TOWARDS AN ENERGY-EFFICIENT ANOMALY DETECTOR: WHEN MULTIMODAL ANOMALY DETECTION MEETS SPIKING NEURAL NETWORK

Anonymous authors
Paper under double-blind review

ABSTRACT

Recent advances in multimodal anomaly detection (MAD) primarily rely on more sophisticated and powerful artificial neural networks (ANNs) to improve detection performance. However, less attention has been paid to the fact that ANN-based paradigms often require high energy consumption to achieve these performance gains. This gap raises a key challenge: Can we significantly reduce energy consumption while maintaining competitive performance? Spiking neural networks (SNNs), which employ binary spikes and event-driven computation, have emerged as a promising energy-efficient alternative to conventional ANNs. Motivated by this, we propose SpikingMAD, a spike-driven MAD framework designed to achieve both low power consumption and high performance. Specifically, we develop a spike-driven multimodal reconstruction network that restores abnormal features to normal ones through spike-sparse computation and multiplication-free operations. Furthermore, given that dot-product attention fails to measure the similarity between RGB and depth features encoded by spiking neurons, we introduce a spike-driven Hamming cross-attention mechanism to capture cross-modal feature similarity and facilitate multimodal feature fusion. Extensive experiments on MVTec 3D-AD, Real-IAD D3, and Eyecandies demonstrate that SpikingMAD achieves competitive performance in terms of accuracy, memory usage, frame rate, and energy consumption, while also showing strong potential in few-shot scenarios. Code: `https://anonymous.4open.science/r/SpikingMAD`.

1 INTRODUCTION

Limitation. Multimodal anomaly detection (MAD) is expected to accurately pinpoint anomalous regions in industrial products by exploiting multi-source information (such as RGB and depth images) like human inspectors. To achieve this purpose, most existing efforts primarily leverage increasingly sophisticated and continuously evolving artificial neural networks (ANNs) to advance the development of MAD. For example, M3DM (Wang et al., 2023), LSFA (Tu et al., 2024), G^{2}SF Tao et al. (2025) and CFR (Guo et al., 2026), memory-bank approaches, aim to store normal features extracted by pre-trained models and detect anomalies via feature discrepancies. 3D-ADNAS (Long et al., 2025) and PIRN (Li et al., 2026b), reconstruction-based methods, seek to empower reconstruction networks with the ability to restore synthetic anomalous

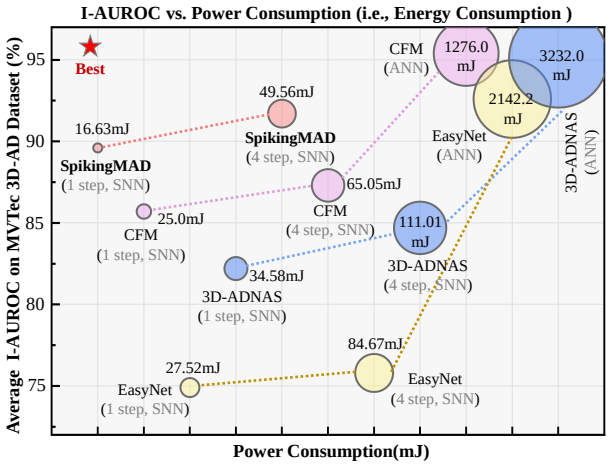

Figure 1: Comparison of I-AUROC and energy/power consumption between SpikingMAD and other SNN- and ANN-based baselines.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

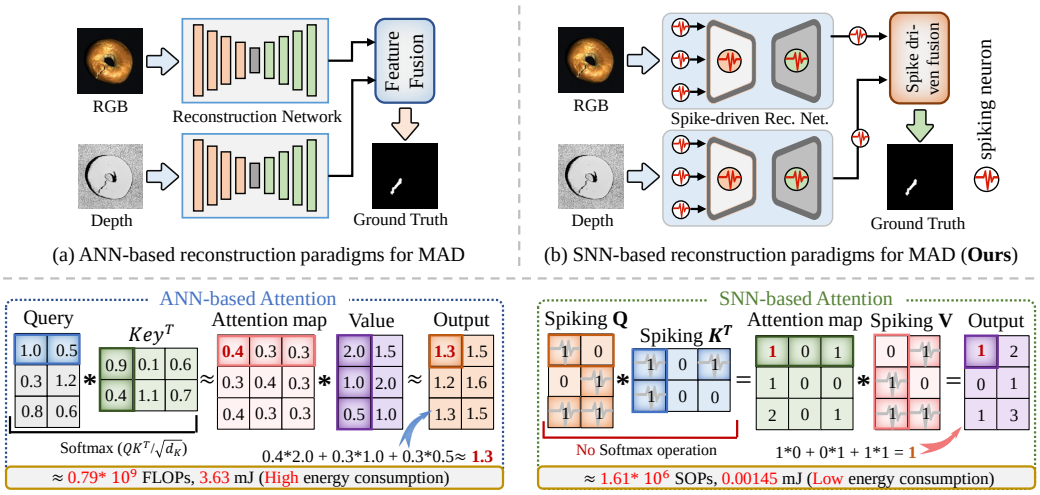

Figure 2: (***Top***) Illustration of the ANN- and SNN-based paradigms for multimodal anomaly detection. (***Bottom***) Example of energy consumption estimation for SNN- and ANN-based MAD methods, where SOPs indicate the theoretical synaptic operations and FLOPs denote the floating-point operations. More details about energy consumption and SOPs are provided in Appendix B.

samples as normal ones for anomaly detection. However, these ANN-based schemas overlook a critical issue: **energy/power consumption**. As previewed in Figure 1, extensive experiments on ANN-based MAD baselines reveal that superior I-AUROC comes at the cost of continuously growing energy consumption. For example, CFM (Costanzino et al., 2024) achieves 95.4% I-AUROC with a power consumption of 1276 mJ, EasyNet (Chen et al., 2023) gives 92.6% with 2142.2 mJ, and 3D-ADNAS (Long et al., 2025) reaches 95.1% with 3232 mJ. This surge stems from the massive dense floating-point multiply-accumulate (MAC) operations, which introduce heavy computational and energy burdens and worsen with increased model depth and complexity.

Motivation. Naturally, the aforementioned predicament leads us to a challenge: ***Can we significantly reduce energy consumption while maintaining competitive detection performance*?** To bridge this gap, spiking neural networks (SNNs), as the third generation of neural networks, have emerged as a promising energy-efficient alternative to conventional ANNs, owing to their binary spike mechanism and event-driven computation. As shown in Figures 1 and 2, we can observe that integrating SNNs into existing MAD frameworks substantially reduces energy consumption. For instance, the power cost of CFM and 3D-ADNAS decreases from 1276 mJ to 65.05 mJ and from 3232 mJ to 111.01 mJ, respectively, demonstrating the potential of spike-driven computation. Thus, this study focuses on the exploration of the energy-efficient multimodal anomaly detection task.

Solution. However, naive aggregation of SNNs to the existing MAD architecture poses two key challenges. The primary one is the substantial degradation in detection accuracy, particularly for the reconstruction-and-segmentation paradigms like EasyNet and 3D-ADNAS, which even fail to restore the original images. The second is that dot-product attention struggles to effectively measure the similarity between RGB and depth features in spike-driven multimodal fusion. To address these issues, we propose SpikingMAD, a spike-driven MAD architecture that achieves a superior trade-off between detection accuracy and energy consumption. Specifically, to address the former, we develop a spike-driven multimodal reconstruction network with a residual topology (discarding the segmentation branch) that restores abnormal RGB/depth images to normal ones through spike-sparse computation and multiplication-free operations. To address the latter, guided by theoretical analysis, we design a spike-driven Hamming cross-attention mechanism to capture cross-modal feature similarity and facilitate multimodal feature fusion. A series of experiments on the MVTec 3D-AD, Real-IAD D3, and Eyecandies datasets demonstrates that SpikingMAD achieves competitive performance with the ANN-based MAD baselines while significantly reducing energy consumption.

Contributions. In summary, the main contributions of this work are as follows:

- To the best of our knowledge, this work is the pioneering effort to utilize the SNN to shape an energy-efficient MAD architecture. To achieve this, we propose a novel SpikingMAD method to steer the architecture with superior performance and lower energy consumption.

- We design a spike-driven multimodal reconstruction network (SMRN) with a residual topology for abnormal image restoration. It consistently outperforms the SNN-based reconstruction-and-segmentation method by $\sim 6\%$ in accuracy across multiple datasets.
- We propose a spike-driven Hamming cross-attention (SHCA) to overcome the limitation of dot-product attention in measuring similarity between RGB and depth features encoded by spiking neurons, thereby facilitating cross-modal interaction and feature fusion.

Remark. Extensive experiments across multiple datasets demonstrate that SpikingMAD significantly outperforms SNN-based baselines. Specifically, under unimodal setting, SpikingMAD surpasses 3D-ADNAS by 6.9% in I-AUROC while reducing power by 51.0% (16.96 mJ vs. 8.31 mJ, Table 1), validating SMRN's superiority over reconstruction-and-segmentation paradigms. Table 8 further indicates SHCA's effectiveness in fusing binary multimodal features. Furthermore, SpikingMAD maintains competitive I-AUROC with ANN-based methods while drastically reducing energy consumption (91.7% vs. 92.6% I-AUROC and 49.56 mJ vs. 2142.2 mJ against EasyNet), together with lower memory usage and higher frame rate. It also shows strong potential for few-shot MAD.

2 RELATED WORK

Multimodal Anomaly Detection. Existing efforts primarily concentrate on the advancement of unimodal AD, i.e., recognizing defects in RGB (Zavrtanik et al., 2021; Guo et al., 2025a; Cao et al., 2024; Gao et al., 2026; Chen et al., 2026a; Zhu et al., 2026) or point cloud data (Li et al., 2024; Ye et al., 2025; Cheng et al., 2026b), involving feature-embedding-based schemes (Liao et al., 2026), teacher-student paradigms (Safarov et al., 2026), and reconstruction-based approaches (Zhang et al., 2025; Liang et al., 2025). Recently, with the emergence of the MVTec 3D-AD (Bergmann et al., 2022), Eyecandies (Bonfiglioli et al., 2022), and Real-IAD D3 (Zhu et al., 2025) datasets, MAD has received widespread attention as a promising paradigm (Chu et al., 2023; Zavrtanik et al., 2024; Miao et al., 2025; Mao et al., 2025; Lee et al., 2026; Zhao et al., 2026b; Chen et al., 2026b; Liang et al., 2026). Approaches such as M3DM-NR (Wang et al., 2025a), UniMMAD (Zhao et al., 2026a), and GPFlow (Li et al., 2026a) have shown remarkable advances in anomaly detection by fusing RGB and depth images. However, existing studies have primarily focused on achieving higher detection performance, often overlooking the substantial energy costs associated with these performance gains. To bridge this gap, this work aims to substantially reduce energy consumption while maintaining competitive detection performance, and proposes an energy-efficient MAD framework.

Spiking Neural Network. SNNs, initially proposed by Hodgkin & Huxley (1952), operate with discrete binary spikes and event-driven computation, offering a more biologically plausible paradigm than conventional ANNs that rely on continuous signals (Zhou et al., 2023a; Shi et al., 2024; Zou et al., 2025). Since their inception, most efforts have been devoted to improving SNN performance and expanding their applications, giving rise to a wide range of architectures and neuron models (Su et al., 2023; Yao et al., 2025; Guo et al., 2025b; Xiao et al., 2025). Notably, the Leaky Integrate-and-Fire (LIF) model has become one of the most widely used foundational models in SNNs, owing to its simple yet effective simulation of neuronal behavior. For clarity, we illustrate the LIF membrane potential dynamics across time steps (Figure 3) and give its definition as follows:

$$\begin{aligned} I[t] &= M[t-1] + f[t], \\ S[t] &= Hea(I[t] - R_{th}), \\ M[t] &= \sigma * I[t] * (1 - S[t]) + R_{reset} * I[t], \end{aligned} \tag{1}$$

where $f[t]$ denotes the input at time step t, and $I[t]$ represents the membrane potential. $Hea(\cdot)$ is the Heaviside step function. When $I[t]$ exceeds R_{th} (firing threshold), the spike neuron will generate a spike $S[t]$, i.e., $S[t] = 1$; $S[t] = 0$ otherwise. The membrane potential $M[t]$ is then reset to R_{reset} (reset potential) when $S[t] = 1$ and $M[t] = \sigma * I[t]$ otherwise, σ is a hyperparameter (Chowdhury et al., 2021). Note that, although applying SNNs to MAD significantly reduces energy consumption, maintaining competitive detection performance and reconstruction quality remains a challenge.

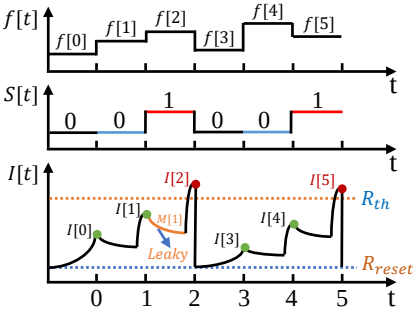

Figure 3: Illustration of membrane potential variations across time steps.

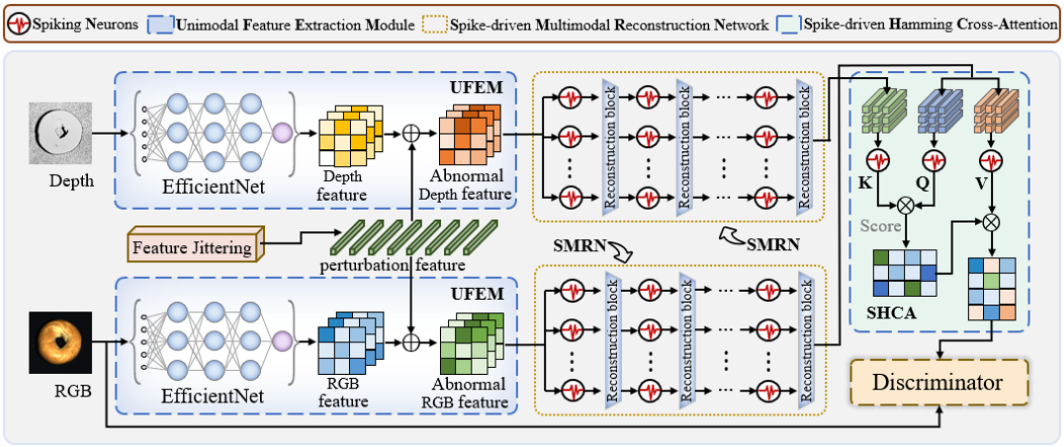

Figure 4: The overall architecture of SpikingMAD, where the spike-driven multimodal reconstruction network (SMRN) and spike-driven Hamming cross-attention (SHCA) are the core of this work.

3 SpikingMAD Method

In this section, we introduce SpikingMAD, a spike-driven approach for achieving energy-efficient multimodal anomaly detection. Subsequently, we outline the overall architecture, detailing the designed spike-driven modules and the theoretical analysis.

3.1 Overall Architecture

Figure 4 illustrates the skeleton of SpikingMAD, which comprises Unimodal Feature Extraction Modules (UFEM), Spike-driven Multimodal Reconstruction Networks (SMRN), a Spike-driven Hamming Cross-Attention (SHCA), and a Discriminator. Specifically, UFEM employs EfficientNet to extract features from RGB and depth images, respectively, and synthesizes anomalous features via feature jittering. Subsequently, SMRN restores these abnormal features to normal ones via spike-driven reconstruction blocks, drastically reducing MAC-induced computational overhead and energy consumption while maintaining reconstruction performance. Finally, we identify that dot-product attention is not well-suited for measuring cross-modal similarity between RGB and depth features within spike-driven fusion paradigms; thus, SHCA is designed to effectively fuse the multimodal features. Particularly, the SMRN and SHCA serve as the core components of the SpikingMAD architecture; we detail their designs in the following subsections.

3.2 Spike-driven Multimodal Reconstruction Network

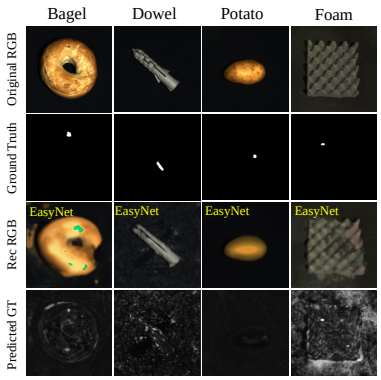

Figure 5: Reconstruction example after integrating SNNs into EasyNet.

As previewed in Figure 1 and discussed above, in MAD tasks, given that the multimodal reconstruction network (MRN) comprises a series of stacked reconstruction blocks, whose dense floating-point MAC operations are the primary factor causing high computational cost and energy overhead. To advance MAD, this raises a core question: ***How to seek a simple yet energy-efficient MAD architecture?*** A natural way is the spiking neural network, which employs binary spikes and event-driven computation, offering a promising alternative. However, naively integrating SNNs into MRN, particularly in reconstruction-and-segmentation architectures like EasyNet and 3D-ADNAS, results in significant performance degradation and even fails to reconstruct the original image (Figure 5). To address this issue, inspired by Tang et al. (2024), we synthesize anomalous features via feature jitter in UFEM (discarding the segmentation branch and mask training) and restore them to the normal space using spike-driven reconstruction blocks with residual topology. This design constitutes a simple yet energy-efficient spike-driven multimodal reconstruction network that

Definition 1. *The network is spike-driven if the input currents of all neurons satisfy the form:*

$$I_m[t] = \sum_n \omega_{m,n} * s_n[t] = \sum_{n, s_n[t] \neq 0} \omega_{m,n},$$

where $I_m[t]$ denotes the input current received by the m-th postsynaptic neuron at time step t, $\omega_{m,n}$ denotes the synaptic weight from n-th to m-th neuron, and $s_n[t] \in \{0, 1\}$ denotes the binary spike emitted by the n-th presynaptic neuron.

performs feature extraction through sparse, event-driven computation, as follows:

$$F_i^{R/D} = \mathbf{BN}(Conv2d(\mathcal{SN}(F_{i-1}^{R/D} + \mathbf{BN}(Conv2d(\mathcal{SN}(F_{i-1}^{R/D})))))), \tag{2}$$

where $F_i^{R/D}$ is the output of the i-th reconstruction block in SMRN. $F_i^{R/D} \in R^{T*C*H*W}$; T, C, H, and W denote time step, channel, height, and width, respectively. Specifically, when i = 1, $F_i^{R/D}$ represents the UFEM output. $\mathbf{BN}$ and $\mathcal{SN}$ denote the batch normalization and spiking neuron layers, respectively. Meanwhile, to encourage the normal RGB and depth images reconstructed by SMRN to be highly consistent with the original ones, we impose an SSIM loss between the input images and SMRN outputs. The reconstruction loss is

$$\mathcal{L}_{rec} = \mathcal{L}_{SSIM}^{R}(F_{org}^{R}, F_{rec}^{R}) + \mathcal{L}_{SSIM}^{D}(F_{org}^{D}, F_{rec}^{D}), \tag{3}$$

where F_{org}^{R} and F_{org}^{D} denote the original RGB and depth images, respectively, while F_{rec}^{R} and F_{rec}^{D} represent the output features of the final reconstruction block in SMRN. Furthermore, we investigate the spike-driven characteristics of SMRN. As stated in ***Definition* 1**, this characteristic requires that input current accumulation is exclusively triggered by sparse spikes emitted by presynaptic neurons. Clearly, both commonly used convolutional and linear layers satisfy this definition; therefore, SMRN naturally exhibits the spike-driven property. For more details, please refer to Appendix C.

3.3 Spike-driven Hamming Cross-Attention

Once the reconstructed RGB and depth features are obtained via SMRN, a multimodal fusion strategy is required to fuse both modalities for anomaly discrimination. To alleviate the substantial computational and energy consumption associated with multimodal fusion, inspired by Huang et al. (2025); Wang et al. (2025c), we employ spike-driven dot-product attention (SDA) for multimodal feature fusion, as follows:

$$Q_R = V_R = R_{eshape}(\mathcal{SN}(F_{rec}^{R})), \tag{4}$$

$$K_D = R_{eshape}(\mathcal{SN}(F_{rec}^{D})), \tag{5}$$

$$F_f^{'} = SDA(Q_R, K_D, V_R), \tag{6}$$

where Q_R, K_D, and $V_R \in \{0, 1\}^{\mathcal{O}*\mathcal{C}}$ are binary spike tensors serving as the Query, Key, and Value, respectively, with $\mathcal{C}$ and $\mathcal{O}$ denoting channel dimension and token length. and $F_f^{'}$ represents the fused features.

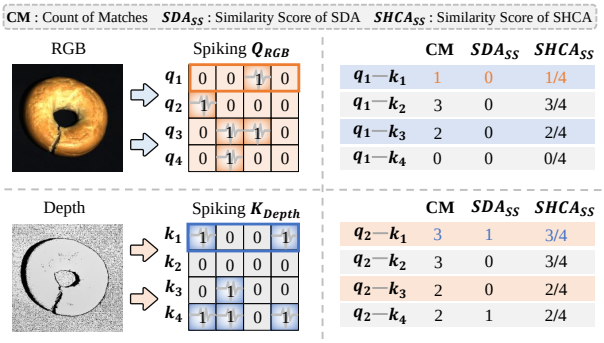

Figure 6: Comparison of similarity measurements between SDA and SHCA. For SDA, the CM (count of matches) for pairs (q_1, k_1) and (q_1, k_2) are 1 and 3, but all similarity scores are zero. This occurs because zero entries in Q_{RGB} cause SDA to indiscriminately disregard the corresponding elements in K_{Depth}, showing its limitation in measuring similarity between spike matrices.

However, as illustrated in Figure 6, when SDA is employed to perform multimodal feature fusion, zero elements in the spike matrix Q_R cause the corresponding elements in K_D to be indiscriminately ignored. This makes SDA ineffective at measuring the similarity between Q_R and K_D, and thus ill-suited for multimodal feature fusion. To resolve this issue, motivated by the Johnson–Lindenstrauss (JL) lemma for binary embedding (Johnson & Lindenstrauss, 1984; Yi et al., 2015; Jacques et al., 2013; Zou et al., 2025; Xiao et al., 2025; Guo et al., 2025b), we convert the real-valued attention mechanism into a spike-driven formulation, offering a feasible solution for measuring similarity between spike matrices. Thus, we give the following proposition:

Proposition 1. *Let X_R and $X_D \in R^{\mathcal{N}}$ denote two real-valued vectors. Following the JL lemma, their binary embeddings are obtained as $Q_{X_R} = \mathrm{sign}(\mathbb{B} * X_R)$, $K_{X_D} = \mathrm{sign}(\mathbb{B} * X_D)$, where $Q_{X_R}, K_{X_D} \in \{0,1\}^{\mathcal{C}}$ and the projection matrix $\mathbb{B} \in R^{\mathcal{N}*\mathcal{C}}$ has entries independently and identically sampled from a standard normal distribution. Accordingly, we have:*

$$f_{dot}(X_R, X_D) = \frac{(X_R)^T * X_D}{\parallel X_R \parallel \parallel X_D \parallel}, \quad f_{ham}(Q_{X_R}, K_{X_D}) = 1 - \frac{1}{\mathcal{C}} \sum_{i=1}^{\mathcal{C}} \mathbf{1}(Q^i_{X_R} \neq K^i_{X_D}), \quad (7)$$

where $f_{dot} \in [0,1]$ denotes the cosine similarity used in real-valued attention, and $f_{ham} \in [0,1]$ denotes the normalized Hamming similarity. Given any $\varepsilon > 0$ and $\mathcal{C} > \frac{\log \Lambda}{\varepsilon^2}$, we have:

$$\mathbb{P}(|f_{ham}(Q_{X_R}, K_{X_D}) - \mathbb{G}[f_{dot}(X_R, X_D)]| \leq \varepsilon) \geq 1 - 2e^{-\varepsilon^2 \mathcal{C}}, \quad (8)$$

where the function $\mathbb{G}[x] = \frac{\pi - \arccos(x)}{\pi}$ is continuous and monotonic, and x with $\mathbb{G}[x]$ in $[0,1]$, Λ denotes the total number of possible queries and keys in the training set, and $\mathbb{P}(\cdot)$ is probability.

The ***Proposition* 1** shows that $f_{ham}(\cdot)$ closely approximates $f_{dot}(\cdot)$ when the channel dimension $\mathcal{C}$ is sufficiently large, demonstrating that Hamming similarity is an effective manner to measure query-key similarity in spike-driven paradigms. However, to overcome the difficulties posed by the non-differentiability, nonlinear complexity, and floating-point MAC operations when applying Hamming similarity to spike-driven MAD, the function $f_{ham}(Q_{X_R}, K_{X_D})$ can be reformulated as:

$$f_{ham}(Q_{X_R}, K_{X_D}) = \frac{1}{2} + \frac{1}{2\mathcal{C}}(2Q_{X_R} - \mathbf{1})^T(2K_{X_D} - \mathbf{1}). \quad (9)$$

Given that the constant factor 1/2 does not affect the similarity evaluation, we employ $f_{ham}(\cdot)$ to SDA and formulate the spike-driven Hamming cross-attention (SHCA) mechanism as follows:

$$F_{SHCA} = \mathcal{SN}([\frac{1}{2\mathcal{C}}(2Q_R - \mathbf{1}) * (2K_D - \mathbf{1})^T]V_R) = \mathcal{SN}_{2\mathcal{C}}((2Q_R - \mathbf{1})[(2K_D - \mathbf{1})^T * V_R)]), \quad (10)$$

where F_{SHCA} denotes the fused feature obtained by SHCA. Due to page limitations, more details on the spike-driven characteristics of SHCA (for instance, to preserve the multiplication-free computation of SHCA, the scaling factor $1/2\mathcal{C}$ can be integrated into the spike-neuron threshold) and the theoretical analysis of ***Proposition* 1** are provided in Appendices C and D.

Finally, we formulate the overall training objective of SpikingMAD, which consists of the reconstruction loss and classification loss, as follows:

$$\mathcal{L}_{All} = \mathcal{L}_{rec} + \mathcal{L}^{mse}_{fusion} = \mathcal{L}_{rec} + \frac{1}{W * H} ||F^R_{org} - F_{SHCA}||^2. \quad (11)$$

4 Evaluation of the Proposed SpikingMAD Architecture

Dataset and Experimental Setting. We evaluate SpikingMAD on three datasets—MVTec 3D-AD (Bergmann et al., 2022) and Real-IAD D3 (Zhu et al., 2025), which are collected from real-world scenes, and Eyecandies (Bonfiglioli et al., 2022), which is synthetic virtual data—and compare it with ANN- and SNN-based baselines (BTF (Horwitz & Hoshen, 2023), EasyNet (Chen et al., 2023), M3DM (Wang et al., 2023), MMRD (Gu et al., 2024), CFM (Costanzino et al., 2024), Cflow (Rudolph et al., 2022), SimpleNet (Liu et al., 2023), PointMAE (Pang et al., 2023), PatchCore (Roth et al., 2022) and AST (Rudolph et al., 2023)) across time steps in terms of I-AUPRC, P-AUROC, AUPRO, and power under unimodal and multimodal settings. In our experimental protocol, we resize the input images to 224×224 and train the model for 1,000 epochs. All other hyperparameters adhere to the settings of IUF (Tang et al., 2024). For more implementation details, see Appendix E.

Evaluations on MVTec 3D-AD. Table 1 reports comparison results of SpikingMAD with both ANN- and SNN-based baselines under unimodal and multimodal input settings, with T (time steps) set to 1 and 4. It is clear that SpikingMAD consistently achieves superior performance while consuming lower power in most test cases. Specifically, when trained with RGB images at T=1, SpikingMAD achieves I-AUROC improvements of 13.6% and 6.9% over the EasyNet and 3D-ADNAS (SNN-based), respectively, while reducing energy consumption by 39.5% and 51.0%; when T increases to 4, these performance gains rise to 14.5% and 7.10%, with energy savings of 41.5%

Table 1: I-AUROC on MVTec 3D-AD across time steps under unimodal and multimodal settings.

	Method	Year	Spike	Types	Time Step	OPs (G)	Power (mJ)	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	BTF	CVPR23	✗	ANN	1	- -	- -	85.4	84.0	82.4	68.7	97.4	71.6	71.3	59.3	92.0	72.4	78.5
	EasyNet	MM23	✗	ANN	1	215.5	991.3	98.2	99.2	91.7	95.3	91.9	92.3	84.0	78.5	98.6	74.2	90.4
	M3DM	CVPR23	✗	ANN	1	- -	- -	94.4	91.8	89.6	74.9	95.9	76.7	91.9	64.8	93.8	76.7	85.0
	MMRD	AAAI24	✗	ANN	1	- -	- -	98.7	93.7	94.3	77.0	98.1	84.7	91.3	75.3	99.3	85.3	89.8
	3D-ADNAS	AAAI25	✗	ANN	1	267.3	1229.6	98.1	98.8	92.7	95.6	94.2	92.8	85.3	79.1	97.7	85.8	92.0
	EasyNet	MM23	✓	SNN	1	11.50	13.74	73.7	84.2	69.5	74.2	77.9	67.0	74.4	64.1	82.8	64.2	73.2
	3D-ADNAS	AAAI25	✓	SNN	1	15.07	16.96	80.4	88.2	83.7	78.8	77.6	82.8	76.5	72.1	85.9	73.3	79.9
	SpikingMAD	Ours	✓	SNN	1	**6.71**	**8.31**	**88.1**	**90.3**	**89.6**	**85.4**	**89.0**	**83.9**	**86.5**	**85.2**	**87.7**	**82.4**	**86.8**
	EasyNet	MM23	✓	SNN	4	43.31	42.33	74.5	84.8	70.6	74.5	79.1	67.3	76.3	64.7	83.0	65.5	74.0
	3D-ADNAS	AAAI25	✓	SNN	4	57.54	55.18	82.5	90.1	84.4	80.5	78.2	83.9	77.4	73.7	87.9	75.1	81.4
	SpikingMAD	Ours	✓	SNN	4	**25.01**	**24.78**	**90.0**	**92.1**	**91.6**	**86.6**	**92.4**	**85.3**	**88.9**	**87.0**	**88.2**	**83.2**	**88.5**
Depth	BTF	CVPR23	✗	ANN	1	- -	- -	69.6	55.3	82.4	69.6	79.5	77.3	57.3	74.6	93.6	55.3	71.4
	EasyNet	MM23	✗	ANN	1	215.5	991.3	73.5	67.8	74.7	86.4	71.9	71.6	71.3	72.5	88.5	68.7	74.7
	M3DM	CVPR23	✗	ANN	1	- -	- -	94.1	65.1	96.5	96.9	90.5	76.0	88.0	97.4	92.6	76.5	87.4
	MMRD	AAAI24	✗	ANN	1	- -	- -	82.9	68.6	93.7	80.4	97.2	86.5	94.7	80.6	96.7	84.9	86.6
	3D-ADNAS	AAAI25	✗	ANN	1	267.3	1229.6	79.4	85.7	69.9	94.6	69.5	68.6	70.5	87.3	95.3	66.7	78.8
	EasyNet	MM23	✓	SNN	1	11.50	13.74	56.0	76.3	60.7	59.2	69.2	**62.3**	57.8	58.3	66.8	50.1	61.7
	3D-ADNAS	AAAI25	✓	SNN	1	15.07	16.96	58.3	72.1	66.3	63.7	68.0	57.6	61.0	68.4	74.5	65.2	65.5
	SpikingMAD	Ours	✓	SNN	1	**6.71**	**8.31**	**63.7**	**83.3**	**70.6**	**64.3**	**76.9**	62.0	**62.8**	**81.6**	**86.1**	**68.9**	**72.0**
	EasyNet	MM23	✓	SNN	4	43.31	42.33	56.3	77.4	61.3	62.2	69.6	**64.5**	58.8	60.0	69.1	54.2	63.3
	3D-ADNAS	AAAI25	✓	SNN	4	57.54	55.18	60.8	77.1	69.2	64.3	71.5	58.3	61.7	72.9	78.7	59.2	67.4
	SpikingMAD	Ours	✓	SNN	4	**25.01**	**24.78**	**65.7**	**85.9**	**74.5**	**69.7**	**77.2**	63.8	**66.1**	**84.4**	**87.8**	**69.5**	**74.5**
RGB + Depth	BTF	CVPR23	✗	ANN	1	- -	- -	93.8	76.5	97.2	88.8	96.0	66.4	90.4	92.9	98.2	72.6	87.3
	EasyNet	MM23	✗	ANN	1	465.7	2142.2	99.1	99.8	91.8	96.8	94.5	94.5	90.5	80.7	99.4	79.3	92.6
	M3DM	CVPR23	✗	ANN	1	- -	- -	99.4	90.9	97.2	97.6	96.0	94.2	97.3	89.9	97.2	85.0	94.5
	CFM	CVPR24	✗	ANN	1	277.4	1276.0	99.4	88.8	98.4	99.3	98.0	88.8	94.1	94.3	98.0	95.3	95.4
	MMRD	AAAI24	✗	ANN	1	- -	- -	99.9	94.3	96.4	94.3	99.2	91.2	94.9	90.1	99.4	90.1	95.0
	3D-ADNAS	AAAI25	✗	ANN	1	702.6	3232.0	99.7	100.0	97.1	98.6	96.6	94.8	89.7	87.3	100.0	86.7	95.1
	EasyNet	MM23	✓	SNN	1	23.02	27.52	75.8	84.7	72.2	74.8	80.3	70.6	76.7	65.6	83.9	64.8	74.9
	CFM	CVPR24	✓	SNN	1	17.37	25.00	86.4	81.2	90.0	**88.7**	89.6	75.1	88.9	**87.6**	85.3	84.2	85.7
	3D-ADNAS	AAAI25	✓	SNN	1	30.87	34.58	83.3	89.4	86.0	82.1	79.8	82.6	77.4	76.5	88.9	75.6	82.2
	SpikingMAD	Ours	✓	SNN	1	**13.43**	**16.63**	**90.4**	**93.8**	**92.3**	**88.7**	**91.3**	**86.2**	**89.2**	87.2	**91.6**	**85.7**	**89.6**
	EasyNet	MM23	✓	SNN	4	86.63	84.67	77.1	85.2	72.7	76.3	82.5	71.9	76.9	66.8	84.2	64.6	75.8
	CFM	CVPR24	✓	SNN	4	61.87	65.05	87.1	82.7	90.9	**89.6**	91.4	76.5	**92.8**	90.0	87.4	84.2	87.3
	3D-ADNAS	AAAI25	✓	SNN	4	115.80	111.01	86.7	91.3	88.6	83.3	78.2	86.1	80.8	79.6	92.7	79.4	84.7
	SpikingMAD	Ours	✓	SNN	4	**50.02**	**49.56**	**92.8**	**95.0**	**96.1**	89.4	**92.1**	**88.8**	91.5	**90.9**	**93.7**	**86.2**	**91.7**

Table 2: I-AUROC on Real-IAD D3 dataset across time steps under unimodal settings (RGB).

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
RGB	Cflow	WACV22	✗	ANN	1	94.3	27.1	83.9	1.8	66.1	96.7	52.9	46.2	85.3	99.6	78.1
	SimpleNet	CVPR23	✗	ANN	1	97.3	71.7	79.5	37.2	66.1	98.1	40.8	41.6	56.4	93.3	73.7
	EasyNet	MM23	✓	SNN	1	61.5	62.0	58.8	61.7	59.1	57.9	59.6	60.4	61.3	63.7	62.0
	3D-ADNAS	AAAI25	✓	SNN	1	65.4	68.3	63.7	66.0	**63.5**	67.6	63.1	63.8	72.7	71.9	69.7
	SpikingMAD	Ours	✓	SNN	1	**65.8**	**83.6**	**70.1**	**73.7**	60.6	**71.1**	**71.2**	**69.9**	**70.3**	**86.6**	**74.8**
	EasyNet	MM23	✓	SNN	4	62.8	65.1	60.6	64.0	63.5	60.1	61.8	63.4	62.2	66.2	63.8
	3D-ADNAS	AAAI25	✓	SNN	4	**66.5**	69.7	64.2	66.8	**64.0**	68.3	64.5	64.9	**73.1**	72.6	70.3
	SpikingMAD	Ours	✓	SNN	4	66.2	**84.7**	**71.0**	**74.9**	60.8	**72.3**	**71.2**	**71.4**	70.5	**88.1**	**75.3**

	Method	Year	Spike	Types	Time Step	konb cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
RGB	Cflow	WACV22	✗	ANN	1	63.7	83.3	82.8	61.5	84.6	40.2	35.4	34.3	57.5	64.5
	SimpleNet	CVPR23	✗	ANN	1	67.2	79.0	85.7	93.9	82.3	40.2	17.6	34.3	62.7	65.9
	EasyNet	MM23	✓	SNN	1	57.9	64.1	60.2	62.7	59.5	64.4	63.5	58.4	57.7	60.8
	3D-ADNAS	AAAI25	✓	SNN	1	63.8	68.6	63.7	69.4	68.8	65.0	64.4	**68.5**	66.8	66.7
	SpikingMAD	Ours	✓	SNN	1	**68.6**	**70.3**	**76.8**	**76.1**	**81.1**	**72.3**	**69.7**	66.2	**75.3**	**72.7**
	EasyNet	MM23	✓	SNN	4	60.4	65.8	61.9	64.7	62.1	66.0	65.8	60.0	58.6	62.9
	3D-ADNAS	AAAI25	✓	SNN	4	64.3	69.1	63.9	70.3	69.2	65.8	65.1	**69.7**	67.5	67.5
	SpikingMAD	Ours	✓	SNN	4	**69.8**	**71.6**	**77.0**	**77.1**	**82.4**	**72.9**	**70.5**	66.5	**76.6**	**73.5**

and 55.1%. When trained with RGB and depth images at T=1, SpikingMAD surpasses CFM and EasyNet (SNN-based baselines) by 3.9% and 14.7%, respectively, with energy savings of 33.5% and 39.6%; When T=4, compared to EasyNet and 3D-ADNAS (ANN-based baselines), SpikingMAD achieves energy savings of 97.7% (49.56 vs 2142.2) and 98.5% (49.56 vs 3232.0) while maintaining comparable performance. These experimental and visualization results (as shown in Figure 7) demonstrate that the spike-driven paradigm plays a crucial role in advancing MAD.

Table 3: I-AUROC on Real-IAD D3 dataset across time steps under unimodal settings (point cloud).

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
3D point cloud	PointMAE	ECCV22	✗	ANN	1	76.3	72.5	95.8	29.1	84.9	100.0	63.4	42.2	30.9	60.7	64.4
	EasyNet	MM23	✓	SNN	1	56.6	50.3	55.9	53.2	52.7	51.1	49.7	54.5	50.2	53.2	55.7
	3D-ADNAS	AAAI25	✓	SNN	1	61.6	52.5	63.1	59.4	58.7	56.9	62.4	63.0	59.8	61.8	61.2
	SpikingMAD	Ours	✓	SNN	1	**78.0**	**57.1**	**71.6**	**68.7**	**77.2**	**86.4**	**73.3**	**69.4**	**60.5**	**72.9**	**68.3**
	EasyNet	MM23	✓	SNN	4	56.8	51.2	57.1	53.8	53.0	52.1	50.8	54.8	50.2	53.7	56.4
	3D-ADNAS	AAAI25	✓	SNN	4	62.0	52.8	64.2	60.7	59.4	57.7	63.1	63.6	61.1	62.2	62.0
	SpikingMAD	Ours	✓	SNN	4	**78.7**	**59.0**	**73.6**	**70.5**	**78.1**	**87.8**	**75.2**	**70.8**	**64.4**	**74.6**	**72.7**

	Method	Year	Spike	Types	Time Step	knob cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
3D point cloud	PointMAE	ECCV22	✗	ANN	1	65.6	76.9	36.1	34.8	76.3	71.7	43.3	86.9	77.1	64.4
	EasyNet	MM23	✓	SNN	1	52.4	52.1	56.6	55.5	51.4	56.2	52.9	53.7	50.8	53.2
	3D-ADNAS	AAAI25	✓	SNN	1	57.7	56.5	60.0	64.1	62.8	**62.5**	58.2	61.1	59.9	60.2
	SpikingMAD	Ours	✓	SNN	1	**73.1**	**71.6**	**66.5**	**78.8**	**72.0**	**62.5**	**69.8**	**71.3**	**60.2**	**70.5**
	EasyNet	MM23	✓	SNN	4	53.1	53.9	58.5	56.6	52.9	58.6	54.2	55.7	52.4	54.3
	3D-ADNAS	AAAI25	✓	SNN	4	58.4	57.1	61.4	64.2	63.3	63.0	58.9	61.3	60.3	60.8
	SpikingMAD	Ours	✓	SNN	4	**73.5**	**72.2**	**68.3**	**79.1**	**73.8**	**65.0**	**70.9**	**73.1**	**64.2**	**72.3**

Table 4: I-AUROC on Real-IAD D3 dataset across time steps under multimodal settings.

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
RGB + 3D point cloud	PatchCore	CVPR22	✗	ANN	1	92.6	52.3	87.0	71.3	72.0	94.7	91.3	76.9	73.6	91.9	68.9
	AST	WACV23	✗	ANN	1	86.0	89.9	91.4	48.5	99.5	100.0	89.4	59.5	59.7	66.0	56.5
	M3DM	CVPR23	✗	ANN	1	98.1	58.0	92.0	74.9	71.5	98.3	96.5	78.0	77.0	98.2	71.7
	EasyNet	MM23	✓	SNN	1	61.2	63.1	59.6	62.3	61.7	59.5	60.1	62.5	61.7	64.8	62.8
	3D-ADNAS	AAAI25	✓	SNN	1	67.5	70.4	65.5	67.6	65.9	68.7	63.8	64.9	**75.9**	74.0	71.3
	SpikingMAD	Ours	✓	SNN	1	**75.9**	**88.1**	**73.4**	**76.4**	**73.8**	**80.3**	**75.5**	**71.1**	70.4	**85.7**	**75.6**
	EasyNet	MM23	✓	SNN	4	63.7	65.4	61.0	65.2	64.3	60.9	62.8	64.1	62.0	67.2	63.5
	3D-ADNAS	AAAI25	✓	SNN	4	68.1	71.0	66.4	68.2	65.9	69.3	65.0	65.7	**76.2**	74.4	72.1
	SpikingMAD	Ours	✓	SNN	4	**77.3**	**89.6**	**74.8**	**78.7**	**74.8**	**81.2**	**76.2**	**72.1**	72.7	**87.5**	**76.1**

	Method	Year	Spike	Types	Time Step	knob cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
RGB + 3D point cloud	PatchCore	CVPR22	✗	ANN	1	90.3	91.1	66.2	54.0	82.2	94.8	98.1	92.1	82.7	81.2
	AST	WACV23	✗	ANN	1	91.9	84.2	84.7	47.1	80.4	76.6	56.4	63.5	95.1	69.3
	M3DM	CVPR23	✗	ANN	1	92.5	91.7	68.1	53.0	86.3	97.5	99.6	94.4	85.6	84.1
	EasyNet	MM23	✓	SNN	1	57.9	64.6	61.5	63.9	60.7	64.8	64.8	59.2	58.3	61.8
	3D-ADNAS	AAAI25	✓	SNN	1	65.1	68.8	64.9	70.6	69.4	67.1	65.5	70.9	69.5	68.4
	SpikingMAD	Ours	✓	SNN	1	**72.9**	**72.1**	**74.0**	**83.5**	**82.3**	**74.3**	**75.4**	**76.0**	**70.1**	**76.3**
	EasyNet	MM23	✓	SNN	4	59.7	65.8	62.2	64.8	63.4	66.5	67.0	61.9	59.4	63.5
	3D-ADNAS	AAAI25	✓	SNN	4	67.4	69.1	65.4	72.0	68.8	68.6	66.7	73.1	69.8	69.2
	SpikingMAD	Ours	✓	SNN	4	**74.0**	**73.8**	**74.9**	**86.1**	**83.7**	**75.6**	**75.5**	**76.7**	**73.8**	**77.8**

Evaluations on Real IAD D3. Tables 2, 3, and 4 report the I-AUROC scores of SpikingMAD on the Real-IAD D3 dataset. It can be observed that SpikingMAD achieves superior performance in both unimodal and multimodal settings. Specifically, in the multimodal setting, SpikingMAD consistently outperforms SNN-based baselines; its I-AUROC exceeds that of EasyNet and 3D-ADNAS by 14.3% and 8.6% when T=4, respectively, and even surpasses the ANN-based baseline (AST) by 7.0% when T=1. These results again validate the effectiveness of the proposed spike-driven MAD architecture. Notably, more experimental results on the Real-IAD D3 dataset are provided in Appendix H.

Table 5: Experimental results on Eyecandies.

Methods	1 Time Step (SNN) I-AUROC	1 Time Step (SNN) P-AUROC	4 Time Step (SNN) I-AUROC	4 Time Step (SNN) P-AUROC
EasyNet	62.1	62.9	62.5	63.5
CFM	60.7	62.8	62.8	65.1
3D-ADNAS	66.2	71.3	66.9	71.9
SpikingMAD	**72.0**	**76.3**	**73.1**	**77.6**

Evaluations on Eyecandies. Table 5 presents a performance comparison between SpikingMAD and other SNN-based methods on the Eyecandies dataset under the multimodal setting. It is evident that SpikingMAD consistently outperforms the baselines; for example, SpikingMAD surpasses CFM by 10.3% and 12.5% in terms of I-AUROC and P-AUROC, and 3D-ADNAS by 6.2% and 5.7% when T=4, further demonstrating the advantages of the proposed method. Due to page limitations, please refer to Appendix G for additional experiments and result analyses.

Table 6: I-AUROC, P-AUROC, and AUPRO scores on MVTec 3D-AD under the few-shot setting.

	Method	Year	5-shot			10-shot			50-shot			full-shot		
			I-AUROC	P-AUROC	AUPRO	I-AUROC	P-AUROC	AUPRO	I-AUROC	P-AUROC	AUPRO	I-AUROC	P-AUROC	AUPRO
R&D	EasyNet	MM23	57.6	63.3	41.4	61.4	67.9	45.9	68.2	74.8	56.2	74.9	81.0	61.3
R&D	3D-ADNAS	AAAI25	64.7	72.4	43.9	69.7	77.2	48.4	75.5	82.7	60.1	82.2	88.4	67.0
R&D	SpikingMAD	Ours	**75.8**	**76.8**	**50.9**	**77.9**	**79.6**	**55.5**	**85.5**	**86.4**	**67.3**	**89.6**	**90.6**	**72.5**

Experimental results on few-shot settings. To further validate the effectiveness of SpikingMAD, we report its experimental results under few-shot settings (Li et al., 2025b; Lin et al., 2026; Zhai et al., 2026; Dong et al., 2026). As shown in Table 6, we can observe that SpikingMAD achieves consistent performance gains over baselines. Specifically, under 5-shot, 10-shot, and 50-shot settings, SpikingMAD outperforms 3D-ADNAS by 11.1%, 8.2%, and 10.0%, and surpasses EasyNet by 18.2%, 16.5%, and 17.3% in I-AUROC when the time step is set to 1, respectively. Similarly, SpikingMAD consistently outperforms both 3D-ADNAS and EasyNet in P-AUROC and AUPRO. These results demonstrate that the spike-driven paradigm is indeed a promising approach for advancing MAD.

Table 7: Comparison results in memory usage, frame rate, Power, and I-AUROC.

Methods	MVTec 3D-AD			
	Memory	Frame Rate	Power	I-AUROC
BTF	381.06	3.197	- -	86.5
M3DM	6526.12	0.514	- -	94.5
CFM	437.91	21.755	1276.0 mJ	**95.4**
3D-ADNAS	269.33	24.693	3232.0 mJ	95.1
SpikingMAD	**128.76**	**27.465**	**16.63 mJ**	89.6

Table 8: The impact of fusion operations.

Fusion types	MVTec 3D-AD		
	I-AUROC	P-AUROC	AUPRO
Addition	87.8	89.2	71.8
ConcatFC	86.2	87.0	69.3
SDA	86.9	88.5	70.4
SHCA (Ours)	**89.6**	**90.6**	**72.5**

Ablation Study of Multimodal Fusion Operations. To reveal the effect of SHCA on the performance of SpikingMAD, we conducted a series of ablation experiments on MVTec 3D-AD under the multimodal setting, comparing SHCA with other fusion operations, including SDA, Addition, and ConcatFC. As shown in Table 8, SHCA consistently achieves superior performance. These results further validate the effectiveness of the proposed SHCA strategy in enhancing multimodal feature fusion in MAD tasks.

Computational Efficiency. To further demonstrate the advantages of SpikingMAD, we report its memory usage, power consumption, I-AUROC, and frame rate results on the MVTec 3D-AD dataset. As shown in Table 7, compared to ANN-based methods, SpikingMAD achieves comparable performance while reducing energy consumption by approximately 194× (3232.0 /16.63 mJ) when $T = 1$; Moreover, it significantly shrinks memory usage and substantially increases the frame rate. These results again reveal the significant advantages of our approach for the MAD task.

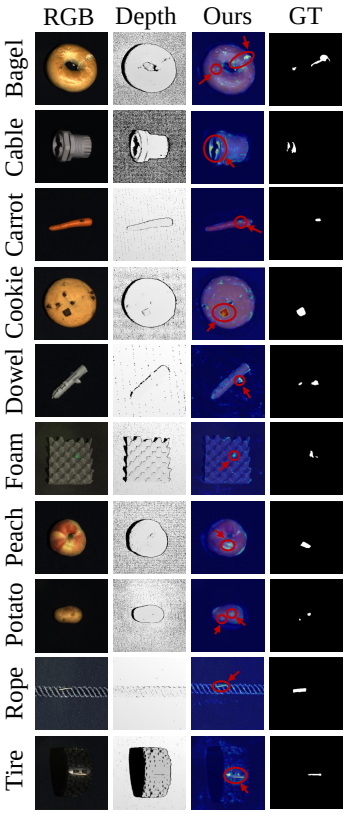

Figure 7: Visualizations results.

5 Conclusion

In this paper, we propose SpikingMAD, a spike-driven multimodal anomaly detection framework that addresses the high energy consumption of existing ANN-based paradigms arising from massive dense floating-point MAC operations. Specifically, SpikingMAD leverages SMRN to reconstruct anomalous images into normal ones by exploiting the sparse computation and multiplication-free characteristics of SNNs. Meanwhile, we introduce the spike-driven Hamming cross-attention mechanism to effectively capture cross-modal feature similarities and facilitate multimodal feature fusion. Extensive experiments on multiple datasets demonstrate the effectiveness of our proposed method. We hope this work will inspire further explorations from the spike-driven perspective, opening up new possibilities for energy-efficient MAD.

AI USE STATEMENT

In this work, we only used generative AI tools to assist with grammar checking and language polishing. We have not used generative AI tools for retrieval and discovery (e.g., finding related work), for research ideation and execution, for drafting sections of the paper, for generating synthetic datasets, or for proving mathematical claims.

In summary, we only used generative AI tools to assist with grammar checking and language polishing. We have reviewed all AI-assisted work. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We release the anonymous code link at `https://anonymous.4open.science/r/SpikingMAD`, which includes all code for data extraction, model training, and evaluation for the proposed SpikingMAD model. The implementation details are provided in Appendix E, and theoretical analyses of the spike-driven multimodal reconstruction network and spike-driven Hamming cross-attention are provided in Appendices C and D.

REFERENCES

Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In *Proceedings of the 17th VISIGRAPP*, 2022.

Luca Bonfiglioli, Marco Toschi, Davide Silvestri, Nicola Fioraio, and Daniele De Gregorio. The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In *Proceedings of the ACCV*, pp. 3586–3602, 2022.

Junhao Cai, Junyu Chen, Deyu Zeng, Junhao Pang, Qiwei Liang, Xiaopin Zhong, and Zongze Wu. Cmds-ad: Cross-modal dual-stream decoupling for few-shot anomaly detection. In *ECCV*, pp. 431–448, 2026.

Yunkang Cao, Jiangning Zhang, Luca Frittoli, Yuqi Cheng, Weiming Shen, and Giacomo Boracchi. Adaclip: Adapting clip with hybrid learnable prompts for zero-shot anomaly detection. In *ECCV*, pp. 55–72, 2024.

Qiyu Chen, Zhen Qu, Wei Luo, Haiming Yao, Yunkang Cao, Yuxin Jiang, Yinan Duan, Huiyuan Luo, Chengkan Lv, and Zhengtao Zhang. Cops: Conditional prompt synthesis for zero-shot anomaly detection. In *CVPR*, pp. 8554–8563, 2026a.

Ruitao Chen, Guoyang Xie, Jiaqi Liu, Jinbao Wang, Ziqi Luo, Jinfan Wang, and Feng Zheng. Easynet: An easy network for 3d industrial anomaly detection. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 7038–7046, 2023.

Xintao Chen, Xiaohao Xu, Bozhong Zheng, Yun Liu, and Yingna Wu. Unsupervised multi-view visual anomaly detection via progressive homography-guided alignment. In *AAAI*, volume 40, pp. 3065–3073, 2026b.

Yuqi Cheng, Yunkang Cao, Haiming Yao, Wei Luo, Cheng Jiang, Hui Zhang, and Weiming Shen. A comprehensive survey for real-world industrial surface defect detection: Challenges, approaches, and prospects. *Journal of Manufacturing Systems*, 84:152–172, 2026a.

Yuqi Cheng, Yihan Sun, Hui Zhang, Weiming Shen, and Yunkang Cao. Towards high-resolution 3d anomaly detection: A scalable dataset and real-time framework for subtle industrial defects. In *AAAI*, volume 40, pp. 3327–3334, 2026b.

Sayeed Shafayet Chowdhury, Chankyu Lee, and Kaushik Roy. Towards understanding the effect of leak in spiking neural networks. *Neurocomputing*, 464:83–94, 2021.

Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *ICML*, pp. 6185–6194, 2023.

Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, and Luigi Di Stefano. Multimodal industrial anomaly detection by crossmodal feature mapping. In *CVPR*, pp. 17234–17243, 2024.

Zhewei Dai, Shilei Zeng, Haotian Liu, Xurui Li, Feng Xue, and Yu Zhou. Seas: few-shot industrial anomaly image generation with separation and sharing fine-tuning. In *ICCV*, pp. 23135–23144, 2025.

Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *International conference on pattern recognition*, pp. 475–489, 2021.

Le Dong, Qinzhong Tan, Chunlei Li, Jingliang Hu, Yilei Shi, Weisheng Dong, Xiaoxiang Zhu, and Lichao Mou. Dual distillation for few-shot anomaly detection. In *International Conference on Learning Representations*, pp. 135274–135289, 2026.

Bin-Bin Gao, Yue Zhou, Jiangtao Yan, Yuezhi Cai, Weixi Zhang, Meng Wang, Jun Liu, Yong Liu, Lei Wang, and Chengjie Wang. Adaptclip: Adapting clip for universal visual anomaly detection. In *AAAI*, volume 40, pp. 4095–4103, 2026.

Zhihao Gu, Jiangning Zhang, Liang Liu, Xu Chen, Jinlong Peng, Zhenye Gan, Guannan Jiang, Annan Shu, Yabiao Wang, and Lizhuang Ma. Rethinking reverse distillation for multi-modal anomaly detection. In *Proceedings of the AAAI*, volume 38, pp. 8445–8453, 2024.

Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, and Hongen Liao. Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In *CVPR*, pp. 20405–20415, 2025a.

Yuan Guo, Wanqi Zhang, and Xu Wang. Remove the ambiguity: Few-shot multimodal anomaly detection using crossmodal feature replacers. In *ICML*, 2026.

Yufei Guo, Xiaode Liu, Yuanpei Chen, Weihang Peng, Yuhan Zhang, and Zhe Ma. Spiking transformer: Introducing accurate addition-only spiking self-attention for transformer. In *CVPR*, pp. 24398–24408, 2025b.

Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *NIPS*, pp. 71162–71187, 2024a.

Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *AAAI*, volume 38, pp. 8472–8480, 2024b.

Allan L Hodgkin and Andrew F Huxley. Currents carried by sodium and potassium ions through the membrane of the giant axon of loligo. *The Journal of physiology*, 116(4):449, 1952.

Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *CVPR*, pp. 2968–2977, 2023.

Teng Hu, Jiangning Zhang, Ran Yi, Yuzhen Du, Xu Chen, Liang Liu, Yabiao Wang, and Chengjie Wang. Anomalydiffusion: Few-shot anomaly image generation with diffusion model. In *AAAI*, pp. 8526–8534, 2024.

Wei Huang, Qinying Gu, and Nanyang Ye. Decision spikeformer: Spike-driven transformer for decision making. In *CVPR*, pp. 19241–19250, 2025.

Laurent Jacques, Jason N Laska, Petros T Boufounos, and Richard G Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. *IEEE transactions on information theory*, 59(4):2082–2102, 2013.

William B Johnson and Joram Lindenstrauss. Extensions of lipschitz mappings into a hilbert space. *Contemporary mathematics*, 26(189-206):1, 1984.

Soopil Kim, Sion An, Philip Chikontwe, Myeongkyun Kang, Ehsan Adeli, Kilian M Pohl, and Sang Hyun Park. Few shot part segmentation reveals compositional logic for industrial anomaly detection. In *AAAI*, pp. 8591–8599, 2024.

Souvik Kundu, Massoud Pedram, and Peter A Beerel. Hire-snn: Harnessing the inherent robustness of energy-efficient deep spiking neural networks by training with crafted input noise. In *ICCV*, pp. 5189–5198, 2021.

Yujin Lee, Sewon Kim, Daeun Moon, Seoyoon Jang, and Hyunsoo Yoon. Bidirectional multimodal prompt learning with scale-aware training for few-shot multi-class anomaly detection. In *CVPR*, pp. 35577–35586, 2026.

Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, and Yingna Wu. Towards scalable 3d anomaly detection and localization: A benchmark via 3d anomaly synthesis and a self-supervised learning network. In *CVPR*, pp. 22207–22216, 2024.

Xiaofan Li, Xin Tan, Zhuo Chen, Zhizhong Zhang, Ruixin Zhang, Rizen Guo, Guanna Jiang, Yulong Chen, Yanyun Qu, Lizhuang Ma, et al. One-for-more: Continual diffusion model for anomaly detection. In *CVPR*, pp. 4766–4775, 2025a.

Yiting Li, Fayao Liu, Jingyi Liao, Sichao Tian, Chuan-Sheng Foo, and Xulei Yang. Find: Few-shot anomaly inspection with normal-only multi-modal data. In *ICCV*, pp. 1–10, 2025b.

Yiting Li, Xulei Yang, Jingyi Liao, Jing Zhang, and Fayao Liu. Gpflow: Gaussian prototype probability flow for unsupervised multi-modal anomaly detection. In *CVPR*, pp. 43103–43112, 2026a.

Yiting Li, Jing Zhang, Sichao Tian, Jingyi Liao, Fayao Liu, et al. Pirn: Prototypical-based intra-modal reconstruction with normality communication for multi-modal anomaly detection. In *International Conference on Learning Representations*, volume 2026, pp. 152723–152741, 2026b.

Ziqi Li, Tao Gao, Yisheng An, Ting Chen, Jing Zhang, Yuanbo Wen, Mengkun Liu, and Qianxi Zhang. Brain-inspired spiking neural networks for energy-efficient object detection. In *CVPR*, pp. 3552–3562, 2025c.

Hanzhe Liang, Jie Zhang, Tao Dai, Linlin Shen, Jinbao Wang, and Can Gao. Taming anomalies with down-up sampling networks: Group center preserving reconstruction for 3d anomaly detection. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7133–7141, 2025.

Hanzhe Liang, Chenxi Hu, Yejin Tang, Linlin Shen, Jinbao Wang, and Can Gao. Mff-m3ad: A unified reconstruction method with multi-scale feature fusion for multi-category 3d anomaly detection. *Neural Networks*, pp. 109131, 2026.

Kailun Liao, Jianfeng Yang, Tao Tao, Wenfei Wu, Jiaming Jiang, and Jinsheng Xiao. Multi-prototype compactness and boundary-aware synthesis for unsupervised anomaly detection. In *CVPR*, pp. 28524–28533, 2026.

Yuxuan Lin, Hanjing Yan, Xuan Tong, Yang Chang, Huanzhen Wang, Ziheng Zhou, Shuyong Gao, Yan Wang, and Wenqiang Zhang. Commonality in few: few-shot multimodal anomaly detection via hypergraph-enhanced memory. In *Proceedings of the AAAI*, volume 40, pp. 7015–7023, 2026.

Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pp. 20402–20411, 2023.

Kaifang Long, Guoyang Xie, Lianbo Ma, Jiaqi Liu, and Zhichao Lu. Revisiting multimodal fusion for 3d anomaly detection from an architectural perspective. In *AAAI*, pp. 12273–12281, 2025.

Kai Mao, Ping Wei, Yiyang Lian, Yangyang Wang, and Nanning Zheng. Beyond single-modal boundary: Cross-modal anomaly detection through visual prototype and harmonization. In *CVPR*, pp. 9964–9973, 2025.

Bingchen Miao, Wenqiao Zhang, Juncheng Li, Wangyu Wu, Siliang Tang, Zhaocheng Li, Haochen Shi, Jun Xiao, and Yueting Zhuang. Robust modality-incomplete anomaly detection: A modality-instructive framework with benchmark. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pp. 7317–7326, 2025.

Yatian Pang, Eng Hock Francis Tay, Li Yuan, and Zhenghua Chen. Masked autoencoders for 3d point cloud self-supervised learning. *World Scientific Annual Review of Artificial Intelligence*, 1: 2440001, 2023.

Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pp. 14298–14308, 2022.

Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Fully convolutional cross-scale-flows for image-based defect detection. In *WACV*, pp. 1829–1838, 2022.

Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *WACV*, pp. 2592–2602, 2023.

Sirojbek Safarov, Jaewoo Park, Yoon Gyo Jung, Kuan-Chuan Peng, Wonchul Kim, Seongdeok Bang, and Octavia Camps. Memory-distilled selection for noise-robust anomaly detection. *arXiv preprint arXiv:2605.26676*, 2026.

Hannah M Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization. In *ECCV*, pp. 474–489, 2022.

Xinyu Shi, Zecheng Hao, and Zhaofei Yu. Spikingresformer: Bridging resnet and vision transformer in spiking neural networks. In *CVPR*, pp. 5610–5619, 2024.

Qiaoyi Su, Yuhong Chou, Yifan Hu, Jianing Li, Shijie Mei, Ziyang Zhang, and Guoqi Li. Deep directly-trained spiking neural networks for object detection. In *CVPR*, pp. 6532–6542, 2023.

Jiaqi Tang, Hao Lu, Xiaogang Xu, Ruizheng Wu, Sixing Hu, Tong Zhang, Tsz Wa Cheng, Ming Ge, Ying-Cong Chen, and Fugee Tsung. An incremental unified framework for small defect inspection. In *ECCV*, pp. 307–324, 2024.

Chengyu Tao, Xuanming Cao, and Juan Du. G2sf: Geometry-guided score fusion for multimodal industrial anomaly detection. In *ICCV*, pp. 20551–20560, 2025.

Yuanpeng Tu, Boshen Zhang, Liang Liu, Yuxi Li, Jiangning Zhang, Yabiao Wang, Chengjie Wang, and Cairong Zhao. Self-supervised feature adaptation for 3d industrial anomaly detection. In *ECCV*, pp. 75–91, 2024.

Chengjie Wang, Haokun Zhu, Jinlong Peng, Yue Wang, Ran Yi, Yunsheng Wu, Lizhuang Ma, and Jiangning Zhang. M3dm-nr: Rgb-3d noisy-resistant industrial anomaly detection via multimodal denoising. *IEEE TPAMI*, 2025a.

Fuyun Wang, Tong Zhang, Yuanzhi Wang, Yide Qiu, Xin Liu, Xu Guo, and Zhen Cui. Distribution prototype diffusion learning for open-set supervised anomaly detection. In *CVPR*, pp. 20416–20426, 2025b.

Shuai Wang, Malu Zhang, Dehao Zhang, Ammar Belatreche, Yichen Xiao, Yu Liang, Yimeng Shan, Qian Sun, Enqi Zhang, and Yang Yang. Spiking vision transformer with saccadic attention. In *International Conference on Learning Representations*, volume 2025, pp. 24148–24169, 2025c.

Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, pp. 8032–8041, 2023.

Yichen Xiao, Shuai Wang, Dehao Zhang, Wenjie Wei, Yimeng Shan, Xiaoli Liu, Yulin Jiang, and Malu Zhang. Rethinking spiking self-attention mechanism: Implementing α-xnor similarity calculation in spiking transformers. In *CVPR*, pp. 5444–5454, 2025.

Ke Xu, Xinle Wang, Yanning Hou, Xueliang Ma, Juan Xie, and Jianfeng Qiu. Cogeoad: Hierarchical color-geometric fusion with multi-view attention for zero-shot 3d anomaly detection. *arXiv preprint arXiv:2606.25273*, 2026.

Qi Xu, Jie Deng, Jiangrong Shen, Biwu Chen, Huajin Tang, and Gang Pan. Hybrid spiking vision transformer for object detection with event cameras. In *ICML*, volume 267, 2025.

Man Yao, Guangshe Zhao, Hengyu Zhang, Yifan Hu, Lei Deng, Yonghong Tian, Bo Xu, and Guoqi Li. Attention spiking neural networks. *IEEE TPAMI*, 45(8):9393–9410, 2023.

Man Yao, JiaKui Hu, Tianxiang Hu, Yifan Xu, Zhaokun Zhou, Yonghong Tian, Bo Xu, and Guoqi Li. Spike-driven transformer v2: Meta spiking neural network architecture inspiring the design of next-generation neuromorphic chips. In *International conference on learning representations*, volume 2024, pp. 52885–52907, 2024.

Man Yao, Xuerui Qiu, Tianxiang Hu, Jiakui Hu, Yuhong Chou, Keyu Tian, Jianxing Liao, Luziwei Leng, Bo Xu, and Guoqi Li. Scaling spike-driven transformer with efficient spike firing approximation training. *IEEE TPAMI*, 47(4):2973–2990, 2025.

Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, and Kaizhu Huang. Po3ad: Predicting point offsets toward better 3d point cloud anomaly detection. In *CVPR*, pp. 1353–1362, 2025.

Xinyang Yi, Constantine Caramanis, and Eric Price. Binary embedding: Fundamental limits and fast algorithm. In *ICML*, volume 37, pp. 2162–2170, 2015.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dræm–a discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pp. 8310–8319, 2021.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Dsr–a dual subspace re-projection network for surface anomaly detection. In *ECCV*, pp. 539–554, 2022.

Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Cheating depth: Enhancing 3d surface anomaly detection via depth simulation. In *WACV*, pp. 2153–2161, 2024.

Guangyao Zhai, Yue Zhou, Xinyan Deng, Lars Heckler-Kram, Nassir Navab, and Benjamin Busam. Foundation visual encoders are secretly few-shot anomaly detectors. In *International Conference on Learning Representations*, volume 2026, pp. 97353–97376, 2026.

Jiangning Zhang, Chengjie Wang, Xiangtai Li, Guanzhong Tian, Zhucun Xue, Yong Liu, Guansong Pang, and Dacheng Tao. Learning feature inversion for multi-class anomaly detection under general-purpose coco-ad benchmark: J. zhang et al. *International Journal of Computer Vision*, 134(5):235, 2026.

Zhe Zhang, Mingxiu Cai, Hanxiao Wang, Gaochang Wu, Tianyou Chai, and Xiatian Zhu. Costfilter-ad: Enhancing anomaly detection through matching cost filtering. In *ICML*, volume 267, 2025.

Yuan Zhao, Youwei Pang, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Huchuan Lu, and Xiaoqi Zhao. Unimmad: Unified multi-modal and multi-class anomaly detection via moe-driven feature decompression. In *CVPR*, pp. 28502–28511, 2026a.

Yuan Zhao, Xiaoqin Zhang, Huchuan Lu, and Lihe Zhang. Complementary prototype mapping for efficient multimodal anomaly detection. In *CVPR*, pp. 14178–14187, 2026b.

Chenlin Zhou, Liutao Yu, Zhaokun Zhou, Zhengyu Ma, Han Zhang, Huihui Zhou, and Yonghong Tian. Spikingformer: Spike-driven residual learning for transformer-based spiking neural network. *arXiv e-prints*, pp. arXiv–2304, 2023a.

Zhaokun Zhou, Yuesheng Zhu, Chao He, Yaowei Wang, Shuicheng Yan, Yonghong Tian, and Li Yuan. Spikformer: When spiking neural network meets transformer. In *The Eleventh International Conference on Learning Representations*, 2023b.

Wenbing Zhu, Lidong Wang, Ziqing Zhou, Chengjie Wang, Yurui Pan, Ruoyi Zhang, Zhuhao Chen, Linjie Cheng, Bin-Bin Gao, Jiangning Zhang, et al. Real-iad d 3: a real-world 2d/pseudo-3d/3d dataset for industrial anomaly detection. In *CVPR*, pp. 15214–15223, 2025.

Wenbing Zhu, Chengjie Wang, Bin-Bin Gao, Jiangning Zhang, Guannan Jiang, Jie Hu, Zhenye Gan, Lidong Wang, Ziqing Zhou, Jianghui Zhang, et al. Real-iad variety: Pushing industrial anomaly detection dataset to a modern era. *Pattern Recognition*, pp. 113354, 2026.

Shihao Zou, Qingfeng Li, Wei Ji, Jingjing Li, Yongkui Yang, Guoqi Li, and Chao Dong. Spikevideoformer: An efficient spike-driven video transformer with hamming attention and O(T) complexity. In *ICML*, volume 267, 2025.

A Overview of the Appendix

This appendix provides supplementary material supporting the main manuscript and is organized as follows:

- **Appendix A**: Overview of the appendix
- **Appendix B**: Energy consumption calculation and SOPs.
- **Appendix C**: Theoretical analysis of the spike-driven characteristics of SMRN and SHCA.
- **Appendix D**: Theoretical analysis of Proposition 1 and the derivation of SHCA.
- **Appendix E**: Implementation details and evaluation metrics.
- **Appendix F**: P-AUROC and AUPRO scores on the MVTec 3D-AD dataset.
- **Appendix G**: I-AUROC and P-AUROC scores on the Eyecandies dataset.
- **Appendix H**: P-AUROC scores on the Real-IAD D3 dataset.
- **Appendix I**: Detailed results of the few-shot setting on the MVTec 3D-AD dataset.

B Energy consumption calculation and SOPs

FLOPs denote the number of floating-point operations, i.e., multiply-accumulate (MAC) operations in ANNs, whereas SOPs represent the number of spike-based accumulate (AC) operations in SNNs. Following Kundu et al. (2021); Yao et al. (2023); Zhou et al. (2023b); Xu et al. (2025); Yao et al. (2024); Li et al. (2025c), a single MAC operation theoretically consumes 4.6 pJ, while an AC operation consumes only 0.9 pJ. Based on these energy costs, we estimate the energy consumption of ANN-based MAD methods from their FLOPs as follows:

$$Power(\mathcal{M}_{ann}) = 4.6pJ * FLOPs(\mathcal{M}_{ann}), \tag{12}$$

where $FLOPs(\mathcal{M}_{ann})$ represents the FLOPs of the ANN-based MAD method (i.e., $\mathcal{M}_{ann}$). For SNN-based MAD methods (i.e., $\mathcal{M}_{snn}$), energy consumption is estimated based on synaptic operations (SOPs), which we first calculate as follows:

$$SOPs(\mathcal{M}_{snn}) = FLOPs(\mathcal{M}_{snn}) * \Upsilon * T, \tag{13}$$

where T denotes the time steps of spiking neurons, and Υ denotes the firing rate of the input spike to the SNN block. $FLOPs(\mathcal{M}_{snn})$ indicates the FLOPs of the spiking block. However, in SNN-based

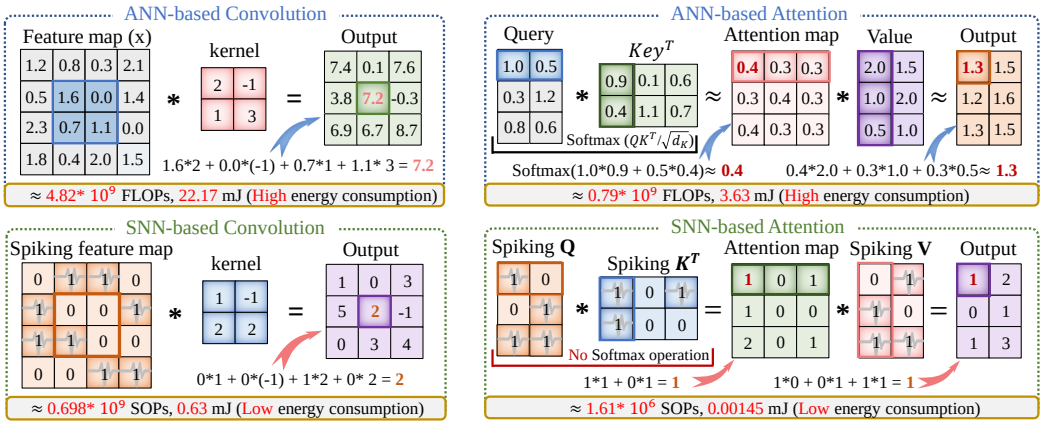

Figure 8: Example of energy consumption estimation for SNN- and ANN-based MAD methods, where SOPs indicate theoretical synaptic operations and FLOPs denote floating-point operations.

MAD methods, the first spiking block typically involves MAC operations, which are necessary to effectively balance detection performance and energy efficiency. Accordingly, the overall energy consumption of an SNN-based MAD method can be estimated by accounting for the MAC operations in the first block (i.e., $\mathcal{M}_{snn}^{mac}$) and the spike-driven AC operations in the subsequent blocks (i.e., $\mathcal{M}_{snn}^{ac}$), formulated as:

$$Power(\mathcal{M}_{snn}) = 4.6pJ * FLOPs(\mathcal{M}_{snn}^{mac}) + 0.9pJ * SOPs(\mathcal{M}_{snn}^{ac}), \tag{14}$$

Furthermore, to provide an intuitive understanding of the energy consumption estimation for SNN- and ANN-based MAD methods, we present an illustrative example in Figure 8.

C Theoretical analysis of the spike-driven characteristics of SMRN and SHCA

Definition 1. *The network is spike-driven if the input currents of all neurons satisfy the form:*

$$I_m[t] = \sum_n \omega_{m,n} * s_n[t] = \sum_{n, s_n[t] \neq 0} \omega_{m,n},$$

where $I_m[t]$ denotes the input current received by the m-th postsynaptic neuron at time step t, $\omega_{m,n}$ denotes the synaptic weight from n-th to m-th neuron, and $s_n[t] \in \{0, 1\}$ denotes the binary spike emitted by the n-th presynaptic neuron.

As stated in ***Definition* 1**, the spike-driven characteristic requires that input current accumulation is exclusively triggered by sparse spikes emitted by presynaptic neurons. Clearly, both commonly used convolutional and linear layers satisfy this definition. Specifically, for the linear layer $\mathbb{Y} = \omega * \chi + b$, ω denotes the weight, b denotes the bias, χ denotes the input features, and $\mathbb{Y}$ denotes the output features. In SMRN, since the linear layer input $\chi \in \{0, 1\}$, the linear layer involves no floating-point MAC operations and thus satisfies the spike-driven property defined in ***Definition* 1**. Similarly, a convolutional layer is essentially a sparse linear transformation that differs from a linear layer only in its weight matrix; consequently, convolution operations likewise satisfy the spike-driven characteristic. Therefore, SMRN naturally exhibits the spike-driven property.

Regarding the spike-driven characteristics of SHCA, we first analyze the properties of spike-driven dot-product attention (SDA). Given spiking matrices Q_R and $K_D \in \{0, 1\}$, the similarity score is calculated as $I = Q_R * W^T * (K_D)^T$. Accordingly, the input current ($I_{m,n}[t]$) at time step t between the spike value at u-th dimension at the m-th position of Q_R (i.e., $q_{m,u}[t]$) and the spike value at v-th dimension at the n-th position of K_D (i.e., $k_{n,v}[t]$) in SDA is given by:

$$I_{m,n}[t] = \sum_{u=1}^{z} \sum_{v=1}^{z} q_{m,u}[t] * \omega_{v,u} * k_{n,v}[t] = \sum_{(q_{m,u}[t] \wedge k_{n,v}[t]) \neq 0} \omega_{v,u}, \tag{15}$$

It is clear that SDA satisfies the spike-driven property stated in ***Definition* 1**. Since SHCA is a variant of SDA and satisfies the properties of no floating-point operations, no nonlinear complexity, and differentiability (see Appendix D), SHCA likewise conforms to the spike-driven property.

D Theoretical analysis of Proposition 1 and the derivation of SHCA

In this section, we conduct a theoretical analysis of ***Proposition* 1**. Since ***Proposition* 1** is a variant of the Johnson–Lindenstrauss Lemma (Johnson & Lindenstrauss, 1984; Yi et al., 2015; Jacques et al., 2013) for binary embeddings, we given the original Lemma: Let $\mathbb{X} = \{x_i\}_{i=1}^{\Lambda}$ be a set of Λ real-valued points. Then its one-bit quantization under the projection is given by:

$$b(\mathbb{X}) = sign(\mathbb{B} * \mathbb{X}), \tag{16}$$

where $b(\mathbb{X}) \in \{0, 1\}^{\mathcal{C}}$ is the binary embedding of $\mathbb{X} \in R^{\mathcal{N}}$ and the projection matrix $\mathbb{B} \in R^{\mathcal{N} * \mathcal{C}}$ has entries independently and identically sampled from a standard normal distribution. Given any $\varepsilon > 0$ and $\mathcal{C} > \frac{\log \Lambda}{\varepsilon^2}$, we have:

$$\mathbb{P}(|d_{Ham}(b_i, b_j) - d_{Cos}(x_i, x_j)| \leq \varepsilon) \geq 1 - 2e^{-\varepsilon^2 \mathcal{C}}, \tag{17}$$

where $d_{Cos}(x_i, x_j)$ denotes the cosine distance, which is formulated as:

$$d_{Cos}(x_i, x_j) = \frac{\arccos(x_i^T * x_j / \parallel x_i \parallel \parallel x_j \parallel)}{\pi}, \tag{18}$$

and the normalized Hamming distance (i.e., $d_{Ham}(b_i, b_j)$) is defined as:

$$d_{Ham}(b_i, b_j) = \frac{1}{\mathcal{C}_\eta} \sum_{\eta=1}^{\mathcal{C}} \mathbf{1}(b_i^\eta \neq b_j^\eta). \tag{19}$$

Based on the Johnson–Lindenstrauss lemma for binary embeddings described above, we derive a variant, as stated in ***Proposition* 1**. Specifically, By replacing x_i and x_j in $d_{Cos}(x_i, x_j)$ with X_R and X_D from the real-valued attention mechanism, and substituting b_i and b_j in $d_{Ham}(b_i, b_j)$ with Q_{X_R} and K_{X_D} obtained via the binary embeddings, we obtain:

$$|d_{Ham}(b_i, b_j) - d_{Cos}(x_i, x_j)| = |d_{Ham}(Q_{X_R}, K_{X_D}) - d_{Cos}(X_R, X_D)| \leq \varepsilon. \tag{20}$$

Next, based on the normalized Hamming similarity defined in ***Proposition* 1** (i.e., $f_{ham}(Q_{X_R}, K_{X_D}) = 1 - \frac{1}{\mathcal{C}} \sum_{i=1}^{\mathcal{C}} \mathbf{1}(Q_{X_R}^i \neq K_{X_D}^i)$) and the normalized Hamming distance (i.e., Eq.(19)), we have:

$$d_{Ham}(Q_{X_R}, Q_{X_D}) = 1 - f_{ham}(Q_{X_R}, K_{X_D}). \tag{21}$$

Then, given the continuous and monotonic function $\mathbb{G}[x] = \frac{\pi - \arccos(x)}{\pi}$, and $f_{dot}(X_R, X_D) = \frac{(X_R)^T * X_D}{\|X_R\| \|X_D\|}$ we have:

$$\begin{aligned} &1 - \mathbb{G}[f_{dot}(X_R, X_D)] = 1 - [1 - \frac{1}{\pi} \arccos(f_{dot}(X_R, X_D))] \\ &= \frac{1}{\pi} \arccos(f_{dot}(X_R, X_D)) = \frac{1}{\pi} \arccos(\frac{(X_R)^T * X_D}{\parallel X_R \parallel \parallel X_D \parallel}) = d_{Cos}(X_R, X_D). \end{aligned} \tag{22}$$

Therefore, substituting Eq.(21) and Eq.(22) into Eq.(20), we obtain:

$$\begin{aligned} |d_{Ham}(Q_{X_R}, K_{X_D}) - d_{Cos}(X_R, X_D)| &= |(1 - f_{ham}(Q_{X_R}, K_{X_D})) - (1 - \mathbb{G}[f_{dot}(X_R, X_D)])| \\ &= |f_{ham}(Q_{X_R}, K_{X_D}) - \mathbb{G}[f_{dot}(X_R, X_D)]| \leq \varepsilon. \end{aligned} \tag{23}$$

Finally, substituting Eq.(23) into Eq.(17), we obtain:

$$\mathbb{P}(|f_{ham}(Q_{X_R}, K_{X_D}) - \mathbb{G}[f_{dot}(X_R, X_D)]| \leq \varepsilon) \geq 1 - 2e^{-\varepsilon^2 \mathcal{C}}. \tag{24}$$

The Eq.(24) shows that $f_{ham}(\cdot)$ closely approximates $f_{dot}(\cdot)$ when the channel dimension $\mathcal{C}$ is sufficiently large, demonstrating that Hamming similarity is an effective manner to measure query-key similarity in spike-driven paradigms. However, directly using $f_{ham}(Q_{X_R}, K_{X_D})$ to measure the similarity score between the spike matrices Q_{X_R} and K_{X_D} results in a non-differentiable computation. To address this, we rewrite it as:

$$\begin{aligned} &f_{ham}(Q_{X_R}, K_{X_D}) = 1 - \frac{1}{\mathcal{C}} \sum_{i=1}^{\mathcal{C}} \mathbf{1}(Q_{X_R}^i \neq K_{X_D}^i) \\ &= \frac{\mathcal{C} - [(Q_{X_R})^T(\mathbf{1} - K_{X_D}) + (\mathbf{1} - Q_{X_R})^T K_{X_D}]}{\mathcal{C}} = \frac{\mathcal{C} - [(Q_{X_R})^T \mathbf{1} + \mathbf{1}^T K_{X_D} - 2(Q_{X_R})^T K_{X_D}]}{\mathcal{C}} \\ &= \frac{\mathcal{C} - [(Q_{X_R})^T \mathbf{1} + \mathbf{1}^T K_{X_D} - 2(Q_{X_R})^T K_{X_D} - \mathcal{C}/2 + \mathcal{C}/2]}{\mathcal{C}} \\ &= 1 - \frac{1}{\mathcal{C}}[(Q_{X_R})^T \mathbf{1} + \mathbf{1}^T K_{X_D} - 2(Q_{X_R})^T K_{X_D} - \frac{\mathcal{C}}{2}] - \frac{1}{2} \\ &= \frac{1}{2} + \frac{4(Q_{X_R})^T K_{X_D} - 2(Q_{X_R})^T \mathbf{1} - 2 * \mathbf{1}^T K_{X_D} + \mathcal{C}}{2\mathcal{C}} = \frac{1}{2} + \frac{1}{2\mathcal{C}}(2Q_{X_R} - \mathbf{1})^T (2K_{X_D} - \mathbf{1}). \end{aligned} \tag{25}$$

Given that the constant factor 1/2 does not affect the similarity evaluation, we employ $f_{ham}(\cdot)$ to SDA and formulate the spike-driven Hamming cross-attention (SHCA) mechanism as follows:

$$F_{SHCA} = \mathcal{SN}([\frac{1}{2\mathcal{C}}(2Q_R - \mathbf{1}) * (2K_D - \mathbf{1})^T] V_R) = \mathcal{SN}_{2\mathcal{C}}((2Q_R - \mathbf{1})[(2K_D - \mathbf{1})^T * V_R)]), \tag{26}$$

where F_{SHCA} denotes the fused feature obtained by SHCA. Meanwhile, to preserve the multiplication-free computation of SHCA, the scaling factor $1/2\mathcal{C}$ can be integrated into the spike-neuron threshold.

Table 9: P-AUROC on MVTec 3D-AD across time steps under unimodal and multimodal settings.

	Method	Year	Spike	Types	Time Step	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	EasyNet	MM23	✓	SNN	1	80.3	87.1	82.5	76.4	85.4	73.3	81.7	77.6	**87.4**	77.2	80.9
	3D-ADNAS	AAAI25	✓	SNN	1	85.7	91.4	93.8	84.1	91.3	85.4	85.6	91.3	84.9	83.2	87.8
	SpikingMAD	Ours	✓	SNN	1	**87.2**	**92.7**	**95.7**	**88.1**	**91.6**	**87.8**	**88.2**	**91.4**	**87.4**	**85.2**	**89.5**
	EasyNet	MM23	✓	SNN	4	78.4	86.8	82.7	77.1	83.8	74.2	81.5	78.4	87.0	76.6	80.7
	3D-ADNAS	AAAI25	✓	SNN	4	86.2	93.1	95.4	85.0	91.0	86.5	86.6	**94.8**	85.7	82.1	88.6
	SpikingMAD	Ours	✓	SNN	4	**87.8**	**93.4**	**96.2**	**87.8**	**92.2**	**88.5**	**89.1**	93.3	**87.6**	**85.5**	**90.1**
Depth	EasyNet	MM23	✓	SNN	1	53.2	75.5	72.2	56.1	69.8	59.4	56.7	68.8	80.4	56.2	64.8
	3D-ADNAS	AAAI25	✓	SNN	1	59.9	84.3	71.8	58.9	70.4	61.8	59.2	81.7	82.6	68.5	69.9
	SpikingMAD	Ours	✓	SNN	1	**61.1**	**85.6**	**73.4**	**60.8**	**72.9**	**63.7**	**63.6**	**84.5**	**87.2**	**70.9**	**72.4**
	EasyNet	MM23	✓	SNN	4	52.4	76.3	74.1	57.5	69.8	62.7	58.8	70.2	79.8	59.2	66.1
	3D-ADNAS	AAAI25	✓	SNN	4	59.3	85.1	70.6	56.1	71.4	63.6	57.1	82.5	83.1	69.1	69.8
	SpikingMAD	Ours	✓	SNN	4	**62.4**	**87.8**	**73.6**	**61.6**	**73.7**	**64.0**	**62.5**	**86.2**	**87.0**	**73.1**	**73.2**
RGB + Depth	EasyNet	MM23	✓	SNN	1	79.8	86.5	83.0	78.1	84.7	73.8	82.4	74.4	**88.5**	78.9	81.0
	CFM	CVPR24	✓	SNN	1	86.7	92.8	95.3	**88.1**	90.6	87.8	**89.2**	94.4	86.7	85.2	89.7
	3D-ADNAS	AAAI25	✓	SNN	1	84.4	**93.9**	94.5	83.9	91.8	86.8	87.1	92.1	85.6	84.3	88.4
	SpikingMAD	Ours	✓	SNN	1	**87.9**	93.1	**96.9**	87.5	**92.5**	**88.6**	89.0	**95.7**	**88.5**	**86.1**	**90.6**
	EasyNet	MM23	✓	SNN	4	79.8	87.2	83.3	78.6	84.5	74.6	83.6	75.2	87.7	78.6	81.3
	CFM	CVPR24	✓	SNN	4	87.5	93.6	96.7	**89.2**	91.7	89.0	**90.1**	95.3	87.5	**86.1**	90.7
	3D-ADNAS	AAAI25	✓	SNN	4	83.6	**94.1**	95.8	84.7	92.6	88.1	87.8	94.0	86.3	83.5	89.1
	SpikingMAD	Ours	✓	SNN	4	**88.3**	93.9	**97.4**	87.8	**93.0**	**89.9**	88.7	**96.9**	**89.1**	85.9	**91.1**

Table 10: AUPRO on MVTec 3D-AD across time steps under unimodal and multimodal settings.

	Method	Year	Spike	Types	Time Step	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
RGB	EasyNet	MM23	✓	SNN	1	41.9	66.2	78.0	44.5	68.5	57.2	55.5	73.4	61.8	61.1	60.8
	3D-ADNAS	AAAI25	✓	SNN	1	43.9	71.2	86.1	52.9	74.6	**64.4**	53.6	77.3	63.5	66.0	65.4
	SpikingMAD	Ours	✓	SNN	1	**51.2**	**78.8**	**88.1**	**65.9**	**77.5**	62.3	**63.4**	**84.6**	**63.9**	**67.7**	**70.3**
	EasyNet	MM23	✓	SNN	4	42.5	67.7	77.5	45.3	68.5	58.9	56.0	73.7	62.4	63.5	61.6
	3D-ADNAS	AAAI25	✓	SNN	4	45.0	72.8	86.9	54.2	75.5	**64.0**	55.7	78.5	**64.6**	**68.2**	66.5
	SpikingMAD	Ours	✓	SNN	4	**53.8**	**80.6**	**90.0**	**66.7**	**78.1**	**64.0**	**64.6**	**87.2**	63.5	68.0	**71.7**
Depth	EasyNet	MM23	✓	SNN	1	20.5	21.1	23.4	26.7	25.6	24.0	23.7	**22.8**	29.4	26.7	24.4
	3D-ADNAS	AAAI25	✓	SNN	1	23.5	19.8	28.6	27.0	22.2	29.8	31.4	21.8	33.2	30.0	26.7
	SpikingMAD	Ours	✓	SNN	1	**28.2**	**21.3**	**33.4**	**31.4**	**26.5**	**34.8**	**32.7**	17.8	**57.4**	**34.9**	**31.8**
	EasyNet	MM23	✓	SNN	4	21.2	23.2	23.8	27.5	24.9	23.4	24.4	**23.5**	30.1	27.2	24.9
	3D-ADNAS	AAAI25	✓	SNN	4	25.1	20.2	28.6	30.2	24.3	31.4	30.0	22.5	29.7	28.3	27.0
	SpikingMAD	Ours	✓	SNN	4	**30.1**	**23.6**	**34.3**	**32.7**	**28.6**	**37.0**	**34.5**	19.8	**60.3**	**36.4**	**33.7**
RGB + Depth	EasyNet	MM23	✓	SNN	1	41.3	67.3	77.2	45.8	67.9	58.3	56.8	73.1	62.5	62.3	61.3
	CFM	CVPR24	✓	SNN	1	**50.7**	76.4	90.1	64.2	75.1	65.9	60.4	86.5	65.8	66.2	70.1
	3D-ADNAS	AAAI25	✓	SNN	1	44.6	72.4	85.2	55.3	74.9	**66.3**	56.8	78.8	66.9	68.6	67.0
	SpikingMAD	Ours	✓	SNN	1	50.2	**80.3**	**90.2**	**66.5**	**81.3**	65.1	**64.9**	**88.7**	**67.3**	**70.7**	**72.5**
	EasyNet	MM23	✓	SNN	4	42.6	68.5	77.6	47.4	70.0	59.1	57.7	75.2	63.3	64.1	62.6
	CFM	CVPR24	✓	SNN	4	52.5	77.2	90.0	65.6	78.7	66.6	62.1	86.5	66.3	67.5	71.3
	3D-ADNAS	AAAI25	✓	SNN	4	46.5	73.3	86.5	55.4	75.6	**67.1**	58.2	79.4	65.8	68.6	67.6
	SpikingMAD	Ours	✓	SNN	4	**54.7**	**82.1**	**90.8**	**67.7**	**80.4**	65.9	**66.2**	**89.4**	**67.9**	**71.5**	**73.7**

E IMPLEMENTATION DETAILS AND EVALUATION METRICS

In this section, we provide the implementation details and evaluation metrics of our experiments.

Firstly, we adopt EfficientNet as the backbone encoder for both RGB and depth images, and take the multi-scale features from layers 1 to 4 of EfficientNet as its output features. Building on this, we introduce a feature jittering technique to perturb the normal features extracted by EfficientNet, thereby generating the corresponding anomalous features (Schlüter et al., 2022). Specifically, the jitter magnitude scale is set to 20.0, and the application probability is set to 1.0, ensuring that the jittering operation is performed on every sample. For the input sizes of the RGB and depth images, following the IUF (Tang et al., 2024) method, we resize the inputs to 224×224 to ensure consistency across modalities. During training, we use the AdamW optimizer to train the model for 1000 epochs on the MVTec 3D-AD (Bergmann et al., 2022), Real-IAD D3 (Zhu et al., 2025), and Eyecandies

Table 11: I-AUROC on Eyecandies across time steps under unimodal and multimodal settings.

	Method	Year	Spike	Types	Time Step	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwich	Lollipop	Marshmallow	Peppermint Candy	Mean
RGB	EasyNet	MM23	✓	SNN	1	56.1	63.8	64.5	62.4	57.0	60.8	59.5	60.7	66.1	68.1	61.9
	3D-ADNAS	AAAI25	✓	SNN	1	60.8	66.2	65.0	**67.8**	**60.4**	59.4	64.5	65.1	**67.0**	**67.2**	64.3
	SpikingMAD	Ours	✓	SNN	1	**68.6**	**70.8**	**65.1**	67.5	57.9	**65.6**	**78.7**	**68.7**	63.2	64.4	**67.1**
	EasyNet	MM23	✓	SNN	4	57.6	64.7	65.5	62.8	57.5	61.4	60.0	61.2	67.0	**68.5**	62.6
	3D-ADNAS	AAAI25	✓	SNN	4	62.1	67.3	65.5	69.0	**61.5**	60.9	66.2	67.0	**67.4**	68.4	65.5
	SpikingMAD	Ours	✓	SNN	4	**69.2**	**70.5**	**66.4**	**69.6**	58.2	**67.3**	**77.3**	**69.8**	64.0	65.5	**67.8**
Depth	EasyNet	MM23	✓	SNN	1	55.4	54.1	56.0	52.9	50.3	55.2	54.7	51.8	49.6	50.5	53.1
	3D-ADNAS	AAAI25	✓	SNN	1	47.6	53.3	61.2	56.8	56.5	51.6	**60.1**	57.4	**60.4**	59.7	56.5
	SpikingMAD	Ours	✓	SNN	1	**59.4**	**60.6**	**64.7**	**59.6**	**60.5**	**63.8**	58.2	**64.3**	57.7	**67.6**	**61.6**
	EasyNet	MM23	✓	SNN	4	53.7	54.2	56.4	52.1	51.6	55.2	54.5	52.0	50.0	50.9	53.1
	3D-ADNAS	AAAI25	✓	SNN	4	48.5	54.0	61.4	57.6	56.9	52.3	**60.7**	57.8	**60.4**	60.2	57.0
	SpikingMAD	Ours	✓	SNN	4	**60.7**	**61.8**	**66.5**	**61.0**	**62.8**	**64.5**	58.3	**64.9**	59.2	**67.6**	**62.7**
RGB + Depth	EasyNet	MM23	✓	SNN	1	55.2	64.3	65.1	61.7	58.7	61.1	61.3	59.2	66.8	67.6	62.1
	CFM	CVPR24	✓	SNN	1	45.4	62.7	64.3	58.7	60.5	55.7	62.2	63.0	**69.6**	65.3	60.7
	3D-ADNAS	AAAI25	✓	SNN	1	65.4	69.5	67.3	66.9	61.3	62.8	68.4	65.2	67.4	68.0	66.2
	SpikingMAD	Ours	✓	SNN	1	**69.7**	**70.2**	**71.5**	**76.1**	**64.4**	**74.6**	**77.4**	**72.3**	67.5	**75.8**	**72.0**
	EasyNet	MM23	✓	SNN	4	55.6	63.5	65.8	62.4	58.7	61.6	62.5	59.4	66.8	68.3	62.5
	CFM	CVPR24	✓	SNN	4	49.6	64.1	65.8	59.5	63.3	57.4	64.8	63.7	**71.0**	69.1	62.8
	3D-ADNAS	AAAI25	✓	SNN	4	65.4	70.4	67.7	68.1	62.5	63.4	69.3	65.8	67.7	68.6	66.9
	SpikingMAD	Ours	✓	SNN	4	**70.8**	**71.1**	**72.4**	**78.2**	**65.3**	**75.4**	**77.4**	**74.9**	68.0	**77.6**	**73.1**

Table 12: P-AUROC on Eyecandies across time steps under unimodal and multimodal settings.

	Method	Year	Spike	Types	Time Step	Candy Cane	Chocolate Cookie	Chocolate Praline	Confetto	Gummy Bear	Hazelnut Truffle	Licorice Sandwich	Lollipop	Marshmallow	Peppermint Candy	Mean
RGB	EasyNet	MM23	✓	SNN	1	56.8	65.6	63.3	62.5	58.4	61.3	62.1	59.4	66.8	67.1	62.3
	3D-ADNAS	AAAI25	✓	SNN	1	72.5	**73.6**	**71.4**	68.9	**67.4**	65.3	71.6	68.8	73.0	**72.8**	70.5
	SpikingMAD	Ours	✓	SNN	1	**88.6**	69.3	70.0	**77.1**	66.4	**73.6**	**75.3**	**87.9**	**74.3**	72.4	**75.5**
	EasyNet	MM23	✓	SNN	4	58.5	67.0	64.2	63.4	59.6	62.0	63.1	60.4	68.1	67.9	63.4
	3D-ADNAS	AAAI25	✓	SNN	4	72.8	**73.5**	**72.1**	69.5	**67.7**	66.0	72.5	69.3	73.6	73.2	71.0
	SpikingMAD	Ours	✓	SNN	4	**90.0**	69.6	71.2	**79.3**	67.5	**74.6**	**76.6**	**88.4**	**75.7**	**73.1**	**76.6**
Depth	EasyNet	MM23	✓	SNN	1	46.8	54.3	55.2	53.1	47.5	50.4	49.8	48.8	**56.9**	56.5	51.9
	3D-ADNAS	AAAI25	✓	SNN	1	55.2	58.3	57.7	56.8	54.6	54.5	**58.0**	52.9	54.4	57.2	56.0
	SpikingMAD	Ours	✓	SNN	1	**59.1**	**61.2**	**60.7**	**59.8**	**55.7**	**59.0**	57.1	**69.1**	53.3	**62.8**	**59.8**
	EasyNet	MM23	✓	SNN	4	46.8	54.8	55.7	54.0	48.1	50.9	50.3	49.6	**57.4**	56.9	52.5
	3D-ADNAS	AAAI25	✓	SNN	4	55.5	59.0	58.2	57.3	55.1	54.8	**58.3**	53.5	55.1	57.6	56.4
	SpikingMAD	Ours	✓	SNN	4	**59.8**	**61.4**	**62.5**	**60.8**	**57.0**	**59.6**	57.5	**69.0**	53.6	**63.2**	**60.4**
RGB + Depth	EasyNet	MM23	✓	SNN	1	57.7	65.6	64.8	62.3	59.2	62.0	61.7	59.9	67.5	68.0	62.9
	CFM	CVPR24	✓	SNN	1	49.8	63.3	65.0	60.8	62.9	60.0	64.7	65.5	69.8	66.6	62.8
	3D-ADNAS	AAAI25	✓	SNN	1	73.2	**75.0**	**71.2**	70.3	**67.2**	66.4	72.4	69.5	73.4	74.4	71.3
	SpikingMAD	Ours	✓	SNN	1	**89.4**	68.7	69.4	**80.2**	67.1	**74.3**	**75.7**	**88.4**	**74.7**	**75.4**	**76.3**
	EasyNet	MM23	✓	SNN	4	58.1	65.8	65.4	62.7	60.0	62.4	62.3	62.0	67.5	68.7	63.5
	CFM	CVPR24	✓	SNN	4	53.2	65.4	66.8	64.1	64.3	62.2	65.6	67.5	71.7	69.8	65.1
	3D-ADNAS	AAAI25	✓	SNN	4	73.9	**75.4**	**72.0**	70.3	**68.3**	67.1	73.3	70.2	74.0	74.7	71.9
	SpikingMAD	Ours	✓	SNN	4	**91.4**	69.4	70.8	**82.1**	67.1	**75.5**	**78.3**	**89.1**	**76.6**	**75.8**	**77.6**

(Bonfiglioli et al., 2022) datasets, with a learning rate of 0.0001, first- and second-order momentum decay rates of 0.9 and 0.999, respectively, and a weight decay rate of 0.0001.

Then, we evaluate the performance of SpikingMAD adopting several metrics (He et al., 2024b;a; Zavrtanik et al., 2022; Zhang et al., 2026; Cheng et al., 2026a), such as I-AUROC (Image-level Area Under the Receiver Operating Characteristic curve), P-AUROC (Pixel-level Area Under the Receiver Operating Characteristic curve), and AUPRO (Area Under the Per-Region Overlap curve) (Wang et al., 2025b; Li et al., 2025a; Guo et al., 2025a; Defard et al., 2021; Xu et al., 2026).

F P-AUROC AND AUPRO SCORES ON THE MVTEC 3D-AD DATASET

Tables 9 and 10 report the comparison results of P-AUROC and AUPRO between SpikingMAD and SNN-based baselines on the MVTec 3D-AD dataset under unimodal and multimodal settings (with the time step T set to 1 and 4, respectively). The results demonstrate that SpikingMAD consistently maintains superior performance across most test cases. Specifically, when trained with both RGB

Table 13: P-AUROC on Real-IAD D3 dataset across time steps under unimodal settings (RGB).

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
RGB	Cflow	WACV22	✗	ANN	1	94.4	84.7	92.1	44.2	72.6	85.3	91.4	65.7	86.1	91.4	83.6
	SimpleNet	CVPR23	✗	ANN	1	92.6	82.2	89.1	74.5	72.5	86.6	80.6	94.5	95.7	87.9	89.0
	EasyNet	MM23	✓	SNN	1	59.8	61.4	63.7	**60.8**	57.0	57.6	58.4	57.6	58.2	61.3	61.9
	3D-ADNAS	AAAI25	✓	SNN	1	54.5	64.9	67.7	55.1	59.3	62.9	64.2	67.9	62.1	66.7	67.8
	SpikingMAD	Ours	✓	SNN	1	**64.8**	**74.3**	**68.9**	58.5	**64.4**	**67.9**	**65.8**	**79.6**	**63.3**	**80.2**	**77.6**
	EasyNet	MM23	✓	SNN	4	61.0	62.4	64.3	**61.5**	58.8	58.6	59.6	57.7	59.5	62.0	63.1
	3D-ADNAS	AAAI25	✓	SNN	4	56.4	65.1	69.1	55.8	61.2	63.7	65.3	68.5	63.6	67.0	67.8
	SpikingMAD	Ours	✓	SNN	4	**63.5**	**74.3**	**69.4**	58.9	**65.1**	**68.3**	**65.8**	**79.7**	**64.2**	**80.8**	**78.0**

	Method	Year	Spike	Types	Time Step	knob cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
RGB	Cflow	WACV22	✗	ANN	1	89.3	85.2	87.7	73.9	95.0	79.9	66.4	57.1	91.0	80.8
	SimpleNet	CVPR23	✗	ANN	1	87.9	89.8	94.7	79.9	79.0	76.0	48.9	93.8	91.6	84.3
	EasyNet	MM23	✓	SNN	1	58.4	63.5	59.6	60.2	56.0	59.1	60.4	55.8	58.5	59.5
	3D-ADNAS	AAAI25	✓	SNN	1	68.0	62.8	64.6	63.7	65.6	66.7	63.5	66.8	56.3	63.6
	SpikingMAD	Ours	✓	SNN	1	**76.3**	**67.7**	**73.9**	**76.9**	**76.0**	**72.5**	**68.5**	**78.6**	**62.0**	**70.9**
	EasyNet	MM23	✓	SNN	4	59.2	63.5	61.7	61.4	56.8	59.5	61.0	56.6	59.1	60.4
	3D-ADNAS	AAAI25	✓	SNN	4	68.4	64.2	65.7	64.4	66.2	67.3	63.9	69.4	57.0	64.5
	SpikingMAD	Ours	✓	SNN	4	**77.5**	**68.5**	**74.1**	**76.7**	**76.6**	**73.1**	**68.5**	**79.4**	**62.5**	**71.2**

Table 14: P-AUROC on Real-IAD D3 across time steps under unimodal settings (point cloud).

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
3D point cloud	PointMAE	ECCV22	✗	ANN	1	65.5	68.7	42.8	36.3	50.7	65.6	71.7	62.0	60.5	63.3	56.2
	EasyNet	MM23	✓	SNN	1	50.0	52.5	54.8	52.6	54.1	50.3	48.6	47.3	49.5	51.4	53.3
	3D-ADNAS	AAAI25	✓	SNN	1	53.1	59.6	61.1	48.5	53.4	58.4	50.6	60.5	57.7	59.3	56.2
	SpikingMAD	Ours	✓	SNN	1	**59.8**	**71.6**	**62.7**	**57.3**	**70.3**	**68.4**	**65.2**	**77.4**	**61.0**	**78.6**	**73.6**
	EasyNet	MM23	✓	SNN	4	52.6	53.8	55.6	53.5	56.2	52.2	50.3	49.0	52.6	53.4	55.0
	3D-ADNAS	AAAI25	✓	SNN	4	54.0	60.6	61.5	49.3	54.1	58.8	51.2	60.7	58.6	60.0	56.2
	SpikingMAD	Ours	✓	SNN	4	**61.7**	**72.5**	**63.3**	**58.7**	**72.2**	**69.5**	**65.5**	**79.8**	**64.6**	**79.3**	**74.5**

	Method	Year	Spike	Types	Time Step	knob cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
3D point cloud	PointMAE	ECCV22	✗	ANN	1	42.5	77.6	48.2	62.0	54.5	43.5	68.7	27.1	41.3	55.4
	EasyNet	MM23	✓	SNN	1	51.2	54.4	46.5	49.8	47.4	50.6	51.8	48.8	49.2	50.7
	3D-ADNAS	AAAI25	✓	SNN	1	50.5	57.5	52.1	56.0	55.9	57.4	56.6	54.9	46.8	55.3
	SpikingMAD	Ours	✓	SNN	1	**53.6**	**69.4**	**72.0**	**69.8**	**73.5**	**65.2**	**74.0**	**65.5**	**59.8**	**67.4**
	EasyNet	MM23	✓	SNN	4	51.8	56.8	47.2	50.4	49.5	53.2	52.5	49.3	50.8	52.3
	3D-ADNAS	AAAI25	✓	SNN	4	50.6	58.0	52.9	56.6	55.5	58.1	57.3	55.2	47.4	55.8
	SpikingMAD	Ours	✓	SNN	4	**54.0**	**71.1**	**72.8**	**70.6**	**73.8**	**66.0**	**74.4**	**66.3**	**60.2**	**68.5**

and depth images at T=4, SpikingMAD achieves 2.0% and 6.1% improvements compared to 3D-ADNAS (SNN-based) in terms of P-AUROC and AUPRO scores, respectively (0.4% and 2.4% superior to CFM and 9.8% and 11.1% higher than EasyNet, respectively). These experimental results again demonstrate that our proposed spike-driven architecture not only effectively maintains detection performance but also significantly reduces energy consumption.

G I-AUROC AND P-AUROC SCORES ON THE EYECANDIES DATASET

To validate the effectiveness of the proposed approach, we report the I-AUROC and P-AUROC of SpikingMAD on the Eyecandies dataset. As shown in Tables 11 and 12, our proposed method achieves highly competitive performance compared with SNN-based approaches. Specifically, when trained solely with RGB images, SpikingMAD surpasses EasyNet (SNN-based baseline) by 5.2% and 13.2% in I-AUROC and P-AUROC, respectively (T=1). It also outperforms 3D-ADNAS by 2.8% and 5.0% in I-AUROC and P-AUROC, respectively (T=1). When trained on both RGB and depth images, SpikingMAD exceeds CFM (SNN-based baseline) by 10.3% and 12.5% in I-AUROC and P-AUROC, respectively (T=4). Extensive experiments on the Eyecandies dataset further demonstrate the effectiveness of our proposed method.

Table 15: P-AUROC on Real-IAD D3 dataset across time steps under multimodal settings.

	Method	Year	Spike	Types	Time Step	audio jack socket	common mode filter	connector housing female	crimp st cable mount box	dc power connector	ethernet connector	ferrite bead	fork crimp terminal	fuse holder	headph-one jack socket	humi-dity sensor
RGB + 3D point cloud	PatchCore	CVPR22	✗	ANN	1	67.3	92.2	91.9	93.1	92.1	95.6	93.2	95.2	92.7	94.2	93.3
	AST	WACV23	✗	ANN	1	59.0	80.2	71.6	58.9	77.0	90.6	81.7	77.3	75.4	69.6	72.3
	M3DM	CVPR23	✗	ANN	1	69.9	93.4	97.9	93.3	95.0	97.8	96.6	96.4	94.8	98.2	95.8
	EasyNet	MM23	✓	SNN	1	60.2	62.1	64.5	61.1	57.6	58.7	59.3	60.0	59.6	62.2	62.5
	3D-ADNAS	AAAI25	✓	SNN	1	57.6	66.7	68.1	56.8	60.8	63.6	65.4	69.8	63.4	68.4	68.8
	SpikingMAD	Ours	✓	SNN	1	**62.6**	**75.7**	**69.3**	**60.4**	**73.8**	**70.0**	**68.7**	**82.4**	**64.4**	**82.1**	**78.0**
	EasyNet	MM23	✓	SNN	4	61.2	63.3	65.7	62.9	58.0	60.1	59.5	61.7	60.5	64.0	63.8
	3D-ADNAS	AAAI25	✓	SNN	4	58.4	67.2	68.8	57.9	61.5	64.1	66.5	70.3	64.0	68.4	69.6
	SpikingMAD	Ours	✓	SNN	4	**63.1**	**76.4**	**71.0**	**62.1**	**75.6**	**72.3**	**79.6**	**83.1**	**64.9**	**82.8**	**79.2**

	Method	Year	Spike	Types	Time Step	knob cap	lattice block plug	lego pin connector plate	lego propeller	limit switch	miniature lifting motor	power jack	purple clay pot	telephone spring switch	Mean
RGB + 3D point cloud	PatchCore	CVPR22	✗	ANN	1	95.8	92.3	75.9	72.7	93.8	96.2	92.3	96.1	94.4	90.5
	AST	WACV23	✗	ANN	1	65.6	91.9	62.9	70.3	64.1	46.7	64.5	44.5	55.1	65.0
	M3DM	CVPR23	✗	ANN	1	93.8	95.8	73.4	77.3	96.6	99.1	90.2	95.3	93.6	92.2
	EasyNet	MM23	✓	SNN	1	59.1	64.3	60.6	61.3	56.6	59.8	61.7	56.9	59.6	60.4
	3D-ADNAS	AAAI25	✓	SNN	1	69.8	63.6	64.3	64.2	67.1	68.6	64.0	66.7	57.7	64.8
	SpikingMAD	Ours	✓	SNN	1	**77.5**	**72.6**	**75.1**	**77.1**	**77.6**	**73.4**	**82.1**	**81.0**	**62.2**	**73.3**
	EasyNet	MM23	✓	SNN	4	60.4	66.2	63.1	62.1	58.2	60.5	62.4	57.4	61.4	61.6
	3D-ADNAS	AAAI25	✓	SNN	4	71.3	63.9	65.2	64.9	68.0	69.5	64.7	67.3	57.9	65.5
	SpikingMAD	Ours	✓	SNN	4	**77.9**	**73.5**	**76.4**	**78.8**	**77.9**	**75.2**	**83.8**	**81.6**	**63.0**	**74.9**

H P-AUROC SCORES ON THE REAL-IAD D3 DATASET

Tables 13, 14, and 15 show the comparison of P-AUROC scores between SpikingMAD and ANN-/SNN-based baselines on the Real-IAD D3 dataset. It is obvious that SpikingMAD consistently outperforms the SNN-based baselines while maintaining comparable performance to the ANN-based methods. Specifically, SpikingMAD achieves a 70.9% P-AUROC when trained solely on RGB images with the time step (i.e., T) set to 1. With the time step set to 4, SpikingMAD achieves a 74.9% P-AUROC when trained using both RGB and depth images. These results demonstrate that the spike-driven architecture can effectively advance the development of MAD in terms of both energy consumption and detection performance.

I DETAILED RESULTS OF FEW-SHOT SETTING ON THE MVTEC 3D-AD DATASET

Table 16 reports the experimental results of SpikingMAD under the few-shot setting (Kim et al., 2024; Cai et al., 2026; Dai et al., 2025; Hu et al., 2024) on the MVTec 3D-AD dataset. Specifically, we randomly select 5, 10, and 50 images (including both RGB and depth images) from each category as training data, and evaluate the performance of SpikingMAD on the full test set. As shown in Table 16, our proposed method exhibits great potential under the few-shot setting. For example, under the 50-shot setting, SpikingMAD achieves I-AUROC, P-AUROC, and AUPRO of 85.5%, 86.4%, and 67.3% on the MVTec 3D-AD dataset, respectively; under the 10-shot setting, its I-AUROC, P-AUROC, and AUPRO scores are 77.9%, 79.6%, and 55.5%, respectively; and under the 5-shot setting, its I-AUROC, P-AUROC, and AUPRO scores are 75.8%, 76.8%, and 50.9%, respectively. This indicates that the spike-driven MAD architecture is still effective for few-shot multimodal anomaly detection tasks.

Table 16: I-AUROC, P-AUROC, and AUPRO scores on MVTec 3D-AD under the few-shot setting.

	Metric	Method	Year	Bagel	Cable Gland	Carrot	Cookie	Dowel	Foam	Peach	Potato	Rope	Tire	Mean
5-shot														
RGB + Depth	I-AUROC	EasyNet	MM23	58.7	62.6	57.3	56.8	62.4	55.6	57.5	52.8	60.7	51.3	57.6
		3D-ADNAS	AAAI25	64.4	65.7	62.3	63.4	69.5	68.8	66.7	60.2	63.0	62.5	64.7
		SpikingMAD	Ours	**78.2**	**80.3**	**75.8**	**71.1**	**79.6**	**76.4**	**77.8**	**79.6**	**74.2**	**65.4**	**75.8**
	P-AUROC	EasyNet	MM23	67.3	69.1	65.7	63.5	64.8	52.6	62.1	57.6	69.6	60.4	63.3
		3D-ADNAS	AAAI25	70.6	74.5	75.6	68.4	72.1	73.8	75.0	76.7	70.5	67.0	72.4
		SpikingMAD	Ours	**74.8**	**78.7**	**80.1**	**75.3**	**77.6**	**76.8**	**79.1**	**80.4**	**77.0**	**68.1**	**76.8**
	AUPRO	EasyNet	MM23	29.6	45.1	46.3	33.3	42.7	41.3	39.4	43.6	45.5	47.2	41.4
		3D-ADNAS	AAAI25	31.9	47.8	50.3	38.2	45.1	44.9	35.4	47.1	48.8	**49.5**	43.9
		SpikingMAD	Ours	**36.5**	**51.9**	**64.5**	**50.0**	**52.6**	**47.8**	**48.3**	**58.2**	**56.5**	42.4	**50.9**
10-shot														
RGB + Depth	I-AUROC	EasyNet	MM23	64.1	66.2	60.5	61.4	66.8	57.7	62.5	56.3	65.1	53.0	61.4
		3D-ADNAS	AAAI25	69.8	71.1	69.2	67.8	76.9	**75.0**	70.4	63.1	67.2	66.7	69.7
		SpikingMAD	Ours	**81.5**	**82.4**	**77.9**	**75.6**	**81.8**	73.8	**76.7**	**78.9**	**79.4**	**71.1**	**77.9**
	P-AUROC	EasyNet	MM23	70.1	73.4	70.4	69.2	69.8	57.3	67.1	61.8	74.2	65.5	67.9
		3D-ADNAS	AAAI25	75.7	80.3	79.8	70.2	77.8	78.4	79.8	81.0	76.5	72.4	77.2
		SpikingMAD	Ours	**77.5**	**85.5**	**81.9**	**76.8**	**83.7**	**79.4**	**80.0**	**82.6**	**78.2**	**70.3**	**79.6**
	AUPRO	EasyNet	MM23	33.4	48.8	50.5	38.4	50.0	46.1	42.1	49.7	48.2	51.9	45.9
		3D-ADNAS	AAAI25	35.4	53.2	57.3	41.4	49.6	47.3	39.7	53.2	52.2	**54.6**	48.4
		SpikingMAD	Ours	**40.4**	**60.8**	**69.3**	**52.3**	**56.5**	**49.2**	**51.6**	**67.1**	**59.2**	48.4	**55.5**
50-shot														
RGB + Depth	I-AUROC	EasyNet	MM23	68.2	79.0	65.5	68.4	74.6	62.5	69.6	60.7	74.9	58.5	68.2
		3D-ADNAS	AAAI25	77.3	79.0	77.8	76.1	71.4	76.4	70.2	69.5	79.2	78.3	75.5
		SpikingMAD	Ours	**87.7**	**90.4**	**88.4**	**84.0**	**88.0**	**80.4**	**83.3**	**82.9**	**86.7**	**82.8**	**85.5**
	P-AUROC	EasyNet	MM23	73.4	79.7	77.9	72.5	77.4	66.8	78.7	68.4	81.1	72.2	74.8
		3D-ADNAS	AAAI25	81.8	85.6	87.8	77.4	84.9	82.1	81.6	86.6	80.2	79.2	82.7
		SpikingMAD	Ours	**83.4**	**90.8**	**94.2**	**78.7**	**85.0**	**85.4**	**87.9**	**93.8**	**83.7**	**81.4**	**86.4**
	AUPRO	EasyNet	MM23	37.5	60.7	71.9	43.5	63.4	52.6	51.9	64.8	58.0	57.6	56.2
		3D-ADNAS	AAAI25	38.2	66.0	78.4	48.8	69.1	58.9	47.5	72.3	59.0	**62.4**	60.1
		SpikingMAD	Ours	**47.5**	**69.9**	**87.7**	**62.3**	**73.6**	**60.6**	**61.4**	**84.6**	**63.2**	62.3	**67.3**
full-shot														
RGB + Depth	I-AUROC	EasyNet	MM23	75.8	84.7	72.2	74.8	80.3	70.6	76.7	65.6	83.9	64.8	74.9
		3D-ADNAS	AAAI25	83.3	89.4	86.0	82.1	79.8	82.6	77.4	76.5	88.9	75.6	82.2
		SpikingMAD	Ours	**90.4**	**93.8**	**92.3**	**88.7**	**91.3**	**86.2**	**89.2**	**87.2**	**91.6**	**85.7**	**89.6**
	P-AUROC	EasyNet	MM23	79.8	86.5	83.0	78.1	84.7	73.8	82.4	74.4	**88.5**	78.9	81.0
		3D-ADNAS	AAAI25	84.4	**93.9**	94.5	83.9	91.8	86.8	87.1	92.1	85.6	84.3	88.4
		SpikingMAD	Ours	**87.9**	93.1	**96.9**	**87.5**	**92.5**	**88.6**	**89.0**	**95.7**	**88.5**	**86.1**	**90.6**
	AUPRO	EasyNet	MM23	41.3	67.3	77.2	45.8	67.9	58.3	56.8	73.1	62.5	62.3	61.3
		3D-ADNAS	AAAI25	44.6	72.4	85.2	55.3	74.9	**66.3**	56.8	78.8	66.9	68.6	67.0
		SpikingMAD	Ours	**50.2**	**80.3**	**90.2**	**66.5**	**81.3**	65.1	**64.9**	**88.7**	**67.3**	**70.7**	**72.5**