

# Wavelet Dynamic Selection Network 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

As attitude and motion sensing components, inertial sensors are widely used in various portable devices, covering consumer electronics, sports health, aerospace, etc. But the severe intrinsic errors of inertial sensors heavily restrain their function implementation, especially the advanced functionality, including motion trajectory recovery and motion semantic recognition, which attracts considerable attention. As a mainstream signal processing method, wavelet is hailed as the mathematical microscope of signal due to the plentiful and diverse wavelet basis functions. However, complicated noise types and application scenarios of inertial sensors make selecting wavelet basis perplexing. To this end, we propose a wavelet dynamic selection network (WDSNet), which intelligently selects the appropriate wavelet basis for variable inertial signals. In addition, existing deep learning architectures excel at extracting features from input data but neglect to learn the characteristics of target categories, which is essential to enhance the category awareness capability, thereby improving the selection of wavelet basis. Therefore, we propose a category representation mechanism (CRM), which enables the network to extract and represent category features without increasing trainable parameters. Furthermore, CRM transforms the common fully connected network into category representations, which provide closer supervision to the feature extractor than the far and trivial one-hot classification labels. We call this process of imposing interpretability on a network and using it to supervise the feature extractor the feature supervision mechanism, and its effectiveness is demonstrated experimentally and theoretically in this paper. The enhanced inertial signal can perform impracticable tasks with regard to the original signal, such as trajectory reconstruction. Both quantitative and visual results show that WDSNet outperforms the existing methods. Remarkably, WDSNet, as a weakly-supervised method, achieves the state-of-the-art performance of all the compared fully-supervised methods.

## Introduction

Inertial sensors can provide information about the position, velocity, attitude, and movement patterns of a moving object in different environments (Shaeffer 2013; Madgwick, Harrison, and Vaidyanathan 2011; Esfahani et al. 2019b). Due to their small size, portable wearing, and low power

consumption, they are widely used in various applications such as navigation (Li et al. 2022), robotics (Gromov et al. 2019), gaming (Ehatisham-ul Haq et al. 2021), and health monitoring (Montesinos, Castaldo, and Pecchia 2018). In addition, inertial sensors are not affected by light, occlusion, noise, and other external environments during data acquisition compared with optical sensors, ultrasonic sensors, and other motion capture equipment (Liu et al. 2020a; Li et al. 2023; Chen et al. 2018).

In the field of localization and navigation, inertial sensors can be used to estimate the location and trajectory of an object or a vehicle, which play a critical role when GPS signals are unavailable indoors, underground, or underwater (Zhuang et al. 2023; Zhang et al. 2020; Ferrera et al. 2019). Inertial sensors can also be combined with other sensors, such as cameras, lasers, or radio signals, to improve the accuracy and robustness of localization and navigation systems (Qin, Li, and Shen 2018; Weber, Gühmann, and Seel 2021). Moreover, inertial sensors can serve to create maps of unknown environments by tracking motion (Esfahani et al. 2019a). In the field of health care and sports engineering, inertial sensors can be worn on the body or attached to medical devices to monitor the physical activity and health status of a person (Qiu et al. 2022). For instance, inertial sensors are employed to measure the step length, cadence, and stride variability of a Parkinson’s patient (Nguyen et al. 2017), and detect falls, injuries, or abnormal behaviors of elderly people or patients (Tunca, Salur, and Ersoy 2019). In the field of gaming and virtual reality, inertial sensors create immersive and interactive virtual reality experiences by tracking the motion and orientation of the user (Caserman, Garcia-Agundez, and Göbel 2019; Wang and Zhao 2024). Inertial sensors are also used to control the movement of virtual characters by mimicking the user’s actions (Pan and Hamilton 2018). Moreover, an inertial measurement unit (IMU) can simulate realistic physical effects such as gravity, inertia, and collisions in virtual environments (Foehn et al. 2022). In summary, inertial sensors are versatile and powerful for object orientation and tracking in various contexts (Liu et al. 2020b; Burri et al. 2016).

However, the major drawback of inertial sensors in practice is the presence of noise, including quantization noise, scale factor error, cross-coupling error, etc., which can severely affect their accuracy and reliability (Chen et al.

\*Corresponding author.

Copyright © 2024, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

2020). Furthermore, these errors can be accumulated in the calculation, thereby making the function realization of inertial signals much more challenging (Wang and Zhao 2023; Saha et al. 2022, 2023). Therefore, signal enhancement is critical to ensure the effective utilization of inertial sensors in diverse application scenes (Caesar et al. 2020).

The inertial sensor signal denoising methods can be broadly classified into model-driven and data-driven categories (Golestani and Moghaddam 2020). Data-driven methods use machine learning techniques, such as k-nearest neighbors (kNN), artificial neural networks (ANN), or convolutional neural networks (CNN) (Engelsman and Klein 2023) to learn the denoising function from data, without relying on explicit models of the signal. However, data-driven methods for IMU denoising require large amounts of labeled data for training, which may be impractical in some cases (Herath, Yan, and Furukawa 2020). Furthermore, deep learning methods may suffer from overfitting or generalization issues when dealing with noisy or inferior signals, especially for low-cost sensors (Yuan and Wang 2023). These defects make the data-driven methods weaker in trustworthiness and interpretability than model-driven methods.

In contrast, model-driven methods can exploit the physical principles or intrinsic properties of the sensor signals (Brossard, Bonnabel, and Barrau 2020) to design accurate and robust denoising algorithms (Min et al. 2021). Relying on prior knowledge of the signal or noise characteristics, Kalman filters, empirical mode decomposition, or wavelet transforms are employed to separate the noise from the signal (Liu et al. 2020c; He, Sun, and Wang 2019). Among them, wavelet-based methods are considered the powerful and effective strategy for signal enhancement, which can adapt to signal characteristics, such as peak widths, frequency bands, or noise levels, and preserve the signal features, such as spikes, edges, or transients, in both time and frequency domains by using different wavelet functions (Saydjari and Finkbeiner 2022). However, the accuracy and efficiency of the wavelet-based methods rely on their wavelet basis functions, which have specification characteristics, including filter length, symmetry, vanishing moment, regularity, and similarity, which determine their suitability to different types of signals. Therefore, it is essential to select wavelet basis functions in the wavelet implementation for signal enhancement.

Currently, although wavelet-based methods are widely used in the field of signal processing, the selection of wavelet basis is subjective or empirical, as there is no universal criterion or rule to choose an appropriate wavelet basis function for a given signal. To this end, we propose a wavelet dynamic selection network (WDSNet), which intelligently selects the appropriate wavelet basis according to the characteristics of the input signal for the best signal enhancement. Specifically, we propose a category representation mechanism that acts on the fully connected (FC) layer, making each column of its weight matrix encodes a category (i.e., a candidate wavelet). In this way, the characteristics of each wavelet are memorized by the network, so the WDSNet can find the most suitable wavelet based on the feature of the input data. The contributions of this paper are as follows:

- This paper proposes a universal wavelet basis selection paradigm, WDSNet, to achieve inertial sensor signal quality enhancement for the first time. The WDSNet has the advantages of not only the flexibility of data-driven methods but also the reliability and interpretability of model-driven methods.
- This paper proposes a category representation mechanism (CRM) that can improve the category awareness ability by constructing category features in the classifier without extra training parameters, thereby capturing the properties of candidate wavelets for suitable selection.
- This paper proposes a feature supervision mechanism (FSM) that uses the network parameters to directly supervise feature extraction by giving the net interpretability. Specifically, the designed CRM turns a fully connected net into a category representation dictionary, providing closer supervision than the far class labels, thus alleviating gradient vanishing, which is verified experimentally and theoretically in this paper.
- We comprehensively compare the existing IMU signal enhancement methods in terms of Allan variance analysis and four downstream tasks. The experimental results show that the proposed WDSNet achieves state-of-the-art performance and has significant superiority compared with all current IMU signal enhancement methods.

## Related Works

### Model-Driven Methods for IMU Signal Denoising

Model-driven methods for IMU denoising aim to exploit the physical properties of the inertial sensors and the motion dynamics to reduce the noise. One common approach was to use various filters to correct the signal (Wu et al. 2021). However, it required accurate sensor performance parameters and noise distribution, which are often difficult to obtain in practice (Toft et al. 2020). Another common approach was the utilization of wavelet-based methods to enhance IMU signal quality. For example, the wavelet thresholding method was widely used to remove the high-frequency noise in the IMU data and preserve the low-frequency signals (Wu et al. 2019), which improved the accuracy of attitude estimation and navigation. However, current wavelet-based methods rely on manually selecting a wavelet basis, which may not be optimal for variable noise types and levels.

### Data-Driven Methods for IMU Signal Denoising

Data-driven methods for enhancing the signal quality of inertial sensors have attracted considerable attention in recent years due to the increasing demand for accurate and robust inertial sensing. Chen *et al.* (Chen, Taha, and Chodavarapu 2022) assumed that IMU signals could be approximated as constants in a short time. Based on this assumption, they divided an IMU signal into multiple short segments, each segment being labeled with a constant as the ideal value, and then used a CNN to predict the ideal value based on these segments. However, this would result in the enhanced signals changing from continuous signals to square wave pulse signals. Moreover, since all the training samples only covered linear and circular motions, this signal enhancement

method became invalid under the scenario of complicated or arbitrary motion. Han *et al.* (Han et al. 2021) fixed the IMU on a high-precision turntable, and collected gyroscope data and real angular velocity data at the same time. Using the ideal data, they built a hybrid deep recurrent neural network (GRU-LSTM), which implemented end-to-end supervised training. This method achieved good signal enhancement results on the training set, but the signal enhancement results on the test data were unsatisfactory. Further, Boronakhin *et al.* (Boronakhin et al. 2022) used Bayesian optimization to optimize the hyperparameters of the GRU-LSTM model, which improved the model performance on the test set, but due to the lack of interpretability of deep learning models, the generated signal has poor trustworthiness and even loses the semantic information of the input signals (Shamwell et al. 2019). Even though the generated signals were better than the input signals in some quantization indicators (such as quantization noise, angle random walk, velocity random walk, and bias instability), they performed worse than the original signals in downstream tasks such as attitude estimation and trajectory reconstruction (Huang et al. 2019). In summary, due to the mechanism that the deep learning method achieves signal enhancement in a generative paradigm, the signal enhancement process is uncontrollable. Albeit the generated IMU signals have improvements in quantization indicators, they are practically unusable due to the lack of authenticity and reliability.

## Methodology

### Framework Overview

Wavelet can extract local spectral and temporal information simultaneously, capture the transient features of signals, and is capable of multi-resolution analysis. It is a powerful signal enhancement tool that meets various denoising requirements (Ma et al. 2020). In this process, the selection of wavelets is crucial. Ideally, the selected wavelet should satisfy orthogonality, high vanishing moments, compact support, and symmetry/anti-symmetry. Practically, such a wavelet does not exist because only the Haar wavelet is symmetric/anti-symmetric, while high vanishing moments and compact support are a pair of contradictory quantities (Wang et al. 2020). Therefore, selecting the appropriate wavelet according to input data is an essential but unsolved problem. To address this problem, we design a wavelet dynamic selection network (WDSNet) for signal quality enhancement, which encodes the features of 16 wavelet bases in the network, thereby achieving the most appropriate wavelet selection. The framework of WDSNet is presented in Fig. 1.

Firstly, a 1D-ResNet is set to extract the input data features, and then an FC layer performs classification (i.e., wavelet selection) based on the extracted features, where each category corresponds to a wavelet. We use the selected wavelet basis to perform wavelet thresholding on the input signal and then obtain the enhanced signal. To supervise the wavelet selection process, we force the enhanced signal to complete two guidance tasks, i.e., attitude prediction and displacement prediction. The smaller error of the guidance tasks, the higher quality of the enhanced signal, indicating

that the wavelet selection is more appropriate. In such a way, these two tasks jointly guide the network to select the most suitable wavelet for signal processing. It is worth noting that compared with the fully supervised deep learning-based signal enhancement methods, the guidance tasks require much less supervision information, which is convenient to deploy.

### Category Representation Mechanism

Although the deep learning classifier excels at mining and extracting input data features, it neglects to capture the underlying features of target categories, which limits the performance of the classifier. Specifically, the lack of awareness for wavelet bases can lead to an inappropriate selection of wavelets. To address this issue, we propose a category representation mechanism (CRM), which comprises a sparsity regularization term  $R_{sparse}$  and an encoding regularization term  $R_{encode}$ , where  $R_{sparse}$  forces the weight matrix of the FC layer to represent category features and  $R_{encode}$  forces category features to be more discriminative and informative.

The sparsity regularization term acts on the output layer  $R_{sparse} = \|\hat{y}\|_1$ , to make the classification vector  $\hat{y}$  sparse, which means that only one element in this vector is close to 1 and the rest are close to 0. Since the elements close to 0 in  $\hat{y}$  have negligible effect in the subsequent calculations, the backward propagation of loss is truncated here, and the corresponding columns in the weight matrix are not updated, which ensures that each element in the output vector matches exactly each column in the weight matrix.

Under the influence of sparse regularization terms, the loss of each wavelet selection is only related to a specific column of the weight matrix, which can be regarded as the vector representation of the corresponding wavelet category. The ideal vector representation of wavelet categories should be discriminative and informative. Specifically, category representation vectors should be orthogonal to each other for reflecting the diversity between different categories as much as possible. Furthermore, the vectors need to store as much information as possible rather than being as trivial as one-hot encoding. We, therefore, design an encoding regularization term that makes category vectors more orthogonal and informative. Its specific operation is as follows.

We first use the  $\alpha$ -order Rényi information entropy  $S_\alpha$  to measure the amount of information in the weight matrix  $W$ , which is calculated by the following equations:

$$S_\alpha(W) = \frac{1}{1-\alpha} \log_2(\text{tr}(\tilde{G}^\alpha)), \quad (1)$$

$$\tilde{G}[i][j] = \frac{1}{16} \frac{G[i][j]}{\sqrt{G[i][i] \cdot G[j][j]}}, \quad (2)$$

$$G[i][j] = \langle W^{(i)}, W^{(j)} \rangle, \quad (3)$$

where  $G$  denotes the Gram matrix of the weight matrix  $W$ ,  $G[i][j]$  is the inner product of the  $i$ -th column and  $j$ -th column of  $W$ , and  $\tilde{G}$  is the trace-normalized  $G$ , i.e.,  $\text{tr}(\tilde{G}) = 1$ .  $\alpha$  is set to 2 according to (Yu et al. 2020, 2021).  $S_\alpha(W)$  measures the amount of information in the category matrix (weight matrix, as shown in the grey blocks in Fig. 1). The larger  $S_\alpha(W)$  is, the more information the matrix  $W$  has,

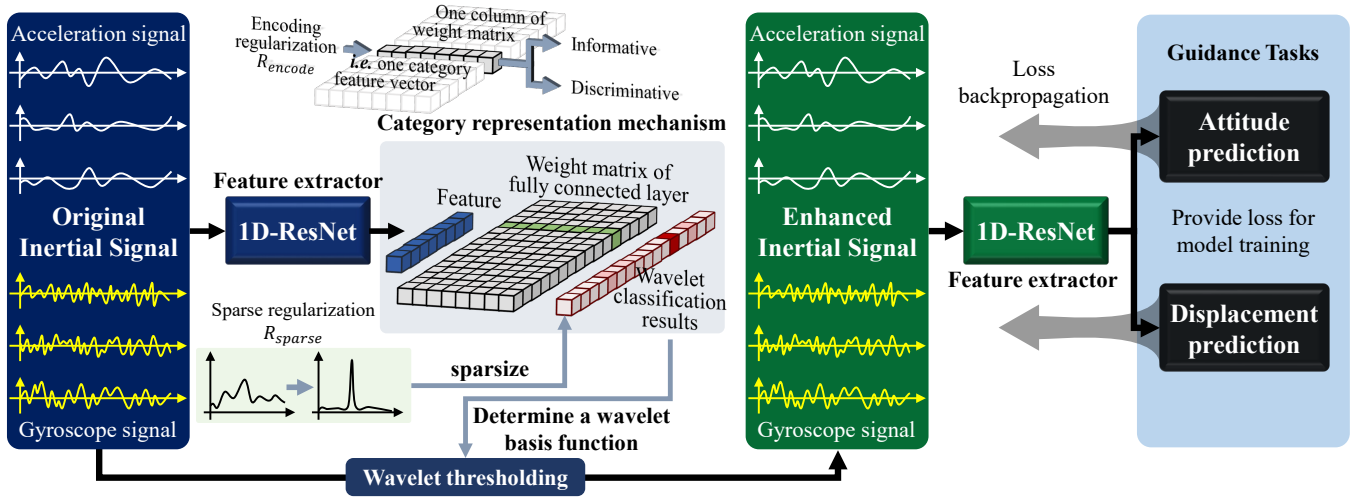


Figure 1: An overview of the WDSNet framework. The input data is fed into a 1D-ResNet for feature extraction, and an FC layer serves as a classifier that performs wavelet selection based on the extracted features. Then the selected wavelet is used to perform wavelet thresholding denoising on the original signal. The enhanced signal is input into another 1D-ResNet for attitude prediction and displacement prediction, which supervises the wavelet selection through the backpropagation of task losses.

indicating that the category representation vectors are more informative. Simultaneously, as  $S_\alpha(W)$  increases, all elements in the Gram matrix  $G$  are forced to descend, thereby indicating that the inner product of different class encoding vectors is smaller and the vectors have stronger orthogonality. Hence, the two requirements for category representation vectors can be satisfied as  $S_\alpha(W)$  increases. Therefore, we propose a regularization term  $R_{encode}$  in equation 4. As this regularization term decreases with training,  $S_\alpha(W)$  gradually increases so we can obtain a discriminative and informative vector representation for each class.

$$R_{encode} = \frac{1}{S_\alpha(W)}. \quad (4)$$

CRM transforms the weakly interpretable FC layer into the encodings of the target categories, which enables the weight matrix to be informative and thus avoids overfitting.

### Feature Supervision Mechanism

CRM enhances the perception ability for target categorization by extracting target category features. We found that these category features can provide direct supervision for deep learning networks. For a general deep learning network, the input data  $x$  is fed into the feature extractor  $f(\cdot)$  to obtain the data feature  $h = f(x)$ , then the feature  $h$  is fed into a task head  $g(\cdot)$  for the final prediction results  $\hat{y} = g(h)$ . The supervision information  $y$  acts on the results  $\hat{y}$ , resulting in a loss function  $L(\hat{y}, y)$ , which is back-propagated to guide the weight updating of feature extractor  $f(\cdot)$  and task head  $g(\cdot)$ . As the network gradually deepens and becomes more complex, the shallow structure, such as the feature extractor, may not be effectively supervised by the far classification labels due to gradient vanishing, gradient explosion, or other unknown reasons.

The proposed CRM provides a feasible solution to this

problem by internalizing category representation in the network, as defined in equation 5.

$$g_{CRM}(h) = \hat{y}_{CRM} \\ \min L(\hat{y}_{CRM}, y) \Rightarrow \max \langle h, W_{CRM}^{(n)} \rangle, \quad (5)$$

where  $g_{CRM}(\cdot)$  and  $\hat{y}_{CRM}$  represent the task head and output results after applying CRM,  $h$  is still the extracted feature, and  $W_{CRM}^{(n)}$  represents the  $n$ -th column weight in the classifier, i.e., the feature vector of the  $n$ -th category.  $\langle \cdot, \cdot \rangle$  represents the inner product, which reflects the similarity between the two vectors. It can be observed that accurate classification requires that the data feature  $h$  has a high similarity with the target category feature  $W_{CRM}^{(n)}$ . Therefore, the supervision of outputs is transformed into the supervision of the features, then transformed into the supervision of the feature extractor  $f(\cdot)$ . This process is named a feature supervision mechanism (FSM), which transforms the fully connected (FC) layer closest to the data features into a category feature layer, allowing it to directly supervise the data features, thus alleviating the gradient vanishing caused by the excessive neural network layers, as shown in Fig. 2.

### Weakly-Supervised Guidance Tasks

Obtaining frame-paired annotations for end-to-end training in the IMU signal enhancement task is extremely challenging (Marchetti et al. 2020). To address this issue, we set up two tasks with readily available labels as guidance tasks, displacement prediction, and attitude prediction, to guide the selection of the wavelet.

IMU provides acceleration and angular velocity data during motion as a motion sensor. The attitude prediction task requires the enhanced signal to predict the changes of three attitude angles: yaw, roll, and pitch. The displacement prediction task requires the enhanced signal to predict the displacement vector corresponding to the motion. These two

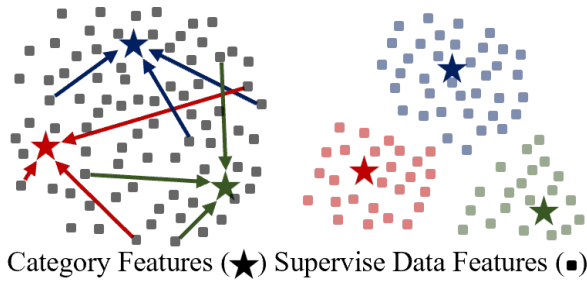


Figure 2: Illustration of feature supervision mechanism. Pentagrams represent category features, and small blocks represent data features. Under the supervision of category features, data features exhibit category characteristics and become more discriminative.

tasks assess the motion information of the enhanced signal in terms of spatial position change and attitude rotation change, so they are configured to guide the improvement of signals.

## Experiments and Results

### Experiment Dataset

The built-in IMUs are the most widely used and representative low-cost inertial sensors. 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

Dataset	Smartphone	IMU	Unit price
Training	HUAWEI Mate30 Pro	ICM20690	\$0.28
	HUAWEI P40	LSM6DSM	\$0.30
	HUAWEI P40 Pro	LSM6DSO	\$0.33
	iPhone 7 Plus	ICM20600	\$0.20
	SAMSUNG Galaxy S7	LSM6DS3	\$0.20
	SAMSUNG Galaxy S8	LSM6DSL	\$0.26
Testing	Realme GT	BMI160	\$0.21
	Xiaomi 11	BHI260AB	\$0.30
	OPPO Reno 6	ICM-40607	\$0.28
	Lenovo Legion Phone	ICM-42605	\$0.20
	VIVO X30	LSM6DSM	\$0.30
	VIVO T2x	LSM6DSO	\$0.33
	iPhone 13	Undisclosed	/
	iPhone 12	Undisclosed	/
	iPhone 11 Pro	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.

Architecture		Acceleration			Angular Velocity		
		QN	VRW	BI	QN	ARW	BI
Raw signal		1.21	1.92	3.12	3.55	5.98	9.06
Model Driven	Kalman filter	0.49	0.58	0.64	0.77	0.85	1.40
	SG filter	0.54	0.71	0.76	0.98	1.10	1.74
Data Driven	CNN	0.56	0.70	0.77	1.06	1.08	1.64
	GRU-LSTM	0.52	0.60	0.71	0.77	0.96	1.53
	Opt-GRU-LSTM	0.40	0.53	0.61	0.65	0.72	0.98
		<u>0.20</u>	<u>0.31</u>	<u>0.47</u>	<u>0.48</u>	<u>0.55</u>	<u>0.65</u>
Model -Data	WDSNet (Ours)	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>	<b>0.09</b>	<b>0.13</b>

Table 2: Performance comparison of mainstream methods in terms of Allan variance analysis. We bold the best and underline the suboptimal results.

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 previous two guidance tasks. All experiments are implemented by Pytorch 1.10.1 with an Nvidia RTX 2080TI GPU and Intel(R) Xeon(R) W-2133 CPU.

### 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 compared WDSNet 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. Some data-driven methods perform relatively poorly since they act as a black box to generate enhanced signals, resulting in unreliable signal enhancement. In contrast, model-driven methods optimize the signal from an interpretative perspective, making the enhanced signal more favorable. However, model-driven methods cannot make flexible adjustments according to the characteristics of the input signal, thus presenting a bottleneck in signal quality enhancement. By combining model-driven and data-driven strategies, WDSNet achieves the best performance.

**Dynamic Evaluation** Allan variance analysis is usually used to evaluate the noise of IMU signals under static conditions, and it is not adequate to evaluate the signal quality enhancement in motion. Therefore, we design four downstream tasks to examine the effectiveness of different methods. Specifically, the four downstream tasks are attitude change estimation, position change estimation (i.e., displacement estimation), motion semantic recognition, and motion trajectory reconstruction. The classic navigation algorithm is employed to perform motion trajectory reconstruction, and a ResNet containing 8 bottleneck blocks is

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	EMD-Kalman filter	5.36 (-49.83%)	<u>0.38 (-65.27%)</u>	85.28% (+8.74%)	<u>0.18 (-76.53%)</u>
	Savitzky Golay filter	6.27 (-41.29%)	0.43 (-59.82%)	84.99% (+8.38%)	0.22 (-72.42%)
Data Driven	CNN	5.85 (-45.31%)	0.48 (-56.13%)	85.52% (+9.05%)	0.29 (-62.24%)
	GRU-LSTM	4.94 (-53.84%)	0.41 (-62.59%)	86.13% (+9.83%)	0.27 (-65.17%)
	Optimized GRU-LSTM	<u>4.43 (-58.53%)</u>	0.39 (-64.82%)	<u>86.22% (+9.95%)</u>	0.25 (-68.39%)
	kNN	6.47 (-39.46%)	0.46 (-57.94%)	85.62% (+9.18%)	0.28 (-63.72%)
Model-Data Driven	WDSNet (Ours)	<b>3.79 (-64.35%)</b>	<b>0.28 (-73.62%)</b>	<b>87.01% (+10.95%)</b>	<b>0.10 (-86.95%)</b>

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 improvement compared with the raw signal in parentheses for convenient comparison. Note that we bold the best and underline the suboptimal results.

Architecture		Acceleration			Angular Velocity			Attitude estimation error (deg) ↓	Position estimation error (m) ↓	Semantic recognition accuracy ↑	Trajectory reconstruction error ↓
CRM	wavelet number	QN	VRW	BI	QN	ARW	BI				
w/o CRM	5	0.46	0.74	0.90	0.94	1.09	1.70	6.24	0.49	84.99%	0.28
	10	0.39	0.67	0.71	0.89	0.93	1.54	5.91	0.43	85.55%	0.24
	16	0.40	0.74	0.87	0.95	0.98	1.55	6.09	0.48	85.21%	0.28
w/ CRM	5	0.42	0.70	0.73	0.74	0.92	1.60	6.18	0.41	85.53%	0.25
	10	<u>0.19</u>	<u>0.21</u>	<u>0.26</u>	<u>0.32</u>	<u>0.40</u>	<u>0.84</u>	<u>4.71</u>	<u>0.35</u>	<u>86.16%</u>	<u>0.20</u>
	16	<b>0.06</b>	<b>0.07</b>	<b>0.08</b>	<b>0.09</b>	<b>0.09</b>	<b>0.13</b>	<b>3.79</b>	<b>0.28</b>	<b>87.01%</b>	<b>0.10</b>

Table 4: Ablation experiments on candidate wavelet number and CRM. The improvement on four downstream tasks compared with the raw signal is given in parentheses for convenient comparison. We bold the best and underline the suboptimal results.

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 WDSNet still achieves the best performance. In contrast, the kNN method, which performs well in Allan variance analysis, does not serve downstream tasks well. We note that signal enhancement has limited assistance in semantic recognition since ill-writing motion and individual writing habits, instead of signal noise, bring about the main challenges for the semantic recognition task.

### Ablation Study

We then conduct ablation experiments to verify the effect of CRM and the number of candidate wavelet bases. The related static and dynamic evaluation results are summarized in Table 4. In comparison with the results of the existing methods in Tables 2 and 3, our wavelet selection framework

without CRM still achieves significant signal enhancement. However, as more candidate wavelets are available for selection, the model is perplexed by massive candidate wavelets to determine the suitable one. Under this scenario, the selected wavelet often is not suitable to adapt to the input signal, such that the performance of signal enhancement does not increase but decreases. The introduction of CRM enhances the wavelet perception ability of the model, thereby enabling it to select the most suitable wavelet from numerous candidates according to signal properties and achieve the best signal enhancement. To visualize the signal enhancement of the proposed method, we reconstruct the concrete trajectories of the handwriting signals. It can be observed from Fig. 3 (a) that the trajectory reconstruction based on the original signal is deplorable, and the written character is almost invisible. Fig. 3 (b) illustrates that the signals enhanced by WDSNet (without CRM) are conducive to improving trajectory reconstruction. However, the lack of CRM leads to inaccurate wavelet selection, so most reconstructed trajectories are not visible enough to identify their motion characters. The model with CRM can allocate the appropriate wavelet basis for arbitrary input signals, so the improved signal can achieve much better trajectory reconstruction, as shown in Fig. 3 (c).

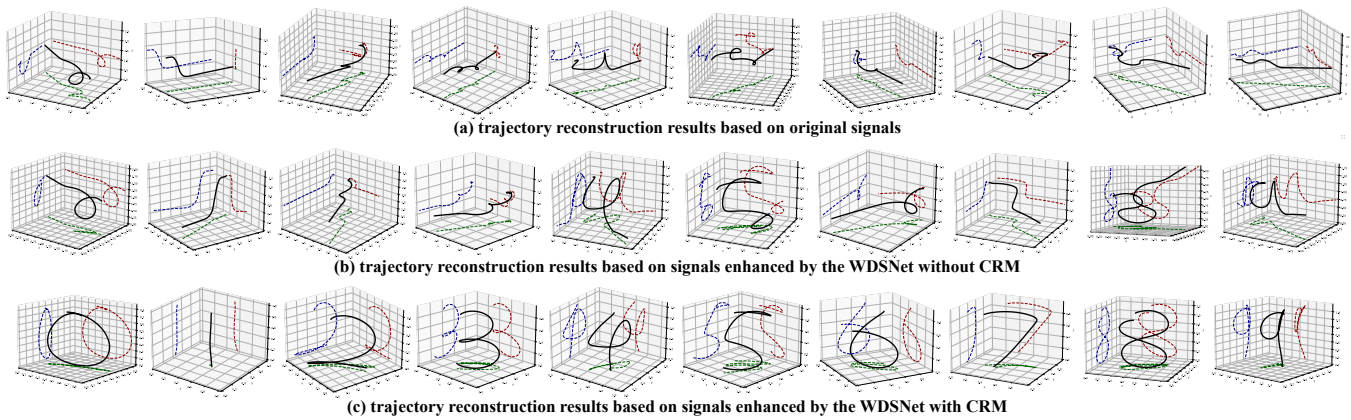


Figure 3: Trajectory reconstruction visualization for ablation study. The dashed lines in each panel indicate the projection of a reconstructed trajectory in the XY, XZ, and YZ planes.

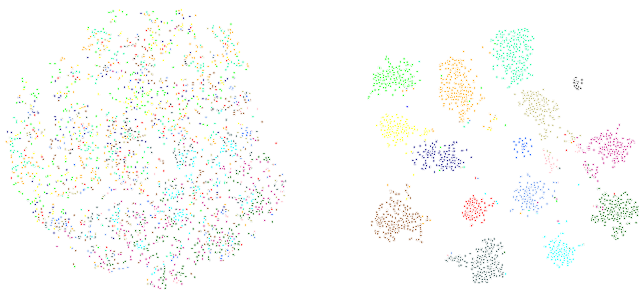


Figure 4: Effect visualization of with (right panel) and without (left panel) FSM. Each point represents the feature of a sample, and 16 colors correspond to the 16 wavelets. The data features are clustered according to wavelet categories under category feature supervision.

### Visualization Analysis on the Effect of FSM

To verify the effect of FSM, we utilize the t-SNE technique to perform dimensionality reduction on the features extracted by 1D-ResNet, and then compare the data feature distributions with and without category feature supervision, as shown in Fig. 4. The left panel presents the distribution of data features supervised by loss backpropagation from the result layer, while the right one presents the distribution of data features directly supervised by category features represented within the network. It can be observed that under the FSM, the extracted data features become well-organized from stochastic distribution. Specifically, the features of the same category have stronger similarity, indicating that the model can extract more discriminative features from the input data, thereby achieving accurate classification.

### Conclusion

In this paper, we propose a signal enhancement method, wavelet dynamic selection network (WDSNet), combining wavelet with deep learning, which holds both the reliability of model-driven approaches and the flexibility of data-driven approaches. To select an appropriate wavelet for signal en-

hancement, we propose a category representation mechanism (CRM) that improves the awareness of wavelet characteristics by learning their category representation. As a plug-and-play module, CRM improves the classification ability of deep learning without increasing trainable parameters. In addition, since the CRM constructs vector representations of target categories within the network, the feature extraction can be directly supervised by these category vectors. This feature supervision mechanism (FSM) is more direct and efficient than the loss backpropagation from the result of the far output layer, which has been verified theoretically and experimentally in this paper. Specifically, the Allan variance analysis is employed to evaluate the signal enhancement effect compared with existing methods, and we also construct four downstream tasks (posture prediction, position prediction, semantic recognition, and trajectory reconstruction) to evaluate the practical effects of enhanced signals. The results show that WDSNet achieves SOTA performance in all comparative experiments. Importantly, signals enhanced by our WDSNet perform satisfactory reconstruction of arbitrary spatial trajectories, which is usually considered an impossible function for a low-cost inertial sensor. On the other hand, this paper provides a universal wavelet selection strategy without requirements of any selection labels, which expands wavelet methods to be qualified for complex and ever-changing application scenarios.

### References

- Boronakhin, A. M.; Shalymov, R. V.; Larionov, D. Y.; Khanh, N. Q.; and Yen, N. T. 2022. 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 (EICon-Rus)*, 1335–1340. IEEE.
- Brossard, M.; Bonnabel, S.; and Barrau, A. 2020. Denoising IMU Gyroscopes With Deep Learning for Open-Loop Attitude Estimation. *IEEE Robotics and Automation Letters*, 5(3): 4796–4803.
- Burri, M.; Nikolich, J.; Gohl, P.; Schneider, T.; Rehder, J.; Omari, S.; Achtelik, M. W.; and Siegwart, R. 2016. The Eu-

- RoC micro aerial vehicle datasets. *The International Journal of Robotics Research*, 35(10): 1157–1163.
- Caesar, H.; Bankiti, V.; Lang, A. H.; Vora, S.; Liong, V. E.; Xu, Q.; Krishnan, A.; Pan, Y.; Baldan, G.; and Beijbom, O. 2020. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 11621–11631.
- Caserman, P.; Garcia-Agundez, A.; and Göbel, S. 2019. A survey of full-body motion reconstruction in immersive virtual reality applications. *IEEE transactions on visualization and computer graphics*, 26(10): 3089–3108.
- Chen, C.; Lu, X.; Markham, A.; and Trigoni, N. 2018. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.
- Chen, C.; Zhao, P.; Lu, C. X.; Wang, W.; Markham, A.; and Trigoni, N. 2020. Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference. *IEEE Internet of Things Journal*, 7(5): 4431–4441.
- Chen, H.; Taha, T. M.; and Chodavarapu, V. P. 2022. Towards improved inertial navigation by reducing errors using deep learning methodology. *Applied Sciences*, 12(7): 3645.
- Ehatisham-ul Haq, M.; Arsalan, A.; Raheel, A.; and Anwar, S. M. 2021. Expert-novice classification of mobile game player using smartphone inertial sensors. *Expert Systems with Applications*, 174: 114700.
- Engelsman, D.; and Klein, I. 2023. Data-driven denoising of stationary accelerometer signals. *Measurement*, 113218.
- Esfahani, M. A.; Wang, H.; Wu, K.; and Yuan, S. 2019a. AbolDeepIO: A novel deep inertial odometry network for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*, 21(5): 1941–1950.
- Esfahani, M. A.; Wang, H.; Wu, K.; and Yuan, S. 2019b. OriNet: Robust 3-D orientation estimation with a single particular IMU. *IEEE Robotics and Automation Letters*, 5(2): 399–406.
- Ferrera, M.; Creuze, V.; Moras, J.; and Trouvé-Peloux, P. 2019. AQUALOC: An underwater dataset for visual-inertial–pressure localization. *The International Journal of Robotics Research*, 38(14): 1549–1559.
- Foehn, P.; Kaufmann, E.; Romero, A.; Penicka, R.; Sun, S.; Bauersfeld, L.; Laengle, T.; Cioffi, G.; Song, Y.; Loquercio, A.; et al. 2022. Agilicious: Open-source and open-hardware agile quadrotor for vision-based flight. *Science Robotics*, 7(67): eabl6259.
- Golestani, N.; and Moghaddam, M. 2020. Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nature communications*, 11(1): 1551.
- Gromov, B.; Abbate, G.; Gambardella, L. M.; and Giusti, A. 2019. Proximity Human-Robot Interaction Using Pointing Gestures and a Wrist-mounted IMU. In *2019 International Conference on Robotics and Automation (ICRA)*, 8084–8091.
- Han, S.; Meng, Z.; Zhang, X.; and Yan, Y. 2021. Hybrid deep recurrent neural networks for noise reduction of MEMS-IMU with static and dynamic conditions. *Micromachines*, 12(2): 214.
- He, J.; Sun, C.; and Wang, P. 2019. Noise Reduction for MEMS gyroscope signal: A novel method combining ACMP with adaptive multiscale SG filter based on AMA. *Sensors*, 19(20): 4382.
- Herath, S.; Yan, H.; and Furukawa, Y. 2020. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 3146–3152. IEEE.
- Huang, X.; Wang, P.; Cheng, X.; Zhou, D.; Geng, Q.; and Yang, R. 2019. The apolloscape open dataset for autonomous driving and its application. *IEEE transactions on pattern analysis and machine intelligence*, 42(10): 2702–2719.
- Li, P.; Zhang, W.-A.; Jin, Y.; Hu, Z.; and Wang, L. 2023. Attitude Estimation using Iterative Indirect Kalman with Neural Network for Inertial Sensors. *IEEE Transactions on Instrumentation and Measurement*.
- Li, Y.; Chen, R.; Niu, X.; Zhuang, Y.; Gao, Z.; Hu, X.; and El-Sheimy, N. 2022. Inertial Sensing Meets Machine Learning: Opportunity or Challenge? *IEEE Transactions on Intelligent Transportation Systems*, 23(8): 9995–10011.
- Liu, S.; Zhang, J.; Zhang, Y.; and Zhu, R. 2020a. A wearable motion capture device able to detect dynamic motion of human limbs. *Nature communications*, 11(1): 5615.
- Liu, W.; Caruso, D.; Ilg, E.; Dong, J.; Mourikis, A. I.; Daniilidis, K.; Kumar, V.; and Engel, J. 2020b. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters*, 5(4): 5653–5660.
- Liu, Y.; Chen, G.; Wei, Z.; Yang, J.; and Xing, D. 2020c. Denoising method of MEMS gyroscope based on interval empirical mode decomposition. *Mathematical Problems in Engineering*, 2020: 1–12.
- Ma, H.; Liu, D.; Yan, N.; Li, H.; and Wu, F. 2020. End-to-end optimized versatile image compression with wavelet-like transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3): 1247–1263.
- Madgwick, S. O.; Harrison, A. J.; and Vaidyanathan, R. 2011. Estimation of IMU and MARG orientation using a gradient descent algorithm. In *2011 IEEE international conference on rehabilitation robotics*, 1–7. IEEE.
- Marchetti, F.; Becattini, F.; Seidenari, L.; and Del Bimbo, A. 2020. Multiple trajectory prediction of moving agents with memory augmented networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Min, J.; Wang, Z.-L.; Li, Y.-P.; Tao, W.-Z.; Li, C.-H.; Lei, J.-G.; Xi, D.-X.; Fan, D.; and Wang, J.-B. 2021. Drop tower tests of Taiji-1 inertial sensor substitute. *npj Microgravity*, 7(1): 25.
- Montesinos, L.; Castaldo, R.; and Pecchia, L. 2018. 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.

- Nguyen, H.; Lebel, K.; Bogard, S.; Goubault, E.; Boissy, P.; and Duval, C. 2017. Using inertial sensors to automatically detect and segment activities of daily living in people with Parkinson's disease. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 26(1): 197–204.
- Pan, X.; and Hamilton, A. F. d. C. 2018. Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3): 395–417.
- Qin, T.; Li, P.; and Shen, S. 2018. Vins-mono: A robust and versatile monocular visual-inertial state estimator. *IEEE Transactions on Robotics*, 34(4): 1004–1020.
- Qiu, S.; Zhao, H.; Jiang, N.; Wang, Z.; Liu, L.; An, Y.; Zhao, H.; Miao, X.; Liu, R.; and Fortino, G. 2022. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, 80: 241–265.
- Saha, S. S.; Du, Y.; Sandha, S. S.; Garcia, L. A.; Jawed, M. K.; and Srivastava, M. 2023. Inertial Navigation on Extremely Resource-Constrained Platforms: Methods, Opportunities and Challenges. In *2023 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 708–723. IEEE.
- Saha, S. S.; Sandha, S. S.; Garcia, L. A.; and Srivastava, M. 2022. Tinyodom: Hardware-aware efficient neural inertial navigation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(2): 1–32.
- Saydjari, A. K.; and Finkbeiner, D. P. 2022. Equivariant wavelets: Fast rotation and translation invariant wavelet scattering transforms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(2): 1716–1731.
- Shaeffer, D. K. 2013. MEMS inertial sensors: A tutorial overview. *IEEE Communications Magazine*, 51(4): 100–109.
- Shamwell, E. J.; Lindgren, K.; Leung, S.; and Nothwang, W. D. 2019. Unsupervised deep visual-inertial odometry with online error correction for RGB-D imagery. *IEEE transactions on pattern analysis and machine intelligence*, 42(10): 2478–2493.
- Toft, C.; Maddern, W.; Torii, A.; Hammarstrand, L.; Stenborg, E.; Safari, D.; Okutomi, M.; Pollefeys, M.; Sivic, J.; Pajdla, T.; et al. 2020. Long-term visual localization revisited. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(4): 2074–2088.
- Tunca, C.; Salur, G.; and Ersoy, C. 2019. Deep learning for fall risk assessment with inertial sensors: Utilizing domain knowledge in spatio-temporal gait parameters. *IEEE journal of biomedical and health informatics*, 24(7): 1994–2005.
- Wang, Y.; Wang, Y.; Hu, G.; Liu, Y.; and Zhao, Y. 2020. Adaptive Skewness Kurtosis Neural Network : Enabling Communication Between Neural Nodes Within a Layer. In *International Conference on Neural Information Processing*.
- Wang, Y.; and Zhao, Y. 2023. Arbitrary Spatial Trajectory Reconstruction Based on a Single Inertial Sensor. *IEEE Sensors Journal*, 23(9): 10009–10022.
- Wang, Y.; and Zhao, Y. 2024. Handwriting Recognition Under Natural Writing Habits Based on a Low-Cost Inertial Sensor. *IEEE Sensors Journal*, 24(1): 995–1005.
- Weber, D.; Gühmann, C.; and Seel, T. 2021. RIANN—A robust neural network outperforms attitude estimation filters. *Ai*, 2(3): 444–463.
- Wu, F.; Luo, H.; Jia, H.; Zhao, F.; Xiao, Y.; and Gao, X. 2021. Predicting the Noise Covariance With a Multitask Learning Model for Kalman Filter-Based GNSS/INS Integrated Navigation. *IEEE Transactions on Instrumentation and Measurement*, 70: 1–13.
- Wu, Y.; Zhu, H.-B.; Du, Q.-X.; and Tang, S.-M. 2019. A survey of the research status of pedestrian dead reckoning systems based on inertial sensors. *International Journal of Automation and Computing*, 16: 65–83.
- Yu, S.; Giraldo, L. G. S.; Jenssen, R.; and Príncipe, J. C. 2020. Multivariate Extension of Matrix-Based Rényi's  $\alpha$ -Order Entropy Functional. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(11): 2960–2966.
- Yu, S.; Wickstrøm, K.; Jenssen, R.; and Príncipe, J. C. 2021. Understanding Convolutional Neural Networks With Information Theory: An Initial Exploration. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1): 435–442.
- Yuan, K.; and Wang, Z. J. 2023. A simple self-supervised imu denoising method for inertial aided navigation. *IEEE Robotics and Automation Letters*, 8(2): 944–950.
- Zhang, X.; He, B.; Li, G.; Mu, X.; Zhou, Y.; and Mang, T. 2020. NavNet: AUV navigation through deep sequential learning. *IEEE Access*, 8: 59845–59861.
- Zhuang, Y.; Sun, X.; Li, Y.; Huai, J.; Hua, L.; Yang, X.; Cao, X.; Zhang, P.; Cao, Y.; Qi, L.; et al. 2023. Multi-sensor integrated navigation/positioning systems using data fusion: From analytics-based to learning-based approaches. *Information Fusion*, 95: 62–90.