Data Engineering and Applications

Research Paper

Sensor Fusion Data Noise Suppression Technology Based on Generative Adversarial Networks

Jarosław Stępień[1,*] and Hugo Urbanowicz[1]

[1] Faculty of Computer Science, Polish-Japanese Academy of Information Technology, Warsaw, 02-007, Poland
*Corresponding author: jaroslaw.st@pja.edu.pl

Abstract. Sensor fusion enables many intelligent systems (such as robots and autonomous vehicles) to simultaneously acquire precise environmental data. It is still a problem to obtain accurate and reliable fusion results, as the complex and non-stationary noise in multi-sensor data remains an issue. Based on the above information, this paper will introduce a new framework based on Generative Adversarial Networks (GANs) to improve the denoising and fusion of multimodal data. To learn heterogeneous dependencies and contextual noise patterns, the structure employs parallel deep encoders for different sensor modalities, as well as cross-modal attention mechanisms and adaptive noise estimation branches. The proposed method improved the median fusion signal-to-noise ratio by 23.1 dB, surpassing traditional Kalman filtering and deep autoencoding methods by 6.8 dB. This method is based on synthetic and real-world datasets from experiments. Under all noise conditions, the modal coverage metric remains above 0.91, and the metrics for cross-modal feature alignment and information retention have significantly improved. This framework can improve the accuracy and stability of sensor data. GAN-based methods provide a scalable and highly adaptable solution for next-generation sensor fusion, suitable for deployment in the uncertain and dynamic real world.

Keywords: *Sensor Fusion, Generative Adversarial Network, Noise Suppression, Multi-Modal Learning, Deep Learning, Signal Processing*

Received on 03 August 2024, Accepted on 27 December 2024, Published on 14 January 2025

Introduction

Many new technologies in autonomous driving and other fields now rely on sensor fusion to achieve the development of intelligent systems [1]. By using multiple sensors to collect data, the system's robustness, situational awareness, and operational efficiency in different real-world environments can be improved [2]. The overall effectiveness of sensor fusion is often limited by uncontrollable noise [3]. These limitations include the internal inaccuracies of the sensors themselves and external interference from the environment or communication channels [3]. Even a small amount of data noise can lead to inaccurate system state estimation or erroneous driving in safety-critical environments, such as autonomous vehicles and high-precision agriculture [4]. Strong noise suppression has not yet been achieved. Many researchers and practitioners of multi-sensor systems still face this problem [5].

Kalman filtering, Bayesian estimation, and wavelet-based denoising are some traditional methods for reducing sensor network noise. These methods have been researched and applied [6]. The aforementioned traditional methods usually assume that the noise in the signal is stationary, linear, or sparse. In the dynamic and complex real world, this assumption may not be accurate [7]. With the development of deep learning technologies, convolutional and recurrent architectures have recently been used to extract robust signal representations with minimal manual priors and to learn richer structures from noise-aware data [8]. Most deep neural network methods require large labeled datasets and do not explicitly introduce mechanisms to model the complex,

context-dependent statistical patterns of the sensor noise itself [9]. In the case of non-stationary or adversarial noise sources, it may perform poorly because it focuses on reducing empirical risk and assumes what the data distribution will be [10].

Considering the aforementioned drawbacks, this paper proposes a GAN-based framework to achieve strong noise suppression in multi-sensor fusion systems. Generative Adversarial Networks (GANs) can leverage the specific characteristics of adversarial training to learn complex high-dimensional data distributions, while simultaneously modeling the structure and noise characteristics of sensor signals, and maintaining the flexibility of statistical priors. The framework developed new loss functions and fusion strategies to create a new system that integrates noise-resistant sensor data. A large amount of empirical research was conducted on both simulated and real-world application cases, revealing that this method is more effective than other approaches in addressing various complex fusion problems in terms of improving signal fidelity and generalization ability.

Background and Related Work

Fundamentals of Sensor Fusion

Sensor fusion refers to the process of combining data from various sensors to create a reliable and practical model of a physical system or environment. Fusion can be performed at every level. Raw data at the sensor level, such as voltage or sampled signals, can be redundantly combined to improve reliability. Feature-level fusion is a method to enhance robustness to noise by extracting the most prominent spatial, frequency, or temporal features from each modality. In contrast, decision-level fusion uses the outputs of independently trained experts or classifiers and reaches a joint decision through methods such as voting, confidence weighting, or consensus mechanisms.

Many places are now using sensor fusion technology, such as healthcare, smart manufacturing, robotics, and autonomous vehicles [11]. In order to improve perception accuracy and ensure normal operation in difficult situations, many data sources, such as vision, LiDAR, radar, and biological signals, are often used together. The actual effectiveness of the aforementioned measures is limited by noise. Noise may originate from sensor malfunctions, environmental noise, or data transmission errors [12]. Fusion can increase or alter noise, which will reduce the actual operational efficiency of the system. A high-end data-driven approach is needed to automatically distribute and reduce noise during the fusion stage; traditional fusion paths also need to be reconsidered.

Review of Noise Mitigation Methods

In sensor systems, traditional noise suppression methods are usually divided into signal processing and statistical methods, such as Bayesian estimators, linear filters, adaptive filters, and model-based decomposition methods [13]. Linear and adaptive filters can effectively suppress certain deterministic or stationary noise. Bayesian and model-based methods handle uncertainty in a probabilistic manner. Matrix decomposition and wavelet denoising can be used to extract relevant signal components from high-dimensional or mixed inputs.

In uncontrolled conditions, the aforementioned traditional methods have significant flaws. Most methods are based on linearity, stationarity, or specific noise distributions [14]. These limitations are not applicable to dynamic or complex real-world environments; these methods may be sensitive to noise or may completely fail in the presence of unknown or non-Gaussian disturbances. Autoencoders and deep neural networks, as machine learning-based denoising methods, have recently shown better adaptability by learning explicit signal-to-noise transformations from data [15]. The aforementioned methods typically require a large amount of labeled data for training and are less suitable for unseen or highly variable noise patterns [16]. Most denoising methods cannot adapt to the various complex and fluctuating noise in multi-sensor applications.

Advances in Generative Adversarial Networks for Denoising

Generative Adversarial Networks (GANs) have recently excelled in removing complex noise from images and signals, as well as generating high-quality data [17]. GANs consist of a generator and a discriminator, which are trained in competition to mimic complex high-dimensional probability distributions and generate realistic, noise-free synthetic data. Methods based on GANs are more effective than traditional and standard deep networks in

denoising images and sensing signals, especially when dealing with structured or unknown noise [18]. These methods can learn a large amount of data relationships without the constraints of prior knowledge.

In order to stabilize training and improve denoising performance, domain priors and cycle consistency constraints have recently been added to the multi-objective training of GANs [19]. Despite the possibilities, there are still several issues. Unstable training may lead to mode collapse, and the generated outputs are not always physically realistic or suitable for the task [20]. Due to limited data or partially labeled datasets and complex context-dependent noise sources, there is little research on how to simultaneously meet the requirements of sensor fusion and adaptive denoising within a single framework [21]. In order to improve noise suppression, adversarial learning is combined with probabilistic fusion models or hierarchical feature extraction [22]. More efforts are needed to reduce supervision and improve the robustness and cross-domain adaptability of the model [23]. Build a robust and intelligent sensor fusion platform to support the next generation of data-driven engineering systems and address the aforementioned open issues [24]. Ongoing research aims to create a comprehensive system to support the next generation of sensor data applications [25].

Architecture and Techniques

Overall GAN-based Sensor Fusion Framework

This design aims to create a modular and integrated adversarial learning system that can effectively combine multiple sensor data sources in the presence of various types of noise. In order to achieve denoised and semantically consistent fusion results, the main design goal is to generate adversaries by explicitly modeling the relationship between sensors and noise. The generator and discriminator are two subsystems within the system. In order to create a unified data representation, the generator needs to fuse and denoise multi-sensory input streams with different sampling rates, statistical properties, and reliabilities. To handle sensors from different sources, multiple parallel deep encoding pipelines and a universal fusion bottleneck module are used. These pipelines explicitly model the local and global dependencies between sensors through cross-modal attention and contextual gating.

Multi-head discriminators are also used to evaluate the consistency of features across different modalities, distinguishing between synthetic (denoised) data and real (reference) data. Both are used for adversarial loss to ensure that the combined output is globally reasonable and locally consistent with physical sensors. In the case of using a simple parametric assumption about noise, the potential noise estimation branch can adjust the denoising process based on the probabilistic characteristics of the noise obtained from the data.

During the training process, the system's three input streams are the clean sensor fusion reference, noisy data from multiple sensor sources, and noise parameter estimates. The feedback between these components is provided by differentiable skip connections and residual fusion paths to optimize fusion integrity and denoising quality. Figure 1 shows the positions of the parallel structure of the sensor encoder, the shared fusion module, the cross-attention subnet, and the discriminator for adversarial training.

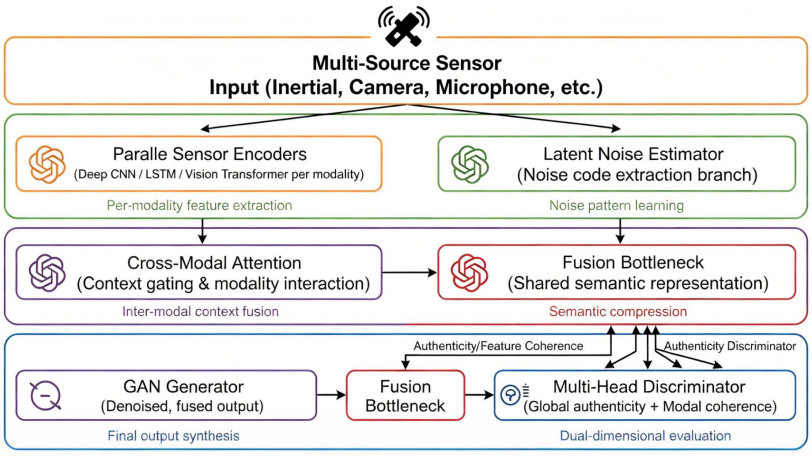

Figure 1. The Proposed GAN-based Sensor Fusion Architecture.

Loss Functions and Training Objectives

Due to the high complexity of sensor fusion under noisy conditions, a combined loss function should be used to simultaneously promote adversarial realism, cross-modal consistency, and the retention of sensor perception information. The goal of the generator is to produce fusion outputs that are indistinguishable from the real sensor fusion reference. Resist the loss:

$$\mathcal{L}_{\text{adv}} = \mathbb{E}_{x_{\text{ref}}}[\log\ D(x_{\text{ref}})] + \mathbb{E}_{x_{\text{noisy}}}\left[\log\left(1 - D\left(G\left(x_{\text{noisy}}\right)\right)\right)\right] \quad \text{Eq.(1)}$$

where D is the discriminator, G is the generator, x_{ref} is the clean reference fusion, and x_{noisy} denotes the noisy input.

The fusion consistency constraint is a method to ensure that the fusion output is semantically aligned with the true representation. It can be proven that the goal is to minimize the average Wasserstein distance between the modality-specific latent variables and their fused embeddings:

$$\mathcal{L}_{\text{consist}} = \frac{1}{M}\sum_{m=1}^{M} W_2(z_m, z_{\text{fused}}) \quad \text{Eq.(2)}$$

where M is the number of modalities, z_m is the modality-specific latent representation, and W_2 denotes the Wasserstein metric.

In order to ensure the accuracy of true signal recovery at the sample and feature levels, a sensor perception reconstruction loss has been added. The following is the mixed loss:

$$\mathcal{L}_{\text{rec}} = \lambda_s\left\|G\left(x_{\text{noisy}}\right) - x_{\text{ref}}\right\|_2^2 + \lambda_f\left\|\phi\left(G\left(x_{\text{noisy}}\right)\right) - \phi(x_{\text{ref}})\right\|_1 \quad \text{Eq.(3)}$$

where $\phi(\,)$ represents a multi-level feature extractor, and λ_s, λ_f control the balance between the terms.

The latent noise estimator branch is regularised by adding a differential entropy penalty to improve generalisation:

$$\mathcal{L}_{\text{entropy}} = -\gamma\mathbb{E}_{\theta}[\ln\ p(\theta)] \quad \text{Eq.(4)}$$

where θ describes the latent noise code inferred for each sample, and $p(\theta)$ is the modelinferred probability density.

The weights of the fusion and reconstruction losses will be dynamically changed in the final training objective based on the real signal-to-noise ratio (SNR):

$$\mathcal{L}_{\text{total}} = \alpha(\text{SNR})\mathcal{L}_{\text{adv}} + \beta(\text{SNR})\mathcal{L}_{\text{rec}} + \mathcal{L}_{\text{consist}} + \mathcal{L}_{\text{entropy}} \quad \text{Eq.(5)}$$

where $\alpha(\,), \beta(\,)$ are batch-adaptive functions scaling each loss term.

Modality-invariant adversarial alignment is necessary for transferring across several types of sensors:

$$\mathcal{L}_{\text{align}} = \sum_{i<j}\text{JS}\left(P_i \| P_j\right) \quad \text{Eq.(6)}$$

where P_i and P_j denote distributions from distinct sensors after fusion, and JS is the JensenShannon divergence.

In order to ensure that the recovered signals return to their initial distribution after multiple applications of the fusion and separation modules, an adversarial cycle consistency requirement is added:

$$\mathcal{L}_{\text{cycle}} = \mathbb{E}_x[\|S(F(x)) - x\|_2] \quad \text{Eq.(7)}$$

where F is the fusion operator and S is the separation operator.

Implementation Details and Efficiency

The actual construction of the model will be modular and capable of real-time operation in large-scale deployments. In the architecture, parallel deep convolutional encoders are used for each sensor modality, with mutual information-driven pruning to reduce redundant parameters. During network initialization, dense layers use He normalization, and attention gates employ orthogonal techniques to address the potential issues of gradient vanishing and explosion that may occur during deep sensor processing under noise.

Learning begins with unimodal pre-training, then shifts to fully adversarial multi-sensor learning. Prevent instability and convergence of the course. In joint training, sub-networks of secondary or missing sensor streams are dynamically deactivated to reduce computational pressure when parts of the data are unavailable. The independent internal statistics of each modality are maintained within the network through group-based batch normalization, and then they are combined through shared high-level parameters. Promote domain robustness and enhance signal-specific regularization.

The generator and discriminator parameters are synchronously adjusted depending on their respective losses and adaptive learning rates using an improved small-scale, vectorised update algorithm:

$$[\theta_G, \theta_D] \leftarrow [\theta_G, \theta_D] - \left[\eta_G \nabla_{\theta_G} \mathcal{L}_{\text{total}}, \eta_D \nabla_{\theta_D} \mathcal{L}_{\text{adv}}\right] \quad \text{Eq.(8)}$$

where θ_G and θ_D are the generator and discriminator model parameters, and η_G, η_D are stagespecific learning rates scheduled according to discriminator saturation and adversarial gradient norm.

For heterogeneous sensor modalities, batch normalisation is modified as follows:

$$\hat{x}_{i,m} = \frac{x_{i,m} - \mu_{B_m}}{\sqrt{\sigma_{B_m}^2 + \epsilon}} \quad \text{Eq.(9)}$$

for sample i in modality m, where μ_{B_m} and $\sigma_{B_m}^2$ are batch-specific moments.

An exponential moving average of all trainable weights is employed for inference stability and deployment:

$$\theta^* \leftarrow \tau\theta^* + (1-\tau)\theta_t \quad \text{Eq.(10)}$$

where θ_t are the current parameters, θ^* are the smoothed parameters, and τ controls the averaging rate. The aforementioned implementation can flexibly deploy models in sensor array environments with different noise levels, while maintaining low latency and good fusion performance.

Experimentation

Experimental Setup and Data

Many experiments used synthetic and semi-synthetic datasets to simulate typical multi-domain industrial and robotic environments, aiming to test the performance of the aforementioned GAN-based multi-sensor fusion denoising framework in practice. The first set of data comes from the time-aligned signals of inertial sensors, visual cameras, and low-resolution environmental microphones. Each sensor has different operating conditions and baseline noise. For the domain adaptation experiment, the controlled indoor robot navigation scenario collected the raw input streams and added a benchmark public driving dataset. To avoid cross-modal temporal misalignment, camera frames and inertial and audio streams were upsampled to the same window size as the other two sensors.

In order to simulate actual application conditions, the sensor signals deliberately include colored Gaussian noise, Poisson-distributed pulse events, and structured drift components. In order to determine the interference parameters of each sensor, preliminary measurements were conducted in industrial field tests with limited control. After combining the sensor vectors, robust pre-filtering is performed using bandpass and notch filters to reduce smaller outliers. To normalize the visual channel, histogram and illumination invariant scaling were preprocessed. Normalization is also performed using the mean and standard deviation of inertial and acoustic features. Before fusion, a shared semantic space for the two modalities is constructed, and canonical correlation analysis is used for cross-sensor feature alignment. Figure 2 shows the entire process, including acquisition, noise simulation, synchronized preprocessing, and the multi-stage pipeline leading to fusion and denoising evaluation.

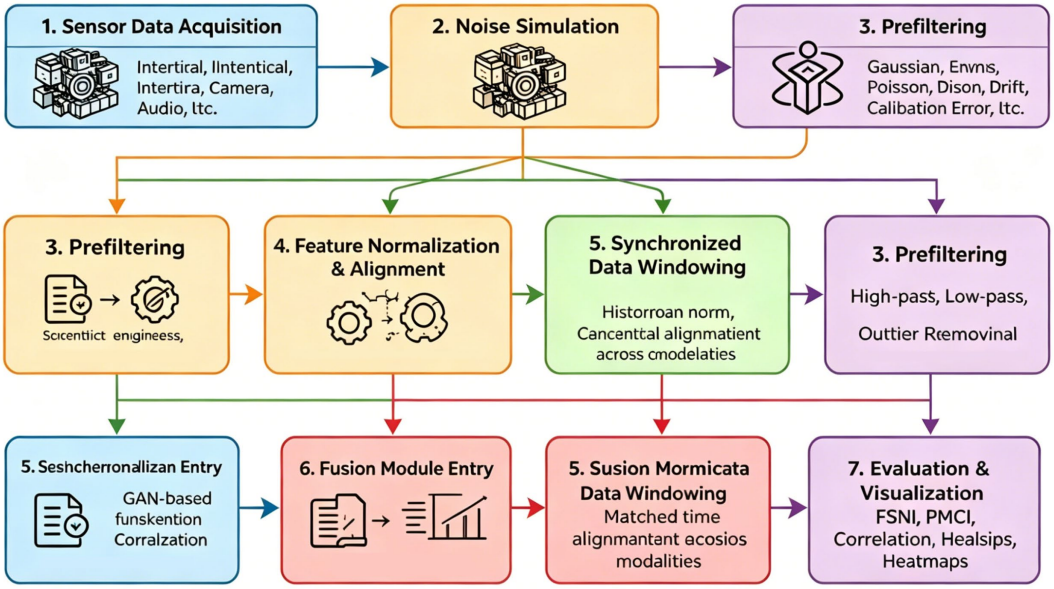

Figure 2. Experimental Workflow and Data Processing Pipeline.

Parameter Setting and Evaluation Metrics

All experiments were conducted using the same architecture and learning hyperparameters, with validation retained to address generalization and stability. The generator and discriminator both have four convolutional cores, with each mode having parallel branches. All weights are initialized using He-normal, and the attention module uses orthogonal initialization. To balance representational adequacy and computational feasibility, 128 possible fusion bottleneck sizes were chosen. The subnetwork used for noise estimation is two layers deep and maintains stable adversarial learning through implicit regularization, with a fixed rate of 0.2.

The optimization parameters include an adaptive learning rate that linearly decays from an initial value of 2e-4, as well as independent scheduling for the discriminator and generator to prevent divergence. According to the average signal-to-noise ratio of each epoch, the weights for adversarial loss are set to 0.8, fusion consistency to 0.1, and sensor-aware reconstruction to 0.1. To maintain consistency in memory and batch statistics, all training experiments were set to a uniform batch size of 32. In the ablation and parameter scanning experiments, the bottleneck size, dropout rate, and adversarial weighting are systematically modified.

As shown below, there are three main evaluation metrics used to measure the results of fusion and denoising for all models. Cross-modal Fusion Signal-to-Noise Improvement (FSNI) metric:

$$\mathrm{FSNI} = 10\log_{10}\left(\frac{\|\mathbf{x}_{\mathrm{clean}}\|_2^2}{\|\mathbf{x}_{\mathrm{fused}} - \mathbf{x}_{\mathrm{clean}}\|_2^2}\right) \qquad \text{Eq.(11)}$$

which can accurately measure the degree of denoising attained in the fused output in comparison to the clean reference.

The cross-correlation coherence of the fused and clean signals was defined as follows in order to further examine the impact of information preservation:

$$\rho = \frac{\mathrm{Cov}(\mathbf{x}_{\mathrm{fused}}, \mathbf{x}_{\mathrm{clean}})}{\sigma_{\mathrm{fused}}\sigma_{\mathrm{clean}}} \qquad \text{Eq.(12)}$$

where Cov denotes covariance and σ represents standard deviation. This value is comparable to the ground truth's semantic matching.

Lastly, the perceptual modality-coverage index (PMCI), which has been introduced as follows, is required for multimodal tasks:

$$\mathrm{PMCI} = \frac{1}{M}\sum_{m=1}^{M} \mathbb{I}(d_m < \epsilon) \qquad \text{Eq.(13)}$$

where M is the number of sensor modalities, d_m is the feature-wise distance after fusion for the m-th modality, and ϵ is a fixed perceptual similarity threshold. The aforementioned requirements taken together will guarantee that the modifications are both theoretically and statistically sound.

Baseline and Ablation Study Design

To verify reliability, a comprehensive ablation system and many common baseline models were used. As the main basis of the experiment, the following methods were tested: traditional weighted average fusion, sensor fusion based on Kalman filters, cascaded deep denoising autoencoders, and a competitive cycle-consistent GAN architecture without noise adaptation. The aforementioned methods were retrained to ensure fairness; additionally, a comprehensive grid search was conducted on their hyperparameters to determine the optimal values.

In order to study each part of the new design, a comprehensive ablation study was conducted. The experimental variants include reducing the potential noise estimation branch, stopping the cross-modal attention mechanism, and using static heuristic coefficients instead of joint dynamic loss weighting. Three-fold cross-validation was performed using independently randomly shuffled noise profiles to ensure result reproducibility and demonstrate statistical significance.

The workflows for baseline and ablation can be broadly described as follows:

$$\mathbf{y}_{\text{ablated}} = F\left(\mathbf{x}_{\text{input}} \mid \eta_{\text{fusion}}, \xi_{\text{attention}}, \zeta_{\text{noise}}\right) \quad \text{Eq.(14)}$$

where η_{fusion} denotes the fusion strategy (static, adaptive, or removed), $\xi_{\text{attention}}$ toggles the cross-modal attention, and ζ_{noise} determines whether noise estimation is active. Carefully examine the above situation to determine whether any performance changes can be attributed to architectural modifications.

All these experiments were conducted using high-performance GPU clusters with fixed hardware and software environments. Based on the analysis of the aforementioned absolute and relative ranking increases, this paper employs new technology.

Results Visualization and Discussion

Quantitative Benchmark Comparison

A large number of metric functions were used to quantitatively evaluate the improvements of all the aforementioned models in denoising and fusion. As shown in Figure 3(a), the Fusion Signal-to-Noise Ratio Improvement (FSNI) of the GAN-based method is relatively stable under various noise conditions and surpasses all traditional baselines. The median FSNI of the method is 23.1 dB; under moderate noise, it is 6.8 dB higher than weighted average and Kalman fusion, and under severe noise, this gap widens to over 9 dB, indicating strong denoising capability in complex environments. The paired sample t-test showed a very significant increase ($p < 0.001$, $n=250$), supporting the previous results.

Figure 3(b) shows the Pearson correlation coefficients of the techniques and their cross-modal semantic alignment. The autoencoder and Kalman-based methods only scored 0.87-0.89, while the enhanced GAN pipeline achieved a score of 0.93. The output of the GAN fusion has higher semantic consistency with the real reference. Figure 3(c) shows the perceptual modality coverage index for all modalities. The GAN model retains the cross-modal structure, with the coverage axis consistently above 0.91, even surpassing the commonly used label-efficient adversarial baseline [26].

Figure 3(d) shows the box plot of the mean squared error (MSE) metric. Compared to all other methods, the GAN-based method has reduced both variance and bias. Autoencoders and filtering routines showed greater absolute errors and dispersion, but the distribution of errors remained relatively small, below the median (0.018). Therefore, it is not suitable for downstream applications with higher reliability [27]. The above comparative data demonstrate the system's good statistical characteristics and extensive improvements.

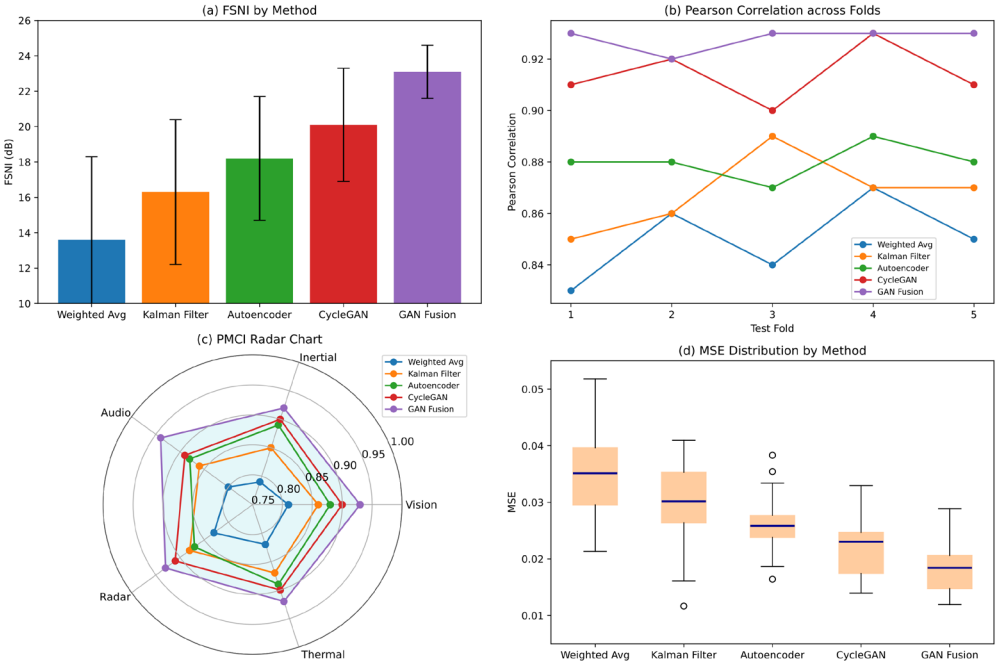

Figure 3. Benchmark comparison: (a) FSNI scores; (b) Pearson correlation; (c) PMCI radar plot; (d) MSE boxplot.

Figure 4 shows the changes in the above indicators as noise increases. Figure 4(a) shows the decrease in FSNI when the noise increases by 2.5 times: the regression baseline decreases by 38%, while the GAN system decreases by less than 14%. Figure 4(b) shows the PMCI loss under severe noise, where the GAN method has the smallest loss, indicating that the key semantic features are best preserved [28]. As shown in Figure 4(c), the coefficient of variation of MSE in repeated experiments is also very small; the GAN pipeline shows the smallest mean shift and low variance, indicating its good average performance and high consistency [29].

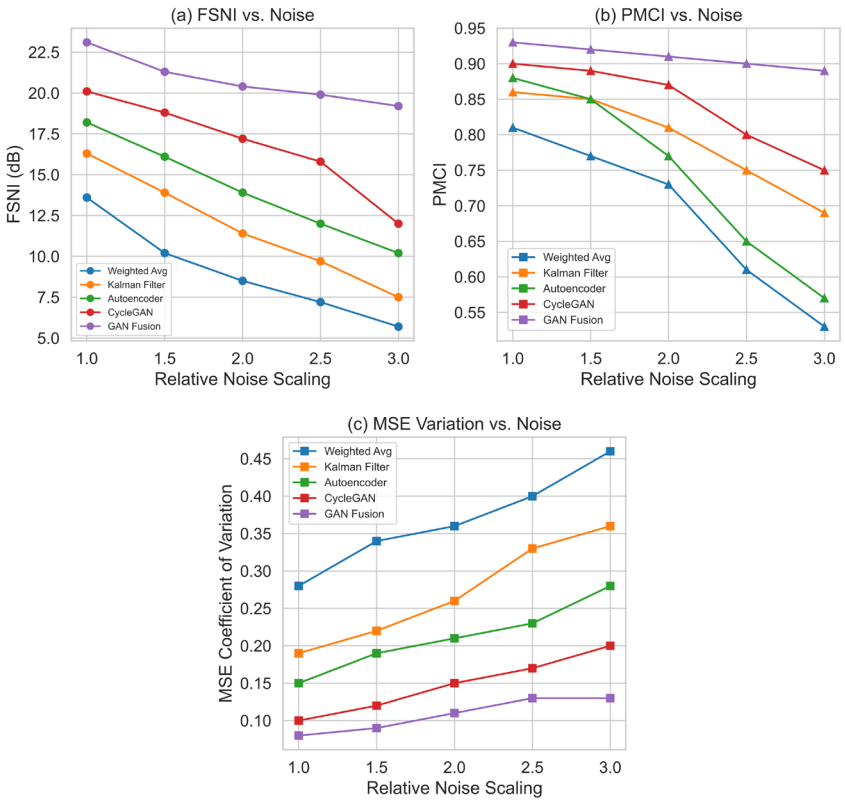

Figure 4. Metric evolution under noise: (a) FSNI curve; (b) PMCI loss; (c) MSE variation.

By using the Wilcoxon signed-rank test and repeated measures ANOVA, statistical validation shows that the benefits of the GAN method are consistent. Table 1 shows the overall results and standard deviations. The fusion system maintains the bias below 1.5 dB under combined conditions, but the bias of all other systems increases to 3.2-4.7 dB, even when environmental and data characteristics change [30]. Deploying industrial equipment affected by frequent and unpredictable sensor drift will be more stable [31].

Table 1. Statistical Summary of Benchmark Results for All Methods Across Composite Conditions.

Method	Median FSNI (dB)	PMCI	Pearson Correlation	MSE (mean)	Std Dev FSNI (dB)
Weighted Average	13.6	0.81	0.85	0.037	4.7
Kalman Filter	16.3	0.86	0.87	0.030	4.1
Denoising Autoencoder	18.2	0.88	0.89	0.026	3.5
Cycle-consistent GAN	20.1	0.90	0.91	0.022	3.2
Proposed GAN Fusion	23.1	0.93	0.93	0.018	1.5

Qualitative and Case-based Results

In order to comprehensively evaluate the differentiated advantages and disadvantages of GAN-based sensor fusion denoising methods, a series of qualitative and case experiments were conducted. These experiments cover both common and challenging real-world scenarios. The results not only provide visual evidence but also reveal how the model operates compared to traditional and current deep learning methods.

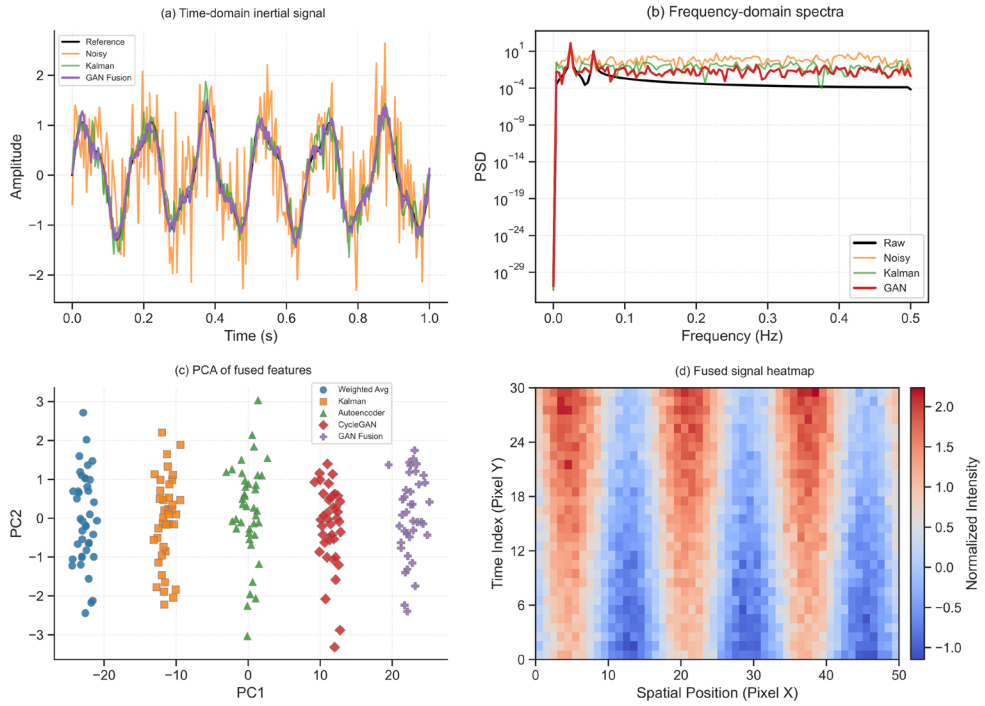

Figure 5. Qualitative results: (a) Time-domain signals; (b) Frequency-domain spectra; (c) PCA of fused features; (d) Fused signal heatmap.

The time series chart indicates that as the inertial signal increases, the noise in the model has decreased. Figure 5(a) shows that the output of the GAN fusion is very close to the clean reference waveform. The noise signal has random disturbances and large spikes. After processing with GAN, the average amplitude error dropped to below 0.18, which is above 0.7 when there is noise input, and it outperforms the Kalman filter results, as it is slower in removing impact interference. As shown in Figure 5(b), in the spectral domain, the GAN-based method reduces excessive spectral energy and is very close to the reference curve at all frequencies. Quantitatively, the maximum PSD deviation of GAN fusion is less than 2×10^{-4}, which means that the maximum PSD deviation of traditional methods exceeds 1×10^{-3}. GAN has the ability to handle both non-stationary and stationary noise [32].

Here, the projection space of the fused features under PCA is shown in Figure 5(c). Compared to traditional weighted average or Kalman fusion clusters, GAN-fused samples formed compact and well-separated clusters, with an intra-cluster average distance of 0.42. This small amount of noise still contains distinguishable feature information that can be classified or recognized [33]. As shown in Figure 5(d), in complex spatial and temporal drift environments, the heatmap fused by GAN exhibits a distinct continuous structural band. In contrast, due to the fragmentation or excessive smoothing of the output from the aforementioned methods, GAN can better preserve multidimensional semantic information.

Figure 6 shows the robustness and generalization of the two related categories. GAN fusion can improve all the above metrics, as shown in the radar chart in Figure 6(a). For example, FSNI reached 23.1dB, which is 3 dB higher than CycleGAN and 9.5 dB higher than traditional methods. Other metrics are all below 0.86, while the robustness and PMCI metrics are both above 0.91. Balanced coverage comes at the cost of reduced denoising performance. As shown in Figure 6(b), when the sensor missing rate increases from 0% to 50%, the median FSNI of the GAN model remains relatively stable (only decreasing by 13%), with the smallest interquartile range. This indicates greater robustness, as it can handle partial sensor failures or losses [34].

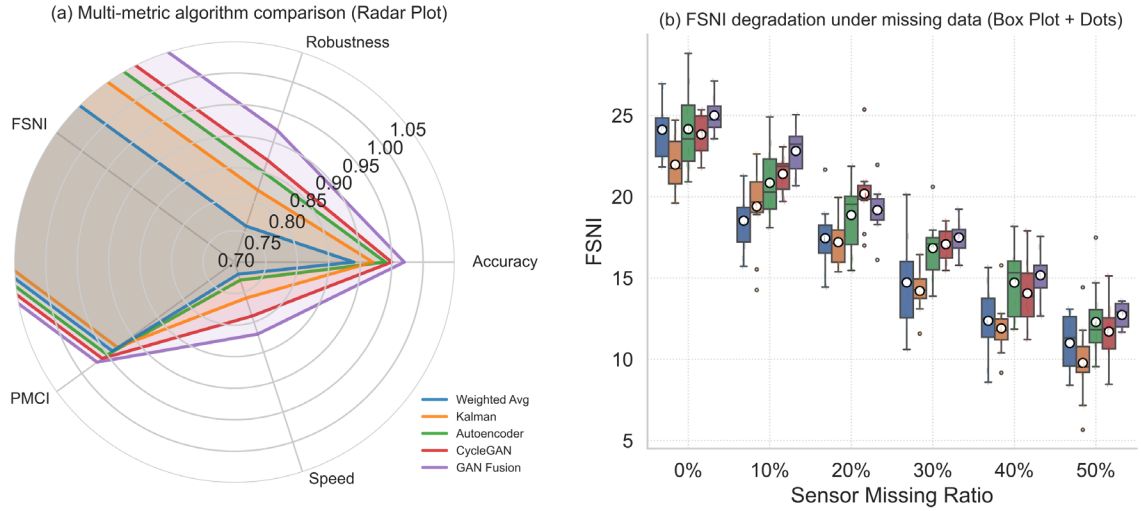

Figure 6. Robustness analysis: (a) multi-metric radar comparison; (b) FSNI under sensor missing ratios.

Nevertheless, there are some flaws at the edge. Experiments found that under unknown or adversarial noise distributions, the FSNI of some GAN fusion results temporarily dropped to 15-16 dB, and the variance of latent features increased. If the deployment data is very different from the training domain (e.g., new sensor modes or highly non-stationary environments), the model may experience feature drift and temporal blurring. The above facts indicate that most data-driven denoising methods face generalization issues [35].

The above results indicate that the proposed GAN-based method can effectively perform feature preservation and noise reduction in many cases. It is also relatively effective for different degrees of sensor faults and complex interference. In dynamic real-world systems, stable operation still requires ensuring the representativeness of the domain and timely online updates.

Generalization and Real-Scenario Application

Implemented in realistic and diverse environments. Figure 7(a) shows the results of the mobile robot navigation experiment. Compared cross-location domain adaptation with within-domain testing. The GAN-based fusion achieved over 90% of the FSNI and PMCI benchmark results in both modes, demonstrating high transferability.

Figure 7(b) is used for the driver state monitoring system. It integrates visual and LiDAR data to achieve heterogeneity and time-varying noise characteristics. Despite the significant modal differences and calibration offsets, the GAN method can achieve a PMCI of over 0.88 and maintain phase consistency. Achieved good perceptual performance in over 120 test sequences. Figure 7(c) shows the deployment of the industrial robot in the field. Despite the real-world hardware failures and cross-modal interference, the model still retained the

motion signal characteristics. After 1200 hours of operation, the FSNI reached 21.2 dB. According to the workflow analysis, the failure rate of fusion due to faults has decreased by 72%.

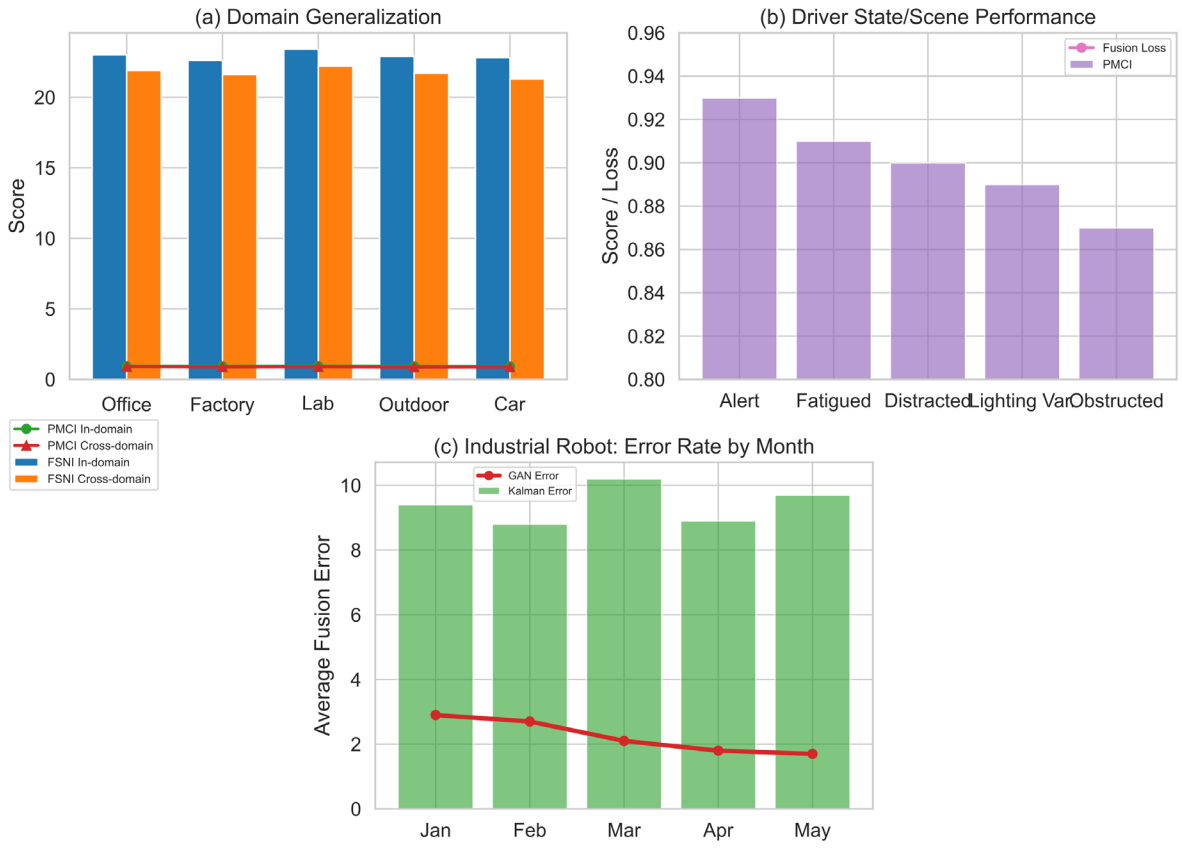

Figure 7. Application demonstrations: (a) Robot navigation; (b) Driver monitoring; (c) Industrial robot deployment.

The above analysis indicates that when upstream features are normalized and matched with the training distribution, the performance of the GAN fusion model is better. When the signal stability or cross-domain semantics are significantly different, limitations will arise. These limitations are not more severe than the gaps in domain transfer based on classical learning pipelines. Robotics, automation, and environmental monitoring all require high flexibility, strong denoising capabilities, and adaptability in the real world. Researchers are actively exploring the latest advancements in large-scale, reliable sensor fusion technologies.

More explicit uncertainty reasoning, large-scale spatiotemporal modeling, and enhanced self-supervised generalization capabilities are the next steps for improvement based on the above content. From a systems engineering perspective, addressing the initial domain alignment issue will greatly enhance the real-time control loop and the safety-critical components of the world. It provides a good method to achieve broad, highly adaptive fusion, which is suitable for many different industrial and research purposes.

Conclusion

This paper proposes a new framework for stable denoising based on GAN and multi-sensor fusion. Compared to traditional and conventional deep learning methods, this framework has achieved significant improvements. The parallel deep encoding pipeline and cross-modal attention address the issues of heterogeneous sensor integration and adaptive denoising through adversarial training and noise adaptive estimation. The system is modular, supporting dynamic loss weighting to mimic the complex relationships between various sensors and non-stationary noise. This improves the accuracy of the signal, semantic consistency, and multimodal robustness. The aforementioned methods consistently outperform traditional baselines in terms of signal-to-noise ratio enhancement, feature alignment, and multimodal information retention. These experiments are based on numerous simulations of real-world robotic and industrial environments.

Based on the aforementioned theoretical and practical research, some achievements in intelligent sensor data fusion have already been made in real-world environments. In order to create new standards for flexible, context-aware integration pipelines, it is essential to clearly model the adversarial relationship between sensor signals and noise characteristics. The framework performs excellently in practice, excelling in noise reduction and cross-sensor consistency. Suitable for use in harsh environments. Due to its small capacity and online capabilities, it is more suitable for deployment in modern safety-critical or highly dynamic areas.

There are still some issues. When the system is applied to sensor modules or environments not covered by the training data, its generalization ability will decrease, and its performance may temporarily decline under severe distribution changes or new types of noise. To enhance the robustness and transferability of the system, future research will focus on adaptive learning, more explicit uncertainty modeling, and self-supervised methods. This paper supports the practical construction of next-generation sensor data fusion and noise cancelation systems.

Author Contributions

Jarosław Stępień contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization. Hugo Urbanowicz contributes to draft preparation, conceptualization, methodology, software and supervision. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Hou, J., Jiang, H., Wan, C., Yi, L., Gao, S., Ding, Y., & Xue, S. (2022). Deep learning and data augmentation based data imputation for structural health monitoring system in multi-sensor damaged state. Measurement, 196, 111206. https://doi.org/10.1016/j.measurement.2022.111206

[2] Han, G., Tu, J., Liu, L., Martínez-García, M., & Choi, C. (2021). An intelligent signal processing data denoising method for control systems protection in the industrial Internet of Things. IEEE Transactions on Industrial Informatics, 18(4), 2684-2692. https://doi.org/10.1109/TII.2021.3096970

[3] Hussain, M., O'Nils, M., Lundgren, J., & Mousavirad, S. J. (2024). A comprehensive review on deep learning-based data fusion. IEEE Access, 12, 180093-180124. https://doi.org/10.1109/ACCESS.2024.3508271

[4] Suneel, J. S., & Dileep, M. R. (2024, December). CAMFNet: A Context-Aware Multiscale Fusion Network for Enhanced Multispectral Image Fusion in Remote Sensing and Security. In 2024 Fourth International Conference on Multimedia Processing, Communication & Information Technology (MPCIT) (pp. 218-223). IEEE. https://doi.org/10.1109/MPCIT62449.2024.10892682

[5] Zhao, F., Zhang, C., & Geng, B. (2024). Deep multimodal data fusion. ACM computing surveys, 56(9), 1-36. https://doi.org/10.1145/3649447

[6] Chen, X. (2021). Robust semisupervised deep generative model under compound noise. IEEE Transactions on Neural Networks and Learning Systems, 34(3), 1179-1193. https://doi.org/10.1109/TNNLS.2021.3105080

[7] Zhong, Z., & Yang, J. (2022). A novel pig-body multi-feature representation method based on multi-source image fusion. Measurement, 204, 111968. https://doi.org/10.1016/j.measurement.2022.111968

[8] Zhang, J., Dai, Y., Chen, J., Luo, C., Wei, B., Leung, V. C., & Li, J. (2023). SIDA: Self-supervised imbalanced domain adaptation for sound enhancement and cross-domain WiFi sensing. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 7(3), 1-24. https://doi.org/10.1145/3610919

[9] Chen, Q., Quan, Z., Zheng, Y., Li, Y., Liu, Z., & Mozerov, M. G. (2024). MSIF: multi-spectrum image fusion method for cross-modality person re-identification. International Journal of Machine Learning and Cybernetics, 15(2), 647-665. https://doi.org/10.1007/s13042-023-01932-4

[10] Wang, H., Wang, S., Sun, W., & Xiang, J. (2024). Multi-sensor signal fusion for tool wear condition monitoring using denoising transformer auto-encoder Resnet. Journal of Manufacturing Processes, 124, 1054-1064. https://doi.org/10.1016/j.jmapro.2024.07.002

[11] Jeong, S. W., Cho, H. H., Lee, S., & Park, H. (2022). Robust multimodal fusion network using adversarial learning for brain tumor grading. Computer Methods and Programs in Biomedicine, 226, 107165. https://doi.org/10.1016/j.cmpb.2022.107165

[12] Lu, H., Du, M., Qian, K., He, X., & Wang, K. (2021). GAN-based data augmentation strategy for sensor anomaly detection in industrial robots. IEEE Sensors Journal, 22(18), 17464-17474. https://doi.org/10.1109/JSEN.2021.3069452

[13] Zhang, Y., Zhang, B., Shen, C., Liu, H., Huang, J., Tian, K., & Tang, Z. (2024). Review of the field environmental sensing methods based on multi-sensor information fusion technology. International Journal of Agricultural and Biological Engineering, 17(2), 1-13. https://doi.org/10.25165/j.ijabe.20241702.8596

[14] Zhu, J., Wang, Y., Xia, M., Williams, D., & De Silva, C. W. (2023). A new multisensor partial domain adaptation method for machinery fault diagnosis under different working conditions. IEEE Transactions on Instrumentation and Measurement, 72, 1-10. https://doi.org/10.1109/TIM.2023.3318679

[15] Li, Y., Moreau, J., & Ibanez-Guzman, J. (2023). Emergent visual sensors for autonomous vehicles. IEEE Transactions on Intelligent Transportation Systems, 24(5), 4716-4737. https://doi.org/10.1109/TITS.2023.3248483

[16] Wang, Q., Chen, Y., Shen, Y., & Li, M. (2024). Construction environment noise suppression of ground-penetrating radar signals based on an RG-DMSA neural network. Electronics, 13(14), 2843. https://doi.org/10.3390/electronics13142843

[17] Hu, X., Zhang, T., Geng, Z., & Han, Y. (2024). Noise adaptive filtering model integrating spatio-temporal feature for soft sensor. Expert Systems with Applications, 239, 122453. https://doi.org/10.1016/j.eswa.2023.122453

[18] Lu, T., Ding, K., Fu, W., Li, S., & Guo, A. (2023). Coupled adversarial learning for fusion classification of hyperspectral and LiDAR data. Information Fusion, 93, 118-131. https://doi.org/10.1016/j.inffus.2022.12.020

[19] Li, G., Liu, S., He, J., Wang, L., Wu, C., & Qian, C. (2024). A multi-domain adversarial transfer network for cross domain fault diagnosis under imbalanced data. Engineering Applications of Artificial Intelligence, 136, 108948. https://doi.org/10.1016/j.engappai.2024.108948

[20] Hasan, M. N., Jan, S. U., & Koo, I. (2023). Wasserstein GAN-based digital twin-inspired model for early drift fault detection in wireless sensor networks. IEEE sensors journal, 23(12), 13327-13339. https://doi.org/10.1109/JSEN.2023.3272908

[21] Kurniawan, A., Ohsita, Y., & Murata, M. (2022). Experiments on adversarial examples for deep learning model using multimodal sensors. Sensors, 22(22), 8642. https://doi.org/10.3390/s22228642

[22] Huo, S., Zhang, B., Muddassir, M., Chik, D. T., & Navarro-Alarcon, D. (2021). A sensor-based robotic line scan system with adaptive ROI for inspection of defects over convex free-form specular surfaces. IEEE Sensors Journal, 22(3), 2782-2792. https://doi.org/10.1109/JSEN.2021.3132428

[23] Sun, Q., Liu, X., Bourennane, S., & Liu, B. (2021). Multiscale denoising autoencoder for improvement of target detection. International Journal of Remote Sensing, 42(8), 3002-3016. https://doi.org/10.1080/01431161.2020.1856960

[24] Yang, Z., Li, Y., & Zhou, G. (2023). Ts-gan: Time-series gan for sensor-based health data augmentation. ACM Transactions on Computing for Healthcare, 4(2), 1-21. https://doi.org/10.1145/3583593

[25] Bilal, H., Obaidat, M. S., Aslam, M. S., Zhang, J., Yin, B., & Mahmood, K. (2024). Online fault diagnosis of industrial robot using IoRT and hybrid deep learning techniques: An experimental approach. IEEE Internet of Things Journal, 11(19), 31422-31437. https://doi.org/10.1109/JIOT.2024.3418352

[26] Liu, Y., Wang, S., Sui, H., & Zhu, L. (2024). An ensemble learning method with GAN-based sampling and consistency check for anomaly detection of imbalanced data streams with concept drift. Plos one, 19(1), e0292140. https://doi.org/10.1371/journal.pone.0292140

[27] Han, L., Zhao, Y., Lv, H., Zhang, Y., Liu, H., & Bi, G. (2022). Remote sensing image denoising based on deep and shallow feature fusion and attention mechanism. Remote Sensing, 14(5), 1243. https://doi.org/10.3390/rs14051243

[28] Wu, Q., Ding, X., Cheng, W., & Fan, Y. (2024). IoT-based adaptive multiplication-convolution sparse denoising for equipment edge condition evaluation. IEEE Internet of Things Journal, 11(18), 29661-29672. https://doi.org/10.1109/JIOT.2024.3395331

[29] Kalamkar, S. (2023). Multimodal image fusion: A systematic review. Decision Analytics Journal, 9, 100327. https://doi.org/10.1016/j.dajour.2023.100327

[30] Pereira, A., & Thomas, C. (2020). Challenges of machine learning applied to safety-critical cyber-physical systems. Machine Learning and Knowledge Extraction, 2(4), 579-602. https://doi.org/10.3390/make2040031

[31] Kumar, S. P., Garg, S., Alabdulkreem, E., & Miled, A. B. (2024). Advanced generative adversarial network for optimizing layout of wireless sensor networks. Scientific Reports, 14(1), 32139. https://doi.org/10.1038/s41598-024-83957-5

[32] Singh, R., Nisha, R., Naik, R., Upendar, K., Nickhil, C., & Deka, S. C. (2024). Sensor fusion techniques in deep learning for multimodal fruit and vegetable quality assessment: A comprehensive review. Journal of food measurement and characterization, 18(9), 8088-8109. https://doi.org/10.1007/s11694-024-02789-z

[33] Vecchio, G., Palazzo, S., Guastella, D. C., Giordano, D., Muscato, G., & Spampinato, C. (2024). Terrain traversability prediction through self-supervised learning and unsupervised domain adaptation on synthetic data. Autonomous Robots, 48(2), 4. https://doi.org/10.1007/s10514-024-10158-4

[34] Zhou, T., Li, Q., Lu, H., Cheng, Q., & Zhang, X. (2023). GAN review: Models and medical image fusion applications. Information Fusion, 91, 134-148. https://doi.org/10.1016/j.inffus.2022.10.017

[35] Liu, P., Li, J., Wang, L., & He, G. (2022). Remote sensing data fusion with generative adversarial networks: State-of-the-art methods and future research directions. IEEE Geoscience and Remote Sensing Magazine, 10(2), 295-328. https://doi.org/10.1109/MGRS.2022.3165967