2024 IEEE 23rd International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom) | 979-8-3315-0620-9/24/$31.00 ©2024 IEEE | DOI: 10.1109/TrustCom63139.2024.00361

A Multi-Stage Spike Stream Processing and Image Reconstruction Method for Industrial Applications

Shuaipeng Wu[1,2], Changhao Yuan[2,3], Kejiang Ye[2,*]
[1]Southern University of Science and Technology
[2]Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences
[3]University of Chinese Academy of Sciences
{sp.wu1, ch.yuan, kj.ye}@siat.ac.cn

***Abstract*—With the advancement of new industrialization, production processes or scenarios characterized by automation and intelligence require monitoring by more powerful visual inspection tools to maintain production continuity, safety, and efficiency. Spike vision sensors, with their high temporal resolution and high dynamic range, can capture subtle changes of high-speed moving objects in complex industrial environments, providing precise data support for equipment maintenance, fault detection, and condition monitoring. However, in industrial scenarios, factors such as mechanical vibrations, electromagnetic interference, and extreme strong light affect the quality of sensor data, leading to higher noise levels and increased probability of data loss and errors. To address this, we propose a multi-stage processing method for spike data processing and image reconstruction. It first performs preliminary feature extraction and denoising through an anti-noise feature encoder. Building on this, it integrates features and enhances consistency through spatiotemporal information utilization. Finally, it reconstructs the images through an fine-tuned spike reconstruction model. Experimental results show that our method can effectively reduce noise and improve data quality, achieving good performance on industrial datasets. This provides higher quality data support for downstream tasks such as predictive maintenance and high-speed production inspection in industrial scenarios.**

***Index Terms*—Industrial Artificial Intelligence, Spike Stream, Data Processing, Deep Learning.**

I. Introduction

With the rapid advancement of new industrialization, modern industrial systems are evolving towards automation and intelligence. In this process, the complexity of production processes and the operational speed of production lines are continuously increasing, creating unprecedented demands for high-precision, high-efficiency monitoring and inspection tools. In modern industry, how to capture equipment operational status and monitor subtle changes in product quality in complex, high-speed environments has become a core element in maintaining production continuity, safety, and efficiency.

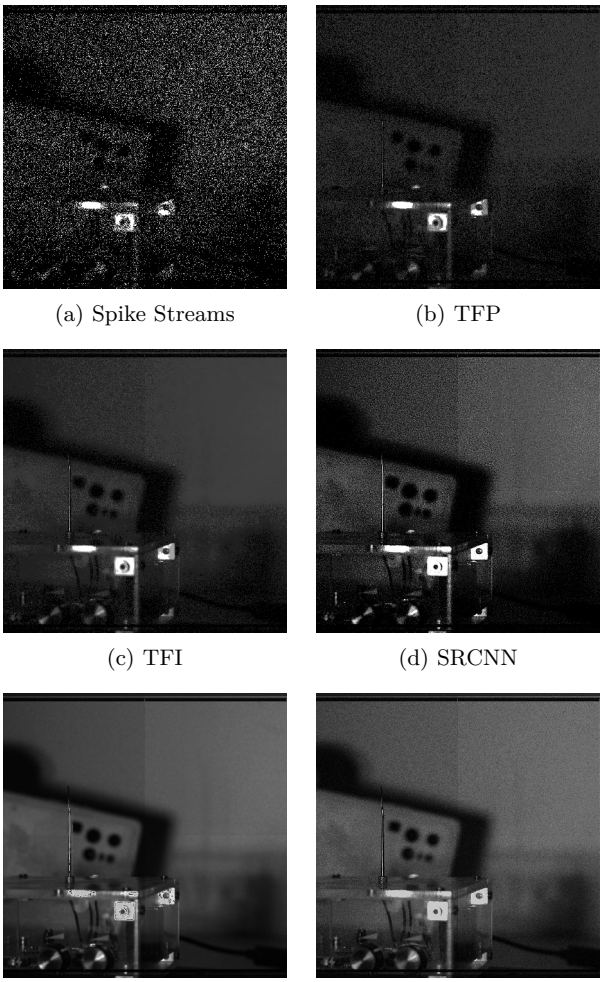
High-speed spike vision sensors [1], [6], as visual inspection tools with high temporal resolution (capable of capturing 20,000 frames per second) and high dynamic range (capture details from extremely dark to extremely bright), can detect subtle changes of high-speed moving objects in various complex environments and are increasingly widely applied in various scenarios. For example, in equipment maintenance processes, spike sensors can capture subtle wear on high-speed rotating or moving parts, detecting potential failures in advance and avoiding production interruptions [2]. In actual scenarios of the power industry, electric arc is a common gas discharge phenomenon. High-intensity or high-frequency electric arc can cause varying degrees of damage to various electrical equipment, posing a significant hidden danger to the reliable operation of power grids. Spike sensors can promptly capture electric arcs and issue risk warnings, providing safety assurance for production processes.

When using spike visual sensors for data collection in industrial environments, although they can capture subtle changes in high-speed scenarios, industrial scenarios are full of factors such as mechanical vibrations and electromagnetic interference [3]. These factors interfere with the sensor's electrical signals, leading to signal distortion or increased noise, and also increasing the probability of data loss and errors. Currently, there are many excellent works that reconstruct spike streams into high-quality images, including traditional basic reconstruction [4], [5] of binary spike streams, as well as image reconstruction using deep learning methods [4], [7]. However, these methods still face some challenges when processing industrial data (see Fig. 1). Therefore, it is necessary to develop more efficient and robust processing methods, which will be elaborated in detail in the next section.

This paper proposed a multi-stage processing method to tackle the challenges of spike stream data processing and high-quality image reconstruction in industrial environments. Specifically, the proposed method consists of three main stages: preliminary feature extraction and denoising, feature integration and consistency enhancement, and image reconstruction. In the preliminary feature extraction and denoising stage, we introduce a noise-resistant feature encoder trained

*Corresponding author.

2324-9013/24/$31.00 ©2024 IEEE
DOI 10.1109/TrustCom63139.2024.00361



(a) Spike Streams (b) TFP

(c) TFI (d) SRCNN

(e) Spk2ImgNet (f) Ours

Fig. 1: Comparison of different methods for spike stream image reconstruction. TFP and TFI are highly susceptible to noise interference. Early DNN-based methods like SRCNN lack the capability to handle low-quality data effectively, while Spk2ImgNet, though more advanced, still faces issues with missing details in the reconstructed images compared to our proposed method.

through multi-representation adversarial robust learning. This encoder effectively suppresses industrial background noise, such as Gaussian and imspike noise, while extracting robust multi-dimensional features from the spike stream data. Next, in the feature integration and consistency enhancement stage, we apply sliding window attention mechanisms [36] to extract both temporal and spatial information from the encoded features. By capturing local spatial details within each frame and dynamic temporal relationships between frames, this stage ensures that the integrated features remain consistent and stable, even when dealing with missing or noisy data. Finally, in the image reconstruction stage, the processed features are passed through an spike stream reconstruction model. This module, fine-tuned on our specific dataset, maps the enhanced features to an image space, generating high-quality reconstructed images that accurately capture the dynamic and spatial characteristics of the original spike stream data.

Overall, our work makes the following contributions:

- We designed a noise-resistant feature encoder based on multi-representation adversarial robust learning to effectively handle high-intensity noise typically found in industrial environments, ensuring robust feature extraction.
- We employ sliding window attention mechanisms to fuse spatial and temporal information, addressing discontinuities and ensuring consistency in spike stream data.
- We fine-tuned a spike stream reconstruction model that can efficiently reconstructs high-quality images and can be adapted to specific industrial scenarios.

The rest of this paper is organized as follows. Section II reviews the related work in denoising, image reconstruction, transfer learning, and model fine-tuning. Section III discusses the spike stream generated in industrial scenarios, including the imaging principles of spike-based sensors and the challenges of data quality in industrial environments. Section IV presents the proposed multi-stage processing method in detail, consisting of feature extraction and denoising, integration of temporal and spatial information, and spike feature reconstruction. Section V reports the experimental results, including the datasets used, evaluation metrics, implementation details, and ablation study. Finally, Section VI concludes the paper and highlights the contributions.

II. Related Work

Denoising. Early works were based on hand-designed filters for image denoising [8], [10]–[12], but this method was overly dependent on hyperparameter settings. Later, deep learning-based approaches emerged, with [8], [13]–[15] introducing the first denoising neural network, followed by numerous other denoising networks [13]–[15] that further improved denoising performance. However, these methods heavily relied on high-quality labeled data. To address the shortcomings of supervised learning methods, some self-supervised denoising models were proposed by [16]–[18] which would not need much well labeled data.

Image Reconstruction. In early works, techniques like TFI [5] infer instantaneous light intensity based on inter-spike intervals, providing primary visual reconstruction for dynamic scenes. However, such simple reconstruction

methods typically fail to meet visual expectations in the presence of significant noise. TFP [4], on the other hand, considers a longer photon accumulation period to stably infer light intensity, but still faces challenges in highly dynamic scenes involving fast motion. With the advancement of deep learning, many approaches based on deep neural networks have emerged [19], [21], [24]. For instance, SRCNN [20] can generate high-resolution images from low-resolution inputs; DnCNN [8] can learn the mapping from noisy images to noise for denoising, thereby improving image reconstruction quality. Other works have incorporated powerful modules such as attention blocks [25] and feedback blocks [26] to further enhance the reconstruction performance of neural networks. Spk2ImgNet [27], while effective, might miss the global context, In addition, SpikeFormer [33] uses an attention module, which requires high computing resources for large images, which is a challenge for resource-constrained industrial edge scenarios.

Transfer Learning and Model Fine-tuning. Transfer Learning and Fine-tuning [28], [29] are core techniques in deep learning aimed at improving model performance while reducing data and computational resource requirements. In transfer learning, base models are typically pre-trained on large-scale datasets to learn general feature representations. These pre-trained models are then applied to specific target tasks, thus avoiding the high cost of training from scratch. As a critical step in transfer learning, fine-tuning involves further optimizing the parameters of the pre-trained model to better adapt it to the specific requirements of the target task.As is shown in Fig. 2, fine-tuning techniques are widely used in the optimization of large language models (such as LLAMA [30] and GPT [31]), enabling them to quickly adapt to various downstream tasks. This technique is equally applicable to smaller-scale deep learning models. Previous research has demonstrated that training reconstruction models using multiple spike datasets, including both real-world and simulated data, can endow the models with certain feature extraction capabilities [22], [32]. Based on this foundation, we use these pre-trained models as the base models and fine-tune them using datasets from industrial scenarios, enabling the models to better adapt to the specific requirements of new tasks.

III. Spike Stream Generated in Industrial Scenarios

In this section, we briefly discuss the imaging principles of spike-based visual sensors and analyze why data quality degrades in industrial environments.

A. Imaging Principles of Spike-Based Sensors

The spike vision sensor directly records light intensity information with spatial and temporal features. It outputs

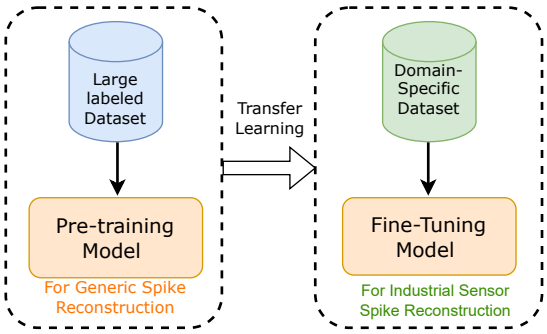


Fig. 2: Transfer learning process from a pre-trained model for generic spike reconstruction to a fine-tuned model for industrial scenarios spike reconstruction.

a binary stream in the form of spikes, where the data is represented solely by 0 or 1. The spike sensor primarily consists of three components: a photosensor, an integrator, and a comparator [34], as illustrated in Fig. 3. The photo-

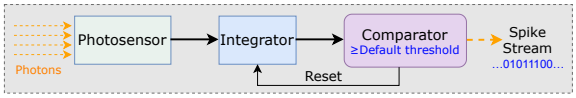


Fig. 3: Sampling Principle of the Spike Vision Sensor.

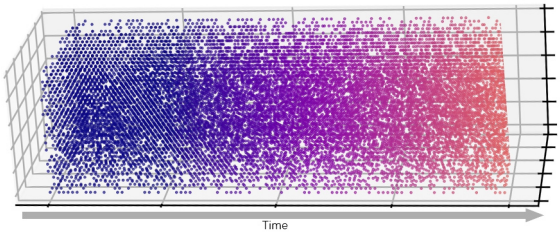
sensitive pixel array is spatially arranged on the sensor's photodetector and continuously captures photons. Subsequently, the integrator converts the light signal into an electrical signal and accumulates the voltage for each cell. A comparator detects whether the accumulated voltage reaches a preset threshold θ. Once the threshold is reached, a spike is triggered, and the voltage is reset to the preset value. The spike generation process for a pixel can be described as:

$$\int_{t_{i-1}}^{t_i} \alpha I(t)\, dt = \theta \tag{1}$$

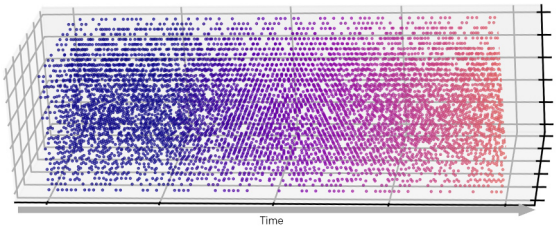
In equation (1), I_t describes the light intensity, where t_i and t_{i-1} represent the trigger times of the i-th and $(i-1)$-th spikes, respectively, and α is the photoelectric conversion rate. Due to hardware circuit limitations, the units in the output circuit read spikes periodically within a fixed interval $\Delta t = 50\,\mu s$ as discrete time signals $s(x, y, n)$, with a sampling frequency of $f_s = 20kHz$. If a pixel at spatial coordinates (x, y) triggers a spike at time t, it will output $s(x, y, n) = 1$ for $n = 1, 2, \ldots$, where $(n-1)\Delta t < t \leq n\Delta t$; otherwise, it will output $s(x, y, n) = 0$. The sensor generates spike frames of size $H \times W$ at each discrete timestamp n using high-speed polling. Within a fixed interval of $\Delta t \cdot T$, the sensor produces a stream of spikes $S = s(x, y, t)_{t=1}^{T}$ with dimensions $H \times W \times T$ [35].

B. Challenges of Data Quality in Industrial Environments

The spike vision sensor may also be affected by noise in industrial scenarios, although its working principle makes it more resistant to interference than traditional frame image sensors in certain situations. The industrial production environment is relatively complex and may be subject to factors such as electromagnetic interference and vibrations.



(a) Spike stream generated in industrial scenarios.



(b) Spike stream generated in normal scenarios.

Fig. 4: Comparison of spike streams generated under different scenarios. Compared with (a), (b) shows sparsity after visualization. This sparse representation usually means that the data is more concise and organized after denoising, while noisy spike stream is denser due to the presence of noise.

As shown in Fig. 4a, the spike points are densely distributed, which typically indicates more noise or that the system is being influenced by a greater number of external factors. In complex industrial scenarios, there are numerous factors that affect the spike stream, such as environmental conditions, the state of mechanical equipment (vibration, friction, etc.), and changes in other operational environments. The combined effect of these factors may lead to an increase in the spike stream density, rather than just being a result of noise. In contrast, the point density in Fig. 4b is significantly lower than in Fig. 4a, suggesting that the spike sensor exhibits better anti-interference capabilities in non-industrial scenarios.The main influencing factors are listed below:

Electromagnetic interference. Electromagnetic interference commonly found in industrial scenarios (such as motors from large equipment, welding machines, and wireless communication systems) may affect spike vision sensors, causing the sensors to respond to non-authentic events. This interference can lead to the generation of additional events, or false signals, thereby reducing the accuracy of the data [8]. Electromagnetic interference introduces extra voltage fluctuations in the sensor circuits, resulting in the generation of false signals by the spike vision sensors [23]. This interference can be represented by induced voltage, the magnitude of which is related to the frequency and distance of the interference source.The electromagnetic induced voltage V_{ind} can be expressed by Faraday's law:

$$V_{ind} = -N\frac{d\Phi}{dt} \tag{2}$$

Rapidly changing electromagnetic fields generated by industrial equipment can cause a large $\frac{d\Phi}{dt}$, which can induce a higher voltage in the sensor circuit and introduce noise. The additional voltage introduced by electromagnetic interference may compete with the threshold of the spike signal, thereby triggering false event occurrences. Although it can be reduced by good electromagnetic shielding and circuit design, it cannot be completely avoided.

Vibration. Vibration can affect the sensor through mechanical coupling, especially in certain situations where vibration may displace the sensor's sensitive components, leading to erroneous signal detection. The impact of vibration can be described using a simple harmonic motion model [8]. Let the displacement amplitude of the sensor be A, and the vibration frequency be ω, then the acceleration of the vibration $a(t)$ can be expressed as:

$$a(t) = -A\omega^2 \cos(\omega t) \tag{3}$$

High-frequency vibrations can cause slight displacements in certain mechanical components of the sensor, which may in turn trigger false events. The noise caused by such physical vibrations can be eliminated through damping measures or by introducing filtering mechanisms in the sensor's algorithms.

In addition to the above noise sources, there is also the influence of sudden bright light.Spike vision sensors are highly sensitive to changes in light, and rapid changes in light intensity (such as the sudden strong light produced by welding equipment) can trigger a large number of events, thereby affecting the data collected by the sensor. In industrial settings, sudden bursts of strong light like those from welding may cause the sensor to generate an excessive number of events, which could increase noise, interfere with normal signal detection, and even lead to data overload.

IV. Multi-Stage Processing Method

A. Overview

As is shown in Fig. 5, our method is divided into three stages. The first stage involves adding noise to the data for

multi-representational adversarial learning [37], allowing the encoder (generator) to learn both noisy and clean features, thereby gaining the ability to denoise noisy data. Next, we use SwinSF [36] to perform further spatiotemporal feature extraction on the initially processed data. Each RSSB consists of multiple SABs, which enable feature extraction based on spatiotemporal consistency. Finally, we employ the fine-tuned reconstruction model for spike reconstruction to obtain high-quality images.

B. Feature Extraction and Denoising

The goal of this phase is to eliminate high-frequency noise and isolated events caused by electromagnetic interference and vibration, thereby reducing noise interference in subsequent processing steps and ensuring the reliability and efficiency of the later stages.

Traditional preprocessing methods [8], [38], [39] in this phase mainly rely on consistency checks in both the spatial and temporal dimensions. Spatially, local neighborhood statistics (e.g., event density or local intensity consistency) are used to determine the validity of isolated events. Isolated events are typically caused by high-frequency electromagnetic interference or vibrations, and they often lack sufficient support from neighboring events. At the same time, temporal filtering mechanisms analyze the time intervals between events to remove pseudo-events with abnormal time gaps. Although spatio-temporal consistency-based filtering is effective for noise reduction, there are cases where it mistakenly identifies valid burst events as noise. To mitigate such misclassifications, context information or adaptive methods based on event sequences can be introduced, reducing the occurrence of false positives.

Noise-Resistant Feature Encoder: Multi-representation adversarial learning [37] is a learning approach that enhances the robustness and performance of the model by introducing multiple feature representations, such as noisy features and clean features. The objective is to capture the characteristics of the data from different perspectives, thereby improving the model's resilience against noise and interference. In the adversarial learning framework, the generator accepts input data with both noisy and clean features and attempts to generate denoised images. These generated images are then passed to the discriminator, whose task is to distinguish whether the images contain noise. Specifically, the discriminator classifies extracted features into noisy features and clean features, and through adversarial training, the generator gradually learns to produce more realistic and noise-free streams. When processing real data, the encoder has learned the characteristic pattern of noise and clean data, so as to obtain the recognition and processing ability of noise.

C. Integration of Temporal and Spatial Information in Spike Streams

In our multi-stage processing framework, the sliding window attention mechanism [36], an improved self-attention mechanism, is employed to extract spatial and temporal information both within and between frames. Spike streams generated by spike sensors at high acquisition frequencies exhibit temporal sparsity, but their continuous triggering within short time intervals can be viewed as dynamic time-series data, making them an ideal scenario for applying the sliding window attention mechanism.

The sliding window attention mechanism is designed to reduce computational overhead while focusing on the extraction of local features. For each event in the spike stream frame, the model first divides the image frame into several local patches, as shown in the figure, and applies the attention mechanism within these local regions. This localized attention operation leverages the strengths of the Swin Transformer [36], which can capture local features in fine detail while also enabling global feature extraction over larger areas. The hierarchical structure and shifted window design of the Swin Transformer not only allow effective extraction of spatial features within frames but also capture temporal dependencies between frames through the sliding window mechanism, addressing the temporal sparsity of spike stream data and improving the model's spatiotemporal modeling capabilities.

D. Spike Feature Reconstruction

After the previous two stages, the spike data has undergone feature extraction and noise suppression. In the first phase, multi-representation adversarial learning was employed to extract key features and effectively suppress noise, while in the second phase, the sliding window attention mechanism captured spatiotemporal information from both within and between frames, making the feature representation more precise and comprehensive. Based on these enhanced features, we proceed with the third phase, which focuses on reconstructing the spike data.

In this stage, we first pre-trained and an existing image reconstruction model [36] using high-quality ground truth simulated data and real-world datasets from non-industrial scenarios. This pre-training enables the models to develop fundamental spike reconstruction capabilities and an initial understanding of spike feature representation. To adapt the model to the characteristics of spike data from industrial scenarios, we fine-tuned the model using real-world spike data collected from industrial scenarios. During fine-tuning, we utilize the features processed in the previous two stages as inputs, enhancing the model's sensitivity to the data characteristics of the target tasks, thus improving its adaptability to spike stream data in specific scenarios. Consequently, the model is not only able to preserve the details of spike data but also enhances

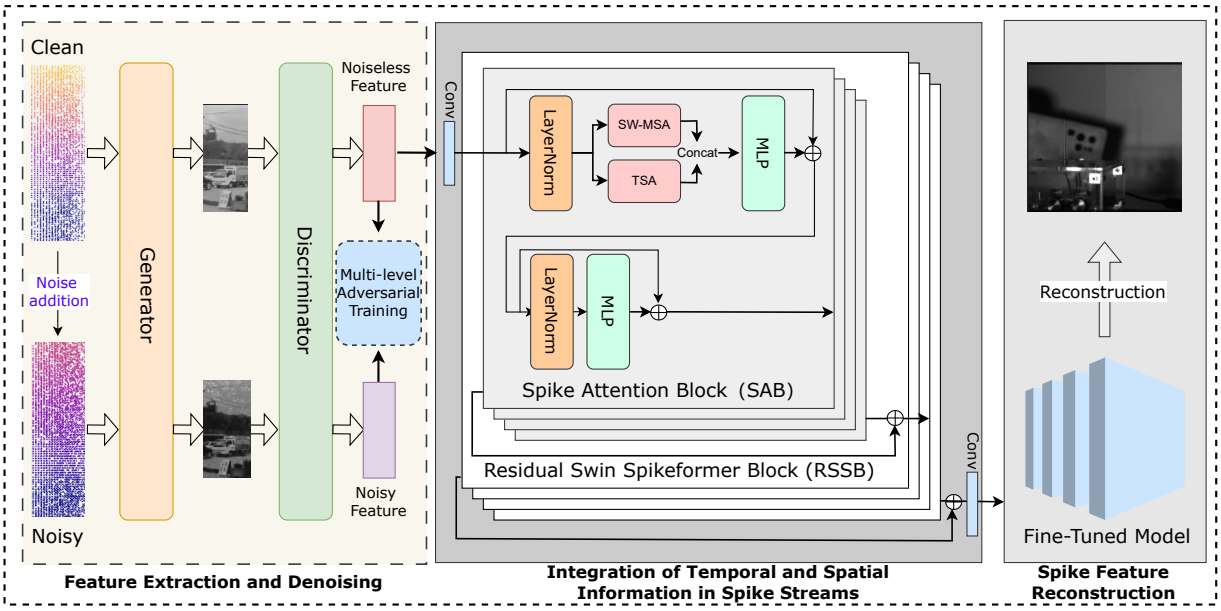


Fig. 5: Overview of the Multi-Stage Processing Method.

its adaptability to the specific characteristics of data in industrial environments, ensuring the quality and accuracy of the reconstruction results.

V. Experiment

A. Datasets

Training Dataset: To train our network, we require large datasets with labeled data. We utilized the Spike-x4K dataset [36], a synthetic dataset containing spike streams paired with ground truth images at a resolution of 1000x1000 pixels. Additionally, we used the real-world dataset Spike40 [27], which has a resolution of 250x400 pixels and covers a variety of everyday scenes. In addition, the verification set is the 15% split from the training set.

Fine-tuning Dataset: To further refine the model's performance in specific scenarios, we fine-tuned the pre-trained model using a dataset collected from power scenario with a resolution of 1000 × 1000. We simulate the discharge in the electric power scenario and add some interference and noise to the data collection, which could enable the model to better capture the characteristics of spike streams in these settings.

B. Metric

The evaluation metrics employed were Structural Similarity Index (SSIM) and Peak Signal-to-Noise Ratio (PSNR). Since no established method currently exists to evaluate spike stream similarity [40], we adopted these two metrics to assess the quality of our reconstructed images, as they have been used in state-of-the-art work [27].

C. Implementation Details

Our experimental platform is Ubuntu 20.04.5 LTS, and the hardware device is NVIDIA Tesla V100 32GB. The epoch of each deep learning method are set to 100, and the batchsize is also set to the same value of 16. We first performed multi-representation adversarial robust learning by training an encoder. Noisy simulated data was used as input for the generator, while the discriminator conducted adversarial learning between the noisy and clean data. To accelerate this process, we segmented the spike stream across multiple scales and employed multi-level adversarial learning. To fully leverage the spatiotemporal information in the data, the features generated by the encoder were further extracted using simulated datasets. Multiple Swin Transformer Blocks [36] were utilized in this step for efficient restoration of both local and global details. Finally, the reconstruction model was trained using real-world datasets to enable its reconstruction capability for spike stream data. After completing these tasks, we fine-tuned the entire architecture with our custom simulated datasets to tailor it for the current task.

D. Result

We first evaluated the reconstruction performance of traditional methods, including TFI and TFP, as well as classical deep learning models such as SRCNN and Spk2ImgNet, using validation datasets split from the training sets to evaluate reconstruction performance under normal scenarios. The experimental results are presented in Table I, which indicate that our method is equipped

with effective reconstruction capabilities and performs well.

TABLE I: Comparison of results on real world normal scenario datasets.

Metric	TFP	TFI	SRCNN	Spk2ImgNet	Ours
PSNR	19.52	25.66	28.76	31.01	**36.77**
SSIM	0.5834	0.7945	0.8498	0.9384	**0.9726**

Subsequently, we conducted experiments using datasets collected from industrial scenarios. The reconstructed images are illustrated in Fig 1. As shown in Table II, the experimental results demonstrate that our method, following multiple processing steps, can achieve higher-quality images during the reconstruction stage of the original industrial spike stream.

TABLE II: Comparison of results on industrial scenario dataset.

Metric	TFP	TFI	SRCNN	Spk2ImgNet	Ours
PSNR	17.06	24.78	25.79	29.48	**36.21**
SSIM	0.5301	0.7489	0.7728	0.8509	**0.9681**

In summary, our approach demonstrates a significant performance improvement when processing complex industrial spike stream data, achieving a 13.77% increase in SSIM and a 22.82% increase in PSNR compared to Spk2ImgNet.

E. Ablation Study

To evaluate the contributions of the proposed Feature Extraction and Denoising (FED) and Integration of Spatial and Temporal (IST) modules, we conducted an ablation study on the processed Spikex4K dataset with manually added noise. Four configurations were compared:

- **Baseline1**: Only the reconstruction part is used, without FED or IST.
- **Baseline2**: Builds upon Baseline1 by incorporating the FED module for denoising, aiming to suppress noise and enhance feature quality.
- **Baseline3**: Adds the IST module to Baseline1, focusing on extracting and integrating spatial and temporal information.
- **Ours**: Combines both FED and IST for a comprehensive improvement.

As shown in Table III, Baseline1 achieved the lowest performance, with a PSNR of 29.50 and SSIM of 0.8710, indicating its vulnerability to noise. Adding the FED module in Baseline2 significantly improved PSNR and SSIM to 32.65 and 0.9319, respectively, demonstrating the denoising module's effectiveness. Baseline3, incorporating the IST module, further enhanced performance to a PSNR of 33.37 and SSIM of 0.9406, highlighting the importance of spatio-temporal feature extraction. Finally, the full model achieved the best results, with a PSNR of 35.65 and SSIM of 0.9660, showing that combining denoising and spatio-temporal integration maximizes reconstruction quality.

This study confirms the complementary benefits of FED and IST in processing high-noise spike data.

TABLE III: Evaluation of the proposed modules.

Method	FED	IST	PSNR	SSIM
Baseline1	✗	✗	29.50	0.8710
Baseline2	✓	✗	32.65	0.9319
Baseline3	✗	✓	33.37	0.9406
Ours	✓	✓	35.65	0.9660

VI. Conclusion

In this paper, we proposed a multi-stage spike stream processing method to deal with the high-noise data generated in complex industrial environments. First, we adopted the multi-representation adversarial robust learning method to conduct preliminary noise reduction on the data, and then used Multiple Swin Transformer Blocks to extract the spatial-temporal characteristics of the data again to restore the details of the spike image as much as possible. Finally, we use the fine-tuned model to reconstruct the spike image, and the experimental results show that our method has obvious advantages in dealing with high noise industrial data further for downstream task application.

Acknowledgment

This work is supported by the National Natural Science Foundation of China (No. 92267105), Guangdong Basic and Applied Basic Research Foundation (No. 2023B1515130002), Guangdong Special Support Plan (No. 2021TQ06X990), Shenzhen Basic Research Program (No. JCYJ20220818101610023, KJZD20230923113800001).

References

[1] G. Gallego, T. Delbruck, G. Orchard, C. Bartolozzi, B. Taba, A. Censi, S. Leutenegger, A. J. Davison, J. Conradt, K. Daniilidis, and D. Scaramuzza, "Event-based vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 154–180, 2022.

[2] S. Elkateb, A. Métwalli, A. Shendy, and A. E. B. Abu-Elanien, "Machine learning and IoT-based predictive maintenance approach for industrial applications," *Alexandria Eng. J.*, vol. 88, pp. 298–309, 2024.

[3] V. G. Biju, A.-M. Schmitt, and B. Engelmann, "Assessing the influence of sensor-induced noise on machine-learning-based changeover detection in CNC machines," *Sensors*, vol. 24, no. 330, 2024.

[4] L. Zhu, S. Dong, T. Huang, and Y. Tian, "A retina-inspired sampling method for visual texture reconstruction," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, 2019, pp. 1432–1437.

[5] L. Zhu, S. Dong, T. Huang, and Y. Tian, "A retina-inspired sampling method for visual texture reconstruction," in *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, Shanghai, China, 2019, pp. 1432–1437.

[6] S. Dong, T. Huang, and Y. Tian, "Spike camera and its coding methods," *arXiv preprint arXiv:2103.15247*, 2021.

[7] L. Zhu, S. Dong, J. Li, T. Huang, and Y. Tian, "Retina-like visual image reconstruction via spiking neural model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 151–160.

[8] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[9] X. Jiang, S. Liu, X. Feng, and L. Zhang, "FOCNet: A fractional optimal control network for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019.

[10] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2017, pp. 486–494.

[11] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2005, pp. 60–65.

[12] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[13] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[14] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1712–1722.

[15] S. Lefkimmiatis, "Universal denoising networks: A novel CNN architecture for image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 3204–3213.

[16] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void: Learning denoising from single noisy images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 2124–2132.

[17] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2019, pp. 6970–6980.

[18] X. Wu, M. Liu, Y. Cao, D. Ren, and W. Zuo, "Unpaired learning of deep image denoising," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020.

[19] Y. Tian, Y. Zhang, Y. Fu, and C. Xu, "TDAN: Temporally-deformable alignment network for video super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2020, pp. 151–160.

[20] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2014, pp. 184–199.

[21] M. Haris, G. Shakhnarovich, and N. Ukita, "Recurrent back-projection network for video super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 151–160.

[22] Z. Liu and K. Ye, "YOLO-IMF: An Improved YOLOv8 Algorithm for Surface Defect Detection in Industrial Manufacturing Field," in *Proc. Metaverse – METAVERSE 2023: 19th International Conference, Held as Part of the Services Conference Federation, SCF 2023*, Honolulu, HI, USA, Sep. 2023, pp. 15–28.

[23] H. Tian and K. Ye, "CEESys: A Cloud-Edge-End System for Data Acquisition, Transmission and Processing Based on HiSilicon and OpenHarmony," in *Proc. Metaverse - METAVERSE 2023: 19th International Conference, Held as Part of the Services Conference Federation, SCF 2023*, Honolulu, HI, USA, Sep. 2023, pp. 3–14.

[24] Y. Jo, S. W. Oh, J. Kang, and S. J. Kim, "Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 151–160.

[25] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 151–160.

[26] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 151–160.

[27] J. Zhao, R. Xiong, H. Liu, J. Zhang, and T. Huang, "Spk2ImgNet: Learning to reconstruct dynamic scene from continuous spike stream," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 151–160.

[28] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, "Exploring the limits of transfer learning with a unified text-to-text transformer," *J. Mach. Learn. Res.*, vol. 21, pp. 1–67, 2020.

[29] J. Howard and S. Ruder, "Universal language model fine-tuning for text classification," in *Proc. 56th Annu. Meeting Assoc. Comput. Linguistics (ACL)*, 2018, pp. 328–339.

[30] H. Touvron et al., "LLaMA: Open and efficient foundation language models," *arXiv preprint arXiv:2302.13971*, 2023.

[31] A. Radford and K. Narasimhan, "Improving Language Understanding by Generative Pre-Training," 2018.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.

[33] C. She and L. Qing, "SpikeFormer: Image reconstruction from the sequence of spike camera based on transformer," in *Proc. 2022 5th Int. Conf. Image Graphics Process.*, 2022.

[34] J. Zhang, Y. Zheng, Z. Yu, and T. Huang, "Research on spike vision for autonomous driving scenarios," *Eng. Sci. China*, vol. 26, no. 1, pp. 160–177, 2024.

[35] J. Zhao, J. Xie, R. Xiong, J. Zhang, Z. Yu, and T. Huang, "Super resolve dynamic scene from continuous spike streams," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 2513–2522.

[36] L. Jiang, C. Zhu, and Y. Chen, "SwinSF: Image reconstruction from spatial-temporal spike streams," *arXiv preprint arXiv:2407.15708*, 2024.

[37] V. Kotariya and U. Ganguly, "Spiking-GAN: A spiking generative adversarial network using time-to-first-spike coding," in *Proc. 2022 Int. Joint Conf. Neural Netw. (IJCNN)*, 2021, pp. 1–7.

[38] P. Arias and J.-M. Morel, "Video denoising via empirical Bayesian estimation of space-time patches," *J. Math. Imaging Vis.*, vol. 60, pp. 70–93, 2017.

[39] M. Tassano, J. Delon, and T. Veit, "FastDVDnet: Towards real-time deep video denoising without flow estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2019, pp. 1351–1360.

[40] L. Hu, L. Ma, Z. Yu, B. Shi, and T. Huang, "Spike Stream Denoising via Spike Camera Simulation," *arXiv preprint arXiv:2304.03129*, 2023. [Online]. Available: https://arxiv.org/abs/2304.03129