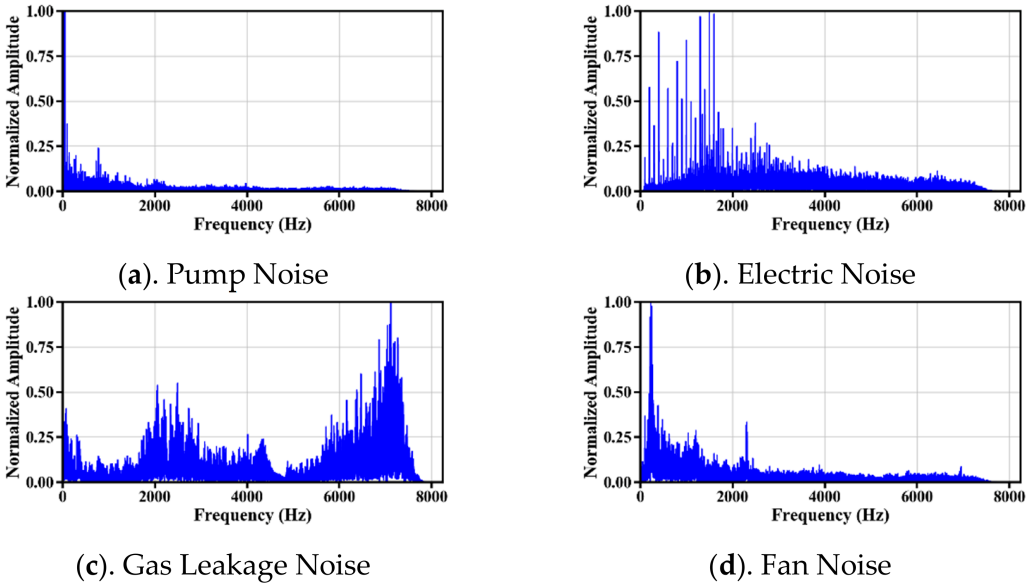
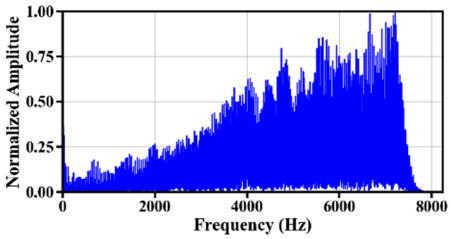
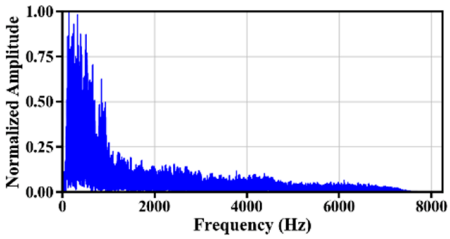


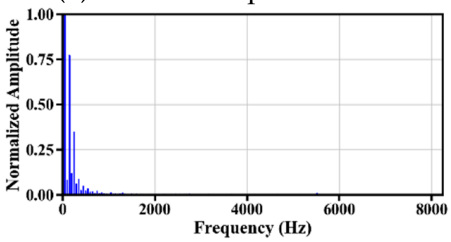
Figure 5. *Cont.*



(**e**). Electrical Spark Noise



(**f**). Footstep Noise



(**g**). Buzzing Noise

Figure 5. Frequency Domain Graphs of Different Noises.

The collected audio data were subsequently preprocessed for voiceprint analysis and converted into MFCC feature maps. The MFCC extraction procedure is shown in Figure 6 and included framing, windowing, short-time Fourier transform (STFT), and Mel filter bank processing. The generated MFCC feature maps were used as input data for subsequent model training and evaluation. The dataset was divided into training and testing sets at a ratio of 7:3. Detailed information regarding dataset composition and sample quantity is provided in Table 2.

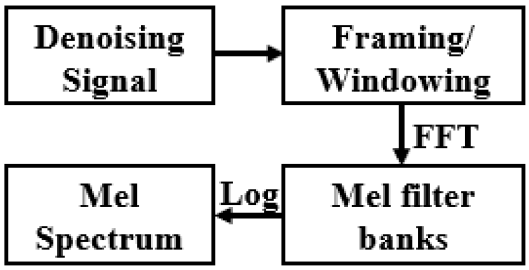


Figure 6. MFCC Feature Extraction Process.

Table 2. Dataset Composition.

Number	Type	Number of Files
1	Water pipeline leakage	1000
2	With Pump Noise	1000
3	With Electric Noise	1000
4	With Gas Leakage Noise	1000
5	With Fan Noise	1000
6	With Electrical Spark Noise	1000
7	With Footstep Noise	1000
8	With Buzzing Noise	1000

Pump noise is mainly concentrated in the low-frequency range below 1000 Hz, and the energy increases as the frequency decreases. Electric noise is primarily distributed below 2000 Hz and shows a clear periodic pattern within this frequency range. Gas leakage noise presents a broad frequency distribution, with energy concentrated in several specific

frequency bands and a prominent high-frequency peak near 7000 Hz. Fan noise is mainly concentrated in the low-frequency range and shows two distinct energy peaks near 300 Hz and 2300 Hz. Electrical spark noise exhibits relatively high energy above 1000 Hz, and the energy gradually increases toward higher frequencies, reaching a maximum near 7200 Hz. Footstep noise is mainly concentrated below 1000 Hz and contains substantial low-frequency energy. In contrast, buzzing noise shows energy concentration within several low-frequency bands and presents a periodic attenuation pattern.

2.5. Model Training Parameters

The training dataset contained eight categories, and the batch size was set to 32. The model was trained for 200,000 iterations. During training, the generator was updated once after every five discriminator updates. The learning rate started to decay after 100,000 iterations, and the Adam optimizer was used for parameter optimization. All training experiments were conducted on an NVIDIA RTX 4080 GPU, and each training session required approximately 7.5 h.

3. Discussion and Result

3.1. Analysis of Mfcc Results

First, the MFCC images of the voiceprints before and after denoising were compared with the MFCC image of the noise-free leakage voiceprint, as shown in Figures 7 and 8.

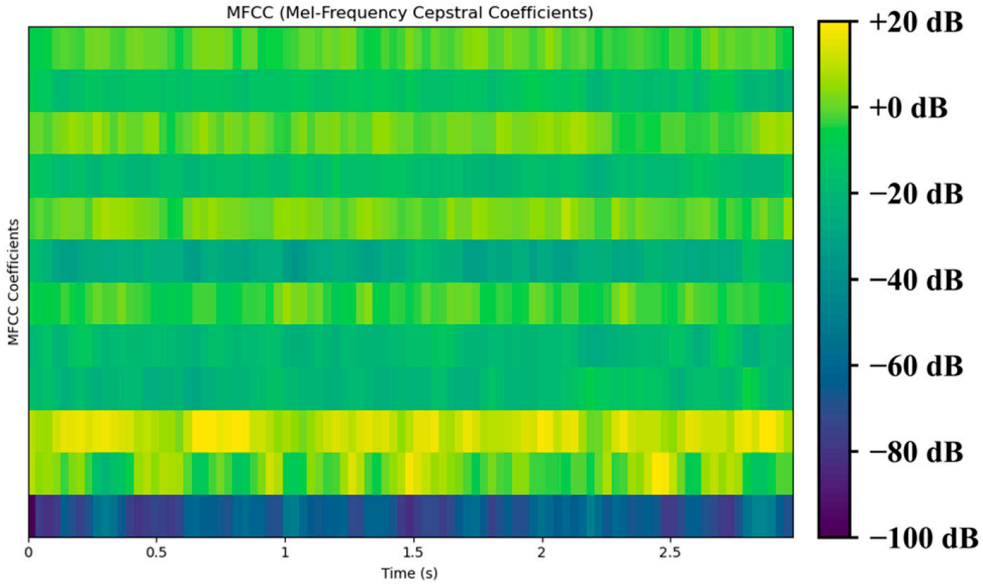


Figure 7. MFCC Spectrogram of Noise-Free Leakage Sound Patterns.

The MFCC representation of the clean leakage voiceprint in Figure 7 shows low energy in the lowest MFCC-coefficient region and relatively high energy in the low-to-intermediate MFCC-coefficient regions. Dominant leakage-related features are concentrated in high-energy horizontal bands highlighted by brighter colors. In the higher-order MFCC-coefficient region, the energy distribution remains relatively uniform and forms a stable texture-like pattern. The clean MFCC representation therefore serves as a reference for evaluating noise-induced feature distortion and denoising effectiveness.

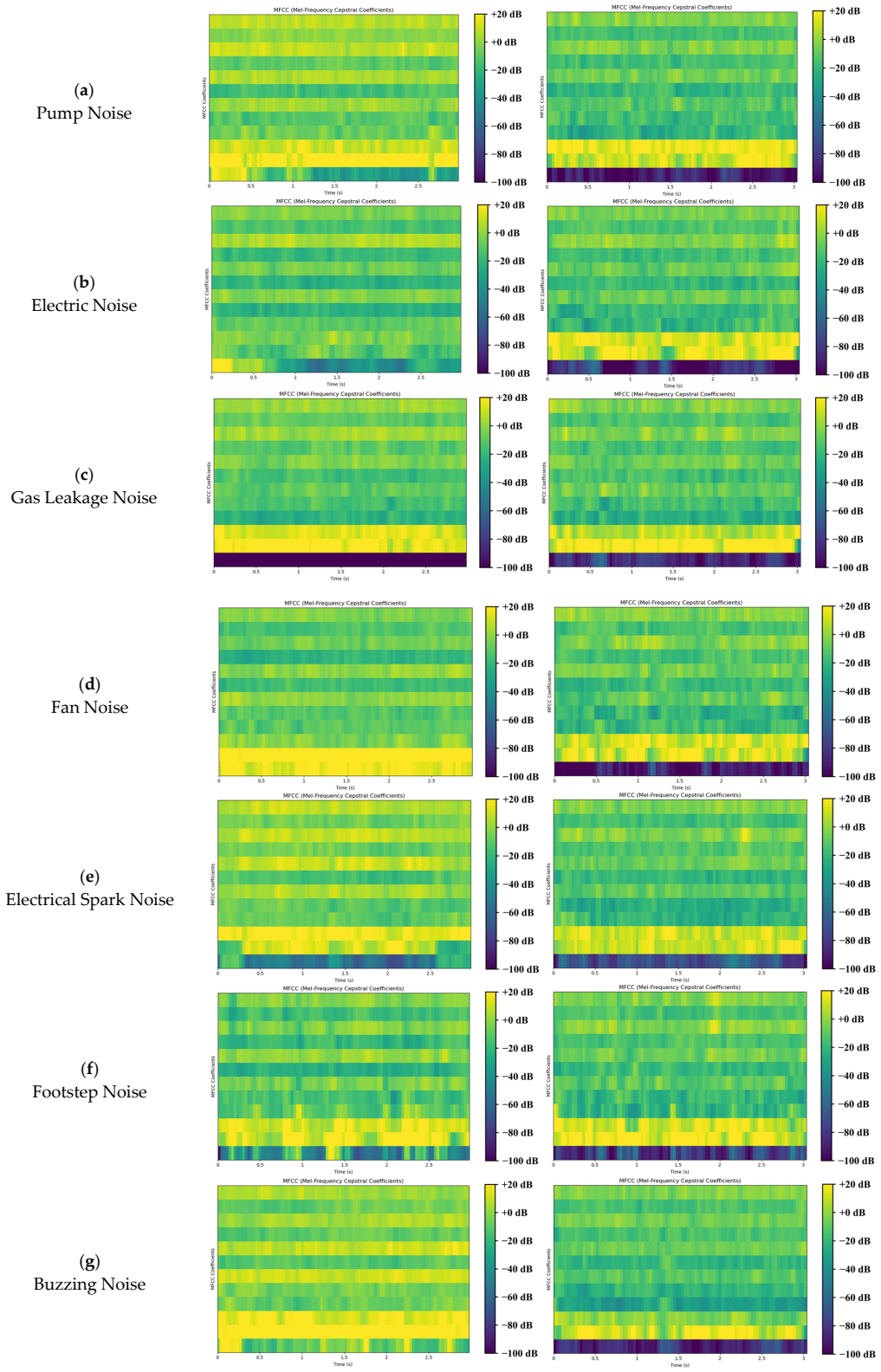


Figure 8. MFCC Spectrograms of Different Types of Sound Patterns Before (**Left**) and After Denoising (**Right**).

As shown in Figure 8, MFCC representations before denoising show noise-dependent distortions. Under pump noise contamination, the low-to-intermediate MFCC-coefficient regions show enhanced energy, whereas the lowest MFCC-coefficient region lacks the clear low-energy band observed in the clean leakage voiceprint. After denoising, the lowest region becomes darker, and the horizontal energy distribution becomes closer to

the clean MFCC pattern. Under electric noise contamination, the original MFCC map shows a relatively uniform energy distribution, indicating reduced contrast in leakage-related features. After denoising, the low-energy band in the lowest region is recovered, and characteristic horizontal bands become more distinguishable. Under gas leakage noise contamination, strong energy appears in the lower MFCC-coefficient region, and the feature distribution deviates from the clean leakage pattern. After denoising, the energy distribution becomes more regular, although residual interference remains in several time frames. Under fan noise contamination, strong continuous high-energy components appear in the lower MFCC-coefficient region. After denoising, excessive low-order energy is suppressed, and the reconstructed MFCC pattern becomes more consistent with the clean leakage voiceprint. Under electrical spark, footstep, and buzzing noise contamination, noisy MFCC maps show enhanced energy, temporal fluctuations, or energy increases in lower- and higher-order MFCC-coefficient regions. After denoising, the disturbances are weakened, and overall MFCC distributions become more stable and closer to the clean leakage pattern.

Overall, the left–right comparison in Figure 8 indicates that the proposed StarGAN-based denoising network suppresses noise-induced MFCC distortions while preserving leakage-related voiceprint structures. After denoising, MFCC maps show a clearer low-energy band in the lowest region, more distinguishable horizontal energy bands in the low-to-intermediate regions, and more stable texture-like patterns in the higher-order region. The denoised MFCC representations therefore provide more reliable feature inputs for subsequent recognition.

3.2. Quantitative Data Results

We selected the changes in SNR, PSNR, and SSR to analyze the denoising results of the voiceprints, as shown in Table 3 and Figure 9. The quantitative results show clear improvements across all noise categories.

Table 3. Denoising Results Data.

Noise	Metrics (SNR (dB)/PSNR (dB)/SSR (-))	Change Result
Pump	SNR	7.66
	PSNR	2.17
	SSR	3.23
Electric	SNR	9.53
	PSNR	3.18
	SSR	4.46
Gas	SNR	4.33
	PSNR	2.14
	SSR	5.85
Fan	SNR	9.60
	PSNR	0.83
	SSR	3.24
Spark	SNR	3.02
	PSNR	2.15
	SSR	4.29
Footstep	SNR	3.32
	PSNR	3.65
	SSR	4.64
Buzzing	SNR	3.42
	PSNR	2.47
	SSR	4.66

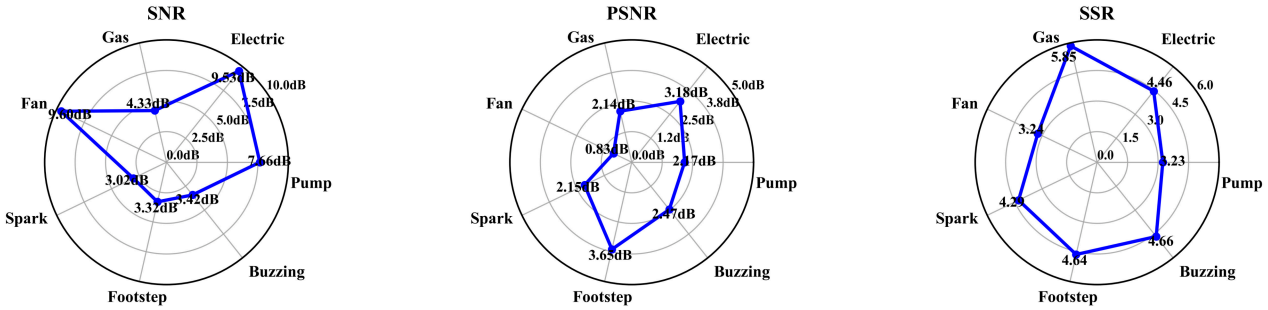


Figure 9. Radar Chart of Denoising Results.

The results in Table 3 and Figure 9 indicate that the proposed StarGAN-based denoising network effectively suppresses multiple types of utility tunnel noise using a single trained model. The SNR results show large improvements for pump noise, electric noise, and fan noise, with increases of 7.66 dB, 9.53 dB, and 9.60 dB, respectively. The relatively large SNR gains may be associated with the strong energy contribution of these noise types, which produces more pronounced changes after denoising. For gas leakage noise, electrical spark noise, footstep noise, and buzzing noise, the SNR improvements range from 3.02 to 4.33 dB, indicating that the proposed network also provides effective noise suppression for different interference patterns. The PSNR results show that, except for fan noise, the improvements range from 2.14 to 3.65 dB. Footstep noise achieves the highest PSNR improvement of 3.65 dB, whereas fan noise obtains the lowest PSNR improvement of 0.83 dB. The lower PSNR improvement for fan noise may be related to stronger overlap between fan-noise components and leakage-related MFCC features, which can increase reconstruction difficulty after denoising. The SSR results further confirm the denoising performance of the proposed network. Gas leakage noise achieves the highest SSR value of 5.85, whereas pump noise and fan noise show the lowest SSR values of 3.23 and 3.24, respectively. The SSR values for electric noise, electrical spark noise, footstep noise, and buzzing noise remain between 4.29 and 4.66, demonstrating that the proposed StarGAN-based network maintains stable denoising performance under different noise conditions.

3.3. Ablation Study and Comparison with Other Denoising Models

To further clarify the contribution of each architectural component and distinguish the proposed model from the original StarGAN-VC2, an ablation study was conducted, as summarized in Table 4. The reported SNR, PSNR, and SSR values represent the average improvements over the seven noisy leakage voiceprint categories, including pump noise, electric noise, gas leakage noise, fan noise, electrical spark noise, footstep noise, and buzzing noise. Compared with the original StarGAN-VC2, the proposed enhanced StarGAN achieves higher average improvements in SNR, PSNR, and SSR, increasing from 2.18 to 2.60 dB, from 1.99 to 2.29 dB, and from 1.82 to 2.19, respectively. These results indicate that the proposed method is not a simple structural modification of StarGAN-VC2, but a task-specific redesign for leakage voiceprint denoising in utility tunnel environments.

Among the ablated variants, removing the U-Net-style reconstruction structure results in the most pronounced degradation, with the average SNR, PSNR, and SSR decreasing to 1.93 dB, 1.82 dB, and 1.89, respectively. This demonstrates that encoder–decoder feature reconstruction is critical for recovering denoised MFCC representations, particularly when leakage-related components and noise components overlap in the time–frequency domain. The removal of skip connections also reduces the three metrics to 2.06 dB, 1.93 dB, and 1.96, confirming that direct feature propagation helps preserve shallow leakage-related acoustic details and maintain reconstruction fidelity. In addition, the variant without the SE block shows lower SNR and PSNR improvements, indicating that channel attention contributes

to the adaptive enhancement of informative MFCC channels and the suppression of noise-dominated responses.

When the lightweight bottleneck is removed, the SNR and SSR are slightly lower than those of the full model, whereas the PSNR is marginally higher. This suggests that increasing bottleneck complexity does not consistently improve denoising performance. Instead, the lightweight bottleneck provides a better balance between reconstruction quality and computational efficiency, which is important for deployable utility tunnel monitoring. Overall, the ablation results demonstrate that the SE block, skip connections, U-Net-style reconstruction, and lightweight bottleneck play complementary roles in noise-aware feature selection, leakage-detail preservation, MFCC reconstruction, and complexity reduction. Therefore, the proposed enhanced StarGAN differs essentially from the original StarGAN-VC2 by reformulating multi-domain voice conversion as multi-noise leakage voiceprint denoising and by introducing a lightweight, feature-preserving generator tailored to utility tunnel acoustic monitoring.

Table 4. Ablation study of the proposed enhanced StarGAN for leakage voiceprint denoising, with metrics averaged over seven noise categories.

Model Variant	Average SNR (dB)	Average PSNR (dB)	Average SSR (-)
Proposed Enhanced StarGAN	2.60	2.29	2.19
Original StarGAN-VC2	2.18	1.99	1.82
w/o SE block	2.31	2.08	2.14
w/o U-net	1.93	1.82	1.89
w/o Skip connection	2.06	1.93	1.96
w/o t Lightweight bottleneck	2.57	2.31	2.14

The denoising performance of the proposed enhanced StarGAN was further compared with representative deep learning-based and signal-decomposition-based denoising methods, including CMGAN [45], HiFi-GAN [36,37], DeepFilterNet2 [46], Denoising Autoencoder (DAE) [47], U-Net [44], CEEMDAN-LMS [48], and CEEMDAN-ICA & Cross-Spectral Analysis [24], as summarized in Table 5. The proposed enhanced StarGAN achieves the best overall performance, with average SNR, PSNR, and SSR improvements of 2.60 dB, 2.29 dB, and 2.19, respectively.

Among the baseline methods, HiFi-GAN obtains the highest SNR improvement of 2.16 dB, CEEMDAN-ICA & Cross-Spectral Analysis achieves the highest PSNR improvement of 1.91 dB, and DeepFilterNet2 shows the most competitive SSR value of 2.13. Compared with these strongest baselines, the proposed method improves the average SNR by 0.44 dB, PSNR by 0.38 dB, and SSR by 0.06, respectively. The two newly added CEEMDAN-based methods also demonstrate effective denoising capability. CEEMDAN-LMS achieves average SNR, PSNR, and SSR improvements of 1.97 dB, 1.80 dB, and 1.71, while CEEMDAN-ICA & Cross-Spectral Analysis performs slightly better, reaching 2.05 dB, 1.91 dB, and 1.88.

Although CEEMDAN-based methods can improve leakage voiceprint quality through adaptive decomposition and noise-related component separation, their performance depends strongly on IMF selection, reference-noise correlation, and frequency-band separation. When leakage-related components overlap with tunnel noise in the time–frequency domain, their denoising stability is limited. In contrast, the proposed enhanced StarGAN provides more balanced improvements in noise suppression, reconstruction quality, and signal preservation, owing to its task-specific multi-noise generation framework, channel attention, skip feature preservation, and U-Net-style reconstruction.

Table 5. Performance comparison between the proposed enhanced StarGAN and representative denoising models for leakage voiceprint denoising.

Denoising Model	Average SNR (dB)	Average PSNR (dB)	Average SSR (-)
Proposed Enhanced StarGAN	2.60	2.29	2.19
CMGAN	1.74	1.63	1.29
HiFiGAN	2.16	1.89	1.86
DeepFilterNet2	2.01	1.84	2.13
DAE	2.12	1.88	1.66
U-Net	1.73	1.47	1.37
CEEMDAN-LMS	1.97	1.80	1.71
CEEMDAN-ICA & Cross-Spectral Analysis	2.05	1.91	1.88

3.4. Analysis of the Model's Generalization Performance

Upon completing the model training and assessing the model's fundamental performance, we proceeded to test its generalization capability for denoising and evaluate its performance at various noise levels. First the noise addition ratio for the data was modified in the following formula.

$$audio_{noisy} = audio_{clean} + l \times Noise \tag{14}$$

To further evaluate generalization under reduced noise intensity, the trained model was tested at 75% and 50% noise levels. Although the improvement values decreased as the noise level decreased, the model still maintained effective denoising performance across different noise categories. The results at the 75% noise level are presented in Table 6 and Figure 10, and those at the 50% noise level are presented in Table 7 and Figure 11.

Table 6. Denoising Results at 75% Noise Level.

Noise	Metrics (SNR (dB)/PSNR (dB)/SSR (-))	Result
Pump	SNR	4.62
	PSNR	1.34
	SSR	1.15
Electric	SNR	3.88
	PSNR	1.50
	SSR	1.27
Gas	SNR	1.48
	PSNR	1.93
	SSR	2.28
Fan	SNR	2.64
	PSNR	2.43
	SSR	2.31
Spark	SNR	1.25
	PSNR	2.78
	SSR	2.99
Footstep	SNR	3.00
	PSNR	3.02
	SSR	2.14
Buzzing	SNR	1.31
	PSNR	3.05
	SSR	3.19

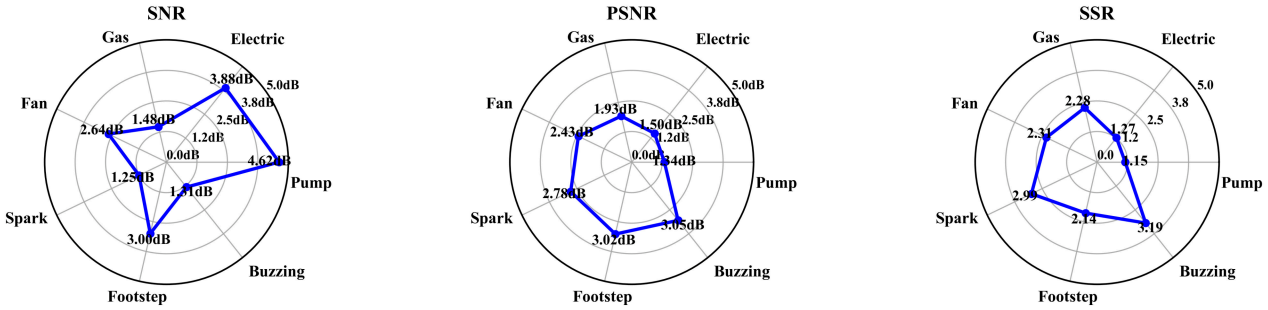


Figure 10. Radar Chart of Denoising Results at 75% Noise Level.

Table 7. Denoising Results at 50% Noise Level.

Noise	Metrics (SNR (dB)/PSNR (dB)/SSR (-))	Result
Pump	SNR	2.93
	PSNR	1.00
	SSR	0.84
Electric	SNR	3.36
	PSNR	0.73
	SSR	1.87
Gas	SNR	0.95
	PSNR	1.79
	SSR	2.96
Fan	SNR	2.92
	PSNR	2.66
	SSR	1.92
Spark	SNR	1.19
	PSNR	2.80
	SSR	1.73
Footstep	SNR	2.52
	PSNR	2.73
	SSR	2.56
Buzzing	SNR	1.69
	PSNR	2.06
	SSR	3.00

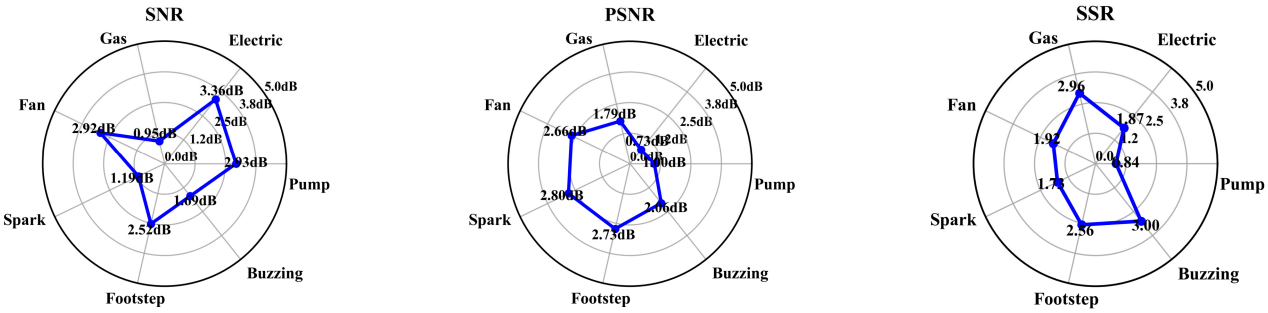


Figure 11. Radar Chart of Denoising Results at 50% Noise Level.

Under the 75% noise condition, the results in Table 6 and Figure 10 show that the proposed StarGAN-based denoising network maintains effective noise suppression across all noise categories. The SNR improvements are highest for pump noise and electric noise, with increases of 4.62 dB and 3.88 dB, respectively, followed by footstep noise and fan noise with increases of 3.00 dB and 2.64 dB. Gas leakage noise, electrical spark noise, and buzzing noise show smaller SNR improvements of 1.48 dB, 1.25 dB, and 1.31 dB, respectively. Compared with the 100% noise condition, the SNR gains decrease because

the reduced noise amplitude leaves less room for improvement after denoising. The PSNR improvements range from 1.34 to 3.05 dB. Buzzing noise and footstep noise achieve the highest PSNR improvements of 3.05 dB and 3.02 dB, whereas pump noise and electric noise show lower PSNR improvements of 1.34 dB and 1.50 dB. The PSNR improvement for fan noise reaches 2.43 dB, indicating improved reconstruction quality under reduced fan-noise intensity. The SSR values range from 1.15 to 3.19, with buzzing noise achieving the highest value and pump noise showing the lowest value. Overall, although the SSR values decrease compared with the 100% noise condition, the proposed network still maintains positive denoising gains under the 75% noise condition.

Under the 50% noise condition, the results in Table 7 and Figure 11 show that the proposed StarGAN-based denoising network still maintains positive denoising gains across all noise categories. The SNR improvements are relatively higher for electric noise, pump noise, fan noise, and footstep noise, with increases of 3.36 dB, 2.93 dB, 2.92 dB, and 2.52 dB, respectively. The average SNR gain of these four noise types is approximately 2.93 dB, indicating that the model still provides effective noise suppression under reduced noise intensity. Gas leakage noise, electrical spark noise, and buzzing noise show smaller SNR improvements of 0.95 dB, 1.19 dB, and 1.69 dB, respectively. The PSNR improvements range from 0.73 to 2.80 dB, with electrical spark noise, footstep noise, and fan noise showing relatively higher reconstruction-quality improvements. The SSR values range from 0.84 to 3.00, with buzzing noise and gas leakage noise achieving the highest values of 3.00 and 2.96, respectively. Overall, although the denoising gains decrease under the lower noise condition, the proposed network continues to provide effective noise suppression across different noise categories, demonstrating robustness and generalization capability under varying noise intensities.

The results under different noise levels indicate that the denoising gains of the proposed network decrease as the added noise amplitude decreases, although effective denoising capability is still maintained. The observed decreasing trend can be explained by the inherent noise components contained in the original pipeline leakage signals. When strong external noise is added, external noise becomes the dominant interference source, resulting in more pronounced metric improvements after noise suppression. In contrast, when the external noise level is low, the inherent noise components in the pipeline leakage signals account for a larger proportion of the total noise energy. Consequently, after external noise suppression, the measured improvement becomes less pronounced.

3.5. Evaluation Under Mixed-Noise Coexistence Conditions

To further evaluate the applicability of the proposed model under more realistic utility tunnel acoustic conditions, mixed-noise experiments were conducted using the three most common interference sources, namely pump noise, electric noise, and fan noise. Four mixed-noise conditions were constructed, including three pairwise combinations and one three-noise combination. Since the proposed enhanced StarGAN performs domain-conditioned denoising for specific noise categories, the corresponding noise-removal mappings were applied sequentially for samples containing multiple noise components.

As shown in Table 8, the proposed model achieves positive improvements in SNR, PSNR, and SSR under all mixed-noise conditions, indicating that the trained model can still suppress coexisting tunnel noises without retraining. Among the pairwise cases, the electric–fan condition obtains the highest SNR improvement of 4.36 dB, suggesting effective suppression of high-energy interference components. The pump–electric condition shows the most balanced overall performance, with SNR, PSNR, and SSR improvements of 3.16 dB, 1.89 dB, and 2.18, respectively. In contrast, the pump–fan condition shows relatively lower SNR and SSR among the pairwise cases, whereas the electric–fan condition exhibits the low-

est PSNR improvement. This may be attributed to the spectral overlap among mechanical noise, fan noise, electrical interference, and leakage-related acoustic components.

For the three-noise condition, the improvements decrease to 1.82 dB, 1.06 dB, and 1.08 for SNR, PSNR, and SSR, respectively. This reduction is reasonable because multiple superimposed noises introduce more severe time–frequency masking and may cause error accumulation during sequential denoising. Nevertheless, the positive gains confirm that the proposed enhanced StarGAN retains a certain degree of robustness under complex mixed-noise conditions and is more consistent with practical utility tunnel monitoring scenarios.

Table 8. Denoising performance of the proposed enhanced StarGAN under mixed-noise coexistence conditions.

Mixed-Noise Condition	Average SNR (dB)	Average PSNR (dB)	Average SSR (-)
Pump & Electric	3.16	1.89	2.18
Pump & Fan	2.62	1.45	1.69
Electric & Fan	4.36	1.31	2.01
Pump & Electric & Fan	1.82	1.06	1.08

3.6. Joint Denoising Using StarGAN Network and Spectral Subtraction

Spectral subtraction is a widely used denoising technique because of the simple implementation and suitability for real-time signal processing. However, spectral subtraction may introduce artifacts and signal distortion, particularly when the noise condition varies substantially. In addition, the performance of spectral subtraction depends strongly on parameter selection, and the resulting denoised signals may still differ from the desired clean target signals. The proposed StarGAN-based denoising network can handle multiple noise categories and exploit the generative reconstruction capability of GAN-based architectures. Therefore, the combination of spectral subtraction and the proposed StarGAN network was further investigated. The proposed network was used to enhance denoising performance and recover voiceprint features that may be degraded during spectral subtraction. Moreover, spectral subtraction reduced the complexity of the input signals, allowing the trained StarGAN network to be directly applied without additional retraining or parameter optimization. For further evaluation, pump noise, electric noise, and gas leakage noise were selected as representative test cases. Figure 12 presents the MFCC representations obtained after spectral subtraction alone in the left column and after the combined spectral subtraction–StarGAN processing in the right column.

When spectral subtraction is applied alone, the denoising performance remains limited, and important leakage-related voiceprint features may be degraded or removed, which can adversely affect subsequent voiceprint recognition. In contrast, when the proposed StarGAN-based network is applied after spectral subtraction, the MFCC representations show stronger noise suppression and improved recovery of partially degraded voiceprint features. The improvement benefits from the generative reconstruction capability of the GAN architecture, which enables the recovery of missing or distorted feature information. The results suggest that the proposed StarGAN-based denoising network provides effective denoising performance across multiple noise categories and also functions as a feature restoration module. Consequently, the proposed network can be integrated with conventional denoising methods to further improve overall denoising quality and reconstruct MFCC representations that are more consistent with the clean leakage voiceprint.

(**a**)
Data with Pump Noise

(**b**)
Data with Electric Noise

(**c**)
Data with Gas Leakage Noise

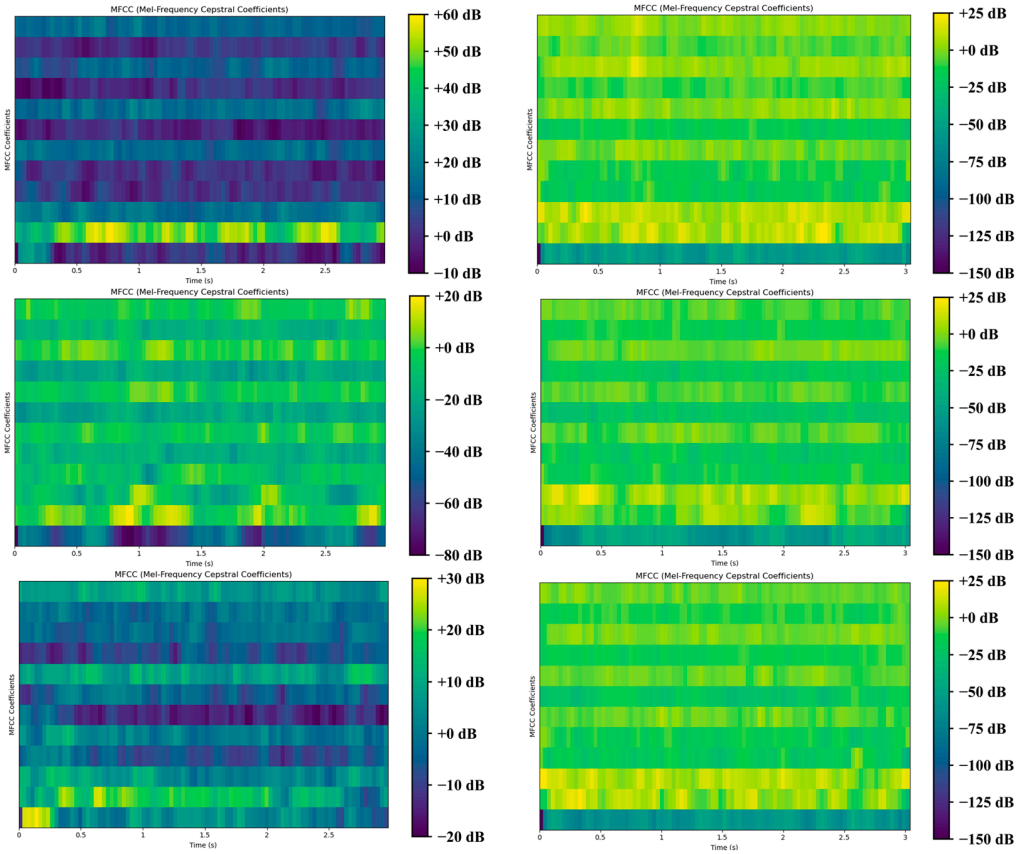


Figure 12. MFCC Images of Different Noisy Data After Using Spectral Subtraction (**Left column**) and After Using Spectral Subtraction and StarGAN Network (**Right column**).

To further evaluate the additional enhancement provided by the proposed StarGAN network after conventional spectral subtraction, three representative noise types, namely pump noise, electric noise, and gas leakage noise, were selected for quantitative analysis. Pump noise, electric noise, and gas leakage noise were selected because these noise types represent typical mechanical, electrical, and leakage-related interference sources in utility tunnel environments, respectively. Instead of repeating the evaluation for all noise categories, the representative-case analysis was conducted to clearly quantify the improvement achieved by applying StarGAN after spectral subtraction. The improvements in SNR, PSNR, and SSR are summarized in Table 9, and the corresponding radar-chart visualization is presented in Figure 13.

Table 9. Improvement Obtained by Applying StarGAN After Spectral Subtraction for Selected Representative Noise Types.

Noise	Metrics (SNR (dB)/PSNR (dB)/SSR (-))	Result
Pump	SNR	3.92
	PSNR	1.34
	SSR	1.65
Electric	SNR	2.34
	PSNR	2.17
	SSR	2.93
Gas	SNR	4.23
	PSNR	2.79
	SSR	1.95

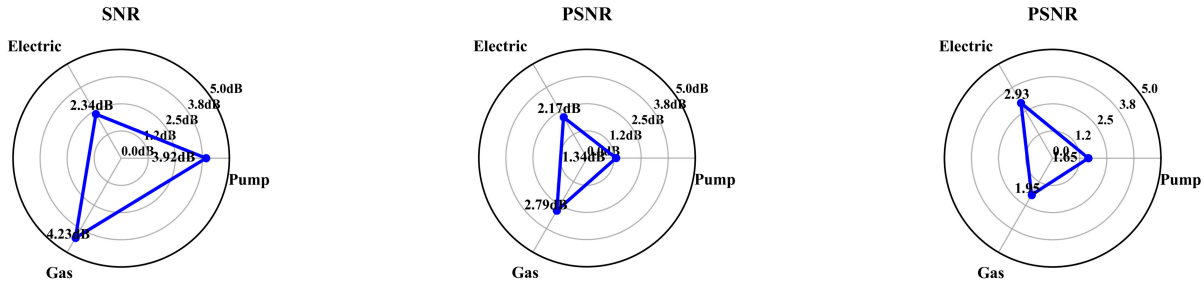


Figure 13. Radar Chart of the Improvement Obtained by Applying StarGAN After Spectral Subtraction for Selected Representative Noise Types.

The quantitative results confirm the additional enhancement achieved by applying the proposed StarGAN network after spectral subtraction. As shown in Table 8 and Figure 13, the combined strategy improves SNR, PSNR, and SSR for the selected representative noise types, namely pump noise, electric noise, and gas leakage noise. For SNR, gas leakage noise shows the largest improvement of 4.23 dB, followed by pump noise and electric noise with 3.92 dB and 2.34 dB, respectively. For PSNR, the improvements are 2.79 dB, 2.17 dB, and 1.34 dB for gas leakage, electric, and pump noise, respectively. For SSR, electric noise achieves the highest improvement of 2.93, followed by gas leakage and pump noise with 1.95 and 1.65. These results indicate that StarGAN can further enhance denoising performance after spectral subtraction.

3.7. Joint Usage of StarGAN Network and Discriminator Network

The primary objective of voiceprint-based utility tunnel monitoring is to determine whether collected acoustic signals contain leakage-related discriminative voiceprint features, thereby enabling real-time identification of potential pipeline leakage events. However, noise contamination can substantially reduce the reliability and accuracy of voiceprint recognition. Therefore, the ability to improve recognition accuracy after denoising represents an important evaluation criterion for the proposed denoising model. To evaluate the recognition-enhancement capability, the proposed StarGAN-based denoising network was assessed in combination with a voiceprint classification network.

The structure of the classification network is shown in Figure 14.

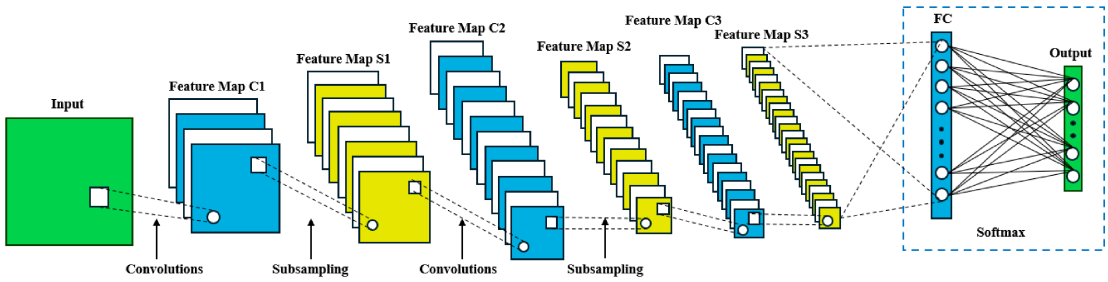


Figure 14. Structure of the Convolutional Neural Network for Discrimination.

The classification network consists of a lightweight binary convolutional neural network designed to determine whether an input voiceprint sample contains leakage-related features. During training, clean leakage voiceprints were treated as positive samples, whereas noisy leakage voiceprints were treated as negative samples.

After model training was completed, 150 samples from each noisy voiceprint category were selected for analysis. The selected samples were first evaluated without denoising and were directly input into the voiceprint classification network. Afterward, the

denoised versions of the selected samples were input into the classification network for further evaluation.

The recognition accuracies before and after denoising were compared. Without denoising, the leakage recognition accuracy is only 42%, whereas the accuracy increases to more than 95% after applying the proposed StarGAN-based denoising network. The comparison results demonstrate that the proposed denoising method substantially improves pipeline leakage recognition performance in utility tunnels. The recognition accuracy curves are presented in Figure 15.

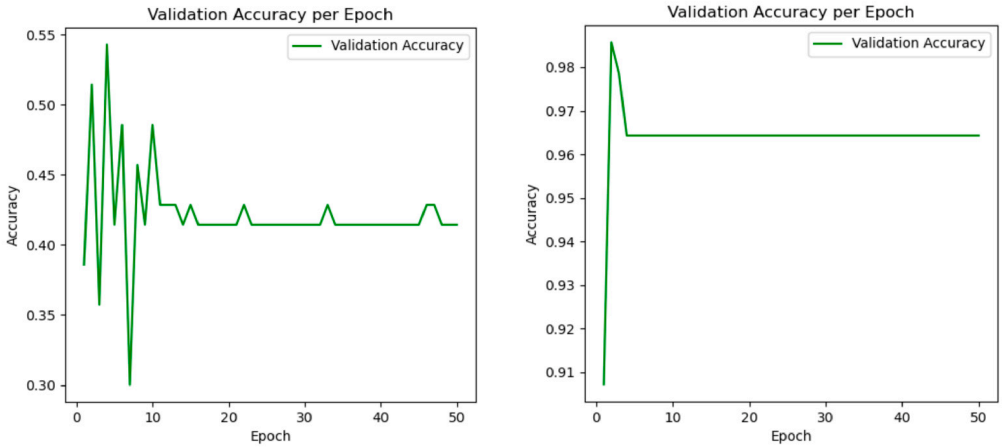


Figure 15. Analysis of Discriminative Results.

The recognition results for different noise categories after denoising are summarized in Table 10.

Table 10. Recognition Results of Voiceprint Data After Denoising.

Number	Type	Test Number	Correct Number
1	With Pump Noise	150	144
2	With Electric Noise	150	142
3	With Gas Leakage Noise	150	148
4	With Fan Noise	150	139
5	With Electrical Spark Noise	150	143
6	With Footstep Noise	150	144
7	With Buzzing Noise	150	141

To further validate the applicability of the proposed method in a realistic utility tunnel environment, field-like experiments were conducted in a pilot utility tunnel, as shown in Figure 16. Because the operating utility tunnel did not contain controllable leakage or fault events, and destructive tests could not be performed without affecting normal tunnel operation and safety, an acoustic playback-based validation scheme was adopted. The playback signals were not synthetic noises but real leakage sounds and real interference noises collected in previous experiments, including pump noise, electric noise, gas leakage noise, fan noise, electrical spark noise, footstep noise, and buzzing noise. The acoustic signals were played using portable Bluetooth speakers, including Sony SRS-XB13 and MIUI ASM11A, which were placed inside the tunnel near the pipeline. The propagated acoustic signals were collected using HS-K10N acoustic sensors connected to an HS-DVR-16L 16-channel recording host through RVVP3 × 0.3 shielded transmission cables. The recording host can support up to 16 sensors, and the acquisition system was configured with a sampling rate of 48 kHz. In the field-like test, sensor points were arranged at 3 m intervals, and three acoustic sensors were deployed at each point to improve spatial

coverage and signal acquisition reliability. This deployment allowed the replayed acoustic signals to propagate through the enclosed utility tunnel before being recorded. Therefore, the recorded signals contained the effects of sound propagation, wall reflection, pipeline reflection, and reverberation in the actual tunnel environment. Although this field-like test could not fully reproduce real-time leakage generation in an operating pipeline, it provided a practical and non-destructive validation condition that better approximated the acoustic characteristics of an actual utility tunnel than laboratory playback alone.

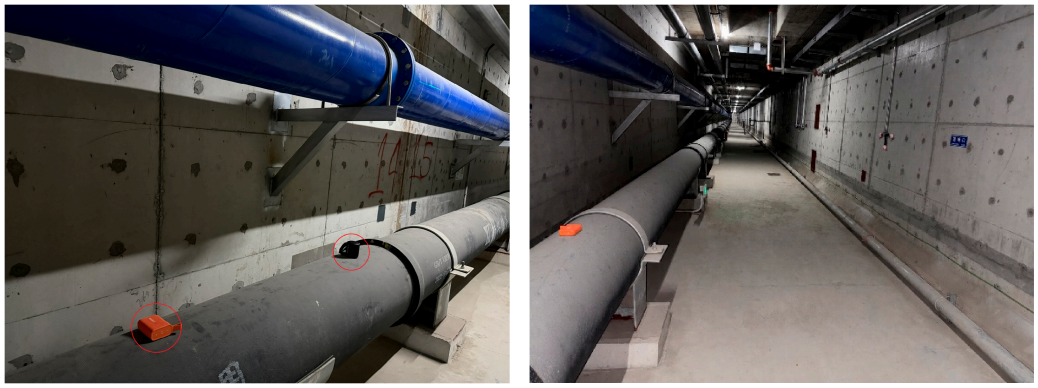


Figure 16. Data Collection in the Pilot Utility Tunnel.

The results are summarized in Table 11. The experimental results show that the recognition accuracy of the denoised audio reaches 94%, indicating that the proposed denoising model has strong potential for practical monitoring applications in utility tunnel environments.

Table 11. Recognition Results of Voiceprint Data in Pilot Utility Tunnel After Denoising.

Number	Type	Test Number	Correct Number
1	With Pump Noise	150	141
2	With Electric Noise	150	143
3	With Gas Leakage Noise	150	144
4	With Fan Noise	150	138
5	With Electrical Spark Noise	150	143
6	With Footstep Noise	150	140
7	With Buzzing Noise	150	138

4. Conclusions

Utility tunnels are complex underground environments characterized by diverse noise sources and complex voiceprint patterns. Therefore, denoising is a critical prerequisite for effective voiceprint-based monitoring and pipeline leakage detection. The present study proposes a StarGAN-based acoustic denoising network for suppressing multiple noise categories in utility tunnel environments. The proposed network can be trained using collected multi-domain voiceprint data and provides effective denoising capability while remaining compatible with conventional denoising methods and voiceprint recognition networks. The experimental results demonstrate that the proposed method substantially improves multiple voiceprint quality metrics after denoising. The MFCC representations of denoised voiceprints become more consistent with clean pipeline leakage voiceprints, indicating effective preservation of leakage-related features. Quantitative results show improvements of 3.02–9.60 dB in SNR, 0.83–3.65 dB in PSNR, and 3.23–5.85 in SSR. In addition, the proposed network demonstrates strong generalization capability and maintains effective denoising performance under different noise levels.

Therefore, the proposed StarGAN-based denoising network enables unified denoising of complex mixed-noise conditions using a single model. The unified processing strategy improves monitoring efficiency and automation in engineering applications while reducing computational burden and deployment cost. In addition, the proposed network serves as a modular plug-and-play component that can be used independently or integrated with other denoising and recognition models. The integration with spectral subtraction further improves overall denoising performance and alleviates voiceprint feature loss caused by spectral subtraction. Furthermore, when coupled with a voiceprint recognition network, the proposed denoising network increases pipeline leakage recognition accuracy from 42% to more than 95%.

The present study has several limitations that require further investigation. First, deep learning-based denoising methods require a large-scale and diverse noise dataset for robust model training. The construction of such a dataset depends on collecting noise samples from different engineering scenarios using recording equipment, which remains labor-intensive and time-consuming. Second, multiple noise sources often coexist in complex engineering environments; therefore, further evaluation is required to assess the denoising performance of the proposed network under mixed-noise interference conditions. In addition, although the MFCC representations indicate effective feature restoration after denoising, the perceptual quality of the reconstructed audio still differs from that of clean pipeline leakage voiceprints.

Author Contributions: Conceptualization, Q.-W.T.; methodology, Q.-W.T.; software, Q.-W.T. and J.G.; validation, Q.-W.T. and Y.-F.C.; formal analysis, Q.-W.T. and Y.-F.C.; investigation, Q.-W.T. and S.-W.Z.; resources, Q.-W.T. and Y.-F.C.; data curation, Q.-W.T. and S-W. Z; writing—original draft preparation, Q.-W.T. and H.-Q.L.; writing—review and editing, Q.-W.T. and H.-Q.L.; visualization, S.-W.Z. and J.G.; supervision, H.-Q.L.; project administration, H.-Q.L.; funding acquisition, H.-Q.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Key Research and Development Program of China "Science and Technology Innovation Program of Xiongan New Area" (No. 2022XAGG0147) and Development Funds of the Central Government for Local Science and Technology (No. 236Z1701G).

Institutional Review Board Statement: This study adheres to all relevant ethical guidelines. All data sets used are publicly available, and no personal sensitive information or private data was involved in the research process. This study does not involve animal or human experiments and therefore does not require approval from an ethics committee. We commit to strictly adhering to academic integrity and ethical standards throughout the research and publication process.

Data Availability Statement: Data will be made available on request.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Canto-Perello, J.; Curiel-Esparza, J. An analysis of utility tunnel viability in urban areas. *Civ. Eng. Environ. Syst.* **2006**, *23*, 11–19. [CrossRef]
2. Zhang, P.; Lan, H.; Yu, M. Reliability evaluation for ventilation system of gas tunnel based on Bayesian network. *Tunn. Undergr. Space Technol.* **2021**, *112*, 103882. [CrossRef]
3. Canto-Perello, J.; Curiel-Esparza, J. Risks and potential hazards in utility tunnels for urban areas. *Proc. Inst. Civ. Eng.-Munic. Eng.* **2003**, *156*, 51–56. [CrossRef]
4. Canto-Perello, J.; Curiel-Esparza, J. Assessing governance issues of urban utility tunnels. *Tunn. Undergr. Space Technol.* **2013**, *33*, 82–87. [CrossRef]
5. Zhang, Y.; Bai, J.; Diao, Y.; Chen, Z.; Wang, C.; Yang, K.; Gao, Z.; Wei, H. Risk and Energy Based Optimization for Fire Monitoring System in Utility Tunnel Using Cellular Automata. *Sustainability* **2024**, *16*, 4717. [CrossRef]

6. Tandel, N.; Prajapati, H.; Dabhi, V. Voice recognition and voice comparison using machine learning techniques: A survey. In *Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS)*; IEEE: New York, NY, USA, 2020; pp. 459–465. [CrossRef]
7. Shan, S.; Liu, J.; Dun, Y. Prospect of voiceprint recognition based on deep learning. *J. Phys. Conf. Ser.* **2021**, *1848*, 012046. [CrossRef]
8. Radha, V.; Vimala, C. A review on speech recognition challenges and approaches. *World Comput. Sci. Inf. Tech. J.* **2012**, *2*, 1–7. [CrossRef]
9. Yao, Y.; Wang, J.; Xu, J.; Hu, Y.; Ren, J.; Pei, C.; Wang, W.; Xiao, K. A brief analysis of the power cable planning and design of the utility tunnel. *J. Phys. Conf. Ser.* **2019**, *1314*, 012050. [CrossRef]
10. Ma, H.; Zhou, X.; Huang, J. Effect of ventilation on thermal and humidity environment of the underground utility tunnel in the plum rain season in southern China: Field measurement and CFD simulation. *Undergr. Space* **2023**, *13*, 301–315. [CrossRef]
11. Wang, R.; Zhang, Z.; Gong, D. Simulation study on gas leakage law and early warning in a utility tunnel. *Sustainability* **2023**, *15*, 15375. [CrossRef]
12. Azarang, A.; Kehtarnavaz, N. A review of multi-objective deep learning speech denoising methods. *Speech Commun.* **2020**, *122*, 1–10. [CrossRef]
13. Shetty, S.; Reeja, S. Audio Noise Removal–The State of the Art. *Int. J. Comput. Eng. Res.* **2014**, *4*, 34–37. [CrossRef]
14. Gul, S.; Khan, M. A survey of audio enhancement algorithms for music, speech, bioacoustics, biomedical, industrial and environmental sounds by image U-Net. *IEEE Access* **2023**, *11*, 144456–144483. [CrossRef]
15. Kumar, A.; Singh, P.; Yadav, J. Language Identification Using Speech Denoising Techniques: A Review. In *Automatic Speech Recognition and Translation for Low Resource Languages*; Wiley: Hoboken, NJ, USA, 2024; pp. 41–49. [CrossRef]
16. Justusson, B. Median Filtering: Statistical Properties. In *Two-Dimensional Digital Signal Processing II, Transform and Median Filters*; Springer: Berlin/Heidelberg, Germany, 2006; pp. 161–196.
17. Buades, A.; Coll, B.; Morel, J. A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* **2005**, *4*, 490–530. [CrossRef]
18. Guo, C.; Wen, Y.; Li, P.; Wen, J. Adaptive noise cancellation based on EMD in water-supply pipeline leak detection. *Measurement* **2016**, *79*, 188–197. [CrossRef]
19. Christiano, L.; Fitzgerald, T. The band pass filter. *Int. Econ. Rev.* **2003**, *44*, 435–465. [CrossRef]
20. Mahmoodi, S. Anisotropic diffusion for noise removal of band pass signals. *Signal Process.* **2011**, *91*, 1298–1307. [CrossRef]
21. Li, J.; Chen, Y.; Qian, Z.; Lu, C. Research on VMD based adaptive denoising method applied to water supply pipeline leakage location. *Measurement* **2020**, *151*, 107153. [CrossRef]
22. Boll, S. Suppression of acoustic noise in speech using spectral subtraction. *IEEE Trans. Acoust. Speech Signal Process.* **2003**, *27*, 113–120. [CrossRef]
23. Hussein, R.; Shaban, K.; El-Hag, A. Denoising different types of acoustic partial discharge signals using power spectral subtraction. *High Volt.* **2018**, *3*, 44–50. [CrossRef]
24. Li, S.; Cai, M.; Han, M.; Dai, Z. Noise Reduction Based on CEEMDAN-ICA and Cross-Spectral Analysis for Leak Location in Water-Supply Pipelines. *IEEE Sens. J.* **2022**, *22*, 13030–13042. [CrossRef]
25. Chen, J.; Benesty, J.; Huang, Y.; Doclo, S. New insights into the noise reduction Wiener filter. *IEEE Trans. Audio Speech Lang. Process.* **2006**, *14*, 1218–1234. [CrossRef]
26. Benesty, J.; Chen, J.; Huang, Y. Study of the widely linear Wiener filter for noise reduction. In *Proceedings of the 2010 IEEE International Conference on Acoustics, Speech and Signal Processing*; IEEE: New York, NY, USA, 2010; pp. 205–208. [CrossRef]
27. Butarbutar, M.; Sachio, K.; Nugroho, M.; David, D.; Saputra, P.A. Adaptive Wiener Filtering Method for Noise Reduction in Speech Recognition System. *TechRxiv* **2023**. [CrossRef]
28. Li, D.; Bai, Y.; Bai, Z.; Li, Y.; Shang, C.; Shen, Q. Decomposed neural architecture search for image denoising. *Appl. Soft Comput.* **2022**, *124*, 108914. [CrossRef]
29. Baptista, M.; Henrique, E. 1D-DGAN-PHM: A 1-D denoising GAN for prognostics and health management with an application to turbofan. *Appl. Soft Comput.* **2022**, *131*, 109785. [CrossRef]
30. Fan, C.; Zhang, Y.; Ma, H.; Ma, Z. A novel deep denoising model integrating transformer and time–frequency loss for gearbox fault diagnosis. *Adv. Eng. Inform.* **2025**, *66*, 103400. [CrossRef]
31. Peng, Z.; Kong, Q.; Yuan, C.; Li, R.; Chi, H.L. Development of acoustic denoising learning network for communication enhancement in construction sites. *Adv. Eng. Inform.* **2023**, *56*, 101981. [CrossRef]
32. Wu, J.; Li, Q.; Yang, G.; Li, L.; Senhadji, L.; Shu, H. Self-supervised speech denoising using only noisy audio signals. *Speech Commun.* **2023**, *149*, 63–73. [CrossRef]
33. Alamdari, N.; Azarang, A.; Kehtarnavaz, N. Improving deep speech denoising by noisy2noisy signal mapping. *Appl. Acoust.* **2021**, *172*, 107631. [CrossRef]
34. Kim, S.; Kim, Y. Enhancing sound-based anomaly detection using deep denoising autoencoder. *IEEE Access* **2024**, *12*, 84323–84332. [CrossRef]

35. Yang, L.; Fomel, S.; Wang, S.; Chen, X.; Chen, W.; Saad, O.M.; Chen, Y. Denoising of distributed acoustic sensing data using supervised deep learning. *Geophysics* **2023**, *88*, WA91–WA104. [CrossRef]
36. Su, J.; Jin, Z.; Finkelstein, A. HiFi-GAN: High-Fidelity Denoising and Dereverberation Based on Speech Deep Features in Adversarial Networks. *arXiv* **2020**, arXiv:2006.05694.
37. Su, J.; Jin, Z.; Finkelstein, A. HiFi-GAN-2: Studio-quality speech enhancement via generative adversarial networks conditioned on acoustic features. In *Proceedings of the 2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*; IEEE: New York, NY, USA, 2021; pp. 166–170. [CrossRef]
38. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2018; pp. 8789–8797.
39. Kameoka, H.; Kaneko, T.; Tanaka, K.; Hojo, N. Stargan-vc: Non-parallel many-to-many voice conversion using star generative adversarial networks. In *Proceedings of the 2018 IEEE Spoken Language Technology Workshop (SLT)*; IEEE: New York, NY, USA, 2018; pp. 266–273. [CrossRef]
40. Kaneko, T.; Kameoka, H.; Tanaka, K.; Hojo, N. StarGAN-VC2: Rethinking Conditional Methods for StarGAN-Based Voice Conversion. *arXiv* **2019**, arXiv:1907.12279.
41. Zheng, F.; Zhang, G.; Song, Z. Comparison of different implementations of MFCC. *J. Comp. Sci. Tech.* **2001**, *16*, 582–589. [CrossRef]
42. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2016; pp. 770–778.
43. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; IEEE: New York, NY, USA, 2018; pp. 7132–7141.
44. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241. [CrossRef]
45. Abdulatif, S.; Cao, R.; Yang, B. Cmgan: Conformer-based metric-gan for monaural speech enhancement. *IEEE/ACM Trans. Audio Speech Lang. Process.* **2024**, *32*, 2477–2493. [CrossRef]
46. Schröter, H.; Maier, A.; Escalante-B, A.N.; Rosenkranz, T. Deepfilternet2: Towards real-time speech enhancement on embedded devices for full-band audio. In *Proceedings of the 2022 International Workshop on Acoustic Signal Enhancement (IWAENC)*; IEEE: New York, NY, USA, 2022; pp. 1–5. [CrossRef]
47. Lu, X.; Tsao, Y.; Matsuda, S.; Hori, C. Speech enhancement based on deep denoising autoencoder. *Interspeech* **2013**, *2013*, 436–440. [CrossRef]
48. Li, S.; Fu, Q.; Shen, P.; He, Z. Adaptive noise cancellation based on CEEMDAN-LMS for pipeline leak location under low SNR. *Process Saf. Environ. Prot.* **2025**, *202*, 107668. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.