Chinese Inertial GAN for Handwriting Signal Generation and Recognition

Authors: Yifeng Wang, Yi Zhao

Date: 2025-05-30T14:08:36+00:00

Abstract

Keyboard-based interaction may not accommodate various needs, especially for individuals with disabilities. While inertial sensor-based writing recognition is promising due to the sensors' small size, wearability, and low cost, accurate recognition in the Chinese context is hampered by the difficulty of collecting extensive inertial signal samples for the vast number of characters. Therefore, we design a Chinese Inertial GAN (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 methods focusing on the meaning of Chinese characters, CGE represents shape and stroke features, providing glyph guidance for writing signal generation. 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. SRA aligns semantic relationships between multiple outputs and their input prompts, ensuring that similar inputs correspond to similar outputs (and vice versa), alleviating model hallucination. 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 inertial writing dataset on GitHub.

Full Text

Preamble

Chinese Inertial GAN for Handwriting Signal Generation and Recognition

Yifeng Wang and Yi Zhao

Harbin Institute of Technology, Shenzhen, China

Abstract

Keyboard-based interaction may not accommodate various needs, especially for individuals with disabilities. While inertial sensor-based writing recognition is promising due to the sensors' small size, wearability, and low cost, accurate recognition in the Chinese context is hampered by the difficulty of collecting extensive inertial signal samples for the vast number of characters. Therefore, we design a Chinese Inertial GAN (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 methods focusing on the meaning of Chinese characters, CGE represents shape and stroke features, providing glyph guidance for writing signal generation. 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. SRA aligns semantic relationships between multiple outputs and their input prompts, ensuring that similar inputs correspond to similar outputs (and vice versa), alleviating model hallucination. 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 inertial writing dataset on GitHub.

Introduction

As efficient motion-sensing components, inertial sensors can measure the acceleration and angular velocity of moving objects [?, ?, ?, ?, ?, ?, ?]. 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 [?, ?, ?, ?, ?, ?, ?], 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 [?, ?]. 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, em-

ployment, 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 [?, ?, ?]. 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 but also ensures consistency between the generated and real signals through a designed 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 on GitHub.

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. [**?**, **?**] utilized the Leap Motion device to propose an RNN-based method for recognizing Chinese characters written in the air. The Leap Motion sensor, consisting of two infrared emitters and two cameras, can accurately capture the motion of hands in three-dimensional (3D) space [**?**, **?**]. 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 accuracy [**?**, **?**]. 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 [**?**, **?**].

As wireless networks become more prevalent, Wi-Fi signals are gradually being applied to motion capture [**?**, **?**, **?**]. Since Wi-Fi signals can penetrate objects and are unaffected by lighting conditions, they have a broader application scope than optical motion capture systems [**?**, **?**, **?**]. [**?**, **?**] used the channel state information (CSI), extracted from Wi-Fi signals reflected by hand movements, to recognize 26 air-written English letters. 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 [**?**, **?**, **?**].

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 [**?**, **?**, **?**, **?**, **?**]. 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.

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
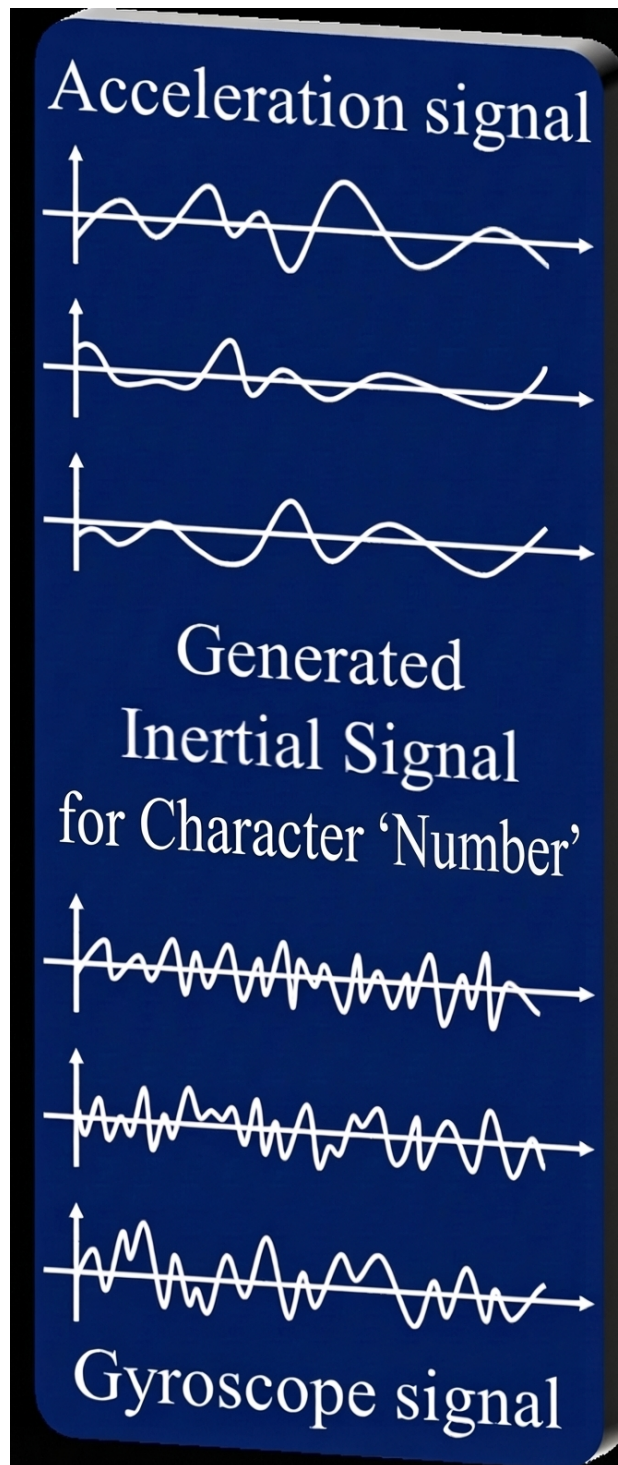
Figure 1: Figure 1

　Machine Translation

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 real 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.

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. Commonly used Chinese character embeddings, while capturing semantic meanings, 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 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:

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

where

$$\tilde{G}_{ij} = \frac{G_{ii} \cdot G_{jj}}{G_{ij}} = \langle W^{(i)}, W^{(j)} \rangle,$$

and 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} equals the inner product of the i-th and j-th rows of W, and $\tilde{G}$ is the trace-normalized G, i.e., $\mathrm{tr}(\tilde{G}) = 1$. 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. Therefore, the glyph encoding regularization R_{encode} is constructed as $R_{\mathrm{encode}} = \frac{1}{S_\alpha(W)}$. 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, representing the differences in glyph shapes among all characters. Thus, this glyph encoding can inject glyph information into GAN, ensuring that the generated signals maintain consistency with the target character glyph. We provide a Chinese glyph encoding visualization in Appendix A.3, which proves that CGE is crucial for guiding GANs in generating writing signals and provides potential tools or perspectives for studying the evolution of Chinese hieroglyphs.

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 dynamic 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 L_{FFM}:

$$L_{\text{FFM}} : \frac{\langle h_G, h_T \rangle + \langle h_G, e \rangle + \langle e, h_T \rangle}{\|h_G\|\|h_T\| + \|h_G\|\|e\| + \|e\|\|h_T\|}.$$

L_{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.

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 [**?**, **?**], which utilizes Wasserstein distance as a constraint to ensure stable gradients, thereby preventing mode collapse and mixing. Combining FFM and OT constraints, we can obtain the forced optimal transport loss $L_{\text{FOT}} = W(P_T, P_G) + \lambda \cdot L_{\text{FFM}}$, where $W(P_T, P_G)$ 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.

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), as shown in Fig. 2 [FIGURE:2], 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 $L_{\text{SRA}} = \|M_h - M_e\|_2^2$ 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. The proposed SRA aligns the relationships between outputs and their corresponding prompts, significantly reducing hallucinations in generative models and enhancing the model's overall practicality and stability.

Module Interaction

CGE, FOT, and SRA not only guide and constrain the generator but also interact with each other, as shown in Fig. 3 [FIGURE: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.

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.

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. Signal generation visualization is provided in the Appendix A.1.

Classifier Comparison 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. 4 [FIGURE:4]. 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). 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.2. In summary, CI-GAN provides an experimental data 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 strategies—encompassing 12 methods for comparison [**?, ?, ?**]. All methods generated the same amount of samples (15,000) for training six classifiers, as shown in Table 3 . Due to the lack of deep learning-based augmentation methods in the sensor field, we introduced the diffusion model-based approach for generating handwriting trajectory, named Diff-Writer [**?, ?**]. Although this approach generates trajectory point sequences rather than the sensor signals required in our study, its ability to produce high-quality and diverse handwriting data makes it highly valuable [**?, ?**]. We adapted this method through modifications and retraining, enabling its application to our inertial signal generation task for a meaningful comparison.

As shown in Table 3, Diff-Writer significantly outperforms all baseline methods except for our CI-GAN, showcasing its strength as a learning-based approach for generating handwriting data. However, as Diff-Writer was not designed for generating inertial sensor signals, it struggles to fully capture the motion dynamics and semantic fidelity required for this task. Consequently, there remains a con-

siderable gap between its performance and that of our CI-GAN, which achieves superior accuracy across all classifiers by addressing the unique challenges of inertial signal generation.

Data Augmentation Comparison

We employed five major categories of data augmentation (DA)—Time Domain, Frequency Domain, Decomposition, Mixup, and Learning-based methods for comparison, as shown in Table 3 . Among the three deep learning models, 1DCNN exhibits the fastest runtime and smallest memory footprint. Its recognition accuracy of 95.7% is slightly lower than the Transformer's but is 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.

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%. Statistical significance analysis is performed to validate the reliability of these results, as shown in Appendix A.4.

Discussion

Chinese characters, as a logographic writing system with a long history, are not random concatenations of symbols but rather embody rich structural information and semantic cues. Unlike phonetic scripts, Chinese characters often exhibit intuitive morphological links between their glyphs and meanings (e.g., 日 depicts the sun, 山 mimics the silhouette of mountains, 火 resembles flames, and 网 represents an intertwined network). This ideographic nature can provide AI with denser information, enabling models to directly decode partial semantics from the glyphs themselves. Studies have shown that the average information entropy of Chinese reaches 9.65 bits, significantly higher than the 4.03 bits for English. This implies that to convey the same semantic meaning, Chinese requires only about 41.7% of the characters needed in English. However, current Chinese vectorization methods essentially treat characters as arbitrary symbols, with learning primarily relying on statistical co-occurrences within character sequences, thereby neglecting the internal structural information and rich prior knowledge inherent in the glyphs themselves. This paper captures Chinese handwriting using sensors, viewing this process as a record of the dynamic formation of glyphs, and consequently designs Chinese Glyph Encoding (CGE) to represent the morphological and structural information of characters from this process. CGE can introduce the structural and stroke features of Chinese characters into deep learning architectures, allowing AI to evolve from merely "recognizing characters" to "understanding character structures." When AI can comprehend that radicals like 氵 are often related to water, 亻 to people, 讠 to speech, 钅 to metal, and 火 to fire, its utilization of Chinese corpora becomes more efficient, and its understanding of the entire Chinese knowledge system deepens. This motion-capture-based representation of Chinese character glyph structure can capture subtle structural differences (e.g., 千 and 干, 天 and 夭, 田 and 甲), enabling AI's language understanding to transcend knowledge representation based solely on statistical regularities from contextual prediction. To some extent, CGE provides AI with a powerful information source for understanding human knowledge, independent of purely text-based statistics, thereby revealing the immense potential for AI in comprehending and utilizing the ancient and sophisticated system of Chinese characters.

The profound implication of this research is that for symbolic systems possessing internal structure and non-arbitrary morphology (especially logographic systems

like Chinese characters), explicitly modeling their "morphological logic" could be an effective pathway to enhancing AI's cognitive capabilities. CGE, as an initial attempt, validates the feasibility of this approach and may have far-reaching impacts on AI's symbol learning and representation learning. Furthermore, the shape of Chinese characters, as a crucial carrier of their meaning, deserves a more central position in future AI research. This focus could be a key path to propelling AI towards higher levels of cognitive intelligence and a more profound understanding of language.

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.

At present, the Chinese Glyph Encoding (CGE) can only represent character categories available in the training data, restricting the model's holistic understanding of the complete Chinese character system. In the future, we plan to extend CI-GAN's comprehension to encompass all Chinese characters by representing fundamental radicals and components. Most Chinese characters are composed of simpler constituent elements. We will represent these basic components and train the model to learn their correct sequential combination according to established writing order, thereby forming entirely new characters, even those not encountered during initial training. Since handwriting is inherently a continuous, temporally ordered action, the model only needs to learn the sequential assembly of components rather than master their complex two-dimensional spatial arrangements. This simplifies the learning objective, making the generation of a vast and diverse range of Chinese characters a more practical and achievable goal.

Moreover, 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.

Acknowledgments

Limitation

Currently, CI-GAN can only generate inertial handwriting signals for Chinese characters that are present in its training data. This limits the model's comprehensive perception of the entire Chinese character system, necessitating the development of a new scheme for understanding Chinese glyph structures.

References

Terrance DeVries. 2017. Improved regularization of convolutional neural networks with cutout. arXiv preprint arXiv:1708.04552.

Georgios Douzas and Fernando Bacao. 2018. Effective data generation for imbalanced learning using conditional generative adversarial networks. Expert Systems with applications, 91:464–471.

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

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

Anibal Flores, Hugo Tito-Chura, and Honorio Apaza-Alanoca. 2021. Data augmentation for short-term time series prediction with deep learning. In Intelligent Computing: Proceedings of the 2021 Computing Conference, Volume 2, pages 492–506. Springer.

Bowen Gao, Minsi Ren, Yuyan Ni, Yanwen Huang, Bo Qiang, Zhi-Ming Ma, Wei-Ying Ma, and Yanyan Lan. 2024. Rethinking specificity in sbdd: Leveraging

delta score and energy-guided diffusion. arXiv preprint arXiv:2403.12987.

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

Julien Audibert, Pietro Michiardi, Frédéric Guyard, Sébastien Marti, and Maria A Zuluaga. 2020. Usad: Unsupervised anomaly detection on multivariate time series. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pages 3395–3404.

Ruiyang Gao, Wenwei Li, Yaxiong Xie, Enze Yi, Leye Wang, Dan Wu, and Daqing Zhang. 2022. 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.

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

Guangyao Chen, Peixi Peng, Li Ma, Jia Li, Lin Du, and Yonghong Tian. 2021. Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 458–467.

Irene Cortes-Perez, Noelia Zagalaz-Anula, Desiree Montoro-Cardenas, Rafael Lomas-Vega, Esteban Obrero-Gaitan, and María Catalina Osuna-Pérez. 2021. 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.

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

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

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

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

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

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

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

Tracey Kah-Mein Lee, HW Chan, KH Leo, Effie Chew, Ling Zhao, and Saeid Sanei. 2022. Improving rehabilitative assessment with statistical and shape preserving surrogate data and singular spectrum analysis. In 2022 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), pages 58–63. IEEE.

Jiaze Li, Yaya Shi, Zongyang Ma, Haoran Xu, Feng Cheng, Huihui Xiao, Ruiwen Kang, Fan Yang, Tingting Gao, and Di Zhang. 2025a. imove: Instance-motion-aware video understanding. arXiv preprint arXiv:2502.11594.

Jiaze Li, Haoran Xu, Shiding Zhu, Junwei He, and Haozhao Wang. 2025b. Multilevel semantic-aware model for ai-generated video quality assessment. arXiv preprint arXiv:2501.02706.

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

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

José Fernando Adrán Otero, Karmele López-de Ipina, Oscar Solans Caballer, Pere Marti-Puig, José Ignacio Sánchez-Méndez, Jon Iradi, Alberto Bergareche, and Jordi Solé-Casals. 2022. Emd-based data augmentation method applied to handwriting data for the diagnosis of essential tremor using lstm networks. Scientific Reports, 12(1):12819.

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

Francesco Pinto, Harry Yang, Ser Nam Lim, Philip Torr, and Puneet Dokania. 2022. Using mixup as a regularizer can surprisingly improve accuracy & out-of-distribution robustness. Advances in Neural Information Processing Systems, 35:14608–14622.

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

Min-Si Ren, Yan-Ming Zhang, and Yi Chen. 2024. Decoupling layout from glyph in online chinese handwriting generation. arXiv preprint arXiv:2410.02309.

Min-Si Ren, Yan-Ming Zhang, Qiu-Feng Wang, Fei Yin, and Cheng-Lin Liu. 2023. Diff-writer: A diffusion model-based stylized online handwritten chinese character generator. In International Conference on Neural Information Processing, pages 86–100. Springer.

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

Swapnil Sayan Saha, Sandeep Singh Sandha, Luis Antonio Garcia, and Mani Srivastava. 2022. Tinyodom: Hardware-aware efficient neural inertial navigation. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 6(2):1–25.

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

Xuanzhi Wang, Kai Niu, Jie Xiong, Bochong Qian, Zhiyun Yao, Tairong Lou, and Daqing Zhang. 2022. 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.

Yifeng Wang, Jiangtao Xu, and Yi Zhao. 2024. Wavelet encoding network for inertial signal enhancement via feature supervision. IEEE Transactions on Industrial Informatics, 20(11):12924–12934.

Yifeng Wang, Shu Zhang, and Yi Zhao. 2025. kan: Reconstructing over-range inertial signals. In ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 1–5. IEEE.

Yifeng Wang and Yi Zhao. 2024a. Scale and direction guided gan for inertial sensor signal enhancement. In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24, International Joint Conferences on Artificial Intelligence Organization, pages 5126–5134.

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

Yifeng Wang and Yi Zhao. 2025a. General pre-trained inertial signal feature extraction based on temporal memory fusion. Information Fusion, 123:103274.

Yifeng Wang and Yi Zhao. 2025b. 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, volume

39, pages 1–8.

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

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

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

Haoran Xu, Jiaze Li, Wanyi Wu, and Hao Ren. 2025. Federated learning with sample-level client drift mitigation. arXiv preprint arXiv:2501.11360.

Zhihan Yue, Yujing Wang, Juanyong Duan, Tianmeng Yang, Congrui Huang, Yunhai Tong, and Bixiong Xu. 2022. Ts2vec: Towards universal representation of time series. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 8980–8987.

Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. 2019. Cutmix: Regularization strategy to train strong classifiers with localizable features. In Proceedings of the IEEE/CVF international conference on computer vision, pages 6023–6032.

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

Appendix

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 [?, ?] to convert both the real and generated signals into corresponding motion trajectories, as shown in the third column of Fig. 5 [FIGURE:5]. 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.

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. 6 [FIGURE:6]. 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.

Performance of Classifiers on Similar Characters

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 "Commonly Used Chinese Characters List" ("八人入大天太", "办为方力万历", "过达这边近还", "认议计许话识") and presented the recognition results of the six classifiers in confusion matrices, as shown in Fig. 7 [FIGURE:7]. 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 recog-

nition classifiers' performance and generalization ability, making it a valuable tool for advancing Chinese handwriting recognition technology.

Visualization Analysis of Chinese Glyph Encoding

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. 8 [FIGURE:8], 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.

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.

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 [FIGURE: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.

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.