Information Fusion 123 (2025) 103274

Contents lists available at ScienceDirect

Information Fusion

journal homepage: www.elsevier.com/locate/inffus

Full length article

General pre-trained inertial signal feature extraction based on temporal memory fusion

Yifeng Wang, Yi Zhao *

School of Science, Harbin Institute of Technology, Shenzhen, 518055, China

ARTICLE INFO

Keywords:
Inertial sensors
Feature extraction
Memory Graph
Small world property
Topology guidance

ABSTRACT

Inertial sensors are widely used in smartphones, robotics, wearables, aerospace systems, and industrial automation. However, extracting universal features from inertial signals remains challenging. Inertial signal features are encoded in abstract, unreadable waveforms, lacking the visual intuitiveness of images, which makes semantic extraction difficult. The non-stationary nature and complex motion patterns further complicate the feature extraction process. Moreover, the lack of large-scale annotated inertial datasets limits deep learning models to learn universal features and generalize them across expansive applications of inertial sensors. To this end, we propose a Topology Guided Feature Extraction (TG-FE) approach for general inertial signal feature extraction. TG-FE fuses time-series information into graph representations, constructing a Memory Graph by emulating the complex network characteristics of human memory. Guided by small-world network principles, this graph integrates local and global information while sparsity constraints emphasize critical feature interactions. The Memory Graph preserves nonlinear relationships and higher-order dependencies, enabling the model to generalize across scenarios with minimal task-specific tuning. Furthermore, a Cross-Graph Feature Fusion mechanism integrates information across stacked TG-FE modules to enhance representation ability and ensure stable gradient flow. With self-supervised pre-training, the TG-FE modules require only minimal fine-tuning to adapt to various hardware configurations and task scenarios, consistently outperforming comparison methods across all evaluations. Compared to the current state-of-the-art method, our TG-FE achieves 11.7% and 20.0% error reduction in attitude and displacement estimation tasks. Notably, TG-FE achieves an order-of-magnitude advantage in stability evaluations, maintaining robust performance even under 20% noise conditions where competing methods degrade significantly. Overall, this work offers a solution for general inertial signal feature extraction and opens new avenues for applying graph-based deep learning to capture and represent sequential signal features.

1. Introduction

Inertial Measurement Units (IMUs), consisting of accelerometers and gyroscopes, are foundational components in a vast array of modern technologies [1–3]. Due to their advantages of low cost, small size, easy integration and low energy consumption, they are widely embedded in smartphones, robotics, wearable devices, and aerospace navigation systems [4–7]. The acceleration and rotation information in three-dimensional space recorded by inertial sensor signals make them invaluable for applications such as human activity recognition, gesture control, motion tracking, indoor navigation, virtual and augmented reality, and biomechanical analysis in sports and healthcare, revolutionizing the way devices interact with the physical world [8–11].

Despite the widespread adoption and versatility of IMUs, robust extraction of features from raw inertial sensor signals remains a significant challenge [12,13]. Traditional feature extraction methods for IMU data often rely heavily on handcrafted features, where domain experts manually design and extract specific attributes from the raw sensor signals [14]. These features typically include statistical measures such as mean, variance, skewness, and kurtosis, as well as temporal characteristics like signal magnitude area, zero-crossing rate, and peak detection. Frequency domain features are also commonly extracted using transformations like the Fast Fourier Transform (FFT) or Wavelet Transform [15]. While these handcrafted features can provide insights into certain aspects of the data, they are inherently constrained to a specific task and limited by skills of the designers [16]. In addition, the manual feature engineering is time-consuming and labor-intensive,

* Corresponding author.
E-mail address: zhao.yi@hit.edu.cn (Y. Zhao).

https://doi.org/10.1016/j.inffus.2025.103274
Received 3 February 2025; Received in revised form 9 April 2025; Accepted 25 April 2025
Available online 11 May 2025

requiring substantial domain expertise to identify which features are most relevant for a given task. These methods lack scalability and adaptability, making it challenging to transfer the features to different applications and sensor configurations. Moreover, handcrafted features may fail to capture the complex, non-linear relationships and high-order dependencies present in inertial signals, leading to poor performance [17–19].

Deep learning models are increasingly applied to IMU tasks without the dependence on manual feature engineering. However, the training of deep learning models always relies on a large number of annotated samples, particularly for training the feature extractor. In practice, many real-world applications suffer from a lack of sufficient labeled datasets, as data collection and annotation is time-intensive and costly, especially for specialized or emerging scenarios [20,21]. This scarcity of labeled data exacerbates the difficulty of training task-specific models, thereby highlighting the necessity for a universal feature extractor that can generalize across diverse IMU tasks and configurations.

One key to achieving such generalization lies in emulating the human memory ability that transforms time-series experiences into graph linkage form through meaningful relationship [22,23]. This graph-based memory allows human to integrate information across different contexts and time scales, enabling adaptability to new tasks and scenarios [24–26]. By leveraging these interconnected memories, humans can recognize patterns, draw inferences, and apply learned knowledge to unfamiliar situations with minimal prior exposure [27–29]. In contrast, traditional deep learning models like CNNs, RNNs, and LSTMs process data either sequentially or locally, lacking the graph-like memory mechanisms to model interactions embedded in complex and abstract signal waveforms [30–32]. As a result, the features extracted by these models tend to be specialized for specific datasets and tasks, limiting their generalizations across diverse applications. This limitation highlights the urgent need for the general and robust feature extraction from IMU signals in a memory-inspired way.

To address these challenges, we propose a Topology Guided Feature Extraction (TG-FE) method, which emulates the memory mechanisms of human by transforming time-series inertial sensor signal into a graph-based feature representation. At the core of this method lies the Memory Graph, which models feature interactions as interconnected nodes, effectively capturing the intricate relationships and dependencies inherent in IMU signals. Inspired by the neural architecture of the human brain, we enhance the Memory Graph with two properties: the small-world property and sparsity. The small-world property organizes related features into densely connected clusters, making information integration at local and global scales. Meanwhile, sparsity ensures that each node maintains a limited number of connections, reducing computational overhead and mitigating the risk of overfitting. To refine the feature extraction process, the proposed model is enhanced with cross-graph feature fusion by means of residual connections to facilitate information flow across modules, improving feature representation and mitigating the vanishing gradient problem.

Overall, the TG-FE module serves as a universal feature extractor for IMU signals by incorporating graph-based associative memory mechanisms with the devised topological guidance. It enable robust features and also high adaptability across diverse tasks, which may pave a way for new advancements in IMU signal processing.

2. Related work

Enhancing model generalization is a critical challenge for applications of deep learning to signal processing, particularly for inertial sensors [33]. The accelerometer and gyroscope data powers applications such as activity recognition, gesture control, and motion tracking [34]. However, achieving robust generalization is much difficult to this type of signals due to several inherent issues. Firstly, the inertial data exhibit significant variability across users, devices, and environments, attributable to heterogeneity in movement patterns, sensor characteristics, and external conditions (e.g., temperature or magnetic interference). In addition, the limited availability of annotated datasets restricts the training of deep learning models, as labeling inertial data is time-consuming and costly. Moreover, the non-stationary, noisy, and dynamic nature of inertial signals further complicates the extraction of consistent, transferable features, making generalization a top priority for their real-world deployment [35].

Self-supervised learning has emerged as a promising approach to address these challenges by learning robust representations without relying on extensive labeled data [36,37]. [38] introduced masked autoencoders to time series by masking portions of the input signal and reconstructing them. This method captures temporal patterns effectively when signals are smooth and predictable. However, inertial signals often exhibit abrupt changes, like sudden jerks or irregular movements, limiting their method applicability. Contrastive learning provides an alternative by training models to differentiate between similar and dissimilar data points, fostering discriminative features. [39] developed Time-Series Temporal and Contextual Contrasting (TS-TCC), blending temporal and contextual contrasts to boost generalization across time-series tasks. Its strength lies in creating representations that distinguish meaningful patterns, but its success depends heavily on tailored data augmentations, such as time warping or noise injection. However, the devised augmentations in TS-TCC may fail to mirror real-world conditions, especially for inertial sensor data, where noise profiles and temporal distortions vary unpredictably, such as sensor drift or erratic user motion. [40] proposed Temporal Neighborhood Coding (TNC), focusing on local temporal smoothness to enhance feature robustness. By emphasizing short-term dependencies, TNC excels at modeling concise, consistent patterns in time series. Yet, this narrow focus struggles to capture long-range interactions, such as extended motion sequences or complex activity transitions, which are common for many inertial sensor applications. [41] introduced TS2Vec, aiming for a universal time-series representation through hierarchical contrastive learning. TS2Vec captures multi-scale contextual information, making it versatile for diverse tasks. However, its reliance on hierarchical augmentation strategies increases computational demands and assumes consistent temporal scales, which misaligns with inertial data's varying timeframes (e.g., ranging from rapid gestures to slow posture shifts), reducing its practical utility. [42] presented Self-Supervised Contrastive Pre-Training for Time Series via Time–Frequency Consistency (TF-C), leveraging consistency between time and frequency domain representations to learn adaptable features. This dual-domain approach offers flexibility but its dependence on frequency domain may be limited by noise and artifacts, like mechanical vibrations or irregular sampling, which often corrupt the frequency spectra, thereby diminishing TF-C's reliability.

In summary, current time-series representation methods share common weaknesses when applied to inertial sensor data while they push time-series analysis forward through exquisite designs. The non-stationary behavior, high variability, and intricate temporal dependencies of inertial signals clash with the assumptions baked into these techniques. Masked autoencoders and TNC presume a level of local predictability that is invalid for erratic motions. Contrastive methods like TS-TCC and TS2Vec demand augmentations that rarely generalize across the diverse noise and distortion patterns in inertial data. TF-C's frequency-based approach falters under noisy conditions inherent to inertial sensors. Compounding these issues, the scarcity of large, annotated datasets limits supervised fine-tuning, amplifying the need for self-supervised methods that can adapt to inertial signals' unique challenges [43]. To our knowledge, no universal pre-trained feature extraction method exists specifically for inertial sensor data.

3. Methodology

3.1. Framework overview

Inertial sensor signals, captured by accelerometers and gyroscopes, provide a time-series record of a motion system's dynamic behavior.

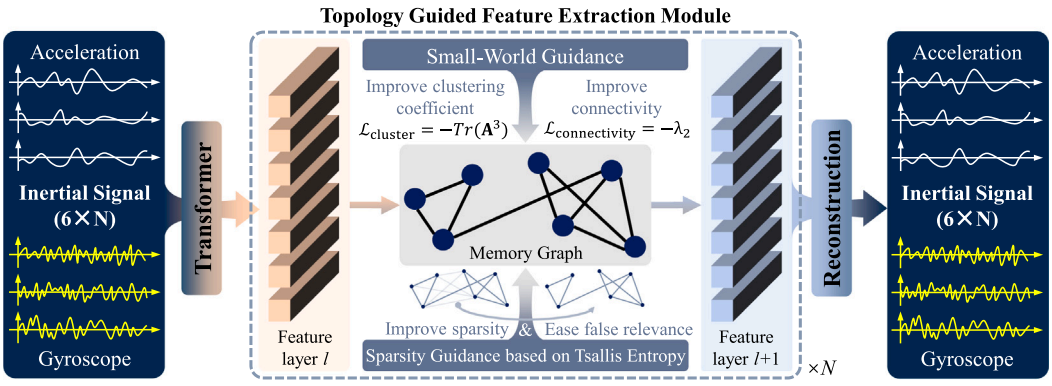

Fig. 1. The pre-training framework of the Topology Guided Feature Extraction modules.

However, this time-series data presents challenges for feature extraction, particularly in capturing complex and abstract semantic dependencies within the incomprehensible waveforms. Therefore, we propose the Topology Guided Feature Extraction (TG-FE) module to convert time-series signals into a graph-based feature representation, which constructs a Memory Graph to model feature interactions. The Memory Graph systematically captures recurring patterns and dependencies within the data through self-supervised reconstruction task during pre-training. This configuration enables the extraction of universally applicable features, allowing the model to adapt to various downstream IMU tasks with only minimal fine-tuning.

The framework of the TG-FE module is shown in Fig. 1. First, a Transformer processes the raw inertial signals, extracting low-level feature representations. These features then serve as input to our TG-FE module, which constructs a Memory Graph to represent intricate feature dependencies. To enhance the Memory Graph's representational capability, we design two types of topology guidance inspired by fundamental mechanisms of human memory: the small-world property and the sparsity. Small-world guidance allows the Memory Graph to cluster related features for seamless local–global integration. Meanwhile, sparsity guidance forces the Memory Graph to retain only the most relevant connections, alleviating false associations and preventing overfitting. By stacking multiple TG-FE modules and progressively refining the graph representation, we obtain semantically rich features that can be leveraged for various IMU-based applications.

To ensure the extracted features encapsulate the fundamental dynamics of the motion system, the cascading TG-FE modules are trained with a reconstruction task that compels the model to reconstruct the original inertial signals from the extracted features. This reconstruction task emulates human memory consolidation by reinforcing the model to recall critical information, which ensures that the model learns a comprehensive and nuanced understanding of the input motion signal, resulting in a robust and informative feature representation that supports various applications with high reliability. Therefore, the features pre-trained in this framework can be directly utilized for various IMU-based tasks, such as activity recognition, gesture analysis, and motion tracking.

In application, the parameters of the TG-FE modules are frozen, meaning they no longer participate in gradient backpropagation. These frozen modules serve as a general feature extractor for a task-specific head that is fine-tuned to adapt the extracted features to the requirements of various specific tasks, as shown in Fig. 2.

3.2. Small-world guidance

Neuroscience research suggests that the human brain's neural networks exhibit small-world characteristics [44–46], combining high local clustering with relatively short average path lengths [47–49]. This topology promotes efficient information dissemination, enables rapid pattern recognition, and enhances global reasoning. Inspired by this, we devise the small-world guidance for our Memory Graph through a dedicated regularization term, $\mathcal{L}_{\text{small-world}}$, that encourages both clustering and connectivity of the graph.

To enhance clustering, we leverage the adjacency matrix $\mathbf{A}$ of the Memory Graph. Specifically, the trace of its cube, $\text{Tr}(\mathbf{A}^3)$, quantifies the number of triangles within the graph, representing tightly connected triplets of nodes. These triangles are essential for capturing meaningful feature co-occurrences and localized relationships because they encode higher-order interactions and structural dependencies within the graph. Unlike simple pairwise connections, which only reveal direct relationships between two nodes, triangles inherently capture mutual interaction among three interconnected features. In fact, triangles represent the smallest closed structure that can capture high-order interactions, which has been regarded as the basic unit to describe high-order structure in a network [50]. Theoretically, triangles directly measure the clustering coefficient, which is a defining characteristic of small-world networks [51]. By encouraging the formation of triangles, higher-order topological structures emerge through triangular composition, as dense triangle regions naturally form cliques and community structures. Therefore, we define a clustering regularization term to facilitate the formation of dense feature clusters:

$$\mathcal{L}_{\text{cluster}} = -\,\text{Tr}(\mathbf{A}^3), \tag{1}$$

where minimizing $\mathcal{L}_{\text{cluster}}$ increases the number of triangles, thereby enhancing the clustering coefficient and localized feature interactions.

In parallel, to improve the overall connectivity of the Memory Graph and enable global information propagation, we employ the spectral properties of the graph Laplacian $\mathbf{L}$, defined as $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D}$ is the degree matrix with diagonal elements $D_{ii} = \sum_j A_{ij}$. The second smallest eigenvalue of $\mathbf{L}$, denoted as λ_2, represents the algebraic connectivity of the graph. A larger λ_2 value indicates a more connected graph with shorter average path lengths, which is critical for efficient information transmission across the entire graph structure. To incorporate this property into the model, we define a connectivity regularization term:

$$\mathcal{L}_{\text{connectivity}} = -\lambda_2, \tag{2}$$

where minimizing $\mathcal{L}_{\text{connectivity}}$ ensures the Memory Graph achieves strong global connectivity while maintaining efficient information flow.

By combining these two aspects, the small-world regularization term is expressed as:

$$\mathcal{L}_{\text{small-world}} = \alpha\mathcal{L}_{\text{cluster}} + \beta\mathcal{L}_{\text{connectivity}}, \tag{3}$$

where the hyperparameters α and β control the relative importance of clustering and connectivity during training, thereby balancing the ability to model detailed feature interactions with the need for efficient global reasoning. Overall, integrating $\mathcal{L}_{\text{small-world}}$ into the training objective endows the Memory Graph with a dual capability: clustering coefficients promote cohesive local clusters for precise feature

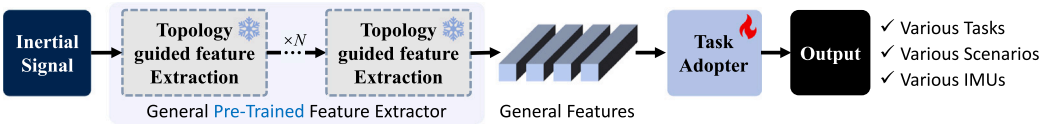

Fig. 2. Fine-tuning and deployment of pre-trained TG-FE modules. The snowflake icon indicates that the module is pre-trained and its parameters are frozen during fine-tuning, enabling general feature extraction. The flame icon represents modules where parameters are fine-tuned for specific downstream tasks, allowing adaptation to different application scenarios.

association, while enhanced algebraic connectivity accelerates global information flow. This synergy ensures robust, generalizable feature representations that efficiently capture both localized and global dependencies, enabling reliable performance across diverse IMU-based applications.

3.3. Sparsity guidance based on Tsallis entropy

The neural networks of the human brain exhibit significant sparsity [52,53]. Despite the vast number of neurons, each neuron connects with only a limited number of others [54]. This sparsity enhances the efficiency of information processing and reduces computational complexity. Furthermore, it increases network plasticity, enabling rapid adaptation to new patterns, while mitigating hallucinations caused by false [55]. Inspired by this principle, we introduce a regularization term based on Tsallis entropy to increase sparsity in the Memory Graph.

We first normalize the adjacency matrix $\mathbf{A}$ of the Memory Graph into a probability distribution $\mathbf{P}$, where its element $P_{ij} = A_{ij} / \sum_{k,l} A_{kl}$. Using this normalized distribution, we define the Tsallis entropy as:

$$S_q(\mathbf{A}) = \frac{1}{q-1}\left(1 - \sum_{i,j} P_{ij}^q\right), \tag{4}$$

where $q < 1$ emphasizes high-probability connections while suppressing weaker ones. Based on this formulation, we then define the sparsity regularization term:

$$\mathcal{L}_{\text{sparse}} = S_q(\mathbf{A}). \tag{5}$$

Minimizing $\mathcal{L}_{\text{sparse}}$ reduces the Tsallis entropy, concentrating the probability mass onto a few dominant connections. This ensures that most P_{ij} values approach zero, thereby inducing a sparse structure in the Memory Graph. The degree of sparsity can be finely controlled by adjusting the hyperparameter q (smaller values of q result in stronger sparsity). Compared to traditional L_1 regularization, which applies independent constraints on each weight, Tsallis entropy offers a global perspective by considering the entire adjacency matrix as a probability distribution. Additionally, the parameter q in Tsallis entropy offers a flexible way to adjust the sparsity level, allowing the model to be tailored to a wide range of scenarios.

By incorporating sparsity guidance into the training objective of the Memory Graph, we increase the plasticity and reduce the computational complexity of the feature extraction. Additionally, the sparse connection structure mitigates the formation of false associations, thereby supporting various IMU-based tasks with improved performance and reliability.

3.4. Cross-graph feature fusion

In our framework, feature extraction relies on a series of Topology Guided Feature Extraction modules, each refining the feature representations from the previous module. However, as the network depth increases, purely sequential architectures tend to encounter two problems: information loss and gradient degradation. Information loss manifests as critical details from early layers being overwritten by subsequent transformations, while gradient degradation occurs as gradients diminish during backpropagation, hampering the optimization of deep networks. In general deep learning architectures, residual connections are commonly used to alleviate these problems [56,57]. However, in the context of cross-graph information interaction, the effectiveness of residual connections may be constrained by the heterogeneity of graph topologies [58,59]. When different graphs exhibit substantial topological differences, such as inconsistent connection patterns or clustering characteristics, residual connections may lead to feature misalignment and even exacerbate misinformation spread. This issue is particularly pronounced in sparse graphs [60,61], where data noise from early layers can be amplified through residual connections, contaminating deeper features, especially for inertial sensor data with severe noise characteristics.

Fortunately, the proposed topology guidance (i.e. sparsity and small-world guidance) exactly provides guarantee for residual connections to achieve cross-graph feature fusion (CGFF). The sparsity guidance, based on Tsallis entropy, induces sparsity in the Memory Graph while preserving its key topological structures. Unlike traditional L1 regularization, Tsallis entropy prioritizes the retention of edges crucial for information transmission through nonlinear constraints, thereby suppressing redundant connections and reducing noise propagation. Meanwhile, small-world guidance enhances the local clustering and global connectivity of the Memory Graph, ensuring that the graph supports efficient local and global information integration while maintaining sparsity. Such configuration reduces structural discrepancies during cross-layer fusion and creating favorable conditions for residual connections. Through the synergistic effects of Tsallis entropy-based sparsification and small-world guidance, residual connections can stably transmit and fuse information without introducing excessive interference due to graph structural heterogeneity.

Specifically, the output of the lth TG-FE module is defined as the summation of the output from the $(l-1)$-th TG-FE module and the transformed features produced by the current TG-FE module:

$$\mathbf{X}_l = \mathbf{X}_{l-1} + \text{TG-FE}_l(\mathbf{X}_{l-1}), \tag{6}$$

where $\mathbf{X}_{l-1}$ represents the output feature from the previous layer, and $\text{TG-FE}_l(\cdot)$ denotes the transformation by the lth TG-FE module. This residual formulation allows each layer to focus on learning the residuals (i.e., the differences between the input and the desired output), rather than attempting to fit a complete mapping from scratch. This simplifies the optimization process, as the network can incrementally refine its representations while preserving the original input features through the direct addition. To rigorously analyze the benefits of residual connections for gradient propagation, we examine the gradient of the loss function $\mathcal{L}$ with respect to the input of the lth layer, $\mathbf{X}_{l-1}$. Using the chain rule, the gradient is expressed as:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}_{l-1}} = \frac{\partial \mathcal{L}}{\partial \mathbf{X}_l} \cdot \frac{\partial \mathbf{X}_l}{\partial \mathbf{X}_{l-1}}. \tag{7}$$

Using the residual connection formulation $\mathbf{X}_l = \mathbf{X}_{l-1} + \text{TG-FE}_l(\mathbf{X}_{l-1})$, we compute the partial derivative of $\mathbf{X}_l$ with respect to $\mathbf{X}_{l-1}$ as:

$$\frac{\partial \mathbf{X}_l}{\partial \mathbf{X}_{l-1}} = \mathbf{I} + \frac{\partial \text{TG-FE}_l(\mathbf{X}_{l-1})}{\partial \mathbf{X}_{l-1}}, \tag{8}$$

where $\mathbf{I}$ is the identity matrix. Substituting this result into the gradient formula, we obtain:

$$\frac{\partial \mathcal{L}}{\partial \mathbf{X}_{l-1}} = \frac{\partial \mathcal{L}}{\partial \mathbf{X}_l} \cdot \left(\mathbf{I} + \frac{\partial \text{TG-FE}_l(\mathbf{X}_{l-1})}{\partial \mathbf{X}_{l-1}}\right). \tag{9}$$

Table 1
The built-in inertial sensors of some smartphones. Data collection was completed in 2023, and usage times were calculated up to the end of that year.

Uses	Device	Usage time	Sensor
Pre-train	Galaxy S8	6 years	LSM6DSL
	iPhone 7 Plus	7 years	ICM20600
	Galaxy S7	7 years	LSM6DS3
Fine-tune	HUAWEI P40	3 years	LSM6DSM
	HUAWEI P40 Pro	3 years	LSM6DSO
	Mate30 Pro	4 years	ICM20690
	Realme GT	2 years	BMI160
	Xiaomi 11	3 years	BHI260AB
	OPPO Reno 6	2 years	ICM-40607
Test	Legion Phone	3 years	ICM-42605
	VIVO X30	4 years	LSM6DSM
	VIVO T2x	1 years	LSM6DSO
	iPhone 13	2 years	Undisclosed
	iPhone 12	3 years	Undisclosed
	iPhone 11 Pro	4 years	Undisclosed

Table 2
Dataset partition for pre-training, fine-tuning, and testing phases. Pre-training samples consist of unlabeled inertial data, while fine-tuning and testing samples include task-specific labels obtained through controlled motion recording facility.

Pre-train	Fine-tune			Test		
	AE	DE	AR	AE	DE	AR
129831	3542	3967	4340	3769	4053	4960

Eq. (9) demonstrates that the gradient propagation consists of two components: a direct path through the identity matrix $\mathbf{I}$, and a contribution from the non-linear transformation TG-FE$_l(\mathbf{X}_{l-1})$. The identity term $\mathbf{I}$ ensures that even if the gradient from the non-linear transformation becomes negligible (e.g., due to activation saturation), the overall gradient remains non-zero. This mechanism effectively mitigates the vanishing gradient problem, enabling stable gradient propagation across cascaded TG-FE modules.

In addition to stabilizing gradients, residual connections also enhance information preservation. In traditional deep networks, information from earlier layers is often lost as it propagates through successive transformations. By adding the input $\mathbf{X}_{l-1}$ to the output, residual connections preserve the original low-level features, allowing the network to integrate both early-stage detailed representations and high-level abstract representations, which is particularly critical for inertial sensor data since inertial signals often contains both low-frequency global trends and high-frequency local variations. Residual connections fuse these complementary components during training, improving the model's ability to capture more informative features from IMU signals.

4. Experiments and results

4.1. Experiment dataset

The built-in IMUs are the most widely used inertial sensors [62]. We take 15 smartphones with built-in IMUs to collect the inertial dataset, of which only three type of smartphone is employed for pre-training the general feature extractor TG-FE, while the data of all others are used for fine-tuning and testing. This setup imposes a significant challenge, requiring the pre-trained feature extractor to generalize across varying hardware specifications. The types of smartphones and their internal IMU specifications are shown in Table 1. Note that since the IMUs in some types of iPhones are customized by the manufacturer, their model and price are not disclosed.

Moreover, we report the sample sizes used in the pre-training, fine-tuning, and testing phases in Table 2. The pre-training data consist of natural and unstructured movements, collected without specific labeling or motion capture requirements, making them easy to obtain while highlighting the model's ability to learn generalizable features from unlabeled irregular signals. In contrast, the fine-tuning and testing samples are collected using a robotic arm and an optical motion capture system to provide precise labels for tasks such as attitude estimation, displacement estimation, and action recognition. Fig. 3 illustrates the experimental scenario. All experiments are implemented by Pytorch with NVIDIA RTX 4090 GPU and Intel(R) Xeon Gold 6330 CPU.

4.2. Comparative results

Given the scarcity of standardized feature extractors specifically designed for IMU data, we adapt some widely used models originally developed for the image domain as baselines for comparison. Inertial sensors are widely used across numerous applications, with their core functions being attitude estimation (AE), displacement estimation (DE), and action recognition (AR). As illustrated in Fig. 4, AE determines the orientation of a device, which is crucial for navigation, robotics, and stabilization systems. DE determines motion displacement, providing foundational support for tasks such as object tracking, localization, and trajectory planning. AR interprets motion patterns, enabling applications like health monitoring, gesture-based interaction, and activity tracking. These three fundamental tasks capture the essential capabilities of inertial sensors and are therefore chosen as evaluation metrics to assess the effectiveness and generalization capability of different pre-trained feature extractors.

The results in Table 3 illustrate the superior performance of our TG-FE module compared to all comparison methods across attitude estimation, displacement estimation, and action recognition tasks. This success stems from its graph-based memory structure, guided by small-world and sparsity properties, which effectively capture both local and global temporal dependencies. Stacking multiple TG-FE modules further refines feature representations, enhancing its versatility across diverse applications. Among all competing methods, TF-C and TS2Vec emerge as the closest rivals. TF-C leverages time–frequency consistency to adapt seamlessly to varying temporal dynamics, making it a strong contender for tasks with diverse requirements, though it falls short in modeling the intricate interactions unique to inertial data. TS2Vec excels through its hierarchical contrastive learning approach, adept at capturing multi-scale temporal contexts critical for tasks involving both short-term and long-term dependencies, yet its dependence on crafted data augmentations may introduce complexity and inconsistency across different scenarios. Other methods, such as TS-TCC and TNC, also demonstrate competitive performance but lag behind TF-C and TS2Vec. TS-TCC employs dual contrastive objectives to boost feature discriminability, but its reliance on high-quality augmentations makes it less robust when augmentation strategies falter. TNC prioritizes local temporal patterns, which restricts its capacity to address long-range dependencies, ultimately weakening its overall effectiveness. Traditional models like LSTM and Bi-LSTM provide reasonable outcomes thanks to their memory mechanisms, which track temporal sequences adeptly, yet their dependence on single-vector representations limits their ability to memorize complex feature interactions, placing them behind TG-FE, TS2Vec, TF-C, and TS-TCC. Meanwhile, CNN-based models such as 1D-CNN, 1D-ResNet, and 1D-DenseNet, along with handcrafted features and MAE, struggle significantly. These approaches lack robust mechanisms to model temporal dynamics, resulting in substantially weaker performance compared to all other methods considered.

4.3. Ablation study

4.3.1. Number of TG-FE modules

To investigate the role of model depth, we conduct an ablation study by varying the number of TG-FE modules from 1 to 8. Fig. 5 shows that when stacking more TG-FE modules, all three tasks, i.e., attitude

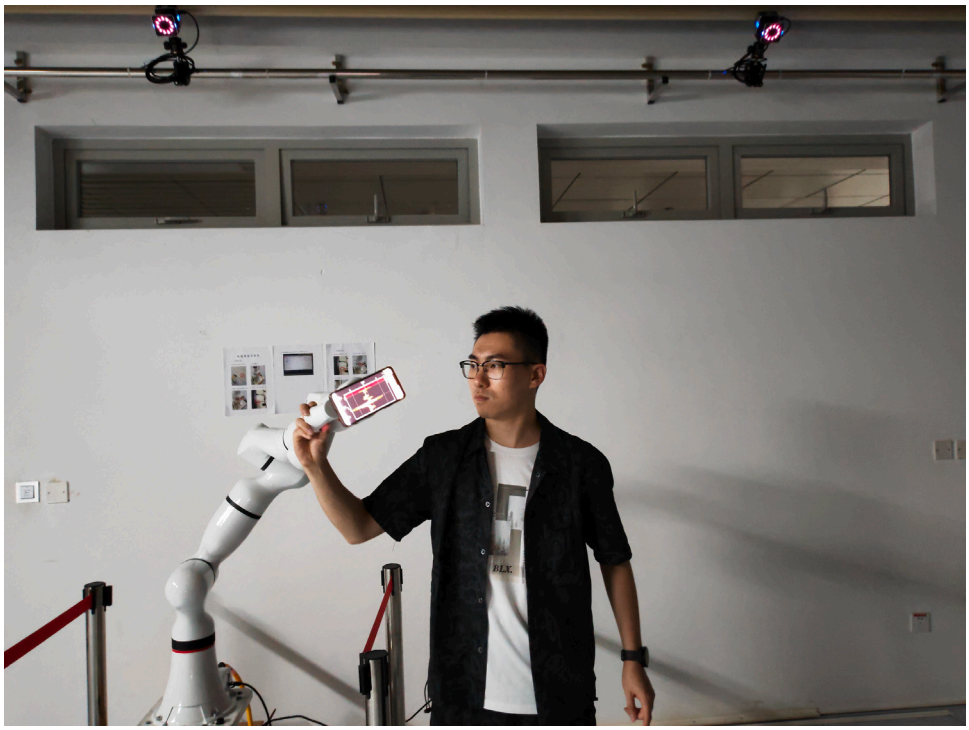

Fig. 3. Experimental setup for data collection of the built-in inertial sensor in the smartphone. This integration of automated control and high-accuracy tracking ensures a diverse and robust dataset, providing reliable ground truth for advancing inertial signal processing.

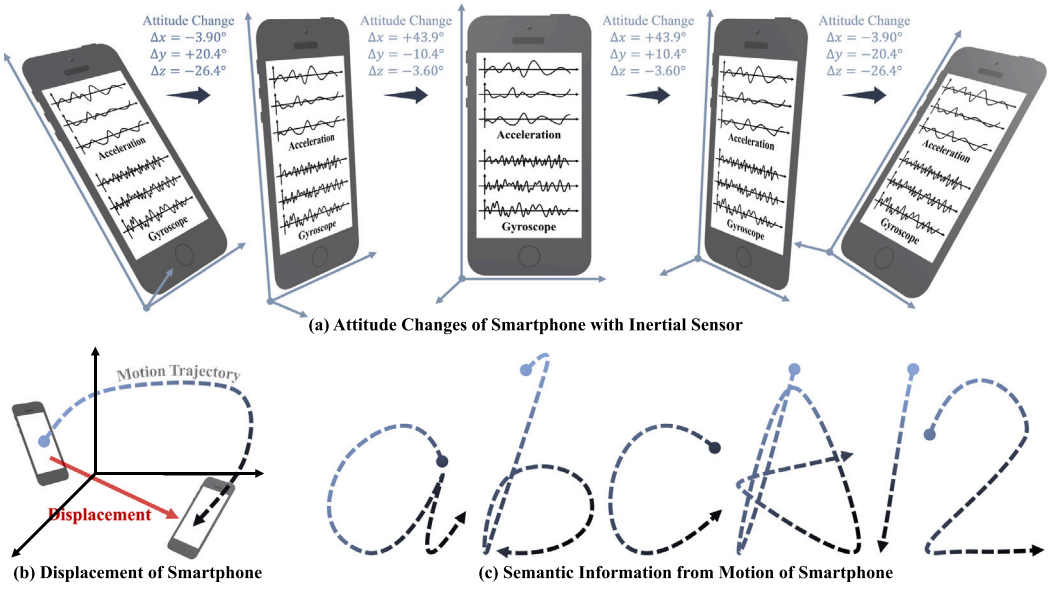

Fig. 4. Illustration of the three core tasks in inertial signal analysis. (a) Attitude estimation: capturing changes in smartphone orientation using accelerometer and gyroscope signals, represented by the angles Δx, Δy, and Δz. (b) Displacement estimation: calculating the smartphone's motion displacement in 3D space. (c) Semantic recognition: interpreting motion patterns, such as handwriting, based on inertial signal trajectories to extract meaningful semantic information.

estimation, displacement estimation, and action recognition, improve substantially until a saturation point is reached.

We notice that the optimal number of modules depends on the complexity of the task, with simpler tasks requiring fewer modules to model straightforward patterns, while more complex tasks demand additional modules to capture intricate and dynamic relationships within the data. For instance, attitude estimation follows a relatively straightforward physical principle: integrating angular velocities to derive orientation. Although measurement noise and sensor drift complicate this process, the underlying relationship is explicitly definable, and a moderate number of modules suffices to capture it effectively. Displacement estimation, inherently more challenging than attitude estimation, requires not only integrating accelerations over time but also accurately leveraging attitude information to resolve directional ambiguities. This dependency on attitude estimation amplifies the complexity of displacement computation, as errors in attitude propagate into the integration process, further compounding the impact of sensor noise and bias accumulation. Therefore, the DE task benefits from a modest increase in modules.

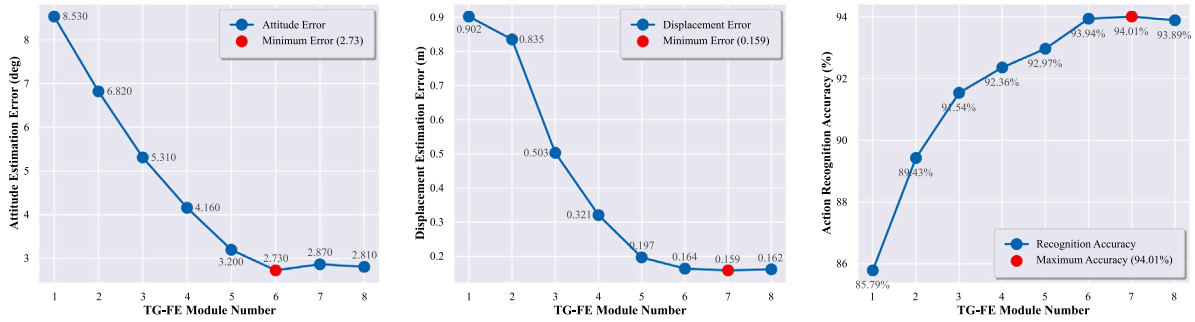

Fig. 5. Ablation study on the number of TG-FE modules. It displays the impact of different numbers of stacked TG-FE modules on the performance of attitude estimation (left), displacement estimation (middle), and action recognition (right). Note that the best-performing configurations for each task are marked in red.

Table 3
Performance comparison of pre-trained feature extractors on IMU-based tasks. We bold the best and underline the 2nd best results.

Pre-trained Feature Extractor	Attitude Estimation Error (deg) ↓	Displacement Estimation Error (m) ↓	Action Recognition Accuracy ↑
Handcrafted	18.74	1.62	66.14%
1D-CNN	17.19	1.55	68.96%
1D-ResNet	13.58	1.46	71.13%
1D-DenseNet	12.35	1.24	79.53%
LSTM	10.85	1.02	86.39%
Bi-LSTM	9.86	0.92	84.26%
MAE	14.38	1.41	73.40%
TNC	4.93	0.31	87.63%
TS2Vec	3.12	0.22	<u>92.39%</u>
TS-TCC	4.07	0.29	90.08%
TF-C	<u>3.09</u>	<u>0.20</u>	91.26%
TG-FE	**2.73**	**0.16**	**93.94%**

Unlike attitude and displacement estimation, action recognition does not rely on explicit formulas to map sensor readings to predefined classes. Instead, meaningful patterns are embedded in the unintelligible signal waveforms and cannot be described by straightforward physical models, requiring the model to extract high-level and abstract features. As we stack additional TG-FE modules, the model improves its ability to identify key patterns by grouping related features and establishing meaningful relationships between them. This progressive refinement gradually transforms raw signals into more related feature representations, supporting the classifier to accurately recognize actions across diverse scenarios. It is also worth noting that while increasing the number of TG-FE modules consistently drives improvements up to a certain depth, the gains do not continue indefinitely. Beyond six or seven modules, performance increments become marginal, and certain metrics may even fluctuate slightly. This suggests that once the model has captured the necessary patterns, additional modules may introduce redundancy.

In summary, this ablation study reveals that different tasks rely on distinct types of features, highlighting the limitations of traditional feature extractors in diverse application scenes. Tasks with simple or straightforward physical rule often rely on shallow, low-level features that can be captured with fewer TG-FE modules. In contrast, tasks involving abstract or highly dynamic patterns demand deeper, high-level representations to capture complex dependencies. Unlike traditional approaches, our modular stacking design provides the flexibility to varying tasks and demonstrates the significant advantage of the TG-FE architecture in addressing diverse and complex feature requirements.

4.3.2. Ablation on TG-FE components

To systematically evaluate the roles and synergies of key components in TG-FE, we conduct a series of ablation studies where we selectively disable small-world guidance, sparsity guidance, and cross-graph feature fusion. The ablation results are summarized in Table 4. Even in a minimal configuration with all three components removed (leaving only the Memory Graph), the model already surpasses most comparison methods, indicating the advantage of the Memory Graph for discovery of underlying motion patterns and contextual dependencies. Nonetheless, the absence of topological guidance and inter-layer fusion leads to a noticeable performance gap compared to the complete TG-FE, indicating that additional topological guidance and layer-wise information fusion are necessary to realize its full potential.

When small-world guidance or sparsity guidance is enabled independently, either mechanism significantly outperforms all other baselines, demonstrating their unique strengths in enhancing model performance. Specifically, small-world guidance increases the clustering coefficient and global connectivity of the Memory Graph, enabling it to rapidly capture localized micro-patterns while facilitating efficient long-range information transmission. This mechanism proves especially effective in tasks requiring integration of both local and global dependencies, as shown by its superior results in Table 4. In contrast, sparsity guidance emphasizes model efficiency by pruning insignificant or low-contribution edges, forcing the Memory Graph to concentrate on salient channels and critical nodes. This selective focus not only reduces noise and overfitting but also improves robustness across diverse tasks. The absence of sparsity guidance leads to excessive connections, introducing redundant or noisy patterns that degrade model generalization. The complementary nature of these mechanisms becomes evident when both are employed simultaneously. Their integration balances local clustering and global connectivity with selective pruning, resulting in a Memory Graph that excels in both capturing critical patterns and maintaining computational efficiency. The joint use of these mechanisms achieves the highest overall performance across all evaluation metrics, highlighting the necessity of emulating mathematical characteristics of human memory association in extracting features from time-series signals.

Further analysis of the performance difference with and without cross-graph feature fusion highlights the complementary roles of small-world and sparsity guidance. Without cross-graph feature fusion, the network faces challenges in combining low-level features and high-level abstractions, making it prone to learning biased or spurious patterns that do not generalize well to unseen data, thereby increasing the risk of overfitting. In such cases, sparsity guidance becomes critical by suppressing redundant connections, focusing the network on essential nodes and reducing overfitting. When cross-graph fusion is introduced, the residual pathways enable seamless interaction between features at different depths, significantly enhancing global modeling capabilities. In this scenario, small-world guidance enhances the Memory Graph's connectivity and clustering, ensuring efficient information propagation between feature clusters.

Finally, combining small-world guidance, sparsity guidance, and cross-graph feature fusion yields the best results across all the three inertial tasks, suggesting their mutual complementarity. Small-world guidance fosters efficient local and global connectivity, sparsity guidance counters overfitting and noise, and cross-graph fusion merges

Table 4
Ablation study on key components in the TG-FE module. We bold the best and underline the 2nd best results.

Architecture		Attitude Estimation Error (deg) ↓	Displacement Estimation Error (m) ↓	Action Recognition Accuracy ↑
w/o all guidance and CGFF		11.73	1.16	81.83%
w/ Small-World Guidance	w/o CGFF	8.07	0.85	87.60%
w/ Sparsity Guidance		7.19	0.63	89.75%
w/ Small-World Guidance	w/ CGFF	<u>3.87</u>	<u>0.22</u>	<u>93.27%</u>
w/ Sparsity Guidance		4.96	0.47	91.98%
w/ all (TG-FE)		**2.73**	**0.16**	**93.94%**

Table 5
Performance comparison of sparsity constraints using L1 regularization and Tsallis Entropy with different q values.

Architecture	Attitude Estimation Error (deg) ↓	Displacement Estimation Error (m) ↓	Action Recognition Accuracy ↑
L1 regularization	5.21	0.43	89.92%
Tsallis Entropy ($q = 0.1$)	3.89	0.27	90.38%
Tsallis Entropy ($q = 0.3$)	3.47	0.24	91.65%
Tsallis Entropy ($q = 0.5$)	**2.73**	**0.16**	**93.94%**
Tsallis Entropy ($q = 0.7$)	3.05	0.19	93.18%
Tsallis Entropy ($q = 0.9$)	3.11	0.21	92.87%

features from multiple depths in the graph. Together, these mechanisms equip the model to precisely capture both local particulars and high-level interdependencies, laying the groundwork for resilient inertial feature extraction and broad real-world applications.

4.4. Ablation on sparsity guidance

To validate the effectiveness of Tsallis entropy in sparsity guidance, we conduct comparative experiments against traditional L1 regularization and analyze the impact of Tsallis entropy parameters q. The results in Table 5 demonstrate that Tsallis entropy consistently outperforms L1 regularization across all evaluation metrics, achieving superior sparsity control and task performance. Specifically, Tsallis entropy with $q = 0.5$ yields optimal results in attitude estimation and displacement estimation, while maintaining the highest action recognition accuracy. This configuration reduces attitude/displacement estimation errors by 47.6%/62.8% compared to L1 regularization.

The performance variation across different q values reveals a critical trade-off between connection sparsity and feature preservation. Lower q values enforce stronger sparsity constraints, which effectively suppress noise interference in estimation tasks but may discard subtle motion patterns crucial for action recognition. Higher q values retain more connections, improving recognition accuracy at the cost of increased estimation errors due to noise propagation. The balanced parameter $q = 0.5$ achieves optimal performance by maintaining sufficient discriminative connections while eliminating redundant edges.

The advantage of Tsallis Entropy over L1 regularization lies in its nonlinear, distribution-aware penalty mechanism. Tsallis Entropy minimizes the generalized entropy of the adjacency matrix, concentrating the probability mass on a few key connections while suppressing weaker ones. This approach inherently adapts to the graph's structure, preserving critical feature interactions essential for task performance. In the context of our Memory Graph, Tsallis Entropy's ability to maintain structural integrity while inducing sparsity is particularly beneficial. In contrast, L1 regularization applies a uniform penalty to all connections, which can inadvertently weaken important feature links, particularly in tasks like action recognition that rely on subtle pattern detection. Furthermore, the tunable parameter q in Tsallis Entropy allows precise control over sparsity, enabling task-specific optimization, whereas L1 regularization's fixed penalty lacks such adaptability.

To clearly illustrate the effects of different sparse regularization methods on the Memory Graph, we plot its adjacency matrix and generated corresponding network structure diagrams to vividly display the topological changes under various sparse regularization approaches, as shown in Fig. 6. Additionally, we compute the clustering coefficient and average shortest path of the Memory Graph to quantitatively assess the impact of sparse regularization on its network structure. Without sparse regularization, the adjacency matrix reveals a dense connection pattern, with the network being highly interconnected. However, this dense structure tends to produce redundant connections, which introduce noise and lead to overfitting issues. In contrast, L1 regularization enhances sparsity by applying uniform penalties to all elements, but this comes at the cost of disrupting the Memory Graph's topological structure, significantly weakening its clustering properties. As a result, the network fragments into multiple isolated subgraphs, losing overall connectivity and rendering the average shortest path uncomputable. Furthermore, L1 regularization disrupts the temporal correlations between features. Visualizations of the adjacency matrix (left column) reveal that both non-regularized (row 1) features and those guided by Tsallis entropy (row 3 and 4) display strong temporal associations, with a notable concentration of elements near the diagonal. However, L1 regularization (row 2) markedly weakens these diagonal patterns.

Overall, Tsallis entropy excels in sparsifying the Memory Graph while preserving its topological structure and connectivity. By minimizing Tsallis entropy, the graph retains critical connections and eliminates redundancy, achieving sparsity without undermining the network's clustering properties or overall integrity. The hyperparameter q provides flexible control over sparsity levels, with smaller values enhancing sparsity and larger values retaining more connections, enabling adaptation to various task demands. This approach ensures robust universal feature extraction from inertial signals and outperforms traditional L1 regularization by effectively balancing sparsity and structural preservation, offering greater adaptability and practical utility in a concise and efficient manner.

4.5. Stability evaluation under noisy conditions

To assess the robustness of different pre-trained feature extractor, we introduce Gaussian noise into the fine-tuning signals at two levels (5% and 20%) and also compare their performance on the three tasks, as detailed in Tables 6 and 7.

In an environment with 5% Gaussian noise, the performance of various pre-trained feature extractors for IMU-based tasks reveals stark differences in robustness. Handcrafted features and CNN-based models like 1D-CNN, 1D-ResNet, and 1D-DenseNet struggle significantly, as they lack mechanisms to filter out noise effectively. MAE also falters due to its self-supervised learning approach not prioritizing noise resilience. In contrast, LSTM and Bi-LSTM show moderate robustness through their memory mechanisms, though their single-vector representations limit their ability to capture complex interactions. More advanced methods such as TNC, TS-TCC, TF-C, and TS2Vec demonstrate stronger resilience. TNC captures short-term dependencies but is disrupted by noise, while TS-TCC's contrastive learning is weakened by distorted signals. TF-C adapts to temporal dynamics but is still affected

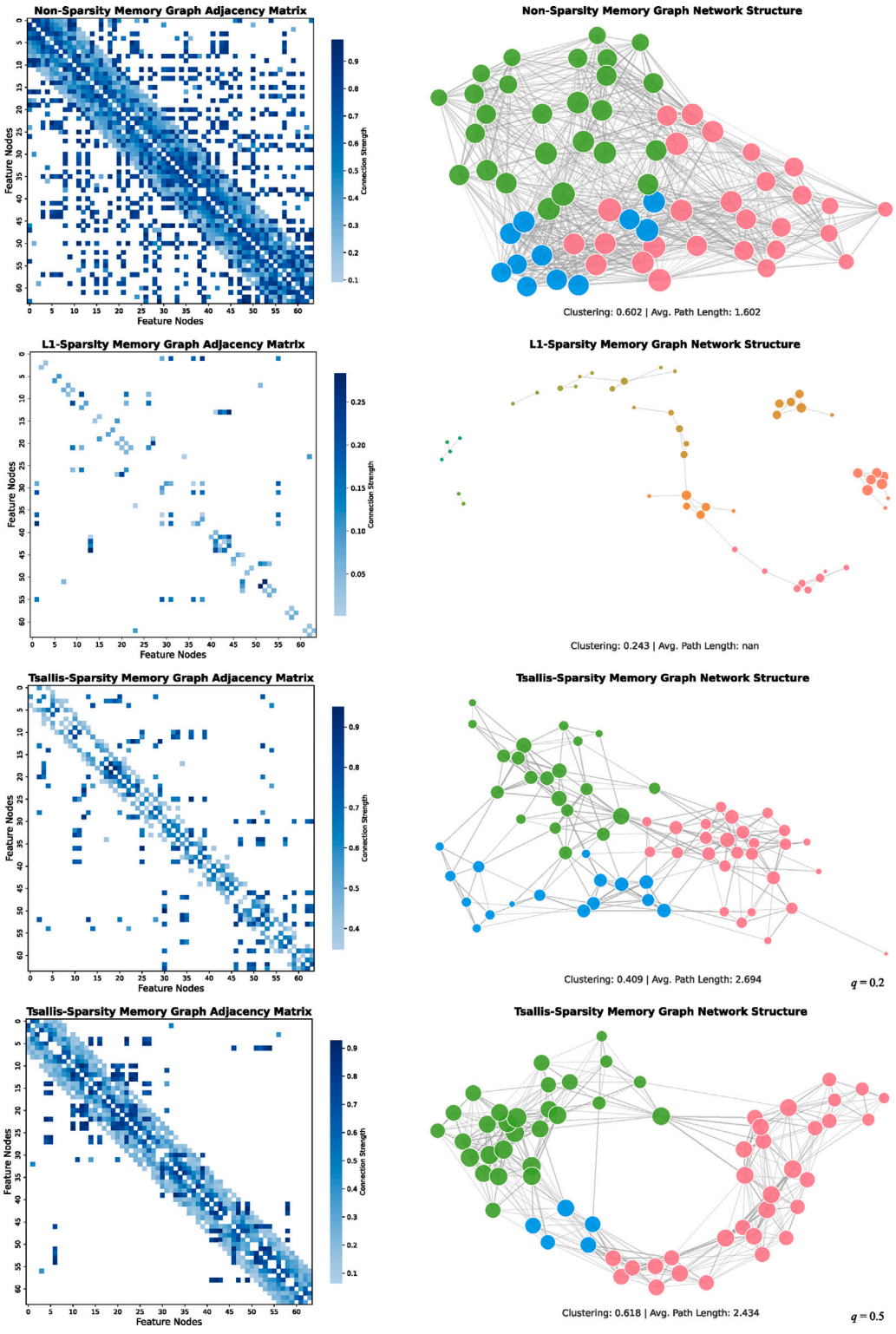

Fig. 6. Visualization of memory graphs under different sparsification strategies. From top to bottom: no sparsity constraint, L1-regularized sparsification, and two Tsallis entropy sparsification implementations with different parameter settings. The left side of each row displays the adjacency matrix visualization of feature node connections, with blue intensity indicating connection strength; the right side shows the corresponding network structure. In the network diagrams, nodes are colored according to communities, node size reflects degree centrality (number of connections), and the thickness of connecting lines indicates the strength of feature correlations. Each network diagram is annotated with its clustering coefficient and average path length, two metrics that together reflect the small-world properties of the network.

by noise in frequency-domain information. TS2Vec excels by capturing multi-scale contexts, offering robust noise resistance. However, the proposed TG-FE outperforms them all, leveraging its graph-structured memory and topology-guided design to model complex interactions while suppressing noise through cross-graph feature fusion, achieving superior stability and accuracy.

When noise escalates to 20%, the challenge intensifies, pushing most methods to their limits. Handcrafted features collapse entirely,

Table 6
Performance comparison of pre-trained feature extractors on IMU-based tasks under 5% Gaussian noise. We bold the best and underline the 2nd best results.

Pre-trained Feature Extractor	Attitude Estimation Error (deg) ↓	Displacement Estimation Error (m) ↓	Action Recognition Accuracy ↑
Handcrafted	28.31	2.19	45.83%
1D-CNN	20.05	1.81	60.55%
1D-ResNet	16.89	1.73	63.58%
1D-DenseNet	15.91	1.59	65.49%
LSTM	11.07	1.18	83.21%
Bi-LSTM	10.12	0.99	81.02%
MAE	18.44	1.67	66.26%
TNC	5.19	0.42	85.35%
TS2Vec	3.27	0.28	<u>91.44%</u>
TS-TCC	4.66	0.33	88.01%
TF-C	<u>3.18</u>	<u>0.25</u>	89.96%
TG-FE	**2.76**	**0.17**	**93.27%**

Table 7
Performance comparison of pre-trained feature extractors on IMU-based tasks under 20% Gaussian noise. We bold the best and underline the 2nd best results.

Pre-trained Feature Extractor	Attitude Estimation Error (deg) ↓	Displacement Estimation Error (m) ↓	Action Recognition Accuracy ↑
Handcrafted	35.49	3.13	22.17%
1D-CNN	26.07	2.88	45.84%
1D-ResNet	24.19	2.71	39.78%
1D-DenseNet	24.04	2.76	42.95%
LSTM	22.57	2.41	46.35%
Bi-LSTM	21.38	2.54	48.22%
MAE	25.77	2.92	44.06%
TNC	12.46	1.14	78.33%
TS2Vec	<u>5.97</u>	<u>0.59</u>	<u>84.05%</u>
TS-TCC	7.42	0.77	82.31%
TF-C	6.83	0.65	83.16%
TG-FE	**3.94**	**0.31**	**89.74%**

their reliance on static measures rendering them nearly ineffective. CNN-based models and MAE also deteriorate sharply, lacking adaptive strategies to cope with such high noise levels. LSTM and Bi-LSTM, while retaining some temporal coherence through their memory units, see significant performance drops due to their inability to handle extreme disruptions in feature interactions. TNC, despite its focus on local temporal patterns, struggles as noise overwhelms these dependencies, undermining its feature extraction capability. TS-TCC, TF-C, and TS2Vec fare better, each benefiting from distinct noise-mitigation strategies. TS-TCC's contrastive learning provides some robustness, though heavy noise weakens its discriminative power. TF-C's time–frequency approach offers adaptability, but its reliance on frequency information becomes a liability under severe distortion. TS2Vec maintains strong performance by exploiting multi-scale contexts, filtering noise through hierarchical learning. TG-FE stands out distinctly, its graph-structured memory and topology-guided framework preserving critical feature interactions even in this harsh environment. By integrating cross-graph feature fusion, TG-FE stabilizes its performance, delivering robustness and precision across all metrics.

4.6. Computational efficiency and complexity analysis

To evaluate the suitability of TG-FE for edge devices such as smartphones, we assess its computational efficiency and complexity through parameter count and inference time, as detailed in Table 8. The small-world and sparsity regularizations incur no additional parameters or inference time, as they shape the model's structure during training without impacting runtime demands. Each TG-FE module, responsible for constructing the memory graph and enabling feature interactions, contributes just 33 thousand parameters and takes 0.23 ms for inference, highlighting its efficiency. The complete model, integrating six TG-FE modules, totals 1.58 million parameters and achieves an inference time of 3.78 ms. These findings demonstrate that TG-FE delivers robust feature extraction while maintaining low computational overhead, making it practical for real-time deployment on resource-limited devices.

5. Discussion

This paper presents several significant contributions to the field of inertial signal processing. The primary innovation lies in our graph-based representation approach, which transforms time-series inertial data into a Memory Graph structure that mimics memory organization. This transformation addresses the fundamental challenge of extracting meaningful features from abstract waveforms lacking visual intuitiveness. The two complementary topology guidance mechanisms further enhance this approach. Small-world guidance promotes efficient information integration by clustering related features while maintaining global connectivity, whereas sparsity guidance based on Tsallis entropy eliminates redundant connections while preserving critical feature interactions. The Cross-Graph Feature Fusion mechanism creates favorable conditions for residual connections in graph structures by ensuring topological compatibility between layers. Moreover, the modular design of TG-FE allows flexible adaptation to task complexities by adjusting the number of stacked modules. These innovations collectively enable robust feature extraction with minimal fine-tuning requirements, demonstrated by the order-of-magnitude performance improvement over competing methods.

Despite its advantages, TG-FE exhibits certain limitations worth acknowledging. The method shows diminishing returns when stacking beyond six or seven modules, suggesting an upper bound on effective depth. Furthermore, it currently lacks the capability for online learning or continuous training during inference phases. Unlike domains such as computer vision or natural language processing, where data is abundant, inertial sensor signals are comparatively scarce, rendering each sample highly valuable. An ideal model should therefore capitalize on every new sample to enhance its performance. To address this, future research will focus on developing an enhanced TG-FE that supports online learning, enabling the model to adapt dynamically to new data in real-world applications while preserving its memory of prior knowledge, thus improving its adaptability in data-scarce scenarios.

6. Conclusion

In many real-world scenarios, inertial measurement units serve as vital sensors for applications. However, traditional single-task or hand-engineered feature extraction approaches often lack broad applicability due to the diverse feature requirements across various tasks. Although deep learning frameworks exhibit strong feature extraction capabilities, they frequently lose their generalization advantage when confronted with limited data or cross-domain scenarios. To address these challenges, we propose the Topology Guided Feature Extraction method, which constructs a graph-based structure to emulate human memory properties, thereby improving the feature extractor's generalization and enabling a unified approach for multiple tasks.

We further enhance the Memory Graph by applying small-world and sparsification principles of human memory. Specifically, small-world guidance moderately increases connectivity among critical nodes, allowing relevant information to propagate efficiently through the network. By contrast, sparsity guidance leverages Tsallis Entropy to eliminate redundant edges, thereby strengthening crucial feature interactions, reducing noise interference, and alleviating overfitting. Considering that TG-FE modules at different depths capture different aspects of the data features, we incorporate cross-graph feature fusion, which unifies shallow signal details with deeper semantic representations, ultimately enhancing the information flow of forward propagation and the gradient flow of backward propagation.

Table 8
Parameter count and inference time of TG-FE components.

Component	Parameters	Inference	Function
Transformer	1.38M	2.4 ms	Initial feature extraction
Small-world regularization	0	0 ms	Guides feature clustering
Sparsity regularization	0	0 ms	Suppresses redundant connections
Single TG-FE module	33k	0.23 ms	Builds memory and feature interaction
Overall model (6 TG-FE)	1.58M	3.78 ms	Universal feature extraction

k: $\times 10^3$, M: $\times 10^6$.

Extensive experiments show that, taken together, these three mechanisms enable the Memory Graph to better handle multiple tasks, producing noticeable gains in performance and generalization. By providing robust and reusable representations, the TG-FE would reduce the reliance on large labeled datasets and extensive model redesign for each new task. Moreover, this general feature extractor would accelerate innovation by lowering the barrier to entry for developing IMU-based applications, fostering advancements in fields where data collection is challenging or expensive.

CRediT authorship contribution statement

Yifeng Wang: Writing – original draft, Visualization, Validation, Project administration, Methodology, Data curation, Conceptualization. **Yi Zhao:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition.

Declaration of competing interest

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

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62473115, Science Center Program of National Natural Science Foundation of China under Grant 62188101, the University Innovative Team Project of Guangdong, China under Grant 2022KCXTD039, and China Scholarship Council under Grant 202306120304. We sincerely appreciate the Education Center of Experiments and Innovations (Analysis and Testing Center) at Harbin Institute of Technology, Shenzhen, China, for their support. Furthermore, we sincerely appreciate the help provided by Professor Hui Ji from the National University of Singapore.

Data availability

Data will be made available on request.

References

[1] Y. Wang, J. Xu, Y. Zhao, Wavelet encoding network for inertial signal enhancement via feature supervision, IEEE Trans. Ind. Inform. 20 (11) (2024) 12924–12934.
[2] M.T.R. Khan, E. Ever, S. Eraslan, Y. Yesilada, Human activity recognition using binary sensors: A systematic review, Inf. Fusion 115 (2025) 102731.
[3] S.S. Saha, Y. Du, S.S. Sandha, L.A. Garcia, M.K. Jawed, M. Srivastava, Inertial navigation on extremely resource-constrained platforms: Methods, opportunities and challenges, in: 2023 IEEE/ION Position, Location and Navigation Symposium, PLANS, IEEE, 2023, pp. 708–723.
[4] P. Li, W.-A. Zhang, Y. Jin, Z. Hu, L. Wang, Attitude estimation using iterative indirect Kalman with neural network for inertial sensors, IEEE Trans. Instrum. Meas. (2023).
[5] M.A. Esfahani, H. Wang, K. Wu, S. Yuan, AbolDeepIO: A novel deep inertial odometry network for autonomous vehicles, IEEE Trans. Intell. Transp. Syst. 21 (5) (2019) 1941–1950.
[6] Y. Wang, Y. Zhao, Handwriting recognition under natural writing habits based on a low-cost inertial sensor, IEEE Sens. J. 24 (1) (2024) 995–1005.
[7] C. Chen, P. Zhao, C.X. Lu, W. Wang, A. Markham, N. Trigoni, Deep-learning-based pedestrian inertial navigation: Methods, data set, and on-device inference, IEEE Internet Things J. 7 (5) (2020) 4431–4441.
[8] S. Herath, H. Yan, Y. Furukawa, Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods, in: 2020 IEEE International Conference on Robotics and Automation, ICRA, IEEE, 2020, pp. 3146–3152.
[9] M. Brossard, A. Barrau, S. Bonnabel, AI-IMU dead-reckoning, IEEE Trans. Intell. Veh. 5 (4) (2020) 585–595.
[10] S.S. Saha, S.S. Sandha, L.A. Garcia, M. Srivastava, Tinyodom: Hardware-aware efficient neural inertial navigation, Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 6 (2) (2022) 1–32.
[11] P. Zhang, Y. Li, Y. Zhuang, J. Kuang, X. Niu, R. Chen, Multi-level information fusion with motion constraints: Key to achieve high-precision gait analysis using low-cost inertial sensors, Inf. Fusion 89 (2023) 603–618.
[12] M. Brossard, S. Bonnabel, A. Barrau, Denoising imu gyroscopes with deep learning for open-loop attitude estimation, IEEE Robot. Autom. Lett. 5 (3) (2020) 4796–4803.
[13] K. Yuan, Z.J. Wang, A simple self-supervised imu denoising method for inertial aided navigation, IEEE Robot. Autom. Lett. 8 (2) (2023) 944–950.
[14] Y. Wang, Y. Zhao, Scale and direction guided GAN for inertial sensor signal enhancement, in: Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24, 2024, pp. 5126–5134, Main Track.
[15] Y. Wang, Y. Zhao, Wavelet dynamic selection network for inertial sensor signal enhancement, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 38, (14) 2024, pp. 15680–15688.
[16] H. Bello, L.A.S. Marin, S. Suh, B. Zhou, P. Lukowicz, InMyFace: Inertial and mechanomyography-based sensor fusion for wearable facial activity recognition, Inf. Fusion 99 (2023) 101886.
[17] D. Weber, C. Gühmann, T. Seel, RIANN—A robust neural network outperforms attitude estimation filters, Ai 2 (3) (2021) 444–463.
[18] C. Chen, X. Lu, A. Markham, N. Trigoni, Ionet: Learning to cure the curse of drift in inertial odometry, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (1) 2018.
[19] Y. Tang, J. Kurths, W. Lin, E. Ott, L. Kocarev, Introduction to focus issue: When machine learning meets complex systems: Networks, chaos, and nonlinear dynamics, Chaos: Interdiscip. J. Nonlinear Sci. 30 (6) (2020).
[20] H. Caesar, V. Bankiti, A.H. Lang, S. Vora, V.E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom, nuscenes: A multimodal dataset for autonomous driving, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 11621–11631.
[21] Z. Gu, Z. Gong, K. Tan, Y. Shi, C. Wu, B. Tao, H. Ding, Climb-Odom: A robust and low-drift RGB-D inertial odometry with surface continuity constraints for climbing robots on freeform surface, Inf. Fusion 117 (2025) 102880.
[22] P. Tacikowski, G. Kalender, D. Ciliberti, I. Fried, Human hippocampal and entorhinal neurons encode the temporal structure of experience, Nature (2024) 1–8.
[23] D. George, R.V. Rikhye, N. Gothoskar, J.S. Guntupalli, A. Dedieu, M. Lázaro-Gredilla, Clone-structured graph representations enable flexible learning and vicarious evaluation of cognitive maps, Nat. Commun. 12 (1) (2021) 2392.
[24] W. Sun, M. Advani, N. Spruston, A. Saxe, J.E. Fitzgerald, Organizing memories for generalization in complementary learning systems, Nature Neurosci. 26 (8) (2023) 1438–1448.
[25] B.M. Lake, M. Baroni, Human-like systematic generalization through a meta-learning neural network, Nature 623 (7985) (2023) 115–121.
[26] H.S. Courellis, J. Minxha, A.R. Cardenas, D.L. Kimmel, C.M. Reed, T.A. Valiante, C.D. Salzman, A.N. Mamelak, S. Fusi, U. Rutishauser, Abstract representations emerge in human hippocampal neurons during inference, Nature 632 (8026) (2024) 841–849.
[27] M. Peer, I.K. Brunec, N.S. Newcombe, R.A. Epstein, Structuring knowledge with cognitive maps and cognitive graphs, Trends Cogn. Sci. 25 (1) (2021) 37–54.
[28] H. Lee, J. Chen, Predicting memory from the network structure of naturalistic events, Nat. Commun. 13 (1) (2022) 4235.
[29] C.W. Lynn, A.E. Kahn, N. Nyema, D.S. Bassett, Abstract representations of events arise from mental errors in learning and memory, Nat. Commun. 11 (1) (2020) 2313.
[30] J. Zhao, F. Huang, J. Lv, Y. Duan, Z. Qin, G. Li, G. Tian, Do RNN and LSTM have long memory? in: International Conference on Machine Learning, PMLR, 2020, pp. 11365–11375.

[31] Z. Shao, X. Chen, L. Du, L. Chen, Y. Du, W. Zhuang, H. Wei, C. Xie, Z. Wang, Memory-efficient CNN accelerator based on interlayer feature map compression, IEEE Trans. Circuits Syst. I. Regul. Pap. 69 (2) (2021) 668–681.
[32] G.-M. Park, S.-M. Yoo, J.-H. Kim, Convolutional neural network with developmental memory for continual learning, IEEE Trans. Neural Netw. Learn. Syst. 32 (6) (2020) 2691–2705.
[33] H. Wu, Z. He, M. Gao, GCEVT: Learning global context embedding for vehicle tracking in unmanned aerial vehicle videos, IEEE Geosci. Remote. Sens. Lett. 20 (2022) 1–5.
[34] H. Wu, H. Sun, K. Ji, G. Kuang, Temporal-spatial feature interaction network for multi-drone multi-object tracking, IEEE Trans. Circuits Syst. Video Technol. (2024).
[35] Y. Wang, Y. Zhao, Heros-gan: honed-energy regularized and optimal supervised gan for enhancing accuracy and range of low-cost accelerometers, in: Proceedings of the AAAI Conference on Artificial Intelligence, 39, (20) 2025, pp. 21348–21356.
[36] J. Sun, H. Sun, L. Lei, K. Ji, G. Kuang, TirSA: A three stage approach for UAV-satellite cross-view geo-localization based on self-supervised feature enhancement, IEEE Trans. Circuits Syst. Video Technol. (2024).
[37] J. Nie, Z. Dong, Z. He, H. Wu, M. Gao, FAML-RT: Feature alignment-based multi-level similarity metric learning network for a two-stage robust tracker, Inform. Sci. 632 (2023) 529–542.
[38] Z. Li, Z. Rao, L. Pan, P. Wang, Z. Xu, Ti-mae: Self-supervised masked time series autoencoders, 2023, arXiv preprint arXiv:2301.08871.
[39] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.K. Kwoh, X. Li, C. Guan, Time-series representation learning via temporal and contextual contrasting, in: Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, 2021, pp. 2352–2359.
[40] S. Tonekaboni, D. Eytan, A. Goldenberg, Unsupervised representation learning for time series with temporal neighborhood coding, in: International Conference on Learning Representations, 2021.
[41] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, B. Xu, Ts2vec: Towards universal representation of time series, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, (8) 2022, pp. 8980–8987.
[42] X. Zhang, Z. Zhao, T. Tsiligkaridis, M. Zitnik, Self-supervised contrastive pre-training for time series via time-frequency consistency, Adv. Neural Inf. Process. Syst. 35 (2022) 3988–4003.
[43] K. Zhang, Q. Wen, C. Zhang, R. Cai, M. Jin, Y. Liu, J.Y. Zhang, Y. Liang, G. Pang, D. Song, et al., Self-supervised learning for time series analysis: Taxonomy, progress, and prospects, IEEE Trans. Pattern Anal. Mach. Intell. (2024).
[44] X. Liao, A.V. Vasilakos, Y. He, Small-world human brain networks: perspectives and challenges, Neurosci. Biobehav. Rev. 77 (2017) 286–300.
[45] R.F. Betzel, L. Byrge, Y. He, J. Goñi, X.-N. Zuo, O. Sporns, Changes in structural and functional connectivity among resting-state networks across the human lifespan, Neuroimage 102 (2014) 345–357.
[46] D.S. Bassett, A. Meyer-Lindenberg, S. Achard, T. Duke, E. Bullmore, Adaptive reconfiguration of fractal small-world human brain functional networks, Proc. Natl. Acad. Sci. 103 (51) (2006) 19518–19523.
[47] D.S. Bassett, O. Sporns, Network neuroscience, Nature Neurosci. 20 (3) (2017) 353–364.
[48] E. Bullmore, O. Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, Nature Rev. Neurosci. 10 (3) (2009) 186–198.
[49] M.P. Van Den Heuvel, O. Sporns, Rich-club organization of the human connectome, J. Neurosci. 31 (44) (2011) 15775–15786.
[50] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world'networks, Nature 393 (6684) (1998) 440–442.
[51] S.H. Strogatz, Exploring complex networks, Nature 410 (6825) (2001) 268–276.
[52] H.A. Elmarakeby, J. Hwang, R. Arafeh, J. Crowdis, S. Gang, D. Liu, S.H. Al-Dubayan, K. Salari, S. Kregel, C. Richter, et al., Biologically informed deep neural network for prostate cancer discovery, Nature 598 (7880) (2021) 348–352.
[53] D.-H. Lim, S. Wu, R. Zhao, J.-H. Lee, H. Jeong, L. Shi, Spontaneous sparse learning for PCM-based memristor neural networks, Nat. Commun. 12 (1) (2021) 319.
[54] A. Aghazadeh, H. Nisonoff, O. Ocal, D.H. Brookes, Y. Huang, O.O. Koyluoglu, J. Listgarten, K. Ramchandran, Epistatic net allows the sparse spectral regularization of deep neural networks for inferring fitness functions, Nat. Commun. 12 (1) (2021) 5225.
[55] A. Ghavasieh, M. De Domenico, Diversity of information pathways drives sparsity in real-world networks, Nat. Phys. 20 (3) (2024) 512–519.
[56] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[57] G. Huang, Z. Liu, L. Van Der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700–4708.
[58] K. Xu, C. Li, Y. Tian, T. Sonobe, K.-i. Kawarabayashi, S. Jegelka, Representation learning on graphs with jumping knowledge networks, in: Proceedings of the 35th International Conference on Machine Learning, in: Proceedings of Machine Learning Research, vol. 80, PMLR, 2018, pp. 5453–5462.
[59] Q. Li, Z. Han, X.-M. Wu, Deeper insights into graph convolutional networks for semi-supervised learning, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 32, (1) 2018.
[60] J. Gasteiger, S. Weißenberger, S. Günnemann, Diffusion improves graph learning, Adv. Neural Inf. Process. Syst. 32 (2019).
[61] K. Oono, T. Suzuki, Graph neural networks exponentially lose expressive power for node classification, in: International Conference on Learning Representations, vol. 8, 2020.
[62] X. Yang, H. Zhang, Y. Zhuang, Y. Wang, M. Shi, Y. Xu, Ulidr: An inertial-assisted unmodulated visible light positioning system for smartphone-based pedestrian navigation, Inf. Fusion 113 (2025) 102579.