Scale and Direction Guided GAN for Inertial Sensor Signal Enhancement

Yifeng Wang , Yi Zhao*
School of Science, Harbin Institute of Technology, Shenzhen
wangyifeng@stu.hit.edu.cn, zhao.yi@hit.edu.cn

Abstract

Inertial sensors, serving as attitude and motion sensing components, are extensively used in various portable devices spanning consumer electronics, sports health, aerospace, etc. However, the severe intrinsic errors of inertial sensors greatly restrict their capability to implement advanced functions, such as motion tracking and semantic recognition. Although generative models hold significant potential for signal enhancement, unsupervised or weakly-supervised generative methods may not achieve ideal generation results due to the absence of guidance from paired data. To address this, we propose a scale and direction-guided generative adversarial network (SDG-GAN), which provides dual guidance mechanisms for GAN with unpaired data across two practical application scenarios. In the unsupervised scenario where only unpaired signals of varying quality are available, our scale-guided GAN (SG-GAN) forces the generator to learn high-quality signal characteristics at different scales simultaneously via the proposed self-supervised zoom constraint, thereby facilitating multi-scale interactive learning. In the weakly-supervised scenario, where additional experimental equipment can provide some motion information, our direction-guided GAN (DG-GAN) introduces auxiliary tasks to supervise signal generation while avoiding interference from auxiliary tasks on the main generation task. Extensive experiments demonstrate that both the unsupervised SG-GAN and the weakly-supervised DG-GAN significantly outperform all comparison methods, including fully-supervised approaches. The combined SDG-GAN achieves remarkable results, enabling unimaginable tasks based on the original inertial signal, such as 3D motion tracking.

1 Introduction

Inertial sensors, which can capture the movement and orientation of objects, are widely used in navigation, robotics, gaming, and health monitoring due to their portable wearing and low power consumption [Li *et al.*, 2022; Ehatishamul Haq *et al.*, 2021; Montesinos *et al.*, 2018]. Compared with optical sensors, ultrasonic sensors, and other motion capture equipment, inertial sensors are not affected by light, occlusion, noise, and other external environments during data acquisition [Zhang *et al.*, 2020; Gromov *et al.*, 2019; Liu *et al.*, 2020a].

In localization and navigation, inertial sensors are indispensable for estimating the position and trajectory of objects or vehicles, especially in areas where GPS signals are weak or non-existent, such as indoors, underground, or underwater [Chen *et al.*, 2018; Ferrera *et al.*, 2019]. They can also be integrated with other sensors to enhance the precision and stability of localization and navigation systems [Madgwick *et al.*, 2011; Caesar *et al.*, 2020]. Inertial sensors also play a crucial role in mapping unknown territories by tracking movements [Esfahani *et al.*, 2019b]. In healthcare and sports engineering, these sensors can be worn or attached to medical equipment to monitor physical activities and health status [Harle, 2013]. They offer valuable feedback for rehabilitation, training, and performance improvement [Herath *et al.*, 2020]. For example, they measure the gait of patients with Parkinson disease and detect falls or unusual behaviors of the elderly or patients [Brossard *et al.*, 2020a]. In gaming and virtual reality, inertial sensors enrich user experiences by tracking their movements and orientations [Weber *et al.*, 2021], allowing for more immersive interactions. Inertial sensors are also used to control the movement of virtual characters by recording the user's actions [Wang and Zhao, 2024a]. Moreover, an inertial measurement unit (IMU) can simulate realistic physical effects such as gravity, inertia, and collisions in virtual environments [Liu *et al.*, 2020b]. In summary, inertial sensors are versatile and powerful for object orientation and tracking in various contexts [Li *et al.*, 2023].

However, the major drawback of inertial sensors in practice is their serious noise, including quantization noise, scale factor error, cross-coupling error, etc., which severely affect their accuracy and reliability [Wang and Zhao, 2024b]. Noise can originate from various sources, such as thermal fluctuations, manufacturing defects, and sensor aging [Saha *et al.*, 2022]. Furthermore, these errors can be accumulated in the calculation, thereby making the function realization of inertial signals much more challenging, especially for low-cost

*Corresponding author.

inertial sensors with lower quality and precision [Esfahani *et al.*, 2019a]. Therefore, signal enhancement is critical to ensure the effective utilization of inertial sensors in diverse application scenes [Brossard *et al.*, 2020b].

The inertial sensor signal enhancement methods can be broadly classified into model-driven and data-driven categories. Model-driven methods leverage the physical principles or intrinsic properties of the sensor signals to design denoising algorithms [Nassar *et al.*, 2004]. For instance, Kalman filters, empirical mode decomposition, or wavelet transforms separate the noise from the signal by relying on prior knowledge of the signal or noise characteristics [Skog and Handel, 2009]. In contrast, data-driven methods learn the denoising function from data without relying on the knowledge about signal characteristics, which can adapt to different sensor types and noise patterns, such as non-stationary and non-Gaussian noise [Yuan and Wang, 2023]. Particularly, those generative deep learning models directly map inferior signals to superior signals, providing a more effective way to enhance signals.

Although generative deep learning models have great potential in the signal enhancement task, their performance relies on strictly paired training data, which is often impractical to obtain for inertial sensors [Wu *et al.*, 2019; Yang *et al.*, 2024]. Without paired data as the label for guidance, unsupervised generative models struggle to simultaneously focus on signal characteristics at different scales, and weakly-supervised generative models integrating auxiliary tasks suffer from the interference of auxiliary tasks on the main generation task. Therefore, there are few studies attempting to apply GAN to improve inertial sensor signals. In this paper, we enhance GAN under unsupervised and weakly-supervised training scenarios by imposing scale and direction guidance to GAN, thereby achieving inertial sensor signal enhancement based on unpaired data. The main contributions of this paper are as follows.

- Considering that unsupervised GAN are prone to losing attention to multi-scale features without the guidance of paired data, we propose a scale guided GAN (SG-GAN), which creates multi-scale generation pipelines and designs zoom supervision to facilitate mutual learning between different scales, thereby avoiding poor generation at certain resolutions.
- Considering that weakly-supervised GAN may be disturbed by auxiliary tasks when receiving assistance from them, we propose the learning direction guided GAN (DG-GAN), which adjusts the impact of the auxiliary task on the main task in real-time during training by evaluating the feature consistency between the auxiliary and main task.
- We comprehensively compare the existing IMU signal enhancement methods in terms of Allan variance analysis and four downstream tasks. Experimental results show that both SG-GAN and DG-GAN surpass all comparison methods in all static indicators and downstream tasks. The fused SDG-GAN achieves an unprecedented effect in improving the inertial signal.

2 Methodology

The primary challenge in enhancing low-cost sensor signals is the inability to obtain strictly paired high and low-cost sensor signals, so the end-to-end methods with fully-supervised training are unusable for converting low-quality signals to high-quality ones. To this end, we utilize CycleGAN [Zhu *et al.*, 2017] as the baseline to construct a mapping between the unpaired signals with varying qualities. Considering that no paired data is available as labels for generative guidance, we propose two learning guidance methods for GAN according to the differences in experimental scenarios.

2.1 Scale Guidance for Unsupervised Scenario

Signal features across different scales contain varied levels of information [Saha *et al.*, 2023]. For instance, lower-resolution features primarily reflect global structures and semantic content [Chen *et al.*, 2020], whereas higher resolutions capture intricate local details and textures [Burri *et al.*, 2016]. Models that focus on a single resolution tend to produce results that lack diversity and realism, often resulting in either overly blurred or excessively sharp outputs [Yang *et al.*, 2023]. This issue is particularly critical in transferring signals from low-cost to high-cost sensors, where capturing multi-scale information is essential. To this end, we design scale guidance for GAN, which creates transfer pipelines for multiple scales, and propose zoom supervision to facilitate mutual learning between scales, as shown in Fig. 1.

SG-GAN adopts cycleGAN as the baseline to cope with unpaired input data, and it consists of two cycles. The cycle with superior signals as input can be summarized as: real superior signal $\rightarrow$ GANS2I $\rightarrow$ virtual inferior signal $\rightarrow$ GANI2S $\rightarrow$ virtual superior signal, corresponding to the front panel in Fig. 1. The cycle with inferior signals as input can be summarized as: real inferior signal $\rightarrow$ GANI2S $\rightarrow$ virtual superior signal $\rightarrow$ GANS2I $\rightarrow$ virtual inferior signal, corresponding to the back panel in Fig. 1. Among them, the generator GANI2S transfers inferior signals to superior signals, and the generator GANS2I transfers superior signals to inferior signals. The two cycles have similar architectures and share all the network parameters. Within each cycle, we simultaneously complete the signal transfer at three scales, 1.0x, 1.2x, and 2.0x.

To force generation pipelines of different scales to learn from each other, we devise a zoom constraint, which provides a supervision that the signal generated at a higher resolution and then downsampled is consistent with the signal generated at a lower resolution. This zoom constraint imposes concurrent demands on the generator across various scales. The zoom loss $\mathcal{L}_{zoom}$ is defined as equation 1.

$$\mathcal{L}_{zoom} = \frac{1}{2}\left\| DS^{(n)}(G(x^{(1)})) - \Theta^{(n)}(G(DS^{(n)}(\overset{\leftrightarrow}{x}^{(n)}))) \right\|_2^2, \tag{1}$$

where $x^{(1)}$ is the input signal, $DS^{(n)}$ denotes $n\times$ downsampling, and G is the generator. The first term, $DS^{(n)}(G(x^{(1)}))$, is the result of input signal $x^{(1)}$ being processed by the generator G and then undergoing downsampling. Within the second term, $\overset{\leftrightarrow}{x}^{(n)}$ denotes a signal centered on $x^{(1)}$ and with a

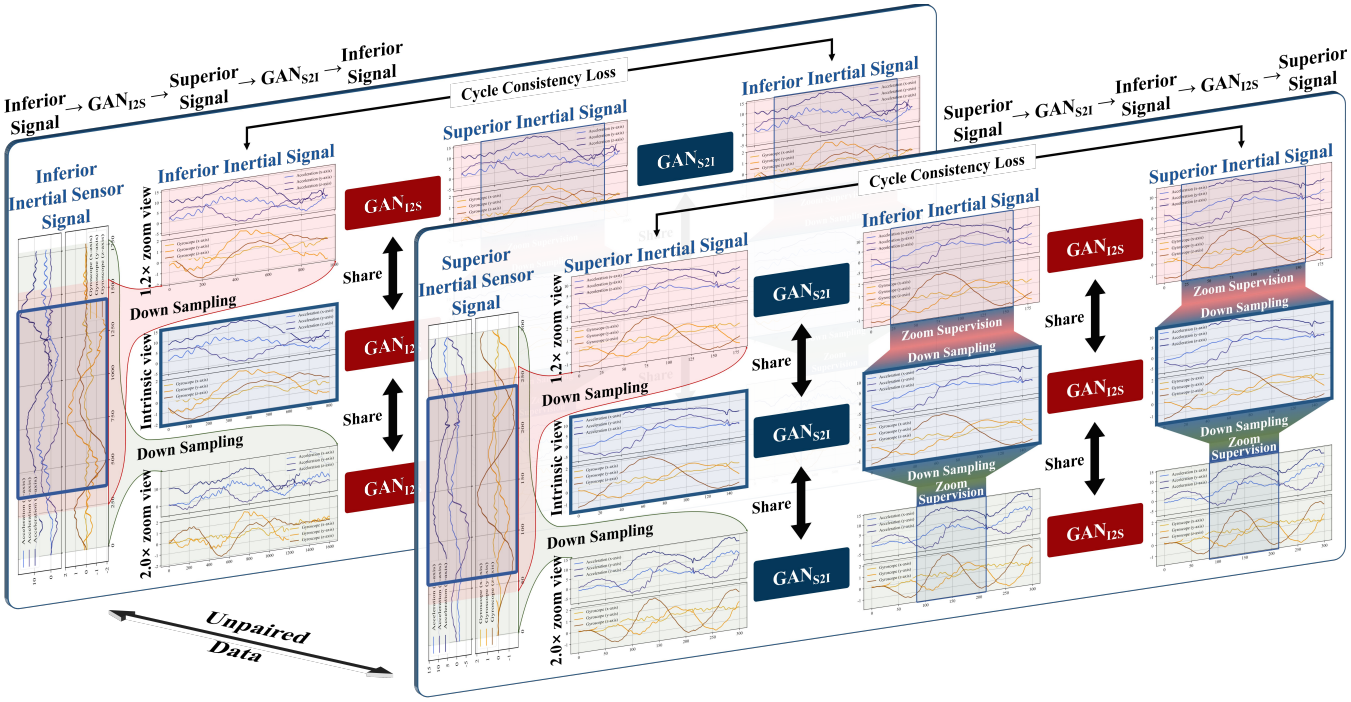


Figure 1: The diagram of scale guidance mechanism. The front and back panels respectively display the two cycles dealing with unpaired input data, which have similar structures and share network structures such as generators and discriminators. In addition, the generation pipelines with different scales also share generators and discriminators. In each cycle, the signals of varying lengths are processed into signals of uniform length but with different resolutions for input into the generator.

length n times that of $x^{(1)}$. After performing n times downsampling $DS^{(n)}$ on $\overleftrightarrow{x}^{(n)}$, the length of $DS^{(n)}(\overleftrightarrow{x}^{(n)})$ meets the input requirement of the generator. $\Theta^{(n)}(sig)$ is a devised signal extraction operator, which extracts the $1/n$ segment of the signal sig centered at sig, e.g. $x^{(1)} = \Theta^{(n)}(\overleftrightarrow{x}^{(n)})$. The expression of $\Theta^{(n)}$ is as follows:

$$\Theta^{(n)}(sig) = sig[\frac{n-1}{2n}len(sig), \frac{n+1}{2n}len(sig)]. \quad (2)$$

where $len(sig)$ denotes the length of the signal sig.

2.2 Direction Guidance for Weakly-Supervised Scenario

When the experimental conditions are upgraded with advanced motion capture instruments such as robotic arms and optical systems, these tools can provide detailed displacement, trajectory, and other motion-related information accompanying inertial signals. While still lacking strictly paired data for generative guidance, the inclusion of this richer supervised information brings the possibility of further signal enhancement. Therefore, we use the robotic arms to record three kinds of motion information during inertial signal collection: attitude, displacement, and motion pattern (such as writing characters), which supports the introduction of three auxiliary tasks: attitude estimation (AE), displacement prediction (DP), and semantic recognition (SR). The GAN is required to synchronously complete these auxiliary tasks during the signal generation process, which forces the generator to provide the signal with higher quality.

However, considering the difference between auxiliary and main generation tasks, auxiliary tasks may also disrupt the learning direction of the generator when providing assistance. To this end, we design a learning direction guidance mechanism. When the learning direction of the auxiliary tasks is consistent with that of the generation task, auxiliary tasks will be given larger weights to better assist the generation task in extracting features from input data. When there is a significant difference between auxiliary tasks and the main task in the learning direction, the weights of auxiliary tasks are suppressed or even zeroed to prevent them from harming the feature extraction of the generation task. The diagram of learning direction guidance is shown in Fig. 2.

The features close to the data end in the generator are generally called low-level features, the features far away from the data end are generally called high-level features, and the features between them are called mid-level features. We set up some auxiliary tasks imposing to different feature layers in the generator consisting of an encoder and a decoder. Specifically, the attitude estimation task is imposed on low-level features in the encoder and decoder, corresponding to losses L_l^e and L_l^d, respectively. The displacement prediction task is imposed on mid-level features in the encoder and decoder, corresponding to losses L_m^e and L_m^d, respectively. The semantic recognition task is imposed on high-level features in the encoder and decoder, corresponding to losses L_h^e and L_h^d, respectively. These auxiliary tasks provide more supervisory information to aid in high-quality signal generation. However, there are differences between the targets of auxiliary and

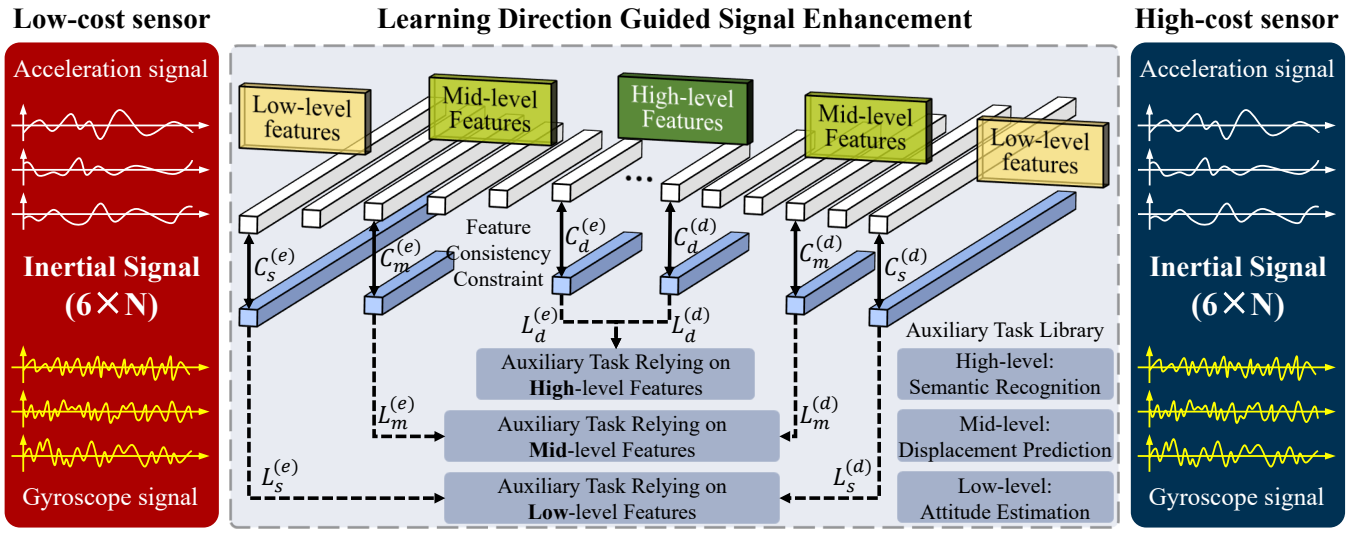


Figure 2: The diagram of direction guidance mechanism. The input low-cost inertial signal is fed into a generator consisting of an encoder and a decoder to generate the high-quality inertial signal. Different auxiliary tasks act on different layer features of the encoder and decoder.

main generation tasks, and naturally, there are differences in the features they rely on. Therefore, the feature extraction process of auxiliary tasks may interfere with the main task, especially in the middle or later stage of training. To this end, we copy the features that three auxiliary tasks rely on. Then, the main task utilizes the original features, and the auxiliary tasks utilize the copied features. We constrain the consistency between the original and copied features to maintain the promotion effect of auxiliary tasks on the main task. The consistency loss of low-level, mid-level, and high-level features at the encoder is C_l^e, C_m^e, and C_h^e. Similarly, The consistency loss of low-level, mid-level, and high-level features at the decoder is C_l^d, C_m^d, and C_h^d. The direction guidance loss $\mathcal{L}_{dir}$ is shown in equation 3.

$$\mathcal{L}_{dir} = \frac{L_l^e}{C_l^e + 1} + \frac{L_m^e}{C_m^e + 1} + \frac{L_h^e}{C_h^e + 1} + \frac{L_l^d}{C_l^d + 1} + \frac{L_m^d}{C_m^d + 1} + \frac{L_h^d}{C_h^d + 1}. \tag{3}$$

In the early stage of training, the main generation task and the auxiliary tasks have strong consistency, and the features they rely on are also highly consistent. Therefore, the feature consistency losses C_l^e, C_m^e, C_h^e, C_l^d, C_m^d, and C_h^d are small. At this time, the auxiliary tasks enjoy large weights since they have a strong promotion effect on the main task. In the middle or later stage of training, the learning direction of the main generation task and the auxiliary tasks diverge, and the features they rely on gradually become inconsistent. The feature consistency losses C_l^e, C_m^e, C_h^e, C_l^d, C_m^d, and C_h^d are large, making the weights of auxiliary tasks small. In summary, this method can evaluate the learning direction of auxiliary and main tasks in real time. When they are consistent (feature consistency is high, feature consistency loss is low), the auxiliary task loss is amplified to enhance the effect of auxiliary tasks. When they are inconsistent (feature consistency is low, feature consistency loss is high), the auxiliary task loss is reduced to suppress the effect of auxiliary tasks.

3 Experiments and Results

3.1 Experiment Dataset

The built-in IMUs are the most widely used and representative low-cost inertial sensors [Jimenez *et al.*, 2009]. We take 15 smartphones with the built-in IMUS to collect the inertial dataset, of which one type of smartphone is employed for collecting the training set, while the data of all the other phones are used for testing. The types of smartphones and their internal IMU specifications are shown in Table 1. It can be found that the price of inertial sensors used in our experiments does not exceed $0.5. Meanwhile, the phone is fixed at the flange of the mechanical arm (ROKAE xMate ER3 Pro), which is employed to accurately record the changes of attitude and position as the labels of the three auxiliary tasks. In addition, an eight-camera optical equipment (Nokov Mars2H) is employed to assist in motion capture. All experiments are implemented by Pytorch 1.10.1 with an Nvidia RTX 2080TI GPU and Intel(R) Xeon(R) W-2133 CPU.

3.2 Comparative Results

Static Evaluation

Allan variance is a classical time-domain technique that provides the quantitative indicators of IMU signal quality, including quantization noise (QN), angle random walk (ARW), velocity random walk (VRW), and bias instability (BI). Based on these indicators, we compare SDG-GAN with the popular IMU signal enhancement methods in recent years, and the results are reported in Table 2. All the comparative methods are implemented strictly following the conditions in their papers or using their open-source codes. Since model-driven methods cannot make flexible adjustments according to the characteristics of the input signal, the data-driven method performs better overall. As the only generative deep learning method in data-driven models, our SDG-GAN achieves unparalleled

Dataset	Smartphone	Release	IMU	Unit price
Training	HUAWEI Mate30 Pro	Sep 2019	ICM20690	$0.28
Testing	HUAWEI P40	Mar 2020	LSM6DSM	$0.30
	HUAWEI P40 Pro	Apr 2020	LSM6DSO	$0.33
	iPhone 7 Plus	Sep 2016	ICM20600	$0.20
	SAMSUNG Galaxy S7	Feb 2016	LSM6DS3	$0.20
	SAMSUNG Galaxy S8	Mar 2017	LSM6DSL	$0.26
	Realme GT	Mar 2021	BMI160	$0.21
	Xiaomi 11	Dec 2020	BHI260AB	$0.30
	OPPO Reno 6	May 2021	ICM-40607	$0.28
	Lenovo Legion Phone	Aug 2020	ICM-42605	$0.20
	VIVO X30	Dec 2019	LSM6DSM	$0.30
	VIVO T2x	May 2022	LSM6DSO	$0.33
	iPhone 13	Sep 2021	Undisclosed	/
	iPhone 12	Oct 2020	Undisclosed	/
	iPhone 11 Pro	Sep 2019	Undisclosed	/

Table 1: The built-in IMU specifications of some smartphones. Note that since the IMUs in some types of iPhones are customized by the manufacturer, the model and price are not disclosed.

superior performance since generative approaches not only eliminate noise but also accurately reconstruct and repair the intrinsic features of the signal. This capability grants generative models remarkable adaptability and efficiency in handling diverse and unknown noise types. Thus, our SDG-GAN trained on one smartphone performs excellently on the test set composed of the remaining 14 smartphones. Hence, the signals of various IMUs with varying usage times can be enhanced by our model without retraining, which brings great convenience to manufacturers considering the rapid iteration of smartphones.

Dynamic Evaluation

Allan variance analysis is usually used to evaluate the noise of IMU signals under static conditions, and it is unavailable to evaluate the signal quality in motion. Therefore, we design four downstream tasks to examine the effectiveness of different methods, which are attitude estimation (AE), position estimation (PE, i.e., displacement prediction), semantic recognition (SR), and trajectory reconstruction (TR). The classic navigation algorithm [Titterton and Weston, 2004] is employed to perform motion trajectory reconstruction, and a ResNet containing 8 bottleneck blocks is constructed for motion semantic recognition.

Considering the issue of inconsistent coordinate systems between the reconstructed trajectory and the real trajectory recorded by the robotic arm, we use the Fréchet spline sliding error (FSSE) [Wang and Zhao, 2023] to calculate the error in the trajectory reconstruction task, which can measure the morphological similarity of two spatial curves even if they are in different coordinate systems. The performance comparison for the four downstream tasks is presented in Table 3. It can be observed that the proposed SDG-GAN still achieves the best performance. Furthermore, we present the visualization results of trajectory reconstruction using different signal enhancement methods, as shown in Fig. 3. Our method is the only one that achieves precise trajectory reconstruction previously deemed impossible with a single inertial sensor.

3.3 Ablation Study

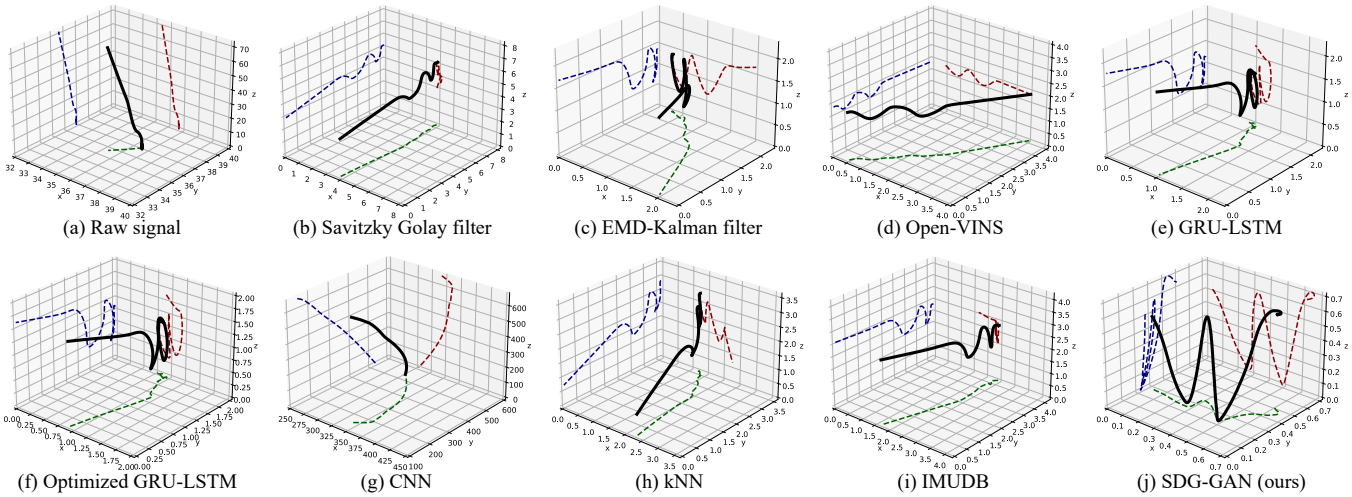
We design ablation experiments for two guidance mechanisms and three auxiliary tasks. From the first row of Table 4, it can be observed that just SG-GAN, without any auxiliary task and direction guidance, surpasses all comparison methods, which demonstrates the effectiveness of the designed scale guidance. We further add auxiliary tasks to SG-GAN (rows 2-4), but the performance fails to improve since the lack of learning direction guidance leads to the interference of auxiliary tasks on the main generation task. When all auxiliary tasks are added to the SG-GAN without the direction guidance (row 5), performance significantly decreases, which proves that additional supervision information may even be a burden if the learning direction cannot be guided. On the other hand, when we adopt direction guidance and remove scale guidance (rows 6-9), introducing any auxiliary tasks can effectively improve the signal, and the more auxiliary tasks we add, the better the performance we obtain. However, there is still a certain gap between the architecture lacking scale guidance and a complete SDG-GAN (row 10), indicating that scale guidance is also indispensable.

3.4 Feature Consistency Analysis

The proposed direction guidance mechanism controls the impact of auxiliary tasks on the main task in real-time during training through the feature consistency between auxiliary and main generation tasks. To intuitively demonstrate the guiding effect, we visualize the feature consistency at each training epoch and apply an exponential weighted moving average to smooth the consistency loss curve to better illustrate the training trend and mode, as shown in Fig. 4. It is observed that the features relied on by the auxiliary tasks gradually diverge from those of the main task, aligning with our motivation. If auxiliary tasks are indiscriminately imposed on the main task, they will cause significant interference during the mid and late stages of training. Furthermore, the impact of different auxiliary tasks on the main task is reflected at various stages. In the early training stage, the attitude estimation task acting on low-level features exhibits strong consistency with the main task, sharing the common goal of improving the signal's apparent quality by eliminating noise, bias, etc. During the mid-stage of training, the displacement prediction task acting on intermediate-layer features aligns closely with the main task, aiming to improve the signal's structural integrity and internal consistency, as well as repairing some basic abstract information. In the later stages of training, semantic recognition task relying on high-level features has strong consistency with the main task, where both strive to mend complex abstract semantic information hidden within the signal, such as deciphering intricate motion patterns or even the identity of the action performer, typically involving a higher-level understanding of the signal. Therefore, adjusting the learning direction of the generator based on the consistency between auxiliary and main generation tasks at different training stages is of utmost importance.

Architecture		Acceleration			Angular Velocity		
		QN	VRW	BI	QN	ARW	BI
Raw signal (No processing)		1.2053	1.9174	3.1186	3.5496	5.9785	9.0625
Model Driven	Savitzky Golay filter [He *et al.*, 2019]	0.5423 (-55.01%)	0.7062 (-63.17%)	0.7595 (-75.65%)	0.9849 (-72.25%)	1.0972 (-81.65%)	1.7411 (-80.79%)
	EMD-Kalman filter [Liu *et al.*, 2020c]	0.4905 (-59.30%)	0.5764 (-69.94%)	0.6429 (-79.38%)	0.7741 (-78.19%)	0.8509 (-85.77%)	1.4032 (-84.52%)
Data Driven	Open-VINS [Geneva *et al.*, 2020]	0.2784 (-76.90%)	0.4805 (-74.94%)	0.5346 (-82.86%)	0.5798 (-83.67%)	0.6635 (-88.90%)	0.8329 (-90.81%)
	GRU-LSTM [Han *et al.*, 2021]	0.5224 (-56.66%)	0.5953 (-68.95%)	0.7097 (-77.24%)	0.7711 (-78.28%)	0.9629 (-83.89%)	1.5301 (-83.12%)
	Optimized GRU-LSTM [Boronakhin *et al.*, 2022]	0.3973 (-67.04%)	0.5296 (-72.38%)	0.6061 (-80.56%)	0.6455 (-81.81%)	0.7197 (-87.96%)	0.9843 (-89.14%)
	Optimized CNN [Chen *et al.*, 2022]	0.5617 (-53.40%)	0.6974 (-63.63%)	0.7726 (-75.23%)	1.0634 (-70.04%)	1.0753 (-82.01%)	1.6425 (-81.88%)
	kNN [Engelsman and Klein, 2023]	0.2034 (-83.12%)	0.313 (-83.68%)	0.4742 (-84.79%)	0.4842 (-86.36%)	0.5533 (-90.75%)	0.6501 (-92.83%)
	IMUDB [Yuan and Wang, 2023]	0.1891 (-84.31%)	0.2743 (-85.69%)	0.4439 (-85.77%)	0.4513 (-87.29%)	0.4957 (-91.71%)	0.5812 (-93.59%)
	SDG-GAN (Ours)	**0.0529 (-95.61%)**	**0.0638 (-96.67%)**	**0.0699 (-97.76%)**	**0.0722 (-97.97%)**	**0.0801 (-98.66%)**	**0.1007 (-98.89%)**

Table 2: Performance comparison of mainstream methods in terms of Allan variance analysis. Considering the units and value ranges of four downstream tasks, we give the enhancement compared with the raw signal in parentheses for convenient comparison. We bold the best and underline the 2nd best results.



(a) Raw signal (b) Savitzky Golay filter (c) EMD-Kalman filter (d) Open-VINS (e) GRU-LSTM

(f) Optimized GRU-LSTM (g) CNN (h) kNN (i) IMUDB (j) SDG-GAN (ours)

Figure 3: Visualization of trajectory reconstruction comparison. The writing motion signal of character 'W' is employed to test 9 signal enhancement methods. The dashed lines in each panel indicate the projection of a reconstructed trajectory in the XY, XZ, and YZ planes.

Architecture		Error of attitude estimation (deg) ↓	Error of position estimation (m) ↓	Accuracy of semantic recognition ↑	FSSE of trajectory reconstruction ↓
Raw signal (No processing)		10.69	1.09	78.42%	0.78
Model Driven	Savitzky Golay filter [He *et al.*, 2019]	6.27 (-41.35%)	0.4279 (-60.74%)	84.99% (+8.38%)	0.2145 (-72.42%)
	EMD-Kalman filter [Liu *et al.*, 2020c]	5.36 (-49.86%)	0.3786 (-65.27%)	85.28% (+8.75%)	0.1826 (-76.53%)
Data Driven	Open-VINS [Geneva *et al.*, 2020]	4.06 (-62.02%)	0.3451(-68.34%)	86.83% (+10.72%)	0.1625 (-79.11%)
	GRU-LSTM [Han *et al.*, 2021]	4.94 (-53.79%)	0.4078 (-62.59%)	86.13% (+9.83%)	0.27095 (-65.17%)
	Optimized GRU-LSTM [Boronakhin *et al.*, 2022]	4.43 (-58.56%)	0.3835 (-64.82%)	86.22% (+9.95%)	0.24595 (-68.39%)
	Optimized CNN [Chen *et al.*, 2022]	5.85 (-45.28%)	0.4782 (-56.13%)	85.52% (+9.05%)	0.2938 (-62.24%)
	kNN [Engelsman and Klein, 2023]	6.47 (-39.48%)	0.4585 (-57.94%)	85.62% (+9.18%)	0.28225 (-63.72%)
	IMUDB [Yuan and Wang, 2023]	3.87 (-63.80%)	0.3179 (-70.83%)	87.69% (+11.82%)	0.1457 (-81.27%)
	SDG-GAN (Ours)	**1.85 (-82.69%)**	**0.1403 (-87.13%)**	**94.74% (+20.81%)**	**0.0512 (-93.42%)**

Table 3: Performance comparison of the proposed method and typical methods for four downstream tasks. Considering the units and value ranges of four downstream tasks, we give the enhancement compared with the raw signal in parentheses for convenient comparison. Note that we bold the best and underline the 2nd best results.

4 Discussion

In this paper, two guidance mechanisms for generative artificial intelligence (AI) models are proposed, which enable the generative model to achieve ideal generation effects without paired training data. This universal guidance of generative AI is not only applicable to the inertial sensor field but is also

Architecture		Acceleration			Angular Velocity			Error of AE (deg) ↓	Error of PE (m) ↓	Acc of SR ↑	FSSE of TR ↓
Guidance	Auxiliary Task	QN	VRW	BI	QN	ARW	BI				
w/ scale w/o direction Guidance	w/o all	0.1657	0.1801	0.1924	0.2065	0.2235	0.2317	3.34	0.2513	87.34%	0.193
	w/ SR	0.1749	0.1915	0.1835	0.1951	0.2159	0.2212	4.12	0.3201	89.24%	0.199
	w/ AE	0.1581	0.1727	0.2081	0.1904	0.2041	0.2383	2.06	0.3159	86.32%	0.189
	w/ DP	0.1626	0.1658	0.1775	0.2138	0.2108	0.2458	3.11	0.1897	85.97%	0.176
	w/ all	0.193	0.2259	0.2306	0.2597	0.2648	0.2794	4.85	0.4033	85.11%	0.203
w/ direction w/o scale Guidance	w/ SR	0.1286	0.1432	0.1733	0.192	0.2064	0.2369	3.76	0.3208	89.03%	0.166
	w/ AE	0.1473	0.1549	0.1629	0.1887	0.1916	0.2292	1.99	0.2742	87.56%	0.152
	w/ DP	0.1419	0.1583	0.1588	0.2031	0.1997	0.2204	2.83	0.1787	88.19%	0.073
	w/ all	0.1093	0.1264	0.1337	0.1547	0.1669	0.1864	1.92	0.1591	91.85%	0.059
SDG-GAN		**0.0529**	**0.0638**	**0.0699**	**0.0722**	**0.0801**	**0.1007**	**1.85**	**0.1403**	**94.74%**	**0.051**

Table 4: Ablation experiments on two guidance mechanisms and four auxiliary tasks. We bold the best and underline the 2nd best results.

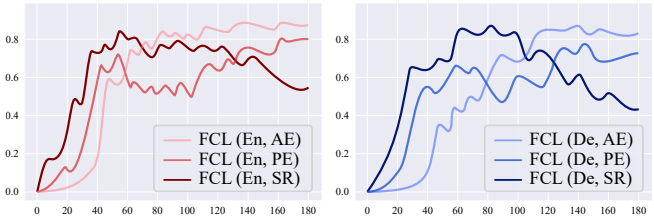


Figure 4: The feature consistency loss (FCL) curves during training. The left and right respectively represent the feature consistency loss curves of the encoder and decoder, where the depth of task dependent features can be reflected by the color depth of the curves.

applicable to areas like audio signal enhancement for communication technologies, equipment signal enhancement for industrial monitoring, biomedical data (such as ECG, EEG, CT, and pathological images) enhancement for accurate diagnosis, and multimodal sensor signal enhancement for autonomous vehicles and intelligent robots.

Deep learning models are often criticized as black-box systems due to their opaque learning processes, which are difficult to explain and control, resulting in low training efficiency and poor learning outcomes. This paper introduces a strategy for guiding the learning direction of deep learning, which may offer insights into the development of interpretable and generalizable artificial intelligence methods.

The experimental results presented in this study reflect the initial performance of the method without extensive hyperparameter tuning. Several reviewers believe that there is substantial room for performance improvement through further hyperparameter optimization and careful training, suggesting that the proposed method can achieve more significant enhancements.

5 Conclusion

In this paper, we propose an SDG-GAN for low-cost inertial sensor signal enhancement based on unpaired data. Considering the lack of paired data as guidance, it is challenging for the generator to achieve signal quality enhancement at different scales simultaneously. Therefore, we propose a scale guidance mechanism that constrains the consistency of signals at different scales through the devised zoom supervision, forcing the generator to realize the transfer of low-cost signals to high-cost signals at all scales, which is a plug-and-play module for GANs with universality.

As the experimental conditions upgrade, researchers may be able to record motion information such as displacement, attitude, and motion pattern while collecting sensor signals, and use them as labels to introduce auxiliary tasks to supervise the signal generation. However, extra supervision information may be harmful to the generation task due to the discrepancy between auxiliary and generation tasks. To this end, we propose a learning direction guidance mechanism that dynamically adjusts the influence of auxiliary tasks on the main task by real-time evaluating the consistency between auxiliary and main generation tasks during the training process, thereby adjusting the learning direction of the generator.

We conduct extensive experiments and compare our method with the latest IMU enhancement methods. The Allan variance analysis is employed for static evaluation, and four downstream tasks are constructed to evaluate the practical effects of improved signals. The comparative and ablation experiments show that either SG-GAN or DG-GAN can surpass all comparison methods on all static indicators and downstream tasks. The signals improved by the completed SDG-GAN achieve unprecedented effects, with an average noise elimination rate of more than 97% and an error elimination rate of more than 85% for downstream tasks. Importantly, signals improved by our SDG-GAN perform satisfactory reconstruction of arbitrary spatial trajectories, which is usually considered an impossible function for a low-cost inertial sensor. Moreover, we prove that the features relied on by the auxiliary tasks indeed gradually diverge from those of the main task. Furthermore, we find the assistance of auxiliary tasks to the main task is reflected in different training stages. Specifically, auxiliary tasks acting on low-level, mid-level, and high-level features optimize generator features in the early, middle, and late training stages, respectively. This finding may bring new insights to generative models and multi-task learning and prove the rationality and necessity of intelligently adjusting the influence of auxiliary tasks on the main task during training.

Acknowledgments

This work was supported by the Science Center Program of National Natural Science Foundation of China under Grant 62188101, the Natural Science Foundation of Guangdong under Grant 2020A1515010812 and 2021A1515011594, and China Scholarship Council under Grant 202306120304. We sincerely appreciate the help provided by Professor Ji Hui from the National University of Singapore.

References

[Boronakhin *et al.*, 2022] Alexander M Boronakhin, Roman V Shalymov, Daniil Yu Larionov, Nguyen Quoc Khanh, and Nguyen Trong Yen. Optimization of an inertial sensor de-noising method using a hybrid deep learning algorithm. In *2022 Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus)*, pages 1335–1340. IEEE, 2022.

[Brossard *et al.*, 2020a] Martin Brossard, Axel Barrau, and Silvère Bonnabel. Ai-imu dead-reckoning. *IEEE Transactions on Intelligent Vehicles*, 5(4):585–595, 2020.

[Brossard *et al.*, 2020b] Martin Brossard, Silvere Bonnabel, and Axel Barrau. Denoising imu gyroscopes with deep learning for open-loop attitude estimation. *IEEE Robotics and Automation Letters*, 5(3):4796–4803, 2020.

[Burri *et al.*, 2016] Michael Burri, Janosch Nikolic, Pascal Gohl, Thomas Schneider, Joern Rehder, Sammy Omari, Markus W Achtelik, and Roland Siegwart. The euroc micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10):1157–1163, 2016.

[Caesar *et al.*, 2020] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[Chen *et al.*, 2018] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[Chen *et al.*, 2020] Changhao Chen, Peijun Zhao, Chris Xiaoxuan Lu, Wei Wang, Andrew Markham, and Niki Trigoni. Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference. *IEEE Internet of Things Journal*, 7(5):4431–4441, 2020.

[Chen *et al.*, 2022] Hua Chen, Tarek M Taha, and Vamsy P Chodavarapu. Towards improved inertial navigation by reducing errors using deep learning methodology. *Applied Sciences*, 12(7):3645, 2022.

[Ehatisham-ul Haq *et al.*, 2021] Muhammad Ehatisham-ul Haq, Aamir Arsalan, Aasim Raheel, and Syed Muhammad Anwar. Expert-novice classification of mobile game player using smartphone inertial sensors. *Expert Systems with Applications*, 174:114700, 2021.

[Engelsman and Klein, 2023] Daniel Engelsman and Itzik Klein. Data-driven denoising of stationary accelerometer signals. *Measurement*, page 113218, 2023.

[Esfahani *et al.*, 2019a] Mahdi Abolfazli Esfahani, Han Wang, Keyu Wu, and Shenghai Yuan. Aboldeepio: A novel deep inertial odometry network for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 21(5):1941–1950, 2019.

[Esfahani *et al.*, 2019b] Mahdi Abolfazli Esfahani, Han Wang, Keyu Wu, and Shenghai Yuan. Orinet: Robust 3-d orientation estimation with a single particular imu. *IEEE Robotics and Automation Letters*, 5(2):399–406, 2019.

[Ferrera *et al.*, 2019] Maxime Ferrera, Vincent Creuze, Julien Moras, and Pauline Trouvé-Peloux. Aqualoc: An underwater dataset for visual–inertial–pressure localization. *The International Journal of Robotics Research*, 38(14):1549–1559, 2019.

[Geneva *et al.*, 2020] Patrick Geneva, Kevin Eckenhoff, Woosik Lee, Yulin Yang, and Guoquan Huang. Openvins: A research platform for visual-inertial estimation. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4666–4672. IEEE, 2020.

[Gromov *et al.*, 2019] Boris Gromov, Gabriele Abbate, Luca M. Gambardella, and Alessandro Giusti. Proximity human-robot interaction using pointing gestures and a wrist-mounted imu. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 8084–8091, 2019.

[Han *et al.*, 2021] Shipeng Han, Zhen Meng, Xingcheng Zhang, and Yuepeng Yan. Hybrid deep recurrent neural networks for noise reduction of mems-imu with static and dynamic conditions. *Micromachines*, 12(2):214, 2021.

[Harle, 2013] Robert Harle. A survey of indoor inertial positioning systems for pedestrians. *IEEE Communications Surveys & Tutorials*, 15(3):1281–1293, 2013.

[He *et al.*, 2019] Jingjing He, Changku Sun, and Peng Wang. Noise reduction for mems gyroscope signal: A novel method combining acmp with adaptive multiscale sg filter based on ama. *Sensors*, 19(20):4382, 2019.

[Herath *et al.*, 2020] Sachini Herath, Hang Yan, and Yasutaka Furukawa. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3146–3152. IEEE, 2020.

[Jimenez *et al.*, 2009] Antonio R Jimenez, Fernando Seco, Carlos Prieto, and Jorge Guevara. A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu. In *2009 IEEE International Symposium on Intelligent Signal Processing*, pages 37–42. IEEE, 2009.

[Li *et al.*, 2022] You Li, Ruizhi Chen, Xiaoji Niu, Yuan Zhuang, Zhouzheng Gao, Xin Hu, and Naser El-Sheimy. Inertial sensing meets machine learning: Opportunity or challenge? *IEEE Transactions on Intelligent Transportation Systems*, 23(8):9995–10011, 2022.

[Li *et al.*, 2023] Peng Li, Wen-An Zhang, Yuqiang Jin, Zihan Hu, and Linqing Wang. Attitude estimation using iterative indirect kalman with neural network for inertial sensors. *IEEE Transactions on Instrumentation and Measurement*, 2023.

[Liu *et al.*, 2020a] Shiqiang Liu, Junchang Zhang, Yuzhong Zhang, and Rong Zhu. A wearable motion capture device able to detect dynamic motion of human limbs. *Nature communications*, 11(1):5615, 2020.

[Liu *et al.*, 2020b] Wenxin Liu, David Caruso, Eddy Ilg, Jing Dong, Anastasios I Mourikis, Kostas Daniilidis, Vijay Kumar, and Jakob Engel. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters*, 5(4):5653–5660, 2020.

[Liu *et al.*, 2020c] Yang Liu, Guangwu Chen, Zongshou Wei, Juhua Yang, and Dongfeng Xing. Denoising method of mems gyroscope based on interval empirical mode decomposition. *Mathematical Problems in Engineering*, 2020:1–12, 2020.

[Madgwick *et al.*, 2011] Sebastian OH Madgwick, Andrew JL Harrison, and Ravi Vaidyanathan. Estimation of imu and marg orientation using a gradient descent algorithm. In *2011 IEEE international conference on rehabilitation robotics*, pages 1–7. IEEE, 2011.

[Montesinos *et al.*, 2018] Luis Montesinos, Rossana Castaldo, and Leandro Pecchia. Wearable inertial sensors for fall risk assessment and prediction in older adults: A systematic review and meta-analysis. *IEEE transactions on neural systems and rehabilitation engineering*, 26(3):573–582, 2018.

[Nassar *et al.*, 2004] Sameh Nassar, KLAUS-PETER SCHWARZ, NASER EL-SHEIMY, and Aboelmagd Noureldin. Modeling inertial sensor errors using autoregressive (ar) models. *Navigation*, 51(4):259–268, 2004.

[Saha *et al.*, 2022] Swapnil Sayan Saha, Sandeep Singh Sandha, Luis Antonio Garcia, and Mani Srivastava. Tinyodom: Hardware-aware efficient neural inertial navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2):1–32, 2022.

[Saha *et al.*, 2023] Swapnil Sayan Saha, Yayun Du, Sandeep Singh Sandha, Luis Antonio Garcia, Mohammad Khalid Jawed, and Mani Srivastava. Inertial navigation on extremely resource-constrained platforms: Methods, opportunities and challenges. In *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pages 708–723. IEEE, 2023.

[Skog and Handel, 2009] Isaac Skog and Peter Handel. In-car positioning and navigation technologies—a survey. *IEEE Transactions on Intelligent Transportation Systems*, 10(1):4–21, 2009.

[Titterton and Weston, 2004] David Titterton and John L Weston. *Strapdown inertial navigation technology*, volume 17. IET, 2004.

[Wang and Zhao, 2023] Yifeng Wang and Yi Zhao. Arbitrary spatial trajectory reconstruction based on a single inertial sensor. *IEEE Sensors Journal*, 23(9):10009–10022, 2023.

[Wang and Zhao, 2024a] Yifeng Wang and Yi Zhao. Handwriting recognition under natural writing habits based on a low-cost inertial sensor. *IEEE Sensors Journal*, 24(1):995–1005, 2024.

[Wang and Zhao, 2024b] Yifeng Wang and Yi Zhao. Wavelet dynamic selection network for inertial sensor signal enhancement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 15680–15688, 2024.

[Weber *et al.*, 2021] Daniel Weber, Clemens Gühmann, and Thomas Seel. Riann—a robust neural network outperforms attitude estimation filters. *Ai*, 2(3):444–463, 2021.

[Wu *et al.*, 2019] Yuan Wu, Hai-Bing Zhu, Qing-Xiu Du, and Shu-Ming Tang. A survey of the research status of pedestrian dead reckoning systems based on inertial sensors. *International Journal of Automation and Computing*, 16:65–83, 2019.

[Yang *et al.*, 2023] Senqiao Yang, Jiaming Liu, Ray Zhang, Mingjie Pan, Zoey Guo, Xiaoqi Li, Zehui Chen, Peng Gao, Yandong Guo, and Shanghang Zhang. Lidar-llm: Exploring the potential of large language models for 3d lidar understanding. *arXiv preprint arXiv:2312.14074*, 2023.

[Yang *et al.*, 2024] Senqiao Yang, Zhuotao Tian, Li Jiang, and Jiaya Jia. Unified language-driven zero-shot domain adaptation. *arXiv preprint arXiv:2404.07155*, 2024.

[Yuan and Wang, 2023] Kaiwen Yuan and Z Jane Wang. A simple self-supervised imu denoising method for inertial aided navigation. *IEEE Robotics and Automation Letters*, 8(2):944–950, 2023.

[Zhang *et al.*, 2020] Xin Zhang, Bo He, Guangliang Li, Xiaokai Mu, Ying Zhou, and Tanji Mang. Navnet: Auv navigation through deep sequential learning. *IEEE Access*, 8:59845–59861, 2020.

[Zhu *et al.*, 2017] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.