



Original Paper

UAV-based pipeline leakage detection via acoustics with owl auditory-inspired OPEI representation

Ying-Han Ma, Hong Zhao^{*}

College of Mechanical and Transportation Engineering, China University of Petroleum (Beijing), Beijing, 102249, China

ARTICLE INFO

Article history:

Received 2 April 2026

Received in revised form

19 May 2026

Accepted 13 July 2026

Available online 17 July 2026

Edited by Xi Zhang

Keywords:

Pipeline leakage detection

Unmanned aerial vehicle

OPEI

Acoustic signal processing

Bionics

ABSTRACT

To address the low sensitivity of visual UAV inspection in leakage detection for above-ground natural gas pipelines, an acoustic detection method inspired by the owl auditory mechanism was developed. An innovative oscillatory positional encoding imaging (OPEI) technique was proposed, through which one-dimensional acoustic signals were converted into two-dimensional image representations incorporating time-domain, frequency-domain, and amplitude information. Recognition accuracy of 97.38% for OPEI images on the testing dataset was achieved by the EfficientNetV2 cascade model. The results demonstrated that the biomimetic design enhanced acoustic signal sensitivity and enabled fine-grained recognition of pipeline leakage apertures (1/5/10 mm). This study proposes a novel acoustic signal processing method, providing a theoretical foundation for ensuring the reliability and safety of natural gas pipeline transportation.

© 2026 The Authors. Publishing services by Elsevier B.V. on behalf of KeAi Communications Co. Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Pipeline networks, which serve as critical infrastructure for oil and gas transportation, require secure operation. Although most pipelines are buried, approximately 20% of above ground systems, including segments located at stations, in mountainous regions, and on offshore platforms, remain exposed to the environment (Korlapati et al., 2022). Elevated risks of physical damage, corrosion, and leakage are faced by these exposed sections, and efficient and reliable inspection methods are therefore urgently required for routine monitoring.

With advancements in unmanned aerial vehicle (UAV) technology and reduced costs, UAVs have emerged as effective carriers for inspecting above ground pipelines, including ground level segments, station piping, mountainous sections, and offshore platform conduits, due to their flexibility and accessibility (Bretschneider et al., 2024). Current UAV based pipeline leak detection primarily relies on three core methodologies, namely visual inspection, infrared thermometry, and gas sensors such as laser methane detectors (Ma et al., 2025; Usamentiaga, 2024; Iwaszenko et al., 2021). However, significant limitations persist in current UAV based inspection methods. Visual inspection is

prone to blind spots, while infrared data accuracy is influenced by seasonal variations, causing deviations between collected and actual measurements. Gas sensors are affected by rotor induced wind fields, which substantially reduces detection precision. To overcome these constraints, integrating acoustic detection with UAV platforms has been recognized as a promising alternative. Nevertheless, strong self-generated noise from UAVs severely contaminates target acoustic signals, fundamentally compromising detection performance and representing a core challenge for this technology (Aljameel et al., 2024).

Inspired by the exceptional auditory perception of owls in nature, an innovative biomimetic UAV based pipeline leak detection method was developed.

- (1) A spatially separated dual-microphone array in a front rear configuration was adopted for audio signal acquisition. This arrangement simulated the asymmetrical ear canal structures of owls, capturing binaural audio signals that exhibited amplitude and phase differentials.
- (2) An innovative oscillatory positional encoding imaging (OPEI) method was conceptualized to enhance temporal, amplitude, and spectral features within raw signals, mimicking the sound amplification and broadband perception afforded by owl facial morphology.
- (3) An EfficientNetV2 cascaded model was constructed to simulate the highdensity auditory neuron processing of

^{*} Corresponding author.

E-mail address: hzhao@cup.edu.cn (H. Zhao).

Peer review under the responsibility of China University of Petroleum (Beijing).

owls, enabling fine grained classification of leakage apertures (1/5/10 mm).

2. Methods

2.1. Auditory anatomy of owls

The acoustic facial disc structure was visualized using high resolution photography (Fig. 1(a)), highlighting a specialized feather formation forming a facial disc that functions acoustically as a parabolic waveguide. When sound waves acted upon the owl's facial feathers, they caused elastic deformation of the feathers. This structural alteration in the feathers correspondingly modified the path of the sound waves, resulting in the generation of a path length difference, which subsequently induced a phase shift. The exposed ear aperture was revealed after retracting auricular feathers (Fig. 1(b)), demonstrating exceptional enlargement relative to head size and encirclement by specialized pre auricular flaps, which provide protection and funnel sound toward inner ears, enhancing temporal and spectral resolution. The schematic diagram of coronal skull section was presented to illustrate vertical positional asymmetry between the left (L) and right (R) ear canals (Fig. 1(c)), generating interaural phase and amplitude differences when sound waves reach both ears and sharpening spatial localization of acoustic signals (Beatini et al., 2018; Surmacki and Minias, 2025; Shalaby et al., 2024).

2.2. OPEI

To simulate the facial disc structure and feather flaps of owls while maximizing retention of temporal, amplitude, and frequency information, an oscillatory positional encoding imaging (OPEI) method is proposed. This method generates 2D images integrating these three signal characteristics by modulating multi-frequency oscillator phases with signal amplitude.

In this section, $s[n]$ is defined as the single-channel input signal, which is a discrete sequence comprising N sample points. The signal amplitude is initially normalized according to Eq. (1).

$$s_{\text{norm}}[n] = \frac{s[n]}{\max |s[n]|} \quad (1)$$

where $\max |s[n]|$ represents the sensor's upper amplitude sampling limit. $s_{\text{norm}}[n] \in [-1, 1]$ is the normalized signal.

For UAV detection, Eq. (2) can be employed since the inherent noise amplitude of UAV typically reaches this sensor limit.

$$s_{\text{norm}}[n] = \frac{s[n]}{\max_{0 < k < N} |s[n]|} \quad (2)$$

Amplitude differences caused by sensor hardware variations can also be normalized using this approach, ensuring numerical stability.

An oscillator group operating at multiple frequencies is constructed, following characteristics of owl auditory systems (Bae et al., 2024). Given a sampling rate f_s , the Nyquist frequency is $f_s/2$. The k logarithmic-scale frequencies are generated over the interval $[f_{\min}, f_{\max}]$.

$$f_i = 10^{\lg(f_{\min}) + \frac{i}{k-1}(\lg(f_{\max}) - \lg(f_{\min}))} \quad (3)$$

For data sampled by full-frequency microphones covering 20 Hz to 20,000 Hz, parameters are set as $f_{\min} = 10\text{Hz}$, $f_{\max} = \min(0.8f_{\text{nyq}}, 20000)\text{Hz}$, $i = 0, 1, \dots, k-1$.

The time vector is defined as $t[n] = n/f_s$. A phase scaling factor is set to $\alpha = \pi/\max |s_{\text{norm}}|$. Each frequency f_i has its corresponding oscillatory function, generating orthogonal components.

$$V_{i,S}[n] = \sin(2\pi f_i t[n] + \alpha s_{\text{norm}}[n]) \quad (4)$$

$$V_{i,C}[n] = \cos(2\pi f_i t[n] + \alpha s_{\text{norm}}[n]) \quad (5)$$

where $\alpha s_{\text{norm}}[n]$ serves as the phase offset.

An instantaneous value matrix $\mathbf{V} \in \mathbb{R}^{N \times 2k}$, containing outputs from all oscillators, is formed for every time point $t[n]$.

$$\mathbf{V} = \begin{bmatrix} V_{0,S}[0] & V_{0,C}[0] & \cdots & V_{k-1,S}[0] & V_{k-1,C}[0] \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ V_{0,S}[N-1] & V_{0,C}[N-1] & \cdots & V_{k-1,S}[N-1] & V_{k-1,C}[N-1] \end{bmatrix} \quad (6)$$

To map this high-dimensional space onto a 2D image without inducing information folding, dual-channel orthogonal projection is utilized. Weights $\omega_i = 1/f_i$ are defined to assign higher priority to lower frequencies, enhancing visualization stability. Projection coordinates are calculated as Eqs. (7) and (8).

$$X[n] = \sum_{i=0}^{k-1} \omega_i V_{i,S}[n] \quad (7)$$

$$Y[n] = \sum_{i=0}^{k-1} \omega_i V_{i,C}[n] \quad (8)$$

Here, $X[n]$ illustrates temporal evolution while $Y[n]$ represents the frequency distribution. Given a target output image size

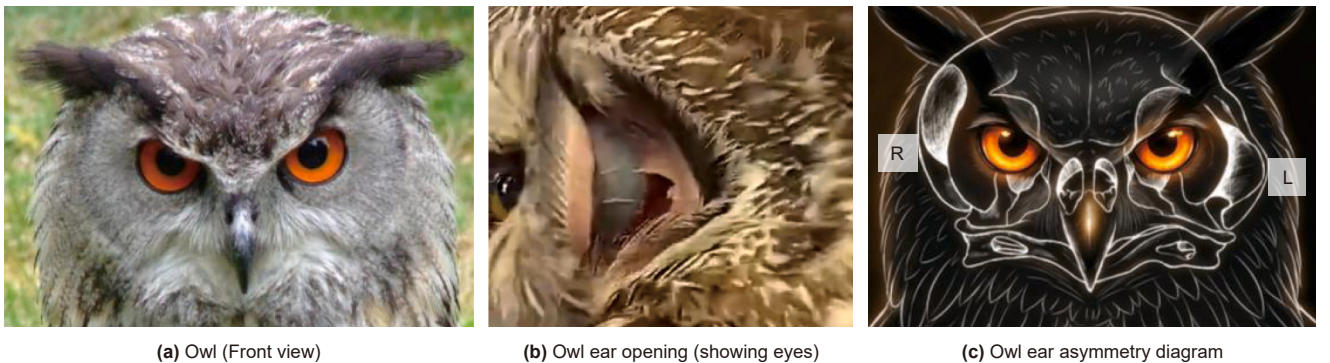


Fig. 1. Morphological characteristics of the auditory system in owls.

$M \times M$, coordinates are mapped linearly to $[0, M - 1]$ per Eqs. (9) and (10).

$$X_{\text{norm}}[n] = (M - 1) \cdot \frac{X[n] - \min X}{\max X - \min X} \quad (9)$$

$$Y_{\text{norm}}[n] = (M - 1) \cdot \frac{Y[n] - \min Y}{\max Y - \min Y} \quad (10)$$

The normalized coordinates are discretized for pixel placement according to Eqs. (11) and (12).

$$x[n] = \lfloor X_{\text{norm}}[n] \rfloor \in \{0, 1, \dots, M - 1\} \quad (11)$$

$$y[n] = \lfloor Y_{\text{norm}}[n] \rfloor \in \{0, 1, \dots, M - 1\} \quad (12)$$

An empty image matrix $\mathbf{I} \in \mathbb{R}^{M \times M}$, $\mathbf{I}_{x,y} = 0$ is initialized. To preserve transient features and prevent information loss caused by overlapping trajectories, pixel intensity at each coordinate is assigned the maximum recorded amplitude at that point, as specified by Eq. (13).

$$\mathbf{I}_{x[n],y[n]} = \max(\mathbf{I}_{x[n],y[n]}, |S_{\text{norm}}[n]|) \quad (13)$$

The raw amplitude value is preserved at non-overlapping points. Image contrast is normalized to enhance visual discriminability according to Eq. (14).

$$\mathbf{I}_{\text{final}} = \frac{\mathbf{I} - \min \mathbf{I}}{\max \mathbf{I} - \min \mathbf{I} + \varepsilon} \quad (14)$$

where $\varepsilon = 10^{-8}$ prevents division by zero.

For audio detection, the OPEI transform for a given single-channel audio signal $s[n]$ at sampling rate f_s is defined as the composite function in Eq. (15).

$$\text{OPEI}(s, f_s) = \Psi \left(\Upsilon \left(\sum_{i=0}^{k-1} \omega_i \mathbf{V}_i[s] \right) \right) \quad (15)$$

where $\mathbf{V}_i[s]$ denotes the dynamic phase modulation component, $\sum \omega_i \mathbf{V}_i$ represents the orthogonal projection operation, Υ corresponds to coordinate mapping, and Ψ denotes the image normalization process.

2.3. Composite OPEI transformation

In the dual-channel stereo detection adopted in this study, the left channel $L(t)$ and right channel $R(t)$ are mapped to two channels within RGB imaging. Specifically, $L(t)$ is mapped to the red channel Eq. (16), while $R(t)$ is mapped to the blue channel Eq. (17).

$$\mathbf{I}_r = \text{OPEI}(L, f_s) \quad (16)$$

$$\mathbf{I}_b = \text{OPEI}(R, f_s) \quad (17)$$

Both channels individually preserve the temporal, amplitude, and frequency information of $L(t)$ and $R(t)$.

To simulate the owl's spatial audio signal analysis capability, particularly its resolution of binaural acoustic disparities. Amplitude and phase differences between left and right channels are encoded and mapped to the green channel. This enhances discriminative features. The amplitude difference is computed as Eq. (18).

$$\Delta A(t) = |L(t)| - |R(t)| \quad (18)$$

The phase difference is derived via the Hilbert transform according to Eqs. (19–21). Let:

$$H[L](t) = \text{Hilbert}(L(t)) \quad (19)$$

$$H[R](t) = \text{Hilbert}(R(t)) \quad (20)$$

$$\Delta \phi(t) = \arg(H[L](t)) - \arg(H[R](t)) \quad (21)$$

A composite differential signal is constructed in the complex domain according to Eq. (22).

$$D(t) = \Delta A(t) \cdot e^{j\Delta \phi(t)} = \Delta A(t) \cos(\Delta \phi(t)) + j\Delta A(t) \sin(\Delta \phi(t)) \quad (22)$$

To concurrently preserve real and imaginary components, the real and imaginary parts are separated and reconfigured into a unitary function according to Eq. (23).

$$D_c(t) = \begin{cases} \text{Re}(D(t/2)), t \in 2n, n \in Z \\ \text{Im}(D((t-1)/2)), t \in 2n+1, n \in Z \end{cases} \quad (23)$$

After restructuring, the signal length expands to $2n$. This restructured signal may be regarded as originating from a channel whose sampling rate corresponds to $2f_s$. Applying the OPEI transformation yields results according to Eq. (24).

$$\mathbf{I}_g = \text{OPEI}(D_c, 2f_s) \quad (24)$$

The final three-channel image is generated according to Eq. (25).

$$\mathbf{I}_{\text{rgb}} = [\mathbf{I}_r, \mathbf{I}_g, \mathbf{I}_b]^T \quad (25)$$

Images generated through the composite OPEI transformation are presented (Fig. 2). Two non-concentric circular patterns are observed near the center in the images corresponding to Fig. 2(a), (c) and (d), which likely correspond to phase differences between dual-channel audio signals. By contrast, distinct circular textures are visually indiscernible at the center of Fig. 2(b). Chromatic blending resulting from multi-channel overlays reduces perceptual clarity for human observation. Nevertheless, the three channels remain independently accessible for computational analysis, where no such limitation is present.

A consistent observation reveals that non-leak regions exhibit a higher proportion of red and blue pixels within the OPEI representation, regardless of signal duration. Conversely, leak points are characterized by elevated proportions of either isolated green pixels or chromatically blended pixels. Furthermore, composite pixel intensity is significantly higher at these leak locations. These distinct visual patterns are attributed to the broader spectral energy distribution across frequency bands at leak points, coupled with lower phase regularity. To avoid oversaturation, long audio is sliced so that the sampling points of the slices do not exceed 80% of the total number of pixels in a single image. To preserve boundary information, there is a 20% overlap in the slices.

2.4. EfficientV2 cascade model

To simulate the owl's neural signal processing capability, a cascaded model based on EfficientNetV2 is proposed. Images generated by the OPEI method contain pixel-level features that may be lost during conventional dimensionality reduction in neural networks. Although dimensionality reduction improves training efficiency, it compromises classification accuracy. EfficientNetV2, a lightweight neural network architecture introduced by Google researchers in 2019 (Tan and Le, 2019), addresses this trade-off. Within lightweight frameworks, greater depth is

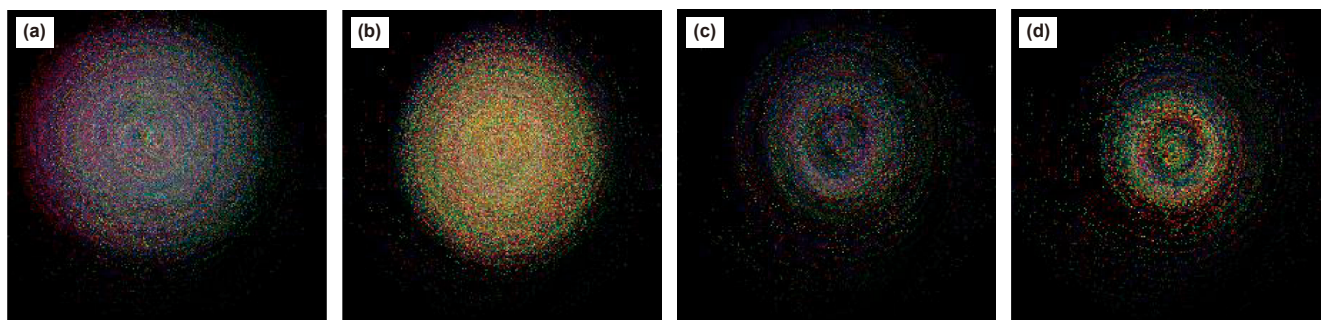


Fig. 2. Composite OPEI transformations. (a) OPEI representation of long-duration non-leak audio. (b) OPEI representation of long-duration leak audio. (c) OPEI representation of short-duration non-leak audio. (d) OPEI representation of short-duration leak audio.

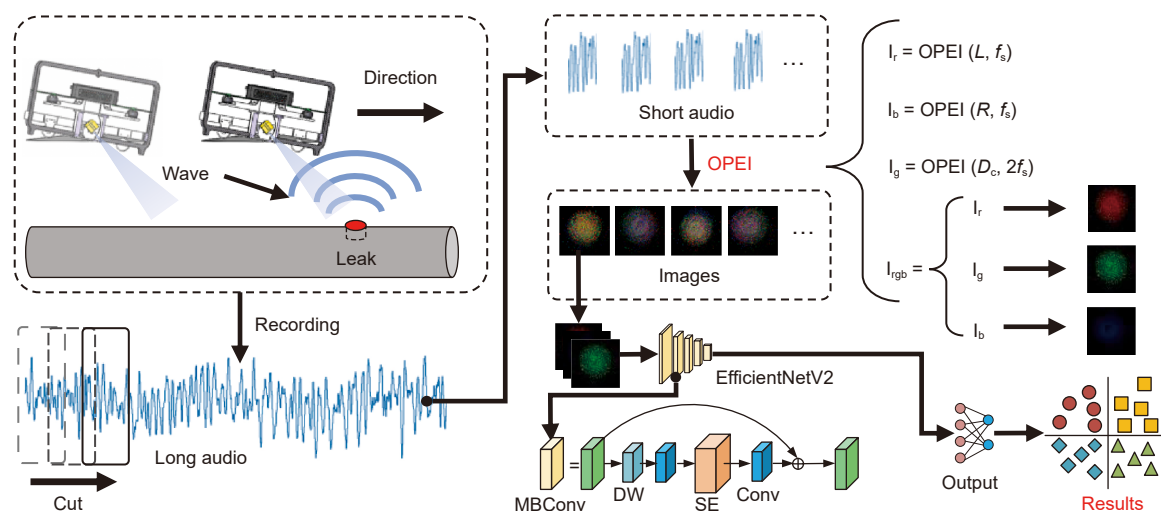


Fig. 3. The overall framework of the proposed method.

incorporated to compensate for reduced receptive fields caused by smaller kernels. Crucially, dynamic regularization replaces dimensionality reduction, ensuring preservation of information. This design enables lightweight deployment on UAVs and facilitates fine-grained pipeline defect identification (no-leak, $D = 1$ mm, $D = 5$ mm, $D = 10$ mm).

3. Model implementation

The overall framework of the model implementation is illustrated in Fig. 3.

Step 1 Audio signals are collected during remotely controlled UAV flight operations, where stereo inputs are acquired and the left and right channels are prepared for subsequent processing.

Step 2 The acquired long-duration audio is segmented into sequential short clips using a 20% overlap ratio to preserve temporal continuity and ensure sufficient data coverage for feature extraction.

Step 3 Each short audio segment undergoes OPEI transformation to generate three-channel images. The left and right channel signals are mapped to the red (R) and blue (B) channels, respectively, while inter-channel phase and amplitude differences are calculated and assigned to the green (G) channel.

Step 4 The transformed images are processed through EfficientNetV2 to produce final classification results for pipeline states, including no-leak, $D = 1$ mm, $D = 5$ mm, and $D = 10$ mm.

4. Results and discussion

4.1. Experimental setup

The experimental setup for UAV-based pipeline leakage detection is illustrated in Fig. 4, where a quadrotor UAV (Fig. 4(a)) serves as the core platform.

The detecting device, designed to emulate owl ear asymmetry, employs two monaural full-frequency microphones positioned asymmetrically at the front and rear of the UAV (Fig. 4(b)).

The steel pipelines tested possess an outer diameter of 220 mm and an inner diameter of 208 mm and are fabricated from St52-grade steel. Circular holes with diameters of 1 mm, 5 mm, and 10 mm were drilled to simulate leaks. Compressed gas, continuously supplied by a 1.0 MPa source located outside the testing area, was used to generate leak sounds while minimizing audio interference.

The UAV was remotely flown approximately 50 cm above the pipelines at a constant velocity of 1 m/s and was manually halted at a 1 m distance from each simulated leak orifice for safety. Audio signals were continuously recorded at a sampling rate of 44,100 Hz for the entire duration of the flight. Data captured while the gas pump was active were designated as audio acquired under specific leak conditions corresponding to the orifice size, whereas audio recorded without gas supply was classified as non-leakage condition data.

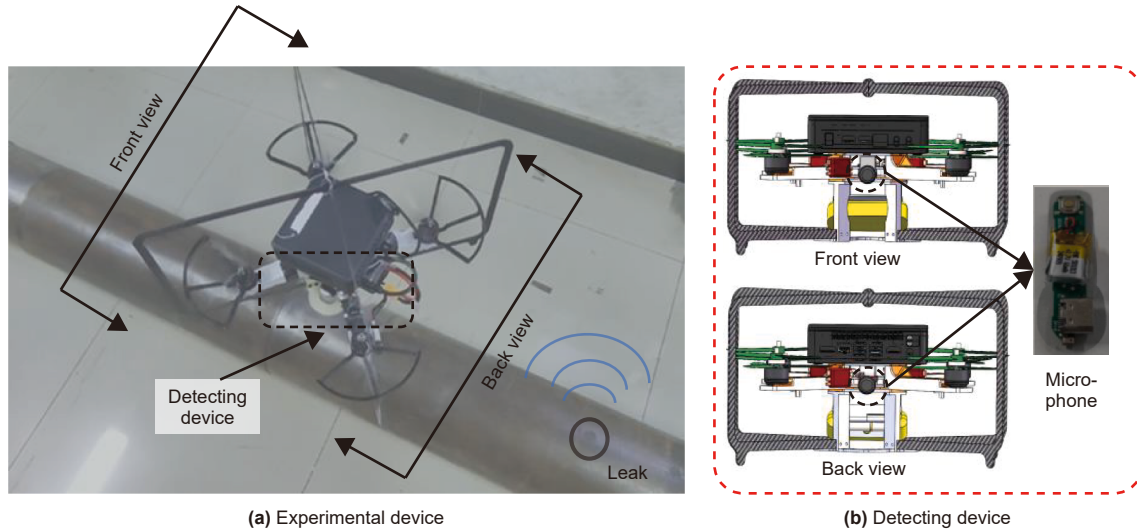


Fig. 4. The experimental setup of UAV-based pipeline leakage detection.

4.2. Quantitative OPEI analysis results

To analyze the importance of features in images, we compared the discrimination scores (Zhao et al., 2026) of these metrics between the non-leakage category and the 10 mm-aperture leakage category (Fig. 5(b)). Additionally, we calculated the normalized mean values of the four features with the highest discrimination scores across all four class: non-leakage, 1 mm-aperture leakage, 5 mm-aperture leakage, and 10 mm-aperture leakage. These results were visualized in Fig. 5(a).

The results demonstrated that the four metrics with the highest discrimination scores were hue mean, sharpness, texture contr, and bright mean. These four metrics exceeded 0.5, indicating that the distance between the centers of the two compared classes surpassed the two ends of the total range. Specifically, Sharpness decreased with increasing leakage aperture size, while hue mean, texture contr, and bright mean increased with larger leakage apertures.

4.3. Preprocessing method comparison

To evaluate classification performance, a dataset of 1000 OPEI-transformed images per category (three leak apertures: 10 mm/ 5 mm/1 mm and no-leak samples) was divided into training/ validation/test sets (8:1:1 ratio) for balanced assessment. Models were trained on an NVIDIA RTX 3090 Suprim GPU with Intel i7-13700k CPU. Each training epoch consisted of 50 steps with a maximum of 30 epochs. Comparative 10-fold k cross-validation under identical conditions assessed eight input types: raw waveform, GASF, WT, MTF, FFT, Composite OPEI transformation (C-OPEI), Dual channel OPEI transformation (D-OPEI), and Single channel OPEI transformation (S-OPEI). Average accuracy, precision, recall, and F1-score metrics were statistically analyzed to evaluate OPEI efficacy and individual component contributions, with the results presented in Fig. 6.

Overall accuracy ranks as follows. C-OPEI achieved 97.38%, D-OPEI achieved 96.62%, S-OPEI achieved 95.12%, WT 96.12%, FFT 94.25%, GASF 93.00%, MTF 92.38%, and raw waveform 88.38%. OPEI

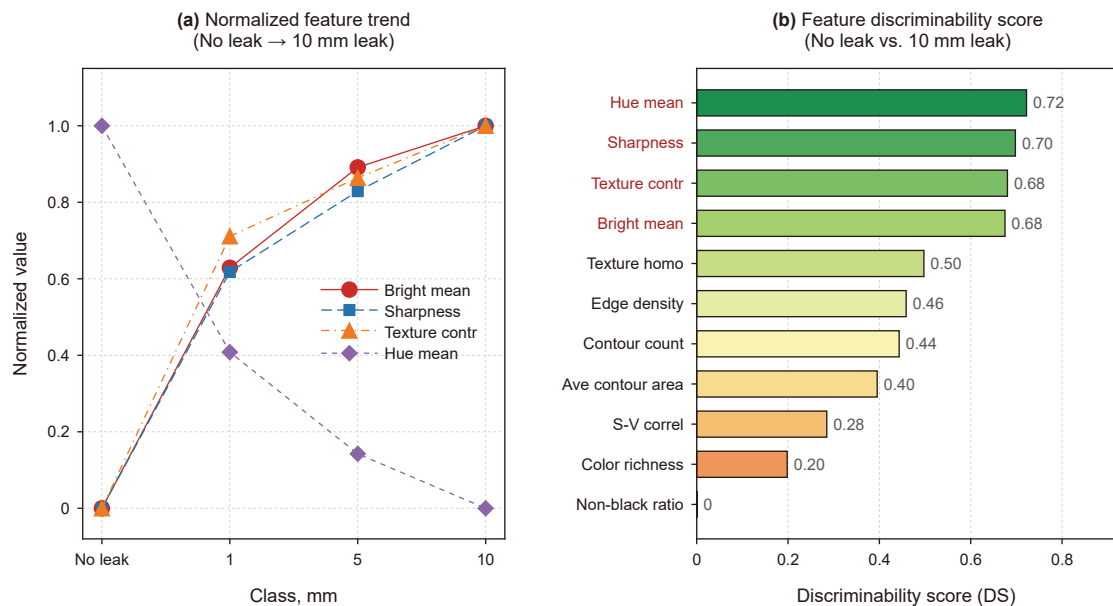


Fig. 5. Feature importance analysis.

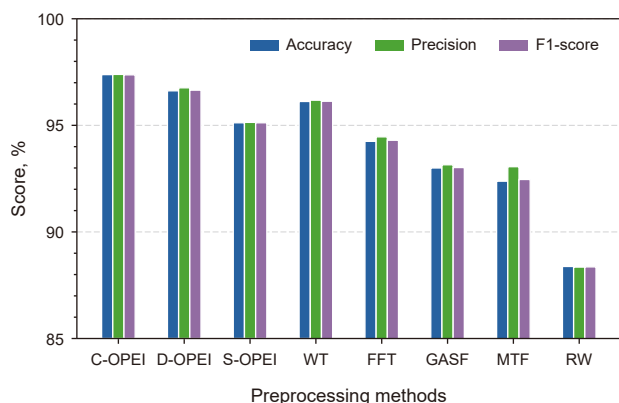


Fig. 6. Classification efficacy of preprocessing methods.

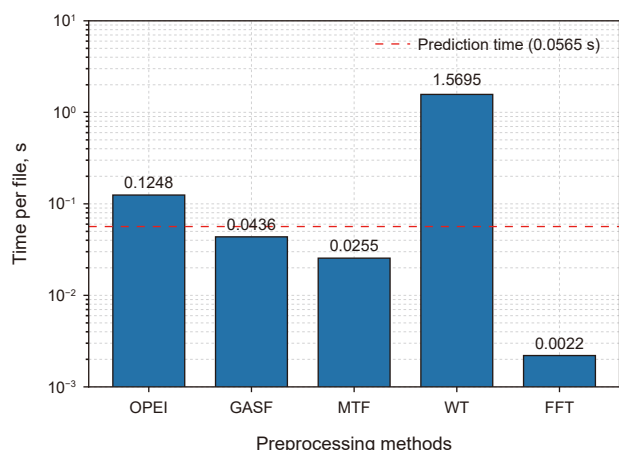


Fig. 7. Processing time comparison.

provides optimal preprocessing by preserving time-frequency characteristics to the greatest extent.

To evaluate the algorithm's real-time performance, we conducted time comparisons between the OPEI method and conventional preprocessing approaches using the identical environment as the training configuration. The average inference time of the trained neural network for single-image prediction was simultaneously recorded Fig. (7).

Empirical results demonstrated that when the pixel density coefficient was 0.8, the duration of a single audio segment measured 0.9102 s. The OPEI method required an average processing time of 0.1248 s per audio segment, while the neural network exhibited an average inference time of 0.0565 s per image. The combined processing time still yielded a computational redundancy of 402.04%, meaning that an onboard computer with merely 1/5 of the training environment's configuration could satisfy the real-time requirements for online detection.

5. Conclusion

This study addressed the limited sensitivity of visual inspection methods using UAVs for detecting natural gas pipeline leaks. A bio-inspired acoustic detection approach was developed, drawing inspiration from the auditory system of owls. Experimental results demonstrated the following.

The biomimetic design substantially enhanced detection sensitivity. OPEI performance exceeded that of traditional

methods. Using the EfficientNetV2 cascade model, classification accuracy for OPEI images reached 97.38%. Significant application potential was demonstrated. This method established a new paradigm for UAV based acoustic pipeline inspection by addressing the persistent challenge of detecting leaks under UAV self generated noise. In addition, its novel image representation technique offers valuable insights for the field of acoustic signal processing.

This biomimetic strategy establishes a transformative paradigm for intelligent pipeline inspection by integrating biological principles with deep learning to enhance industrial safety monitoring.

CRedit authorship contribution statement

Ying-Han Ma: Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Hong Zhao:** Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was financially supported by the National Key Research and Development Program of China (2024YFC3014804).

References

- Aljameel, S.S., Alabbad, D.A., Alomari, D., et al., 2024. Oil and gas pipelines leakage detection approaches: a systematic review of literature. *Int. J. Saf. Secur. Eng.* 14 (3), 773–786. <https://doi.org/10.18280/ijssse.140310>.
- Bae, A.J., Fischer, B.J., Peña, J.L., 2024. Auditory competition and stimulus selection across spatial locations from midbrain to forebrain in barn owls. *J. Neurosci.* 44 (50), e1298242024. <https://doi.org/10.1523/jneurosci.1298-24.2024>.
- Beatini, J.R., Proudfoot, G.A., Gall, M.D., 2018. Frequency sensitivity in Northern saw-whet owls (*Aegolius acadicus*). *J. Comp. Physiol.* 204 (2), 145–154. <https://doi.org/10.1007/s00359-017-1216-2>.
- Bretschneider, L., Bollmann, S., Houssin-Agbomson, D., et al., 2024. Concepts for drone based pipeline leak detection. *Front. Robot. AI* 11, 1426206. <https://doi.org/10.3389/frobt.2024.1426206>.
- Iwaszenko, S., Kalisz, P., Słota, M., et al., 2021. Detection of natural gas leakages using a laser-based methane sensor and UAV. *Remote Sens.* 13 (3), 510. <https://doi.org/10.3390/rs13030510>.
- Korlapati, N.V.S., Khan, F., Noor, Q., et al., 2022. Review and analysis of pipeline leak detection methods. *J. Pipeline Sci. Eng.* 2 (4), 100074. <https://doi.org/10.1016/j.jpse.2022.100074>.
- Ma, Y., Zhao, H., Miao, X., et al., 2025. A UAV-based framework for quick recognition of pipeline defects. *Meas. Sci. Technol.* 36 (1), 015427. <https://doi.org/10.1088/1361-6501/ad9765>.
- Shalaby, W., Kandyel, R., Abumandour, M., et al., 2024. Comparison of anatomical visual features of the eyeball, lens, and retina the diurnal common kestrel (*Falco tinnunculus rupicilaeformis*) and the nocturnal little owl (*Athene noctua glaux*). *BMC Vet. Res.* 20 (1), 541. <https://doi.org/10.1186/s12917-024-04371-7>.
- Surmacki, A., Minias, P., 2025. Evolutionary history reveals information on the functionality of ear tufts in owls (family: strigidae). *Front. Zool.* 22 (1), 35. <https://doi.org/10.1186/s12983-025-00587-x>.
- Tan, M., Le, Q.V., 2019. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, 11946. <https://arxiv.org/abs/1905.11946>.
- Usamentiaga, R., 2024. Semiautonomous pipeline inspection using infrared thermography and unmanned aerial vehicles. *IEEE Trans. Ind. Inf.* 20 (2), 2540–2550. <https://doi.org/10.1109/TII.2023.3295409>.
- Zhao, Z., Duan, M., Lin, Y., et al., 2026. Feature selection for label distribution learning via conditional discrimination index and informational weighted neighborhood rough set model. *Appl. Soft Comput.* 197, 115253. <https://doi.org/10.1016/j.asoc.2026.115253>.