Chinese Inertial GAN for Writing Signal Generation and Recognition

Anonymous authors
Paper under double-blind review

Abstract

Disabled people constitute a significant part of the global population, deserving of inclusive consideration and empathetic support. However, the current human-computer interaction based on keyboards may not meet the requirements of disabled people. The small size, ease of wearing, and low cost of inertial sensors make inertial sensor-based writing recognition a promising human-computer interaction option for disabled people. However, accurate recognition relies on massive inertial signal samples, which are hard to collect for the Chinese context due to the vast number of characters. Therefore, we design a Chinese inertial generative adversarial network (CI-GAN) containing Chinese glyph encoding (CGE), forced optimal transport (FOT), and semantic relevance alignment (SRA) to acquire unlimited high-quality training samples. Unlike existing vectorization focusing on the meaning of Chinese characters, CGE represents the shape and stroke features, providing glyph guidance for GAN to generate writing signals. FOT establishes a triple-consistency constraint between the input prompt, output signal features, and real signal features, ensuring the authenticity and semantic accuracy of the generated signals and preventing mode collapse and mixing. SRA constrains the consistency between the semantic relationships among multiple outputs and the corresponding input prompts, ensuring that similar inputs correspond to similar outputs (and vice versa), significantly alleviating the hallucination problem of generative models. The three modules guide the generator while also interacting with each other, forming a coupled system. By utilizing the massive training samples provided by CI-GAN, the performance of six widely used classifiers is improved from 6.7% to 98.4%, indicating that CI-GAN constructs a flexible and efficient data platform for Chinese inertial writing recognition. Furthermore, we release the first Chinese writing recognition dataset based on inertial sensors in GitHub.

1 Introduction

As efficient motion-sensing components, inertial sensors can measure the acceleration and angular velocity of moving objects Saha et al. (2022); Esfahani et al. (2019a); Zhang et al. (2020b); Liu et al. (2020b). Due to their small size, ease of integration, low power consumption, and low cost, inertial measurement units (IMU) are widely used in electronic devices such as smartphones, smartwatches, and fitness bands Weber et al. (2021); Gromov et al. (2019); Li et al. (2023); Herath et al. (2020), making them particularly suitable for human-computer interaction (HCI) systems. Unlike vision-based HCI systems, IMU-based HCI systems are robust to variations in lighting, environmental conditions, and occlusions, making them an ideal choice for a wide range of applications, such as virtual and augmented reality, healthcare and rehabilitation, education and training, and smart device control Wang et al. (2020). A notable application of IMU-based HCI systems is in assisting disabled individuals. By capturing the subtle movements of a user's hand or other body parts, inertial sensors can translate these motions into written text, enabling effective communication and interaction without the need for a traditional keyboard, even for users with visual impairments or in complete darkness. Providing tailored HCI solutions not only enhances their quality of life but also facilitates their integration into society, enabling greater participation in education, employment, and social activities. Such technological advancements hold profound significance, creating a more inclusive and equitable society.

However, implementing human-computer interaction in the context of Chinese language presents significant challenges due to the complexity and vast number of Chinese characters. For any recognition model aimed at accurately analyzing the complex strokes and structures of Chinese characters, it is crucial to train the model with extensive, diverse writing samples Wang & Zhao (2024). Considering that the collection and processing of Chinese writing samples are laborious and require high data quality and diversity, this task becomes exceedingly challenging and increasingly difficult as the number of characters increases. Therefore, generating realistic Chinese writing signals based on inertial sensors has become a central technological challenge in recognizing Chinese writing.

To acquire high-quality, diverse samples of inertial Chinese writing, we applied GAN for IMU writing signal generation for the first time and proposed CI-GAN, which can generate unlimited inertial writing signals for an input Chinese character, thereby providing rich training samples for Chinese writing recognition classifiers. CI-GAN provides a more intuitive and natural human-computer interaction method for the Chinese context and advances the application of smart devices with Chinese input. The main contributions of this paper are summarized as follows.

- Considering traditional Chinese character embedding methods that only focus on the meaning of characters, we propose a Chinese glyph encoding (CGE), which represents the shape and structure of Chinese characters. CGE not only injects glyph and writing semantics into the generation of inertial signals but also provides new tools for studying the relationships between character structures.
- We propose a forced optimal transport (FOT) loss for GAN, which not only avoids mode collapse and mode mixing during signal generation but also ensures feature consistency between the generated and real signals through a designed forced feature matching mechanism, thereby enhancing the authenticity of the generated signals.
- To inject batch-level character semantic correlations into GAN and establish macro constraints, we propose a semantic relevance alignment (SRA), which aligns the relevance between generated signals and corresponding Chinese glyphs, thereby ensuring that the motion characteristics of the generated signal conform to the Chinese character structure.
- Utilizing the training samples provided by CI-GAN, we increase the Chinese writing recognition performance of six widely used classifiers from 6.7% to 98.4%. Furthermore, we provide the application scenarios and strategies of 6 classifiers in writing recognition according to their performance metrics. For the sake of sharing, we release the first Chinese writing recognition dataset based on inertial sensors in GitHub.

2 Related Work

The technology for recognizing Chinese handwriting movements has the potential to bridge the gap between traditional writing and digital input, providing disabled individuals with a natural way of writing and greatly enhancing their ability to participate in digital communication, education, and employment. It also offers a new human-computer interaction avenue for normal people. Hence, Chinese handwriting movement recognition has garnered significant attention in recent years, leading to numerous related research achievements. Ren et al. utilized the Leap Motion device to propose an RNN-based method for recognizing Chinese characters written in the air Ren et al. (2019). The Leap Motion sensor, consisting of two infrared emitters and two cameras, can accurately capture the motion of hands in three-dimensional (3D) space Guerra-Segura et al. (2021). However, the Leap Motion device is sensitive to lighting conditions, and either too strong or too weak light can interfere with the transmission and reception of infrared rays, affecting the recognition effect Cortes-Perez et al. (2021). Additionally, the detection space of the Leap Motion device is an inverted quadrangular pyramid, limiting its field of view. Movements outside this range cannot be captured. Most importantly, the Leap Motion device is expensive and requires a connection to a computer or VR headset to function, severely limiting its application prospects Ovur et al. (2021).

As wireless networks become more prevalent, Wi-Fi signals are gradually being applied to motion capture Xiao et al. (2021); Wang et al. (2022). Since Wi-Fi signals can penetrate objects and are unaffected by lighting conditions, they have a broader application scope than optical motion capture systems Gao et al. (2023); Regani et al. (2021). Guo et al. used the channel state information (CSI), extracted from Wi-Fi signals reflected by hand movements, to recognize 26 air-written English letters

Guo et al. (2020). However, while Wi-Fi signals do not have visual range limitations and can penetrate obstacles, they are easily disturbed by other signals on the same unlicensed band, severely affecting system performance. Moreover, the sampling frequency and resolution of Wi-Fi signals are very limited, making it difficult to capture detailed information during the writing process and, thus, hard to recognize air-written Chinese characters accurately Gao et al. (2022); Gu et al. (2017).

Despite the advantages of low cost, wearability, and low power consumption offered by inertial sensors, there is currently a lack of large-scale, high-quality public datasets, causing few studies to use inertial sensors for 3D Chinese handwriting recognition Montesinos et al. (2018); Chen et al. (2020); Saha et al. (2023); Esfahani et al. (2019b). To collect data, Zhang et al. employed 12 volunteers, each of whom was asked to write the assigned Chinese characters on paper 30 times Zhang et al. (2020a). The inertial measurement unit built into smartwatches was used to collect the motion signals of the volunteers while writing, ultimately achieving a recognition accuracy of 90.2% for 200 Chinese characters. However, this study aims to identify the signals of normal individuals writing on paper, which is not applicable to people with disabilities. Moreover, this method can only realize desktop-based 2D writing recognition, which reduces the comfort and flexibility of the writing process, inherently limiting the application scenarios of Chinese handwriting recognition. Additionally, this method cannot effectively recognize massive Chinese characters due to the physical and mental limitations of volunteers for data collection. Considering the vast number of Chinese characters, providing large-scale, high-quality writing signal samples for each character is nearly impossible, which has become the most significant bottleneck limiting the development of Chinese handwriting recognition technology based on inertial sensors. Therefore, designing a model for generating Chinese handwriting signals provides researchers with an endless supply of signal samples and a flexible, convenient experimental data platform, accelerating the development and testing of new algorithms and supporting the research and application of Chinese handwriting recognition.

3 Method

To generate inertial writing signals for Chinese characters, we propose the Chinese inertial generative adversarial network (CI-GAN), as shown in Fig. 1. For an input Chinese character, its one-hot encoding is transformed into glyph encoding using our designed glyph encoding dictionary, which stores the glyph shapes and stroke features of different Chinese characters. Thus, the obtained Chinese glyph encoding contains rich writing features of the input character. This glyph encoding, along with a random noise vector, is fed into a GAN, generating the synthetic IMU signal for the character, where glyph encoding provides glyph and stroke features of the input character, while the random noise introduces randomness to the virtual signal generation, ensuring the diversity and variability of the generated signals. To ensure that the GAN learns the IMU signal patterns for each character, we designed a forced optimal transport (FOT) loss, which not only mitigates the issues of mode collapse and mode mixing typically observed in GAN frameworks but also forces the generated IMU signals to closely resemble the actual handwriting signals in terms of semantic features, fluctuation trends, and kinematic properties. Moreover, a semantic relevance alignment (SRA) is proposed to provide batch-level macro constraints for GAN, thereby keeping the correlation between generated signals consistent with the correlation between Chinese character glyphs. Equipped with CGE, FOT and SRA, CI-GAN can provide unlimited high-quality training samples for Chinese character writing recognition, thereby enhancing the accuracy and robustness of various classifiers.

3.1 Chinese Glyph Encoding

In one-hot encoding, each Chinese character is represented by a high-dimensional sparse vector where all characters are equidistant in the vector space, causing the loss of the rich semantic and glyph information inherent in the characters. While commonly used Chinese character embeddings capture semantic meanings, they fail to encode glyph-specific features such as shape, structure, and writing strokes. For example, the characters ”天” (sky) and ”夫” (husband) exhibit similar writing motions but have vastly different meanings. To address this, we propose a Chinese Glyph Encoding (CGE) method that encodes Chinese characters based on their glyph shapes and writing actions.

Since the glyph shapes of Chinese characters are inherently embedded in the writing motions recorded by inertial sensor signals, we design a learnable weight matrix W applied after the one-hot input layer to capture glyph information. When a Chinese character is input, its one-hot encoding is multiplied by

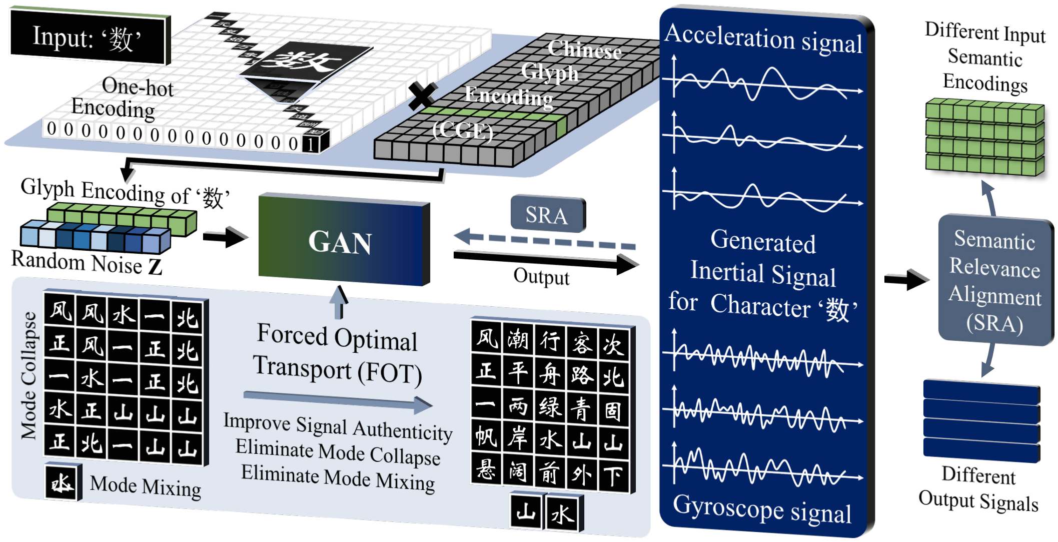

Figure 1: Flowchart of Chinese inertial generative adversarial network. The Chinese character "数" is input into the model, and its one-hot encoding is converted into glyph encoding (green cubes), which is then input into GAN together with random noise (blue cubes of different colors).

W, effectively retrieving the corresponding row of W as the character's glyph encoding. This weight matrix functions as a glyph encoding dictionary for all characters. However, without proper guidance, the dictionary may assign similar glyph encodings to characters with distinct glyphs. To prevent this, we introduce Glyph Encoding Regularization (GER), which enforces orthogonality among encoding vectors and increases their information entropy. This ensures that the encoding preserves as much glyph-specific information as possible, avoiding the triviality of one-hot encoding.

Specifically, we use the α-order Rényi entropy to measure the information content of the glyph encoding dictionary W, calculated as follows:

$$S_\alpha(W) = \frac{1}{1-\alpha}\log_2(tr(\tilde{G}^\alpha)), \text{where } \tilde{G}_{ij} = \frac{1}{N}\frac{G_{ij}}{\sqrt{G_{ii}\cdot G_{jj}}}, G_{ij} = \left\langle W^{(i)}, W^{(j)} \right\rangle. \quad (1)$$

where, N represents the number of Chinese characters, which corresponds to the number of rows in the weight (encoding) matrix W. G is the Gram matrix of W, where G_{ij} equal to the inner product of the i-th and j-th rows of W, and $\tilde{G}$ is the trace-normalized G, i.e., $tr(\tilde{G}) = 1$. In similar problems, α is generally set to 2 for optimal results. $S_\alpha(W)$ measures the information content of the glyph encoding matrix W. A larger $S_\alpha(W)$ indicates more information encoded in W, meaning the glyph encodings are more informative. Meanwhile, as $S_\alpha(W)$ increases, all elements in the Gram matrix G are forced to decrease, indicating that different encoding vectors have stronger orthogonality. It is evident that the improvement of $S_\alpha(W)$ simultaneously enhances the information content and the orthogonality among the encodings. In light of this, the glyph encoding regularization R_{encode} is constructed as $R_{\text{encode}} = \frac{1}{S_\alpha(W)}$. As R_{encode} decreases during training, $S_\alpha(W)$ gradually increases, meaning the glyph encoding dictionary stores more information while enhancing the orthogonality among all Chinese glyph encodings, effectively representing the differences in glyph shapes among all characters. Thus, this glyph encoding can inject sufficient glyph information into GAN, ensuring that the generated signals maintain consistency with the target character's glyph.

3.2 Forced Optimal Transport

Unlike images, the quality of signals cannot be readily assessed through visual inspection. Thus, stringent constraints are essential to ensure the reliability and authenticity of the generated signals, especially in following physical laws and simulating the potential dynamical characteristics of actual motions. To this end, we propose the forced feature matching (FFM), which ensures that the generated signal feature closely matches the real signal feature and the corresponding glyph encoding.

Specifically, we use a pre-trained variational autoencoder to extract the real signal feature h_T and generated signal feature h_G. Then, the consistency of h_T, h_G, and the corresponding glyph encoding e is constrained by $\mathcal{L}_{FFM}$. FFM establishes a triple-consistency constraint for generative models: input prompt, generated signal features, and truth signal features, which not only improves the realism of the generated signals but also ensures their semantic accuracy.

$$\mathcal{L}_{FFM} = 1 - \frac{\langle h_G, h_T \rangle + \langle h_G, e \rangle + \langle e, h_T \rangle}{\|h_G\| \, \|h_T\| + \|h_G\| \, \|e\| + \|e\| \, \|h_T\|}. \tag{2}$$

Another challenge lies in the mode collapse and mode mixing issue inherent to GAN architectures. Mode collapse limits the diversity of generated signal samples, causing GAN to generate signals only for a few Chinese characters, regardless of the diversity of input. On the other hand, mode mixing problems cause the generated signal to contain blend characteristics of multiple modes. Therefore, we introduce the loss function of OT-GAN Salimans et al. (2018), which utilizes Wasserstein distance as a constraint to ensure stable gradients, thereby preventing mode collapse and mixing. Combing FFM and OT constraints, we can obtain the forced optimal transport loss $\mathcal{L}_{FOT} = W(\mathbb{P}_T, \mathbb{P}_G) + \lambda \cdot \mathcal{L}_{FFM}$, where $W(\mathbb{P}_T, \mathbb{P}_G)$ is the optimal transport loss, representing the Wasserstein distance between the distributions of real and generated signals, enhancing the stability and diversity of the samples. λ is a weighting coefficient for the forced feature matching loss $\mathcal{L}_{FFM}$. As $\mathcal{L}_{FFM}$ decreases during training, the generated signals increasingly approximate the characteristics of real signals.

3.3 Semantic Relevance Alignment

As motion records of Chinese writing, the semantic relationships between generated signals should align with the relationships between Chinese character glyphs. To ensure the generated inertial signals accurately reflect the character relationships between Chinese character glyphs, we propose semantic relevance alignment (SRA), which ensures consistency between the glyph encoding relationships and the signal feature relationships, thereby providing batch-level macro guidance for GANs and enhancing the quality of the generated signals. For each batch of input Chinese characters, we compute the pairwise cosine similarities of their Chinese glyph encodings to form an encoding similarity matrix M_e. Simultaneously, the pairwise cosine similarities of generated signal features (extracted by the pre-trained VAE) are computed to form a feature similarity matrix M_h. Then, the loss of semantic relevance alignment $\mathcal{L}_{SRA} = \|M_h - M_e\|_2^2$ is established to minimize the difference between the two matrices, thereby ensuring that the semantic relationships in the input character glyphs are accurately contained in the generated signals.

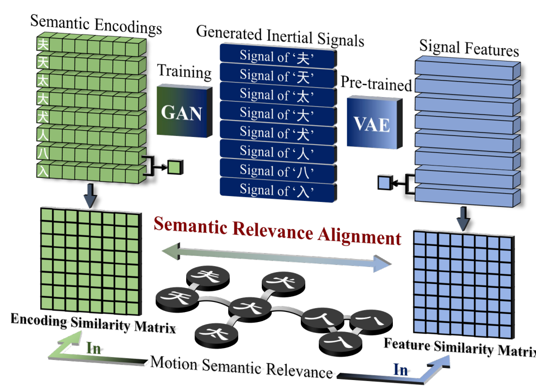

Figure 2: Diagram of SRA.

SRA aligns the relationships between outputs and their corresponding prompts, significantly reducing hallucinations in generative models. A June 2024 Nature paper (after our experiments) Farquhar et al. (2024) shares a similar approach, which demonstrated that ensuring consistent outputs for similar prompts can reduce hallucinations. However, SRA goes a step further by not only aligning individual outputs but also ensuring that the relationships between different prompts are mirrored in the relationships between the outputs. This deeper alignment significantly reduces hallucinations and enhances the model's overall practicality and stability, offering a more robust and reliable solution.

3.4 Module Interaction

CGE, FOT, and SRA not only guide and constrain the generator but also interact with each other, as shown in Fig. 3. The Chinese glyph encoding not only provides semantic guidance to the generator but also supplies the necessary encoding for FOT and SRA, and it is also supervised in the process. FOT and SRA share the VAE and generated signal features, providing different constraints for the generator, with FOT focusing on improving signal authenticity and enhancing the model's cognition of different categories through the semantic information injected by CGE, thereby mitigating mode collapse and mode mixing. SRA ensures consistency between the relationships of multiple outputs and prompts through group-level supervision, which helps alleviate the hallucination problem of generative models. In summary, the three modules proposed in CI-GAN, CGE, FOT, and SRA are innovative and interlinked, significantly enhancing the performance of GANs in generating inertial sensor signals, as evidenced by numerous comparative and ablation experiments. This method is a typical example of deep learning empowering the sensor domain and has been recognized by the industry and adopted by a medical wearable device manufacturer. It has the potential to become a benchmark for data augmentation in the sensor signal processing field.

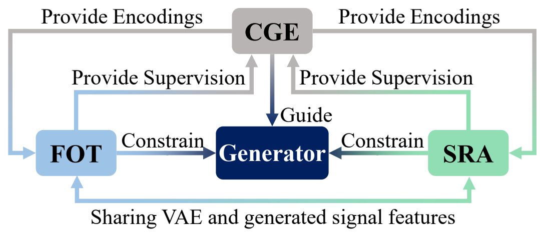

Figure 3: Interaction of three modules and generator in CI-GAN.

4 Experiments and Results

4.1 Data Collection and Experimental Setup

We invited nine volunteers, each using their smartphone's built-in inertial sensors to record handwriting movements. The nine smartphones and their corresponding sensor models are listed in Table 1. Each volunteer held their phone according to their personal habit and wrote 500 Chinese characters in the air (sourced from the "Commonly Used Chinese Characters List" published by the National Language Working Committee and the Ministry of Education), writing each character only once. In total, we obtained 4500 samples of Chinese handwriting signals. We randomly selected 1500 samples from three volunteers as the training set, while the remaining 3000 samples from six volunteers were used as the test set without participating in any training. All experiments are implemented by Pytorch 1.12.1 with an Nvidia RTX 2080TI GPU and Intel(R) Xeon(R) W-2133 CPU.

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.

Dataset	Smartphone	Release Time	IMU	Unit price
Training	iPhone 13 pro	Sep. 2021	Undisclosed	/
	HUAWEI P40	Mar. 2020	LSM6DSM	\$0.30
	HUAWEI P40 Pro	Apr. 2020	LSM6DSO	\$0.33
Testing	iPhone 14	Sep. 2022	Undisclosed	/
	iPhone 15	Sep. 2023	Undisclosed	/
	VIVO T2x	May. 2022	LSM6DSO	\$0.33
	OPPO Reno 6	May. 2021	ICM-40607	\$0.28
	Realme GT	Mar. 2021	BMI160	\$0.21
	Redmi K40	Mar. 2021	ICM-40607	\$0.28

Signal collection and segmentation in Chinese handwriting recognition are exceptionally challenging. Volunteers continuously wrote different Chinese characters, and accurately locating the corresponding signal segments from long streams required substantial effort, please refer to the Appendix B for details. Synchronizing optical motion capture equipment and manually aligning inertial signals frame by frame to extract the start and end points of each character demanded precise and time-consuming work. This meticulous process highlights the difficulty and complexity of data collection, making our achievement of 4,500 signal samples a significant milestone. By contrast, CI-GAN streamlines this process, generating handwriting signals directly from input characters, eliminating the need for laborious segmentation, and offering a far more efficient data collection platform.

4.2 Signal Generation Visualization

To visually demonstrate the signal generation effect of CI-GAN, we visualized the real and generated inertial sensor signals of the handwriting movements for the Chinese characters "科" and "学", respectively. In these figures, the blue curves represent the three-axis acceleration signals, and the yellow curves represent the three-axis gyroscope signals. It can be observed that the generated signals closely follow the overall fluctuation trends of the real signals, indicating that CI-GAN effectively preserves the handwriting movement information of the real signals. To further verify the consistency of the movement characteristics between the generated and real signals, we employed a classical inertial navigation method Grewal et al. (2007) to convert both the real and generated signals into corresponding motion trajectories, as shown in the third column of Fig. 4. It is important to note that the purpose of reconstructing the motion trajectories is not to precisely reproduce every detail of the writing process but to compare the overall shape similarity between the trajectories derived from real and generated signals. The highly similar shapes between the trajectories indicate that the generated signals accurately capture the structural information of different Chinese characters and can effectively simulate the key movement features of the handwriting process, including stroke order, movement direction changes, and velocity variations. Additionally, the obvious differences in details between the real and generated signals demonstrate CI-GAN's capability to generate diverse signals. Since the generated signals maintain the core movement and semantic features of the handwriting process, these differences do not impair the overall recognition of the characters but rather enhance the diversity of the training data.

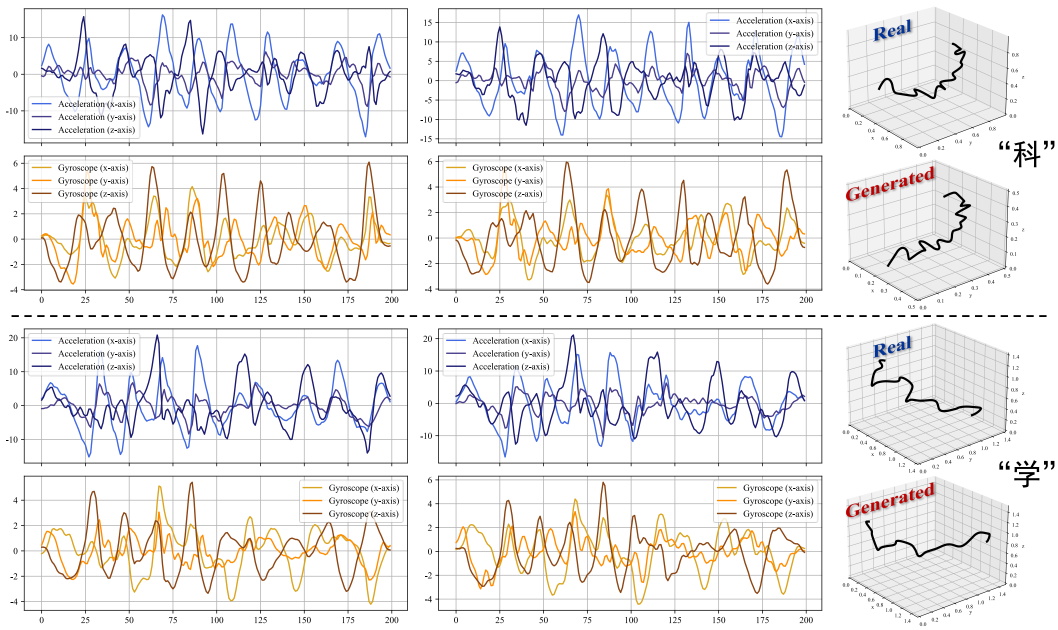

Figure 4: The visualization results of the 6-axis signals recorded by the inertial sensor for different Chinese character writing movements and the corresponding generated signals. The left side is the real signal, the middle is the generated signal, and the right side is the reconstructed writing trajectory.

To demonstrate CI-GAN's ability to generate unlimited high-quality signals, we generated five IMU handwriting signals for the same character ”王” and compared them with a real handwriting signal, as shown in Fig. 5. We chose this character because its strokes are distinctly separated, making it easier to compare the consistency of stroke features between the generated and real signals. It can be observed that the generated signals exhibit similar fluctuation patterns to the real signal in all three axes of acceleration and gyroscope measurements, verifying CI-GAN's precision in capturing dynamic handwriting characteristics. Although the overall trends of the generated signals align with the real signal, the individual features show variations, demonstrating CI-GAN's potential to produce large-scale, high-quality, and diverse IMU handwriting signal samples.

4.3 Comparative Experiments

4.3.1 Classifier Comparison based on CI-GAN

Using the CI-GAN, we generated 30 virtual IMU handwriting signals for each character, resulting in a total of 16500 training samples. To evaluate the impact of the generated signals on handwriting recognition tasks, we trained six representative time-series classification models with these training samples: 1DCNN, LSTM, Transformer, SVM, XGBoost, and Random Forest (RF). We then tested the performance of these classifiers on the test set, as shown in Fig. 6. When the number of training samples is small (1500 real samples), the recognition accuracy of all classifiers is poor, with the highest accuracy being only 6.7%. As the generated training samples are introduced, all classifiers' recognition accuracy improves significantly, whereas deep learning ones such as 1DCNN, LSTM, and Transformer show the most notable improvement. When the number of training samples reaches 15000, the recognition accuracy of 1DCNN can reach 95.7%, improving from 0.87% (without data augmentation).

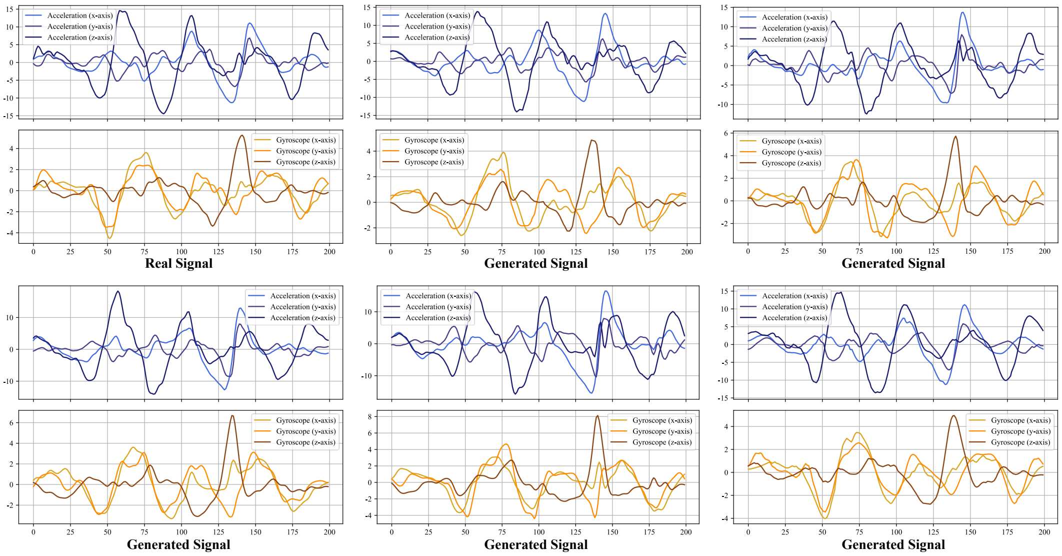

Figure 5: Visualization of the real IMU signal for writing "王" and the virtual signals generated by CI-GAN. The upper left corner is the real signal, and the remaining signals are virtual signals.

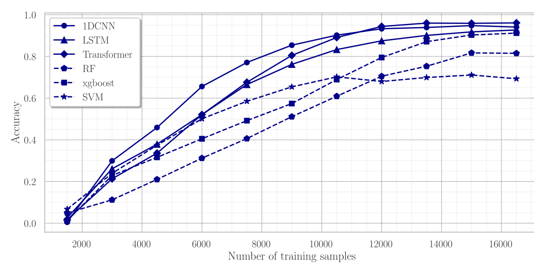

Figure 6: The recognition accuracy of 6 classifiers with varied training samples provided by CI-GAN.

The Transformer captures long-range dependencies in time-series data through its self-attention mechanism, enabling it to understand complex movement patterns. However, its excellent recognition ability relies on large amounts of data, making its performance improvement the most significant as CI-GAN continuously generates training data, improving from 1.7% to 98.4%. Compared to deep learning models, machine learning models also exhibit significant dependence on the amount of training data, highlighting the critical role of sufficient generated signals in handwriting recognition tasks. With the abundant training samples generated by CI-GAN, six classifiers achieve accurate recognition even for similar characters as shown in Appendix A.1.

Table 2: Performance comparison of 6 classfiers.

Classifier	1DCNN	LSTM	Transformer	RF	XGBoost	SVM
Runtime (s)	0.00743	0.13009	0.03439	0.01269	0.00154	0.00173
Memory (MB)	22.153	29.897	52.336	35.418	19.472	3.881
Accuracy	95.7%	93.9%	98.4%	83.5%	93.1%	74.6%

In summary, CI-GAN provides a data experimental platform for Chinese writing recognition, enabling various classifiers to utilize the generated samples for training and improving their recognition accuracy. To help researchers select suitable classifiers for different application scenarios, we further tested the recognition speed and memory usage of different classifiers for a single input sample and summarized their recognition accuracy in Table 2. Among the three deep learning models, 1DCNN has the fastest runtime and the smallest memory usage, with a recognition accuracy of 95.7%, slightly lower than the Transformer but sufficient for most practical applications. It is more suitable for integration into memory and computation resource-limited smart wearable devices such as phones, watches, and wristbands. In contrast, Transformer has the highest accuracy among the six classifiers and the highest memory usage, making it more suitable for PC-based applications. Compared to deep learning classifiers, traditional machine learning classifiers generally have lower accuracy, but with the support of abundant training samples generated by CI-GAN, the XGBoost model still achieves a recognition accuracy of 93.1%, very close to deep learning classifiers. More importantly, XGBoost, as a tree model, has strong interpretability, allowing users to intuitively observe which features significantly impact the model's decision-making process, which is a strength that deep learning models lack. Additionally, XGBoost's runtime and memory usage are better than the three deep learning classifiers, making it outstanding in scenarios requiring a balance between model performance, interpretability, and resource efficiency. For example, XGBoost can be integrated into stationery and educational tools to analyze students' handwriting habits and provide personalized feedback suggestions. Similarly, in the healthcare field, XGBoost can be used to analyze patients' writing characteristics, assisting doctors in evaluating treatment effects or predicting disease risks. Its high interpretability can provide an auxiliary reference for medical decisions and treatment plans, increasing patients' trust in the treatment.

4.3.2 Data Augmentation Method Comparison

We employed five major categories of data augmentation (DA)—Time Domain, Frequency Domain, Decomposition, Mixup, and Learning-based strategies—encompassing 12 methods for comparison Wen et al. (2020). All methods generated the same amount of samples (15,000) for training six classifiers, as shown in Table 3. Notably, except for our proposed augmentation method, the accuracy of classifiers trained using all other data augmentation methods failed to surpass 50%, whereas our method achieved over 90%. Additionally, due to the lack of deep learning-based augmentation methods in the sensor field, we could only compare our approach with cGAN, which performed worse than many non-deep learning methods, underlining the difficulty of designing deep learning models capable of generating accurate and realistic inertial handwriting signals and highlights the value of our CI-GAN. In summary, our method is pioneering in the inertial sensor domain and has been adopted by a wearable device manufacturer.

Table 3: Comparison with competitive DA baselines.

Data Augmentation Methods		1DCNN	LSTM	Transformer	RF	XGBoost	SVM
Time Domain	Cropping	15.7%	9.1%	7.7%	12.8%	16.3%	9.6%
	Noise Injection	17.3%	11.9%	12.2%	8.5%	13.8%	10.1%
	Jittering	20.1%	13.0%	14.4%	9.7%	17.4%	7.5%
Frequency Domain	APP	22.3%	13.6%	19.7%	19.0%	25.1%	16.3%
	AAFT	32.1%	20.7%	25.4%	27.5%	35.9%	19.2%
Decomposition	Wavelet	19.9%	12.1%	10.6%	13.8%	22.6%	9.5%
	EMD	24.4%	17.1%	20.9%	17.9%	23.4%	12.2%
Mixup	CutMix	21.9%	14.8%	15.5%	14.7%	18.9%	13.1%
	Cutout	25.6%	16.4%	16.9%	18.5%	27.1%	16.6%
	RegMixup	<u>41.5%</u>	<u>27.8%</u>	<u>36.8%</u>	<u>38.4%</u>	<u>45.9%</u>	<u>30.3%</u>
Learning based	cGAN	18.5%	14.8%	15.7%	12.4%	20.5%	8.4%
	CI-GAN (ours)	**95.7%**	**93.9%**	**98.4%**	**83.5%**	**93.1%**	**74.6%**

4.4 Ablation Study

Systematic ablation experiments are conducted to evaluate the contributions of the CGE, FOT, and SRA modules in CI-GAN. We generated writing samples using the ablated models and trained the six classifiers on these samples. The results are summarized in Table 4. When no generated data is used (No augmentation), the recognition accuracy of all classifiers is very poor. Employing the Base GAN to generate training samples brings slight improvement but still underperforms, underscoring the critical importance and necessity of data augmentation for accurate recognition. This also indicates that utilizing GAN to improve classifier performance is a challenging task. Introducing CGE, FOT, and SRA individually into the GAN significantly improves its performance, with the introduction of CGE bringing the most noticeable improvement. This demonstrates that incorporating Chinese glyph encoding into the generative model is crucial for accurately generating writing signals. When CGE, FOT, and SRA are simultaneously integrated into the GAN (i.e., CI-GAN), the performance of all six classifiers is improved to above 70%, with four classifiers achieving recognition accuracies exceeding 90%. Notably, the Transformer classifier achieves an impressive accuracy of 98.4%. Furthermore, statistical significance analysis is performed to validate the reliability of these results, as shown in Appendix A.2.

Table 4: Performance comparison of six classifiers trained on samples generated by different ablation models.

Ablation model	1DCNN	LSTM	Transformer	RF	XGBoost	SVM
No augmentation	0.87%	2.6%	1.7%	4.9%	1.2%	6.7%
w/o all (Base GAN)	18.5%	14.8%	15.7%	12.4%	20.5%	8.4%
w/ OT	26.4%	28.6%	27.3%	21.0%	30.9%	20.9%
w/ FOT	39.9%	38.0%	35.3%	31.9%	46.8%	27.3%
w/ CGE	54.6%	51.2%	47.9%	38.6%	57.5%	34.1%
w/ CGE (w/o GER)	35.7%	32.1%	30.9%	33.8%	41.1%	29.0%
w/ CGE (w/o GER)+SRA	61.4%	58.1%	60.2%	51.0%	59.9%	45.2%
w/ CGE (w/o GER)+FOT	59.6%	55.2%	54.0%	53.4%	58.3%	47.5%
w/ CGE+SRA	84.9%	77.4%	86.8%	61.4%	68.9%	56.1%
w/ FOT+CGE	80.7%	80.5%	80.9%	57.2%	70.4%	59.5%
w/ FOT+CGE+SRA (CI-GAN)	95.7%	93.9%	98.4%	83.5%	93.1%	74.6%

4.5 Visualization Analysis of Chinese Glyph Encoding

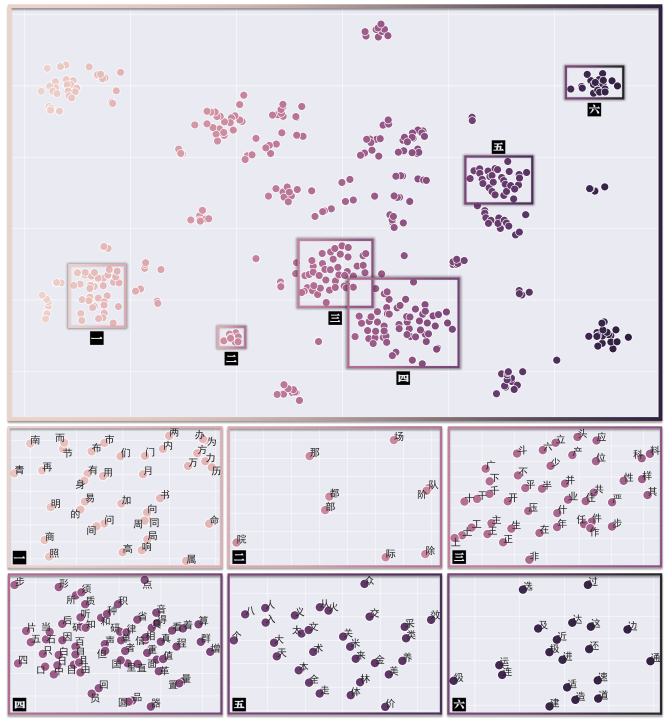

Figure 7: The t-SNE visualization of Chinese glyph encodings.

To demonstrate the effectiveness of the Chinese glyph encoding in capturing the glyph features of Chinese characters, we conducted a visualization analysis using t-SNE, which reduced the dimensionality of the glyph encodings of 500 Chinese characters and visualized the results in a 2D space, as shown in Fig. 7, where each point represents a Chinese character. For the convenience of observation, we selected 6 local visualization regions from left to right and zoomed in on them at the bottom. It can be observed that characters with similar strokes and structure (e.g., ”办-为”, ”目-且”, ”人-入-八”) are close to each other. Additionally, the figure shows several clusters where characters within the same cluster share similar radicals, structures, or strokes, indicating that CGE effectively captures the similarities and differences in the glyph features of Chinese characters. By incorporating CGE into the generative model, CI-GAN can produce writing signals that accurately reflect the structure and stroke features of Chinese characters, ensuring the generated signals closely align with real writing movements. This encoding is not only crucial for guiding GANs in generating writing signals but also potentially provides new tools and perspectives for studying the evolution of Chinese hieroglyphs.

5 Conclusion

This paper introduces GAN to generate inertial sensor signals and proposes CI-GAN for Chinese writing data augmentation, which consists of CGE, FOT, and SRA. The CGE module constructs an encoding of the stroke and structure for Chinese characters, providing glyph information for GAN to generate writing signals. FOT overcomes the mode collapse and mode mixing problems of traditional GANs and ensures the authenticity of the generated samples through the forced feature matching mechanism and OT constraint. The SRA module aligns the semantic relationships between the generated signals and the corresponding Chinese characters, thereby imposing a batch-level constraint on GAN. Utilizing the large-scale, high-quality synthetic IMU writing signals provided by CI-GAN, the recognition accuracy of six widely used classifiers for Chinese writing recognition was improved from 6.7% to 98.4%, which demonstrates that CI-GAN has the potential to become a flexible and efficient data generation platform in the field of Chinese writing recognition. This research provides a novel human-computer interaction, especially for disabled people. Its limitations and impact are discussed in Appendix C.1 and C.2. In the future, we plan to extend CI-GAN to generate signals from other modalities of sensors, constructing a multimodal human-computer interaction system tailored for disabled individuals, which can adapt to the diverse needs of users with different disabilities. Through continuous collaboration with healthcare professionals and the disabled community, we will refine and optimize these multimodal systems to ensure they deliver the highest functionality and user satisfaction. Ultimately, this research aims to foster a society where digital accessibility is a fundamental right, ensuring that all individuals, regardless of physical abilities, can engage fully and independently with the digital world.

References

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

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

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

Irene Cortes-Perez, Noelia Zagalaz-Anula, Desiree Montoro-Cardenas, Rafael Lomas-Vega, Esteban Obrero-Gaitan, and María Catalina Osuna-Pérez. Leap motion controller video game-based therapy for upper extremity motor recovery in patients with central nervous system diseases. a systematic review with meta-analysis. *Sensors*, 21(6):2065, 2021.

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

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

Sebastian Farquhar, Jannik Kossen, Lorenz Kuhn, and Yarin Gal. Detecting hallucinations in large language models using semantic entropy. *Nature*, 630(8017):625–630, 2024.

Ruiyang Gao, Wenwei Li, Yaxiong Xie, Enze Yi, Leye Wang, Dan Wu, and Daqing Zhang. Towards robust gesture recognition by characterizing the sensing quality of wifi signals. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–26, 2022.

Ruiyang Gao, Wenwei Li, Jinyi Liu, Shuyu Dai, Mi Zhang, Leye Wang, and Daqing Zhang. Wicgesture: Meta-motion based continuous gesture recognition with wi-fi. *IEEE Internet of Things Journal*, 2023.

Mohinder S Grewal, Lawrence R Weill, and Angus P Andrews. *Global positioning systems, inertial navigation, and integration*. John Wiley & Sons, 2007.

Boris Gromov, Gabriele Abbate, Luca M. Gambardella, and Alessandro Giusti. Proximity human-robot interaction using pointing gestures and a wrist-mounted imu. In *2019 International Conference on Robotics and Automation (ICRA)*, pp. 8084–8091, 2019. doi: 10.1109/ICRA.2019.8794399.

Yu Gu, Jinhai Zhan, Yusheng Ji, Jie Li, Fuji Ren, and Shangbing Gao. Mosense: An rf-based motion detection system via off-the-shelf wifi devices. *IEEE Internet of Things Journal*, 4(6):2326–2341, 2017.

Elyoenai Guerra-Segura, Aysse Ortega-Pérez, and Carlos M Travieso. In-air signature verification system using leap motion. *Expert Systems with Applications*, 165:113797, 2021.

Zhengxin Guo, Fu Xiao, Biyun Sheng, Huan Fei, and Shui Yu. Wireader: Adaptive air handwriting recognition based on commercial wifi signal. *IEEE Internet of Things Journal*, 7(10):10483–10494, 2020.

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

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

You Li, Ruizhi Chen, Xiaoji Niu, Yuan Zhuang, Zhouzheng Gao, Xin Hu, and Naser El-Sheimy. Inertial sensing meets machine learning: Opportunity or challenge? *IEEE Transactions on Intelligent Transportation Systems*, 23(8):9995–10011, 2022. doi: 10.1109/TITS.2021.3097385.

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

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

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

Salih Ertug Ovur, Hang Su, Wen Qi, Elena De Momi, and Giancarlo Ferrigno. Novel adaptive sensor fusion methodology for hand pose estimation with multileap motion. *IEEE Transactions on Instrumentation and Measurement*, 70:1–8, 2021.

Sai Deepika Regani, Beibei Wang, and K. J. Ray Liu. Wifi-based device-free gesture recognition through-the-wall. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8017–8021, 2021. doi: 10.1109/ICASSP39728.2021.9414894.

Haiqing Ren, Weiqiang Wang, and Chenglin Liu. Recognizing online handwritten chinese characters using rnns with new computing architectures. *Pattern Recognition*, 93:179–192, 2019.

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

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

Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas. Improving gans using optimal transport. *arXiv preprint arXiv:1803.05573*, 2018.

Derek K Shaeffer. Mems inertial sensors: A tutorial overview. *IEEE Communications Magazine*, 51 (4):100–109, 2013.

Jayraj V Vaghasiya, Carmen C Mayorga-Martinez, Jan Vyskočil, and Martin Pumera. Black phosphorous-based human-machine communication interface. *Nature communications*, 14(1):2, 2023.

Xuanzhi Wang, Kai Niu, Jie Xiong, Bochong Qian, Zhiyun Yao, Tairong Lou, and Daqing Zhang. Placement matters: Understanding the effects of device placement for wifi sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 6(1):1–25, 2022.

Xumeng Wang, Xinhu Zheng, Wei Chen, and Fei-Yue Wang. Visual human–computer interactions for intelligent vehicles and intelligent transportation systems: The state of the art and future directions. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 51(1):253–265, 2020.

Yifeng Wang and Yi Zhao. Handwriting recognition under natural writing habits based on a low-cost inertial sensor. *IEEE Sensors Journal*, 24(1):995–1005, 2024. doi: 10.1109/JSEN.2023.3331011.

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

Qingsong Wen, Liang Sun, Fan Yang, Xiaomin Song, Jingkun Gao, Xue Wang, and Huan Xu. Time series data augmentation for deep learning: A survey. *arXiv preprint arXiv:2002.12478*, 2020.

Ning Xiao, Panlong Yang, Yubo Yan, Hao Zhou, Xiang-Yang Li, and Haohua Du. Motion-fi$^+$+: Recognizing and counting repetitive motions with wireless backscattering. *IEEE Transactions on Mobile Computing*, 20(5):1862–1876, 2021. doi: 10.1109/TMC.2020.2971996.

Jian Zhang, Hongliang Bi, Yanjiao Chen, Qian Zhang, Zhaoyuan Fu, Yunzhe Li, and Zeyu Li. Smartso: Chinese character and stroke order recognition with smartwatch. *IEEE Transactions on Mobile Computing*, 20(7):2490–2504, 2020a.

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

Appendix / Supplemental Material

A Additional Experimental Results

A.1 Performance of Classifiers on Similar Characters

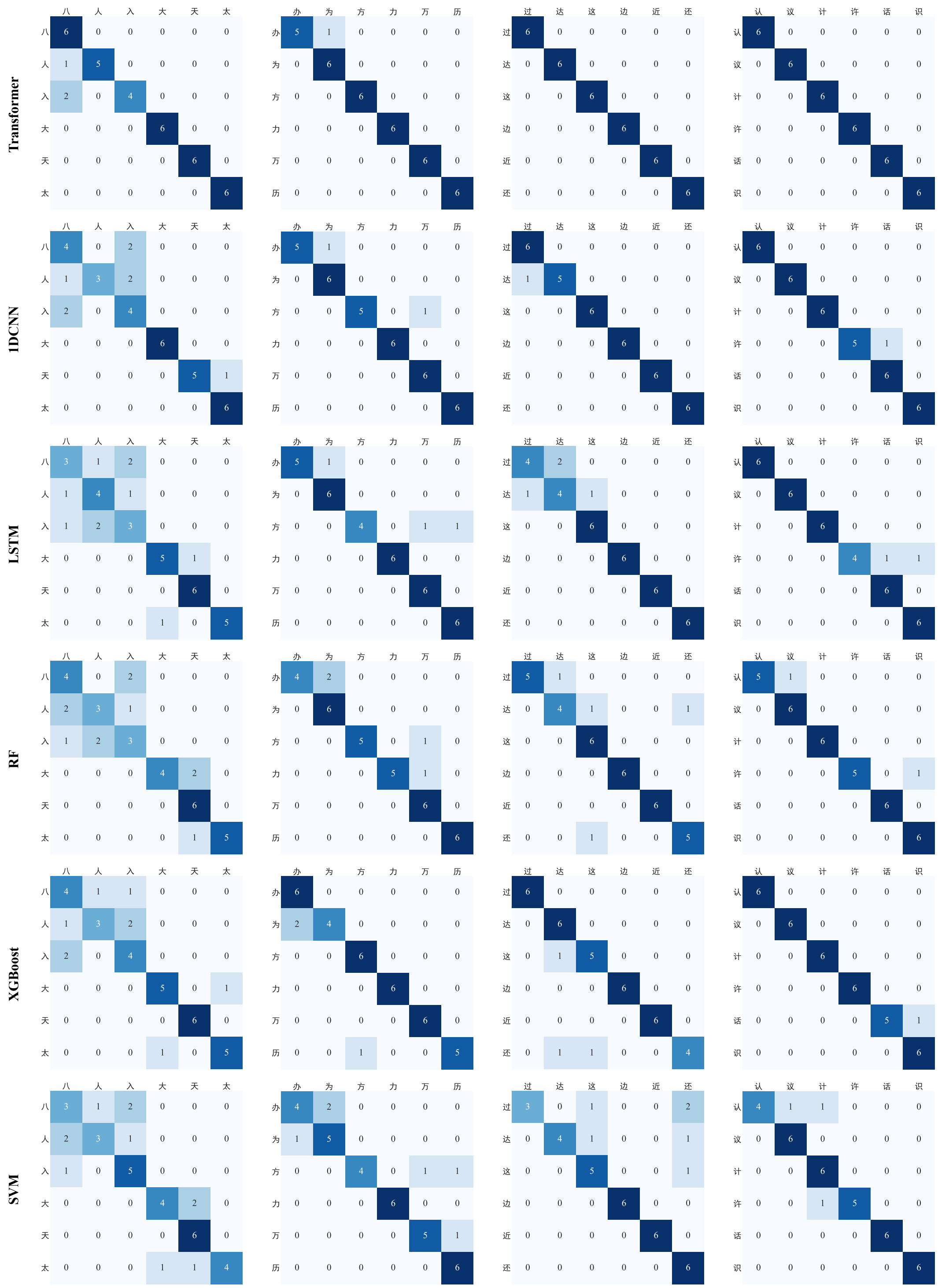

Figure 8: Confusion matrices of different classifiers for recognition results of Chinese characters with similar glyphs.

With the abundant training samples generated by CI-GAN, the handwriting recognition performance of all six classifiers significantly improved. To further verify the recognition performance of different classifiers on characters with similar strokes and glyphs, we selected four groups of characters with similar handwriting movements from the test set (”八人入大天太”, ”办为方力万历”, ”过达这边近还”, and ”认议计许话识”) and presented the recognition results of the six classifiers in confusion matrices, as shown in Fig. 8. It can be observed that the values on the diagonal of all confusion matrices are significantly higher than the non-diagonal values, indicating high recognition accuracy for these similar handwriting characters with the help of samples generated by CI-GAN. However, some characters are still misrecognized. For instance, the characters ”八”, ”人”, and ”入” have extremely similar structures and writing movements, posing challenges even when massive training samples are provided. Moreover, continuous and non-standard writing can also cause recognition obstacles. For instance, although the characters ”过” and ”达” have different strokes in static form, they are very similar in dynamic handwriting. Despite these challenges, the synthetic IMU handwriting samples generated by CI-GAN significantly enhance the classifiers' ability to recognize characters with similar glyph structures and handwriting movements, highlighting the value and significance of the proposed CI-GAN method. By providing diverse and high-quality training samples, CI-GAN improves handwriting recognition classifiers' performance and generalization ability, making it a valuable tool for advancing Chinese handwriting recognition technology.

A.2 Statistical Significance Analysis

The CI-GAN model demonstrates significant performance improvements across multiple classifiers, as shown in Table 4. The Transformer classifier, for instance, achieves a mean accuracy of 98.4%, compared to 15.7% with the traditional GAN and 1.7% without data augmentation. This highlights CI-GAN's ability to generate realistic and diverse training samples that enhance handwriting recognition. Moreover, CI-GAN consistently improves accuracy and stability for all classifiers tested. The 1DCNN's accuracy increases to 95.7% from 18.5% with the traditional GAN and 0.87% without augmentation. Similarly, other models, including LSTM, RandomForest, XGBoost, and SVM, show substantial gains, underscoring CI-GAN's effectiveness across diverse machine-learning contexts. In addition, the narrow 95% confidence intervals, such as [98.2822%, 98.5178%] for the Transformer, validate the statistical significance and reliability of these results. This confirms CI-GAN's potential to consistently enhance classifier performance. In conclusion, CI-GAN represents a major advancement in Chinese handwriting recognition by generating high-quality, diverse inertial signals. This significantly boosts the accuracy and reliability of various classifiers, demonstrating CI-GAN's transformative potential in the field.

Table 5: Performance of different classifiers with CI-GAN generated data

Ablation	Classifier	Mean Accuracy	Standard Deviation	95% Confidence Interval
No data augmentation	1DCNN	0.87%	0.11%	[0.8018%, 0.9382%]
	LSTM	2.61%	0.20%	[2.4761%, 2.7239%]
	Transformer	1.70%	0.13%	[1.6194%, 1.7806%]
	RandomForest	4.89%	0.09%	[4.8439%, 4.9556%]
	XGBoost	1.20%	0.15%	[1.1071%, 1.2929%]
	SVM	6.65%	0.10%	[6.5881%, 6.7119%]
Traditional GAN	1DCNN	18.5%	0.16%	[18.4008%, 18.5992%]
	LSTM	14.8%	0.37%	[14.5707%, 15.0293%]
	Transformer	15.7%	0.15%	[15.6071%, 15.7929%]
	RandomForest	12.4%	0.17%	[12.2948%, 12.5052%]
	XGBoost	20.5%	0.23%	[20.3573%, 20.6427%]
	SVM	8.40%	0.34%	[8.1893%, 8.6107%]
CI-GAN	1DCNN	95.7%	0.24%	[95.5513%, 95.8487%]
	LSTM	93.9%	0.53%	[93.5713%, 94.2287%]
	Transformer	98.4%	0.19%	[98.2822%, 98.5178%]
	RandomForest	83.5%	0.35%	[83.2831%, 83.7169%]
	XGBoost	93.1%	0.46%	[92.8148%, 93.3852%]
	SVM	74.6%	0.38%	[74.3644%, 74.8356%]

B Challenge in Handwriting Sample Collection

Collecting handwriting samples of Chinese characters is not easy. During data collection, volunteers wrote different Chinese characters continuously. We had to accurately locate the signal segments corresponding to each character from long signal streams, as shown in Fig. 9. However, accurately segmenting and extracting signal segments requires synchronizing optical motion capture equipment and then comparing the inertial signals frame by frame with the optical capture results to find all character signal segments' starting and ending frames. Consequently, we expended significant time and effort to obtain 4,500 signal samples in this paper, establishing the first Chinese handwriting recognition dataset based on inertial sensors, which we have made open-source partially. By contrast, our CI-GAN can directly generate handwriting motion signals according to the input Chinese character, eliminating the complex processes of signal segmentation, extraction, and cleaning, as well as the reliance on optical equipment. We believe it provides an efficient experimental data platform for the field.

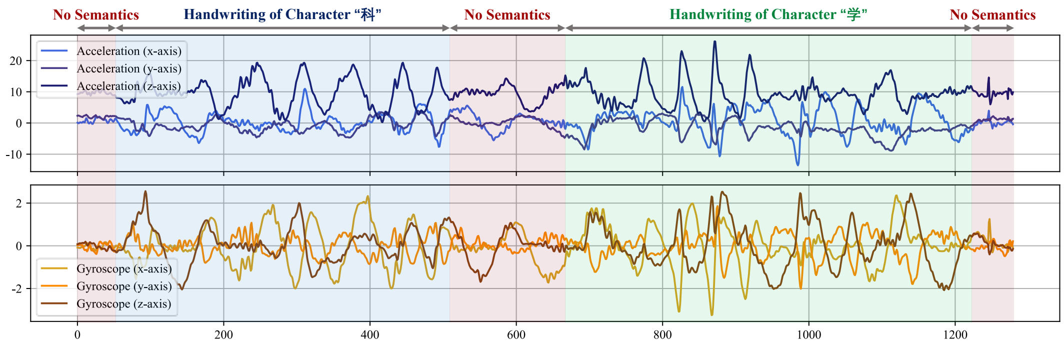

Figure 9: Signal segmentation diagram. Since the raw data contains both meaningful handwriting and extraneous movements, segmenting and extracting the relevant segments corresponding to individual characters from the continuous signal stream is crucial. Reliance on human observation alone is insufficient and prone to errors, thus making optical devices indispensable for accurately segmenting and extracting the signal segment.

Unlike the fields of CV and NLP, many deep learning methods have not yet been applied to the sensor domain. More importantly, unlike image generation, where the performance can be visually judged, it is challenging to identify semantics in waveforms by observation and determine whether the generated signal fluctuations are reasonable, which imposes high requirements on generative model design. Therefore, we had to design multiple guidance and constraints for the generator, resulting in the design of Chinese Glyph Encoding (CGE), Forced Optimal Transport (FOT), and Semantic Relevance Alignment (SRA).

- CGE introduces a regularization term based on Rényi entropy, which increases the information content of the encoding matrix and the distinctiveness of class encodings, providing a new category representation method that can also be applied to other tasks. As far as we know, this is the first embedding targeted at the shape of Chinese characters rather than their meanings, providing rich semantic guidance for generating handwriting signals.
- FOT establishes a triple-consistency constraint between the input prompt, output signal features, and real signal features, ensuring the authenticity and semantic accuracy of the generated signals and preventing mode collapse and mixing.
- SRA constrains the consistency between the semantic relationships among multiple outputs and the corresponding input prompts, ensuring that similar inputs correspond to similar outputs (and vice versa), significantly alleviating the hallucination problem of generative models. Notably, the June 2024 Nature paper "Detecting Hallucination in Large Language Models Using Semantic Entropy," shares a similar idea with our proposed SRA. They assess model hallucination by repeatedly inputting the same prompts into generative models and evaluating the consistency of the outputs. Their approach essentially forces the model to produce similar outputs for similar prompts. Our SRA not only achieves this but also ensures

that the relationships between prompts are mirrored in the relationships between the outputs. This significantly reduces hallucinations and enhances the model's practicality and stability.

C Discussion

C.1 Societal Impact

CI-GAN model significantly improves the accuracy of Chinese writing recognition and offers an alternative means of human-computer interaction that can overcome the limitations of traditional keyboard-based methods, which are often inaccessible to those who are blind or lose their fingers. By providing a more accessible and user-friendly way to interact with digital devices, inertial sensors can facilitate effective communication, enhance the participation of disabled people in education and employment, and promote greater independence. Moreover, by addressing the unique needs of this population, such technological advancements reflect a commitment to inclusivity and social justice, ensuring that everyone, regardless of their physical abilities, has the opportunity to fully participate in and contribute to society.

Furthermore, by releasing the world's first Chinese handwriting recognition dataset based on inertial sensors, this research provides valuable data resources for both academia and industry, facilitating further studies and advancements. Additionally, the technology offers an intuitive and efficient learning tool for Chinese language learners, aiding in preserving and disseminating Chinese cultural heritage and strengthening the global influence of Chinese characters. In summary, the CI-GAN technology achieves not only significant breakthroughs in algorithmic research but also demonstrates extensive practical potential and substantial societal value, thereby being adopted by educational aid device manufacturers. This study provides a solid foundation for future academic research, technological development, and industrial applications, driving technological progress and societal development.

C.2 Limitation

While the CI-GAN model demonstrates significant advancements in Chinese handwriting generation and recognition, some practical limitations could impact its performance in real-world applications. For instance, non-standard or cursive handwriting may pose challenges for accurate signal generation and recognition. Additionally, environmental factors such as external movements or vibrations when using handheld devices could affect the inertial sensor data quality, leading to variations in recognition accuracy. Future work could focus on developing more robust algorithms that account for these real-world variations and improving the model's adaptability to diverse handwriting styles and conditions. These enhancements would ensure that the CI-GAN technology remains effective across a broader range of practical scenarios.

D Mathematical Explanation of FOT for Preventing Mode Collapse

To rigorously demonstrate how Feature Optimal Transport (FOT) mitigates mode collapse, we begin by formalizing the problem within the context of GANs. Let P_{real} represent the true data distribution, and P_{gen} the distribution generated by the generator G, both defined over the data space $\mathcal{X}$. Mode collapse occurs when P_{gen} fails to cover all the modes of P_{real}, resulting in a mismatch where certain regions of the support of P_{real} have no corresponding mass in P_{gen}. To address this, FOT operates in a feature space $\mathcal{F}$, which is defined by a mapping $f : \mathcal{X} \to \mathcal{F}$. This mapping transforms the real and generated distributions into $P_f = f_\# P_{\text{real}}$ and $Q_f = f_\# P_{\text{gen}}$, respectively, where $f_\#$ denotes the pushforward measure induced by f.

The Wasserstein distance between P_f and Q_f in the feature space serves as the basis for FOT. It is defined as:

$$W(P_f, Q_f) = \inf_{\gamma \in \Pi(P_f, Q_f)} \mathbb{E}_{(u,v)\sim\gamma}[d(u, v)],$$

where $\Pi(P_f, Q_f)$ is the set of all joint distributions γ with marginals P_f and Q_f, and $d(u, v)$ is a distance metric in the feature space $\mathcal{F}$. The Wasserstein distance inherently penalizes mismatches between the supports of P_f and Q_f, making it particularly effective for addressing mode collapse.

Mode collapse corresponds to the case where the support of Q_f is a strict subset of the support of P_f, such that there exist regions of $\mathcal{F}$ where P_f assigns positive probability but Q_f assigns none. In such cases, for any coupling $\gamma \in \Pi(P_f, Q_f)$, the Wasserstein distance remains strictly positive. Formally, let:

$$P_f = \sum_{i=1}^{n} p_i \delta_{u_i}, \quad Q_f = \sum_{j=1}^{m} q_j \delta_{v_j},$$

where δ_{u_i} and δ_{v_j} are Dirac measures centered at u_i and v_j, respectively. If there exists $u_k \in \text{supp}(P_f)$ such that $u_k \notin \text{supp}(Q_f)$, then for all couplings γ, we have:

$$W(P_f, Q_f) \geq \epsilon,$$

where $\epsilon > 0$ is the cost associated with transporting mass from u_k to the closest point in $\text{supp}(Q_f)$. This lower bound demonstrates that mode collapse leads to a nonzero Wasserstein distance, which FOT actively penalizes. To mitigate this, FOT incorporates $W(P_f, Q_f)$ into the GAN objective as:

$$\mathcal{L}_{\text{FOT}} = W(P_f, Q_f).$$

Minimizing $\mathcal{L}_{\text{FOT}}$ forces the generator to adjust its output distribution such that Q_f aligns with P_f in the feature space. Specifically, minimizing the Wasserstein distance requires the support of Q_f to expand to fully cover the support of P_f. By construction, the optimal transport plan ensures that all mass in P_f is matched to corresponding mass in Q_f, eliminating regions of the feature space where P_f assigns probability but Q_f does not. Furthermore, the choice of the feature mapping f ensures that the feature space captures semantically meaningful structures and relationships in the data. The metric $d(u, v) = \|u - v\|_2^2$ in the feature space penalizes discrepancies in both spatial and structural characteristics, enabling the generator to learn not just local patterns but also global dependencies between modes. This ensures that the generator produces diverse samples that faithfully represent the underlying data distribution.

In conclusion, by minimizing $\mathcal{L}_{\text{FOT}}$, the generator is guided to align Q_f with P_f, covering all modes of the real distribution and addressing mode collapse. This mathematical framework validates FOT as a principled solution to one of the most persistent challenges in GAN training, ensuring the generation of diverse and high-quality samples.

E Mathematical Explanation of FOT for Preventing Mode Collapse

To thoroughly evaluate the system's robustness under external disturbances, we conducted experiments by introducing varying levels of Gaussian noise to the real inertial signals during training. The Gaussian noise was added at proportions of 0.0%, 5.0%, 10.0%, and 20.0% of the original signal's standard deviation to simulate sensor inaccuracies and environmental interference. Under different levels of Gaussian noise added to the real inertial signals, we trained CI-GAN models to generate 15,000 IMU signals for each noise setting. These generated signals were then used to train six classifiers (1DCNN, LSTM, Transformer, RF, XGBoost, and SVM), and their classification accuracy was evaluated using 5-fold cross-validation. The results, presented in the table below, reflect the accuracy of the classifiers under varying noise conditions.

Noise Ratio	1DCNN	LSTM	Transformer	RF	XGBoost	SVM
0.0%	95.7%	93.9%	98.4%	83.5%	93.1%	74.6%
5.0%	95.2%	94.1%	98.0%	82.9%	93.3%	71.8%
10.0%	94.5%	92.3%	97.1%	81.7%	92.6%	70.7%
20.0%	93.9%	92.5%	95.9%	79.8%	91.0%	69.4%

Table 6: Performance comparison of different models under various noise ratios.

These results demonstrate that the system maintains high performance even under significant noise levels. While performance slightly decreases with higher noise ratios, the overall degradation is minimal. This robustness is attributed to the combined contributions of Glyph Encoding Regularization (GER), Forced Optimal Transport (FOT), and Semantic Relevance Alignment (SRA). CGE introduces a regularization term based on Rényi entropy, which is the first embedding targeted at the shape of Chinese characters rather than their meanings, providing rich semantic guidance for generating handwriting signals. FOT establishes a triple-consistency constraint between the input prompt, output signal features, and real signal features, ensuring the authenticity and semantic accuracy of the generated signals and preventing mode collapse and mixing. SRA constrains the consistency between the semantic relationships among multiple outputs and the corresponding input prompts, ensuring that similar inputs correspond to similar outputs (and vice versa), significantly alleviating the hallucination problem of generative models. Together, these components ensure the system's resilience to external disturbances and its capacity to generate realistic and accurate signals under challenging scenarios.