SICAF: Time-Varying Focus Bottleneck for Self-Supervised Event-Based Optical Flow with Spiking Neural Network

Shuangming Yang*
Tianjin University
Tianjin, China
yangshuangming@tju.edu.cn

Qing He
Tianjin University
Tianjin, China
3020234423@tju.edu.cn

Shangqi Guo*
Tsinghua University
Beijing, China
shangqi_guo@mail.tsinghua.edu.cn

Badong Chen*
Xi'an Jiaotong University
Xi'an, China
chenbd@mail.xjtu.edu.cn

Abstract

Event-based optical flow estimation achieves high temporal resolution and dynamic range by exploiting asynchronous events, while self-supervised learning provides an effective solution to the absence of ground-truth labels. Spiking neural networks (SNNs), as brain-inspired models, process information through asynchronous spikes, naturally aligning with optical flow events. However, the temporal irregularity across event streams hinders the compact extraction of spatiotemporal features, while the absence of explicit labels leads to spurious correlations, thereby increasing redundancy within latent representations. In this paper, we propose a novel "Compactly Compress, Time-varying Focus" training strategy. We present SICAF (**S**pikes **I**nformation **C**ompression **A**nd **F**ocus) framework that formulates the learning process as a constrained optimization with temporal modulation. It encourages self-supervised SNNs to extract compact latent representations while preventing over-compression that would discard temporal saliency. To better enable feature selectivity across time steps, we design an estimator with learnable temporal attention, achieving context awareness with cross-time dependencies. Experimental results on several datasets demonstrate that SICAF achieves state-of-the-art performance in self-supervised SNNs, with a 8.14% improvement in AEE on the MVSEC dataset and a 12.80% increase in robustness under white-box adversarial attacks.

1 Introduction

Event-based optical flow estimation is an advanced technique that utilizes event cameras to measure motion in visual scenes [1–3]. With high temporal resolution and low power consumption, event cameras outperform frame-based sensors by delivering precise, rapid motion detection, enabling energy-efficient optical flow estimation in high-speed and low-light environments [4, 5]. Since ground-truth optical flow labels are difficult to obtain, recent work has adopted self-supervised learning for event-based optical flow without annotation [6–9].

Emerging as a promising candidate for neuromorphic computing, spiking neural networks (SNNs) employ temporal coding mechanisms to efficiently encode time-dependent dynamic information

*Corresponding authors: Shuangming Yang, Shangqi Guo, and Badong Chen.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

[10, 11]. Besides, SNNs operate in an event-driven spiking manner, allowing flexible processing of asynchronous and sparse data [12, 13]. These properties make SNNs highly promising for handling event-based optical flow [1, 14]. Recent work on SNNs has successfully developed self-supervised learning strategies for event-based optical flow estimation [7, 6, 15], and further incorporated learnable membrane-potential parameters to improve their representational capacity [7]. Yang *et al.* [16] propose SeLHIB, a self-supervised high-order information bottleneck training scheme, which improves the robustness of event-based optical flow estimation.

Despite these significant advances, a glaring limitation persists: variations in motion patterns give rise to event streams that are highly irregular along the temporal dimension, while existing optical flow learning methods largely overlook such temporal heterogeneity. This omission hinders the model's ability to capture informative motion cues in the event stream, especially in complex scenarios involving fast motion, adverse illumination conditions, or highly dynamic backgrounds. This brings us to an intriguing inquiry: *Is it feasible to design an SNN learning strategy that intrinsically accounts for the temporal evolution of spike-based latent representations when extracting compact yet informative motion features?*

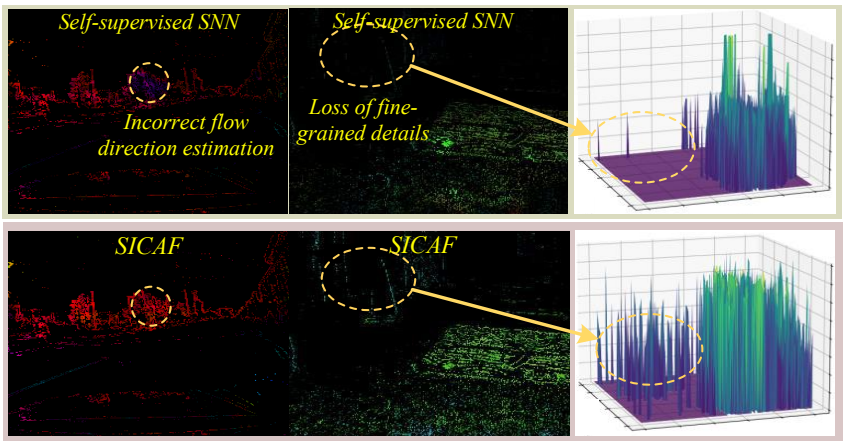


Figure 1: **Visualization of performance evaluation with previous self-supervised SNN (top) and the proposed SICAF (bottom).** Our SICAF enhances the estimation accuracy by emphasizing temporal saliency. Concurrently, our SICAF effectively captures structured yet informative motion features, improving the issues of incorrect flow direction estimation and the loss of fine-grained motion details.

In response to the aforementioned challenge, we propose a pioneering training strategy termed ***Compactly Compress, Time-varying Focus***. Central to our approach is a novel adaptive focus mechanism equipped with a flexible information compression strategy across time steps to achieve temporal feature selectivity. While learning compact latent representations, the proposed approach explicitly takes into account the temporal heterogeneity of the input event streams. As illustrated in Fig. 1, previous self-supervised SNNs often suffer from incorrect optical flow direction estimation and the loss of fine-grained motion details, whereas our approach strengthens the capability of SNNs to represent motion patterns with more accurate optical flow predictions.

In event-based optical flow estimation, general self-supervised SNNs typically adopt contrast maximization, as illustrated in Fig. 2(a). First, such self-supervised SNNs lack a mechanism for compressing redundant and noisy information, and therefore fail to extract compact, task-relevant spatiotemporal representations from event data. Second, contrast maximization drives learning solely based on the spatial alignment of events, without exploiting their timestamps or temporal distribution. SeLHIB [16], as shown in Fig. 2(b), captures more compact representations by formulating information bottleneck (IB) constraint. Nevertheless, this form of IB fails to consider the irregular temporal distribution of event streams. A uniform information bottleneck applied across time is inadequate for capturing temporally salient features. As a result, jointly handling temporal non-uniformity and redundancy compression remains somewhat uncharted.

Targeting the above challenge, we propose **S**pikes **I**nformation **C**ompression **A**nd **F**ocus (**SICAF**), which formulates the learning process as a constrained optimization with temporal modulation. As

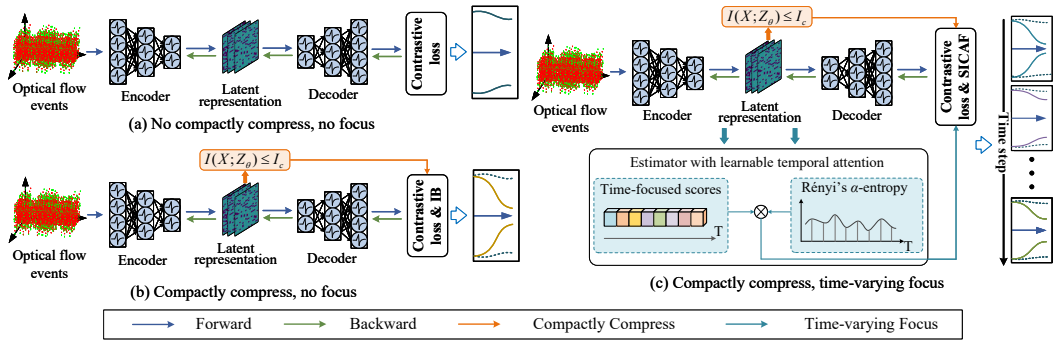


Figure 2: **Comparison between the general self-supervised SNNs [7, 6, 15], SeLHIB [16], and ours for handling optical flow events with self-supervised learning strategy.** (a) General self-supervised SNNs typically rely on contrast maximization. This objective enhances the spatial contrast of motion-compensated events to drive learning, but it neglects the temporal non-uniformity of event data. (b) SeLHIB presents an information bottleneck constraint to encourage compact latent representations. However, because it enforces a uniform compression constraint across time, it struggles to preserve temporally salient features. (c) **SICAF** addresses these issues by formulating a constrained optimization process with temporally modulated focus. Within this framework, a learnable temporal-attention–based estimator allows SICAF to explicitly model temporal heterogeneity, encouraging the formation of compact yet more temporally informative representations.

depicted in Fig. 2(c), SICAF learns a latent and compact space in a time-varying compression manner while capturing temporal salient feature across time steps. To this end, we first design a novel estimator with learnable temporal attention to modulate the compression process. Specifically, to capture temporal dynamics in the representation learning process, we map the spike firing rates of latent representations into queries and keys to compute temporal attention scores. This enables the model to capture feature variations along the temporal dimension and provides a time-varying focusing capability for the input event stream. Building on Rényi's α-entropy, the estimator robustly measures the entropy of high-dimensional latent representations at different time steps. These entropy values, weighted by the temporal attention scores, form a time-varying focusing constraint that modulates the strength of information compression. Extensive experiments demonstrate that SICAF achieves the top performance in self-supervised SNNs. Overall, the main **contributions** of this paper are:

- We present a novel "compactly compress, time-varying focus" paradigm, which learns to enhance the temporal saliency of representations while suppressing redundant event information.
- We propose SICAF, a constrained optimization framework with temporal modulation by neural activities. It enjoys a concise and tractable objective and enables adaptive, time-varying compression of irregular optical-flow events.
- We design a learnable temporal-attention estimator that directly extracts fine-grained temporal cues from spiking neural activities and dynamically guides the compression process.
- Experiments demonstrate that SICAF achieves superior performance compared with existing state-of-the-art SNN-based optical flow models, while significantly enhancing the robustness. SICAF can serve as a strong baseline and inspire further advancements in the field of neuromorphic computing and brain-inspired intelligence.

2 Related Work

2.1 Event-based Optical Flow Estimation with SNNs

Event-based optical flow estimation with SNNs has attracted growing interest. Early work exploited adaptive neurons and spike timing-dependent plasticity (STDP) to develop unsupervised SNN training for efficient high-bandwidth flow estimation [14]. Hybrid architectures such as Spike-FlowNet further

showed that SNN encoders can match artificial neural network (ANN) performance when paired with ANN decoders [8]. Subsequent efforts introduced self-supervised frameworks leveraging temporal cues from short event slices [6], multi-sensor fusion with frame-based inputs [17], and adaptive neural mechanisms for improved spatiotemporal extraction [7]. S-TLLR proposed a three-factor temporal local learning rule inspired by STDP, achieving state-of-the-art performance under self-supervision [15]. ANN-to-SNN conversion approaches have also been explored to enhance estimation accuracy [9]. Directly trained SNNs, however, offer unique advantages by fully exploiting the intrinsic temporal dynamics of spiking neurons, enabling low-latency and efficient spatiotemporal feature learning [18–20]. In this work, we follow this direction and focus on direct training of SNNs to unlock their full potential for event-based optical flow estimation.

2.2 Direct Training of SNNs with the IB Principle

Recent advances have applied the IB principle to directly train SNNs. Methods such as HOSIB and HHO-IB introduce higher-order IB formulations for surrogate-gradient learning, yielding notable gains in generalization and robustness [21, 22]. SeLHIB further presents an IB-based self-supervised framework for event-based optical flow with SNNs, substantially improving robustness under noisy conditions [16]. However, this uniform temporal compression suppresses critical temporal cues. We address this limitation by proposing a strategy that preserves compactness while selectively retaining temporally salient features.

3 Preliminary

Spiking Neuron Model. Spiking neurons are the core computational units of SNNs. They integrate incoming postsynaptic potentials, emit a spike once the membrane potential crosses a firing threshold, and thereby enable efficient event-driven processing of asynchronous data. The mathematical modeling is presented below:

$$u_{pre}^{(t)} = \tau u^{(t-1)} + i^{(t)}, \tag{1}$$

$$u^{(t)} = u_{pre}^{(t)}(1 - o^{(t)}), \tag{2}$$

where $u_{pre}^{(t)}$ and $u^{(t)}$ denote the pre-update and post-update membrane potentials in the leaky integrate-and-fire (LIF) model, respectively. $i^{(t)}$ represents the synaptic input current at timestep t, while $o^{(t)} \in \{0, 1\}$ indicates the binary output spike generated by the spiking neuron. The detailed neuron parameter settings are provided in Appendix A.4.

Contrast Maximization. Contrast maximization is a self-supervised approach for optical flow estimation that exploits the spatiotemporal misalignment of events in a continuous stream [23]. By applying motion compensation, events are warped to a common reference frame, reducing motion-induced blur. Assuming $u(x)$ denotes the optical flow at each pixel location and $\bar{t}(x, y; u \mid t_{\text{ref}})$ the average timestamp at each pixel. The objective is to minimize the normalized sum of squared average timestamps, yielding the following convex contrastive loss:

$$\mathcal{L}_{\text{contrast}}(t_{\text{ref}}) = \frac{\sum_x \bar{t}_{p=+}(x; u \mid t_{\text{ref}})^2 + \bar{t}_{p=-}(x; u \mid t_{\text{ref}})^2}{\sum_x [n(x') > 0] + \epsilon}. \tag{3}$$

4 Method

Fig. 2(c) visually elucidates the overall procedure of SICAF. After being encoded through $E(Z|X)$, the input X is mapped to the latent representation Z, and the upper bound of the mutual information between X and Z, denoted $I(X; Z) \leq I_c$, enforces compact compression during the feature extraction process. We compute time-focused scores from the spike-rate matrices of Z along the temporal dimension, enabling contextual awareness across time dependencies. Moreover, we employ the Rényi's α-entropy estimator to robustly quantify the information content of the latent variables. The entropy is multiplied by the time-focused scores, forming an estimator with learnable temporal attention that maximizes the entropy estimation to preserve the temporal heterogeneity of event streams. The pseudocode of the method is provided in Appendix A.5, and the generalization bound as well as the sample complexit bound theorems and their proofs are presented in AppendixA.7.

4.1 Temporal Attention Scores

Unlike conventional ANNs, SNNs communicate through discrete spike events, allowing temporal information to be encoded directly in spike timing as activity propagates across layers. Motivated by this inherent temporal structure, we compute attention weights over latent representations across time, forming a temporal attention matrix T_A. Specifically, for a latent representation $Z \in \mathbb{R}^{T\times C\times D}$, we first compute its spike rate matrix $T_Z \in \mathbb{R}^{T\times C}$, and then project it into two latent spaces as follows:

$$Z_Q = T_Z \cdot W_Q, \quad Z_K = T_Z \cdot W_K \tag{4}$$

where $Z_Q, Z_K \in \mathbb{R}^{T\times T}$ represent temporal queries and keys, respectively. The temporal attention matrix is computed as:

$$A = \mathrm{softmax}_{\mathrm{dim}=2}\left(\frac{Z_Q \cdot Z_K^T}{\sqrt{d_k}}\right), \tag{5}$$

where d_k is the key dimension. The softmax is applied along the key temporal dimension (the second dimension of the attention matrix), such that $\sum_{j=1}^{T} A_{ij} = 1$ for each query timestep i.We then average the attention weights over all query timesteps to obtain the normalized temporal attention distribution:

$$T_{A,j} = \frac{1}{T}\sum_{i=1}^{T} A_{ij}, \qquad j = 1, \ldots, T. \tag{6}$$

where $T_A \in \mathbb{R}^{T\times 1}$ assigns an attention weight to each timestep and satisfies $T_{A,j} \geq 0$ and $\sum_{j=1}^{T} T_{A,j} = 1$. Since event-based streams exhibit sparse and irregular temporal distributions, T_A explicitly captures this variability and provides a principled basis for the temporally selective accumulation mechanism used in SICAF.

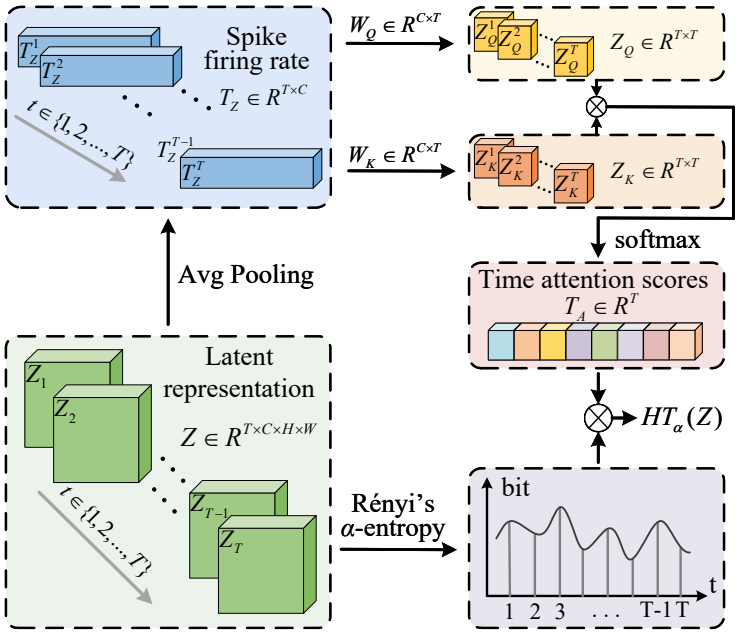


Figure 3: **Overview of the time-sensitive Rényi's α-entropy estimator.** Temporal attention scores are derived from spike firing rates of latent representations to capture temporal heterogeneity. Rényi's α-entropy provides a robust and flexible measure of information content, enabling selective feature compression when combined with temporal attention.

4.2 Time-Sensitive Rényi's α-Entropy Estimator

We construct a time-sensitive Rényi's α-entropy estimator $HT_\alpha(Z)$, as illustrated in Fig 3, and propose a spike-rate-based temporal selection information bottleneck. In information theory, Rényi's

α-entropy can measure the amount of information contained in a variable while exhibiting better numerical stability and scalability compared to Shannon entropy. It thus provides a robust and flexible theoretical tool for estimating high-dimensional temporal information features.

Given a latent activation tensor Z, let $z = \mathrm{vec}(Z) \in \mathbb{R}^M$ denote its flattened representation, where M is the number of entries in Z. We construct a discrete probability distribution over the activation coordinates using softmax normalization:

$$p_i(Z) = \frac{\exp(z_i)}{\sum_{j=1}^{M} \exp(z_j)}, \qquad i = 1, \ldots, M. \tag{7}$$

We define the Rényi α-entropy regularizer of this distribution as

$$H_\alpha(Z) = \frac{1}{1-\alpha} \ln \left(\sum_{i=1}^{M} p_i(Z)^\alpha \right), \qquad \alpha > 0,\ \alpha \neq 1. \tag{8}$$

In the subsection 4.1, we compute the temporal attention scores based on spike-rate temporal information, thereby constructing a time-sensitive Rényi's α entropy estimator

$$HT_\alpha(Z) = \sum_{j=1}^{T} T_{A,j} H_\alpha(Z_j), \tag{9}$$

This estimator flexibly evaluates the information content of the latent variable across the temporal dimension, adaptively assigning different attention weights. Thus, in the entropy estimation process, it preserves the temporal heterogeneity of the latent variable.

4.3 SICAF Learning Strategy

The IB framework achieves redundant information compression by minimizing the mutual information $I(Z; X)$ between the input X and the latent representation Z. However, most existing SNN-based IB approaches fail to consider the intrinsic temporal characteristics of SNNs, leading to limited advantages in temporal sequence tasks such as optical flow estimation. To address this limitation, we construct a time-sensitive Rényi's α-entropy estimator and propose a spike-based temporal attention information bottleneck learning strategy.

We introduce Y as an abstract task-output variable for the theoretical formulation; it denotes neither available ground-truth optical flow labels nor explicitly generated pixel-wise pseudo-labels.For input X and output Y, we use an encoder $E(Z|X)$ that projects the input X into the latent space to obtain a representation Z. The objective of SICAF is to realize **Compactly Compress and Time-varying Focus** during this mapping process. Specifically, **Compactly Compress** imposes an upper-bound constraint on the mutual information between X and Z: $I(X; Z) \leq I_c$, while **Time-varying Focus** enforces a temporal modulation constraint during compression, expressed as: $HT_\alpha(Z) = \sum_{i=1}^{T} T_{A,i} H_\alpha(Z_i) \geq H_c$, where the learning process is described as an optimization problem with temporal modulation constraints.

Assuming that the mapping between the latent representation Z and the output Y is defined by $q(y|z)$, the optimization objective can be expressed as:

$$\begin{aligned} \mathcal{L}(q, E) &= \min_{E,q} \ \mathbb{E}_{x,y\sim p(x,y)} \big[\mathbb{E}_{z\sim E(z|x)} [-\log q(y|z)] \big] \\ s.t. \quad & I(X; Z) \leq I_c, \quad HT_\alpha(Z) \geq H_c, \end{aligned} \tag{10}$$

where the computation of $HT_\alpha(Z)$ is detailed in subsection 4.2. The mutual information $I(X; Z)$ is defined as:

$$\begin{aligned} I(X; Z) &= \int p(x, z) \log \frac{p(x, z)}{p(x)p(z)} \, dxdz \\ &= \int p(x) \mathbb{E}_{E(z|x)} \log \frac{E(z|x)}{p(z)} \, dxdz, \end{aligned} \tag{11}$$

where $p(x)$ and $p(z)$ denote the marginal distributions of the input variable and the latent variable, respectively. Since the distribution $p(z)$ of the latent variable is often intractable, we approximate it

with a variational distribution $r(z)$ that satisfies the KL divergence constraint $\mathrm{KL}[p(z)\|r(z)] \geq 0$. Hence,

$$\int p(z) \log p(z)\, dz \geq \int p(z) \log r(z)\, dz. \tag{12}$$

Using this inequality, we obtain an upper bound for the mutual information $I(X; Z)$:

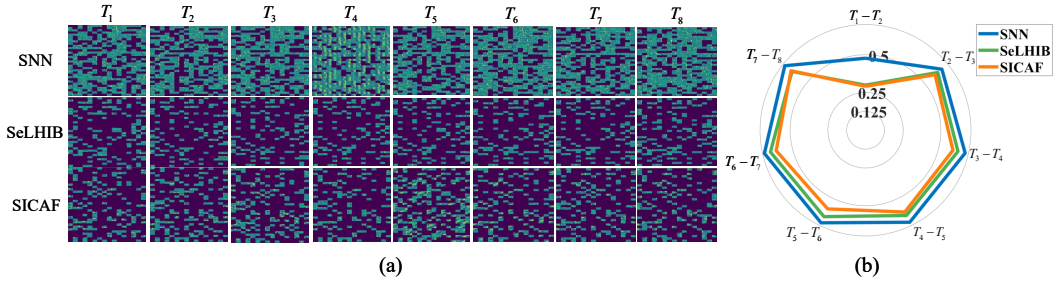
$$\begin{aligned} I(X; Z) &\leq \int p(x) \mathbb{E}_{E(z|x)} \log \frac{E(z|x)}{r(z)}\, dxdz \\ &= \mathbb{E}_{x\sim p(x)}[\mathrm{KL}[E(z|x)\|r(z)]]. \end{aligned} \tag{13}$$

Therefore, the constrained optimization objective can be approximated by the following upper bound:

$$\begin{aligned} \widetilde{\mathcal{L}}(q, E) = \min_{E,q}\ & \mathbb{E}_{x,y\sim p(x,y)}\left[\mathbb{E}_{z\sim E(z|x)}[-\log q(y|z)]\right] \\ s.t.\quad & \mathbb{E}_{x\sim p(x)}[\mathrm{KL}[E(z|x)\|r(z)]] \leq I_c, \\ & HT_\alpha(Z) \geq H_c. \end{aligned} \tag{14}$$

For the model training process, we introduce regularization coefficients β and γ to transform the constrained optimization into an unconstrained training objective:

$$\begin{aligned} J_{\text{SICAF}} = \min_{E,q}\ & \mathbb{E}_{x,y\sim p(x,y)}\left[\mathbb{E}_{z\sim E(z|x)}[-\log q(y|z)]\right] \\ & + \beta \cdot \left(\mathbb{E}_{x\sim p(x)}[\mathrm{KL}[E(z|x)\|r(z)]] - I_c\right) \\ & - \gamma \cdot \left(HT_\alpha(Z) - H_c\right). \end{aligned} \tag{15}$$



(a) (b)

Figure 4: (a) Comparison of feature maps in latent space. Both SeLHIB and our SICAF demonstrates sparser feature maps compared to baseline self-supervised SNN, indicating the acquisition of more structured spatiotemporal features. (b) The radar charts of SSIM for self-supervised SNN, SeLHIB, and SICAF, respectively.

4.4 Mechanistic Analysis of Temporal Heterogeneity in SICAF

To evaluate the effectiveness of SICAF in achieving compact compression and time-varying focusing for irregular event streams, we visualize the latent feature maps of trained models and measure the structural similarity of their neuronal firing patterns across consecutive time steps. As shown in Fig. 4(a), the SNN baseline exhibits the densest activations, whereas both SeLHIB and SICAF produce noticeably sparser patterns. This suggests that both of these two strategies encourage more compact and better-structured feature representations.

To more intuitively examine how the firing patterns of the three methods vary across time steps, we plot radar charts of structural similarity (SSIM) to measure the similarity between adjacent time steps. The SSIM is defined as:

$$\mathrm{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \tag{16}$$

where μ_x, σ_x^2, μ_y, and σ_y^2 denote the mean and variance of images x and y, σ_{xy} is their covariance, and C_1 and C_2 are stabilization constants. As shown in Fig. 4(b), SICAF achieves consistently lower

SSIM scores than both the conventional self-supervised SNN and the SeLHIB baseline. Because SSIM measures structural similarity between temporally adjacent firing maps, lower SSIM indicates stronger temporal decorrelation, implying that the network produces activation patterns with higher temporal resolution. In essence, the lower SSIM values observed in SICAF reflect a theoretically grounded effect: SICAF enforces temporal diversity during minimizing redundant dependencies, thereby promoting a temporally selectivity regime that aligns with the irregular statistics of event-based optical flow.

5 Experiments

5.1 Experimental Setup

We demonstrate the effectiveness of SICAF on event-based optical flow estimation task with different datasets, including multi-vehicle stereo event-camera (MVSEC), high quality frames (HQF) [24], and event camera dataset (ECD) [25]. All experiments were conducted using the PyTorch framework on RTX 4090 and A6000 GPUs. Adam optimizer is used with learning rate of 0.0002 and batch size of 8 for 50 epochs. Nearly 15 minutes of event data are contained in the training sequences, and they are split into 140 sequences with 128×128 size with 500k events in each sequence. Optical flow predictions are produced at each frame timestep with dt=1, i.e., 25% of the window events at a time.

5.2 Performance Evaluation on Event-based Optical Flow Estimation

We evaluated the generalization ability of SICAF in the self-supervised optical flow estimation task across the MVSEC, ECD, and HQF datasets, and compared it with the SOTA methods. The results on the MVSEC dataset are shown in Table 1, while the results on the ECD and HQF datasets are shown in Appendix A.8. The results demonstrate that SICAF significantly improves the accuracy of SNN in self-supervised optical flow estimation tasks. Specifically, compared to the work that applies IB to the SNN self-supervised optical flow task [16], SICAF achieves a 8.16% improvement, which suggests that the integration of compact compression and time-varying focusing enables SICAF to achieve compact yet selective feature extraction, improving the model's generalization ability. Furthermore, we compared the qualitative optical flow estimation results of SICAF and the method without SICAF, as shown in Figure 5. From the comparison of the flow estimation results, it is noticeable that SICAF provides more accurate and consistent motion direction predictions, while preserving finer motion details and producing clearer and more complete motion contours.

Table 1: Experimental results of SICAF on MVSEC compared to previous work. (lower is better, best in red, runner-up in blue).

Methods	Average Endpoint Error (AEE) ↓				
	Indoor (flying1)	Indoor (flying2)	Indoor (flying3)	Outdoor (day1)	Avg.
LIF-EV-FlowNet [6]	0.71	1.44	1.16	0.53	0.96
LIF-FireNet [6]	0.98	1.77	1.50	0.57	1.21
ALIF-EV-FlowNet [6]	1.00	1.78	1.55	0.57	1.23
PLIF-EV-FlowNet [6]	0.75	1.52	1.23	0.60	1.03
XLIF-EV-FlowNet [6]	0.73	1.45	1.17	0.45	0.95
ALIF-FireNet [6]	1.04	1.85	1.58	0.62	1.27
PLIF-FireNet [6]	0.90	1.67	1.41	0.56	1.14
XLIF-FireNet [6]	0.98	1.82	1.54	0.54	1.22
Adaptive-SpikeNet-ANN [7]	0.84	1.59	1.36	0.48	1.07
Adaptive-SpikeNet-SNN [7]	0.79	1.37	1.11	0.44	0.93
SeLHIB [16]	0.74	1.48	1.21	0.50	0.98
FSFN [15]	0.76	1.19	1.00	0.50	0.86
EV-FlowNet + SICAF	0.71	1.37	1.09	0.43	0.90
FSFN + SICAF	0.73	1.09	0.93	0.42	0.79

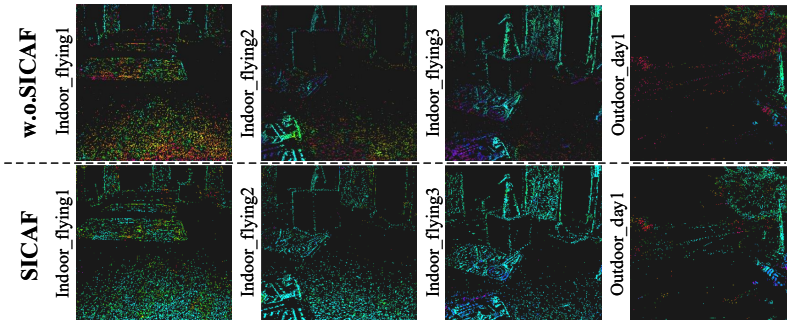


Figure 5: Qualitative evaluation of optical flow estimation results on sequences from the MVSEC dataset.

5.3 Robustness of Event-based Optical Flow Estimation in Noisy Situations

Due to environmental interference, event cameras often generate spurious events caused by non-physical brightness changes. These noisy events can interfere with the model's ability to extract accurate motion features. To further investigate the impact of SICAF on the robustness of SNNs in self-supervised optical flow estimation, we artificially injected various types of noise into the input event streams and compared the results with a baseline model, as shown in Table 2. The results demonstrate that SICAF significantly enhances robustness under both random and adversarial noise, achieving a 13.49% improvement in accuracy under white-box adversarial conditions. Additional experimental results on the ECD and HQF datasets are provided in Appendix A.9. SICAF enhances robustness under white-box adversarial attacks through its time-sensitive adaptive compression mechanism. The time-varying compression framework enables SICAF to dynamically focus on critical events while suppressing irrelevant or noisy data, leading to more robust and structured feature extraction and thus making it less susceptible to adversarial perturbations. It also adjusts the weighting of temporal features, further aids in capturing temporal heterogeneity and reducing the impact of adversarial noise.

Table 2: Experimental results of SICAF on MVSEC under different noise (lower is better, best in red, runner-up in blue).

Noise type	Methods	AEE				
		Indoor (flying1)	Indoor (flying2)	Indoor (flying3)	Outdoor (day1)	Average
SP	LIF-EV-FlowNet [6]	0.91	1.68	1.39	0.56	1.14
	XLIF-EV-FlowNet [6]	0.86	1.76	1.21	0.62	1.11
	SICAF-LIF	0.82	1.44	1.20	0.60	1.02
	SICAF-XLIF	0.75	1.43	1.18	0.44	0.95
Gaussian	LIF-EV-FlowNet [6]	0.95	1.72	1.44	0.62	1.18
	XLIF-EV-FlowNet [6]	0.93	1.70	1.39	0.62	1.16
	SICAF-LIF	0.83	1.49	1.25	0.59	1.04
	SICAF-XLIF	0.94	1.68	1.42	0.53	1.14
Black-box attack	LIF-EV-FlowNet [6]	1.15	1.96	1.72	0.79	1.41
	XLIF-EV-FlowNet [6]	0.93	1.79	1.78	0.73	1.31
	SICAF-LIF	1.03	1.80	1.54	0.69	1.27
	SICAF-XLIF	1.06	1.82	1.58	0.70	1.29
White-box attack	LIF-EV-FlowNet [6]	1.02	1.77	1.52	0.68	1.25
	XLIF-EV-FlowNet [6]	0.91	1.82	1.49	0.71	1.23
	SICAF-LIF	0.86	1.53	1.36	0.59	1.09
	SICAF-XLIF	0.90	1.63	1.39	0.58	1.13

5.4 Ablation on Loss Function Components

To investigate the roles of various components of loss functions, we compare the performance of $L_{contrast}$ and SICAF w.o.L_{smooth}, with w.o.SICAF and SICAF on MVSEC. The results in Table 3 show that adding L_{smooth} to $L_{contrast}$ improves the estimation accuracy, indicating that it helps denoise after contrast enhancement, reducing the impact of irrelevant information, and thus extracting features more accurately. When SICAF is used without smoothing, the model's accuracy decreases, possibly because high-contrast noise is misinterpreted as valid motion information, leading to feature extraction errors. To mitigate this effect, adding L_{smooth} is crucial for SICAF.While the above study

confirms the necessity of all three loss terms, it does not isolate the contributions of the internal mechanisms of SICAF itself. To address this, we further conduct a component-level ablation in Section 5.5.

Table 3: Ablation study on the various components of loss functions (lower is better, best in red, runner-up in blue).

Methods	$L_{contrast}$	L_{smooth}	SICAF	AEE				
				Indoor (flying1)	Indoor (flying2)	Indoor (flying3)	outdoor (day1)	Avg.
(a)	✓	✗	✗	0.77	1.50	1.25	0.56	1.02
(b)	✓	✓	✗	0.81	1.45	1.19	0.57	1.01
(c)	✓	✗	✓	0.83	1.60	1.31	0.54	1.07
(d)	✓	✓	✓	**0.71**	**1.37**	**1.09**	**0.43**	**0.90**

5.5 Ablation on Internal Components of SICAF

To investigate the roles of the key components of SICAF, we compare the performance of SICAF, SICAF under uniform compression, SICAF w.o. compact compression, and SICAF w.o. time-varying focus on MVSEC. The results in Table 4 show that SICAF achieves the lowest AEE across all four sequences through its temporally sensitive information-compression strategy, demonstrating the effectiveness of the proposed method.

When the temporally varying compression mechanism is replaced with uniform compression, the model's accuracy decreases, possibly because a uniform information-compression strategy is unable to extract compact and critical features effectively from event streams with temporal heterogeneity, where excessive compression of temporal segments containing important motion information leads to an increase in prediction error. To mitigate this effect, the temporally varying compression is crucial for SICAF.

Removing either of the two key components of SICAF also degrades the prediction accuracy, indicating that the compact compression mechanism suppresses redundant information, while the temporally sensitive entropy constraint $HT_\alpha(Z)$ preserves critical temporal information. These two components work collaboratively in event-based optical flow estimation and are both indispensable.

Table 4: Ablation study on the key components of SICAF on the MVSEC dataset (lower is better, best in red, runner-up in blue).

Methods	Compact Compression	Time-Varying Focus	Uniform Compression	AEE				
				Indoor (flying1)	Indoor (flying2)	Indoor (flying3)	Outdoor (day1)	Avg.
(a)	✓	✓	✗	**0.72±0.031**	**1.38±0.026**	**1.09±0.03**	**0.47±0.053**	**0.91±0.012**
(b)	✓	✗	✓	0.81±0.047	1.53±0.044	1.28±0.064	0.55±0.041	1.04±0.019
(c)	✗	✓	✗	0.81±0.075	1.51±0.029	1.24±0.010	0.56±0.061	1.03±0.043
(d)	✓	✗	✗	0.81±0.021	1.52±0.059	1.28±0.025	0.61±0.055	1.06±0.036

6 Conclusion

In this paper, we propose SICAF to address the challenges faced by SNNs in self-supervised optical flow estimation. SICAF introduces a novel "Compactly Compress, Time-varying Focus" training strategy and designs an estimator equipped with learnable temporal attention. Through a temporally modulated constrained optimization process, SICAF enables SNNs to extract compact latent representations while preserving the temporal heterogeneity of input events, thereby enhancing their ability to capture irregular spatiotemporal structures in event streams. Overall, SICAF provides a new pathway toward more efficient feature extraction and enhanced robustness in SNNs.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant Nos. 62622610, 62376185 and U25A20540.

References

[1] J. Cuadrado, U. Rançon, B. R. Cottereau, and et al., "Optical flow estimation from event-based cameras and spiking neural networks," *Frontiers in Neuroscience*, vol. 17, p. 1160034, 2023.

[2] Z. Ding, R. Zhao, J. Zhang, and et al., "Spatio-temporal recurrent networks for event-based optical flow estimation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 1, 2022, pp. 525–533.

[3] F. Paredes-Vallés, K. Y. W. Scheper, C. De Wagter, and et al., "Taming contrast maximization for learning sequential, low-latency, event-based optical flow," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 9695–9705.

[4] G. Gallego, H. Rebecq, and D. Scaramuzza, "A unifying contrast maximization framework for event cameras, with applications to motion, depth, and optical flow estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 3867–3876.

[5] Y. Schnider, S. Woźniak, M. Gehrig, J. Lecomte, A. von Arnim, L. Benini, D. Scaramuzza, and A. Pantazi, "Neuromorphic optical flow and real-time implementation with event cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023, pp. 4129–4138.

[6] J. Hagenaars, F. Paredes-Vallés, and G. De Croon, "Self-supervised learning of event-based optical flow with spiking neural networks," in *Advances in Neural Information Processing Systems*, vol. 34, 2021, pp. 7167–7179.

[7] A. K. Kosta and K. Roy, "Adaptive-spikenet: Event-based optical flow estimation using spiking neural networks with learnable neuronal dynamics," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 6021–6027.

[8] C. Lee, A. K. Kosta, A. Z. Zhu, and et al., "Spike-flownet: Event-based optical flow estimation with energy-efficient hybrid neural networks," in *European Conference on Computer Vision*, 2020, pp. 366–382.

[9] H. Sun, J. Wang, W. Cai, D. Chen, Q. Liao, J. He, Y. Cui, D. Yao, and D. Guo, "ST-FlowNet: An efficient spiking neural network for event-based optical flow estimation," *Neural Networks*, vol. 190, p. 107730, 2025.

[10] M. Dampfhoffer, T. Mesquida, A. Valentian, and L. Anghel, "Backpropagation-based learning techniques for deep spiking neural networks: A survey," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 9, pp. 11 906–11 921, 2024.

[11] G. Gallego, T. Delbrück, 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 154–180, 2022.

[12] S. Barchid, J. Mennesson, J. Eshraghian, and et al., "Spiking neural networks for frame-based and event-based single object localization," *Neurocomputing*, vol. 559, p. 126805, 2023.

[13] G. Qiao, N. Ning, Y. Zuo, P. Zhou, M. Sun, S. Hu, Q. Yu, and Y. Liu, "Spatio-temporal fusion spiking neural network for frame-based and event-based camera sensor fusion," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 8, no. 3, pp. 2446–2456, 2024.

[14] F. Paredes-Vallés, K. Y. W. Scheper, and G. C. H. E. D. Croon, "Unsupervised learning of a hierarchical spiking neural network for optical flow estimation: From events to global motion perception," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 8, pp. 2051–2064, 2019.

[15] M. P. E. Apolinario and K. Roy, "S-TLLR: STDP-inspired temporal local learning rule for spiking neural networks," *Transactions on Machine Learning Research*, 2025. [Online]. Available: https://openreview.net/forum?id=CNaiJRcX84

[16] S. Yang, B. Linares-Barranco, Y. Wu, and B. Chen, "Self-supervised high-order information bottleneck learning of spiking neural network for robust event-based optical flow estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 47, no. 4, pp. 2280–2297, 2025, dOI: 10.1109/TPAMI.2024.3510627. [Online]. Available: https://doi.org/10.1109/TPAMI.2024.3510627

[17] C. Lee, A. K. Kosta, and K. Roy, "Fusion-flownet: Energy-efficient optical flow estimation using sensor fusion and deep fused spiking-analog network architectures," in *2022 International Conference on Robotics and Automation (ICRA)*. IEEE, 2022, pp. 6504–6510.

[18] Y. Wu, L. Deng, G. Li, and et al., "Direct training for spiking neural networks: Faster, larger, better," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 1311–1318.

[19] H. Zheng, Y. Wu, L. Deng, and et al., "Going deeper with directly-trained larger spiking neural networks," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 35, no. 12, 2021, pp. 11 062–11 070.

[20] S. Yang, J. Tang, T. Zhou, B. Linares-Barranco, G. Indiveri, X. Zheng, M. Rahimi Azghadi, and B. Chen, "Efficient training of neuromorphic electronics," *Nature Electronics*, vol. 9, pp. 594–610, 2026.

[21] S. Yang and B. Chen, "Effective surrogate gradient learning with high-order information bottleneck for spike-based machine intelligence," *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[22] K. Wu, E. Shunzhuo, N. Yang, and et al., "A novel approach to enhancing biomedical signal recognition via hybrid high-order information bottleneck driven spiking neural networks," *Neural Networks*, vol. 183, p. 106976, 2025.

[23] G. Gallego, M. Gehrig, and D. Scaramuzza, "Focus is all you need: Loss functions for event-based vision," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 12 280–12 289.

[24] T. Stoffregen, C. Scheerlinck, D. Scaramuzza *et al.*, "Reducing the sim-to-real gap for event cameras," in *Computer Vision–ECCV 2020*. Springer, 2020, pp. 534–549.

[25] E. Mueggler, H. Rebecq, G. Gallego *et al.*, "The event-camera dataset and simulator: Event-based data for pose estimation, visual odometry, and slam," *The International Journal of Robotics Research*, vol. 36, no. 2, pp. 142–149, 2017.

[26] A. Z. Zhu, D. Thakur, T. Özaslan *et al.*, "The multivehicle stereo event camera dataset: An event camera dataset for 3d perception," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 2032–2039, 2018.

A Technical appendices and supplementary material

A.1 Broader Impacts

This paper focuses on improving the compression process of traditional IB to achieve both compactness and diversity of features, thereby enhancing the generalization and robustness of SNN models in optical flow estimation tasks. In general, this work has no negative societal impacts.

A.2 Limitation

The proposed SICAF framework assists SNNs in achieving both compactness and diversity in feature extraction for optical flow estimation tasks. It addresses the issue of information loss that arises from traditional IB methods. However, in practical applications, SICAF may require parameter tuning to achieve better accuracy, which could be time-consuming.

A.3 Experiment Settings

All experiments were conducted using the PyTorch framework on RTX 4090 and A6000 GPUs. Adam optimizer is used with learning rate of 0.0002 and batch size of 8 for 50 epochs. Gradients are clipped based on a global norm of 100, and the number of events for each input partition is set to 1000. 10,000 events are used for each training event partition. The scaling weight for L_{smooth} is set as 0.001. Nearly 15 minutes of event data are contained in the training sequences, and they are split into 140 sequences with 128×128 size with 500k events in each sequence. Random horizontal, vertical, and polarity flips are used to augment data. The SNN employs a direct training approach using surrogate gradients. MVSEC dataset [26] is first used for evaluation with the ground-truth data. Optical flow predictions are produced at each frame timestep with *dt*=1, i.e., 25% of the window events at a time. To compare with ground-truth data, we convert predicted optical flow from units of pixels/partitions to pixel displacement by a multiplication factor dt_{gt}/dt_{input}. ECD and HQF datasets are also used for evaluation of SICAF. Since there is no ground-truth data for these two datasets, the quality of the estimated optical flow is evaluated according to metrics based on the contrast maximization framework [4, 23]. The average test accuracy over two runs under different random seeds and is reported as final test accuracy.

For robustness evaluation, we consider both black-box and white-box adversarial attacks. For the black-box attack, following "Adversarially Robust Spiking Neural Networks Through Conversion", we adopt a random-direction zeroth-order gradient estimation strategy. Candidate adversarial inputs are constructed using independent Gaussian random directions $\delta \sim \mathcal{N}(0, I)$ as $x' = x + \epsilon\delta$. The number of queries is set to 5, the perturbation magnitude is set to $\epsilon = 0.1$, and the perturbed input is finally clipped to the range $[0, 1]$. For the white-box attack, following "A Comprehensive Analysis on Adversarial Robustness of Spiking Neural Networks", we employ the FGSM. The perturbation is generated according to the sign of the gradient of the loss with respect to the input, with the perturbation magnitude set to $\epsilon = 0.1$.

A.4 Additional Details on SNN Neurons

In LIF neurons, when the membrane potential V reaches the threshold m, the neuron generates an output spike. The membrane potential threshold can be either a fixed value or can vary with the input. To better handle the complex motion changes in optical flow estimation, we set the membrane potential threshold m and the leakage parameter λ of the LIF neurons as learnable parameters to achieve improved performance. In addition, we adjust the neuron model to reduce adaptation delay and adapt more quickly to the rapidly changing motion characteristics in optical flow estimation. The output current I and membrane potential threshold m of the adjusted neuron are as follows:

$$I_i^k = \sum_i W_{ij}^f \delta_j^k + \sum_l W_{il}^r u_j \delta_l^{k-1} - \kappa P_i^k, \tag{17}$$

$$P_i^k = \eta P_i^{k-1} + \frac{1-\eta}{|R_i|} \sum_{j \in R_i} \delta_j^{k-1}, \tag{18}$$

$$m_i^k = \alpha_0 + \alpha_1 P_i^k, \tag{19}$$

where R_i is the set of receptive fields of neuron i across all channels. For convenience, we list the hyperparameters of the neurons and provide their corresponding initialization value ranges, as shown in Table 5.

A.5 Pseudocode

The overall network in this paper processes event streams through four strided convolutional encoders, two residual blocks, and four decoder layers. SICAF trains the network parameters by maximizing the conditional entropy of intermediate variables, balancing the feature compactness and diversity of latent space.

Table 5: SNN hyperparameter initialization table

Hyperparameter	Initialization range
α_0 (learnable)	0.01
α_1 (learnable)	1
ν (learnable)	$\nu = \frac{1}{1+\exp(-a)}, a \in \mathcal{N}(-4, 0.1)$
δ	$\delta \in \{0, 1\}$
η (learnable)	$\eta = \frac{1}{1+\exp(-c)}, c \in \mathcal{N}(-4, 0.1)$
m	0.8
κ (learnable)	0.01
P	0

Algorithm 1: The training process of the SICAF structure learning framework

Input: Event data (X)
Output: Updated network weights (W); updated microscopic parameters (λ', γ')
Parameters: SNN neuron parameters (λ', γ'); SICAF hyperparameters (λ, γ); batch size (B); training epochs (TE); learning rate (r); network depth (H); neuronal membrane potential V and spike output O;
Forward:;
Initialize the model and optimizer;
Preprocess the input data to obtain batch data;
for $i \leftarrow 1$ **to** TE **do**
  Fetch a mini-batch of size B;
  for $h \leftarrow 1$ **to** $H-1$ **do**
    Calculate V_h and O_h;
  Calculate V_H and $\hat{Y}$;
  `// Calculate the loss`
  $L_{\text{loss}} \leftarrow \text{SICAF}\big(L_{\text{flow}}(X, \hat{Y}), \lambda, I(X; Z), \gamma, HT_\alpha(Z)\big)$;
Backward:;
for $h \leftarrow H$ **to** 1 **do**
  if $h \neq H$ **then**
    `// Update the network weights`
    $\Delta W_h = \sum_k \frac{\partial L_{\text{loss}}}{\partial O_h^k} \frac{\partial O_h^k}{\partial U_h^k} \frac{\partial U_h^k}{\partial W_h^k}$;
    $W_h \leftarrow W_h - r\Delta W_h$;
    `// Update the microscopic parameters`
    $\Delta \lambda'_h = \sum_k \frac{\partial L_{\text{loss}}}{\partial O_h^k} \frac{\partial O_h^k}{\partial U_h^k} \frac{\partial U_h^k}{\partial \lambda'_h}$;
    $\Delta \beta'_h = \sum_k \frac{\partial L_{\text{loss}}}{\partial O_h^k} \frac{\partial O_h^k}{\partial U_h^k} \frac{\partial U_h^k}{\partial \beta'_h}$;
    $\lambda'_h \leftarrow \lambda'_h - r\Delta \lambda'_h$;
    $\beta'_h \leftarrow \beta'_h - r\Delta \beta'_h$;
  else
    $\Delta W_h = \sum_k \frac{\partial L_{\text{loss}}}{\partial O_h^k} \frac{\partial O_h^k}{\partial U_h^k} \frac{\partial U_h^k}{\partial W_h^k}$;
    $W_h \leftarrow W_h - r\Delta W_h$;

Meanwhile, the microscopic parameters in adaptive neurons adjust throughout the training process, further optimizing the network's weight parameters. The general training process is outlined in Algorithm 1.

A.6 Implementation Details of the SICAF Objective

The encoder output is denoted by $\mu_\phi(x)$, and the variational encoding distribution is defined as

$$E_\phi(z \mid x) = \mathcal{N}\left(z; \mu_\phi(x), \sigma^2 I\right), \qquad \sigma^2 = \exp(\text{logvar}_t) + 10^{-10}, \tag{20}$$

where logvar_t is learnable and initialized to -1. During training, we directly use $\mu_\phi(x)$ as the bottleneck representation to reduce estimation variance. For a mini-batch of size B, we approximate the marginal distribution using kernel density estimation (KDE):

$$r(z) = \frac{1}{B}\sum_{j=1}^{B}\mathcal{N}\left(z;\mu_j,\sigma^2 I\right), \tag{21}$$

where $\mu_j = \mu_\phi(x_j)$. With $D_{ij} = \|\mu_i - \mu_j\|_2^2$, we use the following mean-based surrogate for the information-bottleneck term:

$$\widehat{I}(X;Z) = \log B - \frac{1}{B}\sum_{i=1}^{B}\log\left[\sum_{j=1}^{B}\exp\left(-\frac{D_{ij}}{2\sigma^2}\right)\right]. \tag{22}$$

This expression is obtained by evaluating $\log E_\phi(z \mid x_i) - \log r(z)$ at $z = \mu_i$, rather than computing the full expectation over $E_\phi(z \mid x_i)$. The connection to the information-bottleneck objective is given by the identity:

$$\mathbb{E}_x\left[\mathrm{KL}\left(E_\phi(z \mid x)\,\|\,r(z)\right)\right] = I(X;Z) + \mathrm{KL}\left(p(z)\,\|\,r(z)\right). \tag{23}$$

Using KDE to approximate $p(z)$, we neglect the marginal-distribution mismatch term and use the mean-based surrogate above in training. This introduces an approximation rather than an exact KL evaluation. For entropy regularization, we use the Rényi entropy of the softmax-normalized activation distribution defined in Eq. (8), with $\alpha = 2$. We set $\beta = 4.8 \times 10^{-3}$ and $\gamma = 4 \times 10^{-4}$. Moreover, $I_c = H_c = 0$ are fixed constants and therefore do not affect model gradients in the penalized objective with fixed β and γ.

A.7 Theoretical Guarantees of SICAF

In this section, we provide a generalization bound and a corresponding sample-complexity result for SICAF. Let

$$S = \{X_i\}_{i=1}^{n} \sim \mathcal{D}^n$$

be an i.i.d. training sample drawn from the data distribution $\mathcal{D}$. The encoder f_θ maps the input X to a latent representation

$$Z = f_\theta(X),$$

which is subsequently processed by a predictor h. We denote the population and empirical risks by

$$\mathcal{J}_{\mathrm{SICAF}}(h) = \mathbb{E}_{X\sim\mathcal{D}}\left[\ell\big(h(Z)\big)\right], \tag{24}$$

and

$$\widehat{\mathcal{J}}_{\mathrm{SICAF},S}(h) = \frac{1}{n}\sum_{i=1}^{n}\ell\big(h(Z_i)\big), \qquad Z_i = f_\theta(X_i), \tag{25}$$

respectively.

For each input X, let

$$\mathbf{a}(X) = \big(a_1(X),\ldots,a_T(X)\big)^\top \in \mathbb{R}^T$$

denote the temporal-attention vector produced by the temporal-attention estimator. We assume that

$$a_t(X) \geq 0, \qquad \sum_{t=1}^{T} a_t(X) = 1. \tag{26}$$

The corresponding time-sensitive Rényi entropy is

$$H_T^\alpha(Z) = \sum_{t=1}^{T} a_t(X) H_\alpha(Z_t). \tag{27}$$

We define the mean temporal-attention vector as

$$\bar{\mathbf{a}} = \mathbb{E}_{X\sim\mathcal{D}}\left[\mathbf{a}(X)\right], \tag{28}$$

and define the effective temporal-attention fluctuation by

$$\Phi_{\text{eff}} = \mathbb{E}_{X\sim\mathcal{D}}\left[\|\mathbf{a}(X) - \bar{\mathbf{a}}\|_2^2\right]. \tag{29}$$

Moreover, let

$$\Delta_H = \left|H_T^{\alpha}(Z) - \widehat{H}_T^{\alpha}(Z)\right| \tag{30}$$

denote the discrepancy between the population time-sensitive Rényi entropy and its finite-sample estimate.

Theorem 1 (Generalization bound of SICAF)**.** *Let $\mathcal{F}$ be the hypothesis class induced by the encoder f_θ and predictor h. Assume that:*

1. *the latent representation is norm-bounded, i.e.,*

$$\|Z\|_2 \le B \quad \textit{almost surely}; \tag{31}$$

2. *the loss function ℓ is L-Lipschitz with respect to the latent representation;*
3. *the temporal-attention vector satisfies* (26)*;*
4. *the time-sensitive Rényi entropy is bounded as*

$$0 \le H_T^{\alpha}(Z) \le H_{\max}; \tag{32}$$

5. *the loss fluctuation associated with temporal modulation is sub-Gaussian with variance proxy σ^2.*

Then, for any $\delta \in (0,1)$, with probability at least $1-\delta$ over the draw of S, the following bound holds:

$$\begin{aligned}\mathcal{J}_{\text{SICAF}}(h) - \widehat{\mathcal{J}}_{\text{SICAF},S}(h) \le \sqrt{\frac{2L^2B^2}{n}\left(I(X;Z) + \ln\frac{1}{\delta}\right)} \\ + \sqrt{2\sigma^2\Phi_{\text{eff}}} + \gamma\Delta_H.\end{aligned} \tag{33}$$

Here, $I(X;Z)$ denotes the mutual information between the input and the latent representation, Φ_{eff} quantifies the variability of temporal attention across input samples, and Δ_H denotes the finite-sample estimation error of the time-sensitive Rényi entropy.

Proof. We decompose the generalization gap into three components:

$$\mathcal{J}_{\text{SICAF}}(h) - \widehat{\mathcal{J}}_{\text{SICAF},S}(h) = \Delta_{\text{IB}} + \Delta_{\text{temp}} + \Delta_{\text{ent}}, \tag{34}$$

where Δ_{IB} is the representation-complexity term, Δ_{temp} is the additional deviation induced by temporal attention, and Δ_{ent} is the error caused by estimating the entropy regularizer from finitely many samples.

First, since ℓ is L-Lipschitz and $\|Z\|_2 \le B$, the information-theoretic generalization inequality gives, with probability at least $1-\delta$,

$$\Delta_{\text{IB}} \le \sqrt{\frac{2L^2B^2}{n}\left(I(X;Z) + \ln\frac{1}{\delta}\right)}. \tag{35}$$

This term is controlled by the information-bottleneck constraint used in SICAF. In particular, enforcing a smaller information budget for $I(X;Z)$ reduces the complexity contribution associated with the latent representation.

Second, SICAF employs an input-dependent temporal-attention vector $\mathbf{a}(X)$. Its deviation from the population mean $\bar{\mathbf{a}}$ introduces an additional fluctuation in the temporally weighted objective. By the assumed sub-Gaussianity of this fluctuation and the definition in (29), a concentration bound yields

$$\Delta_{\text{temp}} \le \sqrt{2\sigma^2\mathbb{E}_{X\sim\mathcal{D}}\left[\|\mathbf{a}(X) - \bar{\mathbf{a}}\|_2^2\right]} = \sqrt{2\sigma^2\Phi_{\text{eff}}}. \tag{36}$$

Therefore, a more stable temporal-attention distribution leads to a smaller value of Φ_{eff} and hence a tighter bound.

Third, the SICAF objective contains the entropy regularization term $-\gamma H_T^\alpha(Z)$. In practice, this population quantity is replaced by the finite-sample estimate $\widehat{H}_T^\alpha(Z)$. Therefore, the absolute deviation of the entropy contribution is bounded by

$$\begin{aligned}\Delta_{\text{ent}} &= \gamma \left| H_T^\alpha(Z) - \widehat{H}_T^\alpha(Z) \right| \\ &= \gamma \Delta_H.\end{aligned} \tag{37}$$

Under (32), this term is finite; in particular, $\Delta_H \le H_{\max}$ whenever both entropy quantities lie in $[0, H_{\max}]$.

Combining (35), (36), and (37) in (34) gives (33). □

Theorem 2 (Sample complexity of SICAF)**.** *Suppose that the assumptions of Theorem 1 hold and that*

$$I(X;Z) \le I_0. \tag{38}$$

Assume further that the total temporal-modulation contribution is bounded by

$$\sqrt{2\sigma^2 \Phi_{\text{eff}}} + \gamma \Delta_H \le \varepsilon_{\text{temp}}. \tag{39}$$

For any target tolerance ε satisfying

$$\varepsilon > \varepsilon_{\text{temp}}, \tag{40}$$

it is sufficient to choose

$$n \ge \frac{2L^2 B^2}{(\varepsilon - \varepsilon_{\text{temp}})^2} \left(I_0 + \ln \frac{1}{\delta} \right) \tag{41}$$

to guarantee, with probability at least $1-\delta$, that

$$\mathcal{J}_{\text{SICAF}}(h) - \widehat{\mathcal{J}}_{\text{SICAF},S}(h) \le \varepsilon. \tag{42}$$

Proof. By Theorem 1,

$$\begin{aligned}\mathcal{J}_{\text{SICAF}}(h) - \widehat{\mathcal{J}}_{\text{SICAF},S}(h) \le \sqrt{\frac{2L^2B^2}{n}\left(I(X;Z) + \ln \frac{1}{\delta} \right)} \\ + \sqrt{2\sigma^2 \Phi_{\text{eff}}} + \gamma \Delta_H.\end{aligned} \tag{43}$$

Using (38) and (39), we obtain

$$\mathcal{J}_{\text{SICAF}}(h) - \widehat{\mathcal{J}}_{\text{SICAF},S}(h) \le \sqrt{\frac{2L^2B^2}{n}\left(I_0 + \ln \frac{1}{\delta} \right)} + \varepsilon_{\text{temp}}. \tag{44}$$

To ensure that the right-hand side of (44) is at most ε, it suffices to require

$$\sqrt{\frac{2L^2B^2}{n}\left(I_0 + \ln \frac{1}{\delta} \right)} \le \varepsilon - \varepsilon_{\text{temp}}. \tag{45}$$

The right-hand side is positive by (40). Squaring both sides gives

$$\frac{2L^2B^2}{n}\left(I_0 + \ln \frac{1}{\delta} \right) \le (\varepsilon - \varepsilon_{\text{temp}})^2. \tag{46}$$

Rearranging (46) with respect to n yields (41). Therefore, (42) holds with probability at least $1-\delta$. □

Remark. Theorem 2 does not discard the temporal terms. Instead, it explicitly requires the temporal-attention fluctuation and entropy-estimation error to satisfy the error-budget condition in (39). The original mutual-information-only sample-complexity expression is recovered as a special case when $\varepsilon_{\text{temp}} = 0$, or approximately when the temporal contribution is negligible relative to the target generalization tolerance.

A.8 Additional Self-Supervised Optical Flow Estimation Results

In addition to validating the performance of SICAF on the MVSEC dataset, we also evaluated SICAF on the ECD and HQF datasets. For datasets lacking ground-truth labels, we adopted the RSAT metric to measure the quality of optical flow estimation by the model on these datasets. Its mathematical definition is as follows:

$$\text{RSAT} := \frac{\mathcal{L}_{\text{contrast}}(t_{\text{ref}}^{\text{fw}}|\mathbf{u})}{\mathcal{L}_{\text{contrast}}(t_{\text{ref}}^{\text{fw}}|\mathbf{0})}. \tag{47}$$

The contrast loss $\mathcal{L}_{\text{contrast}}(t_{\text{ref}})$ is defined as:

$$\mathcal{L}_{\text{contrast}}(t_{\text{ref}}) = \frac{\sum_x \bar{t}_{p'=+}(\mathbf{x};\mathbf{u}|t_{\text{ref}})^2 + \bar{t}_{p'=-}(\mathbf{x};\mathbf{u}|t_{\text{ref}})^2}{\sum_x \mathbb{I}[n(\mathbf{x}')>0] + \epsilon}, \tag{48}$$

where $\bar{t}_{p'=+}$ and $\bar{t}_{p'=-}$ represent the positive and negative timestamps respectively, $n(\mathbf{x}')$ indicates the number of events at position $\mathbf{x}'$, and ϵ is a small constant added for numerical stability. When RSAT < 1, it indicates that the predicted optical flow outperforms the baseline scenario composed of null vectors.

Table 6 and Figure 6 present the results for the ECD and HQF datasets, respectively. From the experimental results, it can be observed that SICAF also effectively enhances the model's generalization ability in self-supervised optical flow estimation on the ECD and HQF datasets. These results are attributed to the maximum entropy information bottleneck increasing the randomness of the intermediate variable T, allowing the model to extract useful information while giving deterministic prediction results a greater variance, thus improving the model's generalization.

Table 6: Experimental results on ECD and HQF datasets. (Higher FWL and lower RSAT indicate better performance. Best in red, runner-up in blue.)

Methods	ECD		HQF	
	FWL ↑	RSAT ↓	FWL ↑	RSAT ↓
LIF-EV-FlowNet [6]	1.21	0.95	1.24	0.94
ALIF-EV-FlowNet [6]	1.17	0.98	1.21	0.98
PLIF-EV-FlowNet [6]	1.24	0.95	1.28	0.93
XLIF-EV-FlowNet [6]	1.23	0.95	1.25	0.93
SICAF	1.26	0.94	1.30	0.92

A.9 Additional Details on Robustness Experiments

To assess the robustness of the model, we introduced random noise attacks and adversarial noise attacks during the model testing process. Random noise attacks involve using clean datasets without additional noise during model training and adding noise to the datasets during testing to simulate disturbances in real-world scenarios. In this study, random noise attacks mainly consisted of adding two common types of noise, Gaussian noise and salt-and-pepper (SP) noise. Gaussian noise with a mean of $\mu = 0$ and variance of σ can be expressed mathematically as follows:

$$f_{Gaussian}(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{49}$$

where x is a random variable, μ and σ represent its mean and variance, respectively. SP noise is composed of white and black pixels generated by severe disturbances. In addition, we explore adversarial noise attacks in two forms, i.e. white-box and black-box attacks. In white-box attack, the attacker possesses detailed information about the model, including data preprocessing, structure, and parameters, using this internal model information to generate adversarial samples. In our experiments, white-box attacks employed the fast gradient sign method (FGSM), which primarily utilizes the gradient of the loss function with respect to the input data x to generate adversarial noise. The expression for the generated noise is as follows:

$$\eta_{white-box} = \varepsilon \cdot sign(\nabla_x J(\theta, x, y)), \tag{50}$$

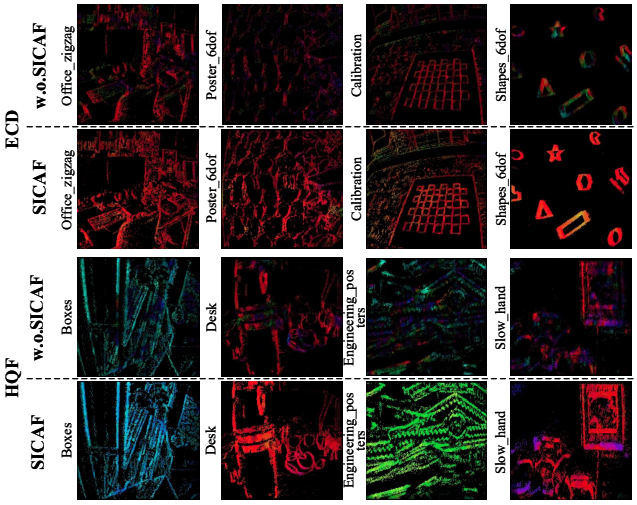


Figure 6: Qualitative evaluation of optical flow estimation results on sequences from the ECD and HQF datasets.

where the parameter ε represents the intensity of the added noise, and $J(\theta, x, y)$is the loss function. $\nabla_x J$ denotes the gradient of the loss function with respect to the input x. Using the FGSM algorithm, we obtain white-box adversarial noise and add it to the input dataset during the testing phase. Black-box attacks refer to scenarios where the attacker is unaware of the internal details and can only access the input and output information. Here, we use zero-order optimization methods to estimate gradients using experimental function values without directly computing or using gradient information. Zero-order optimization methods are a way to generate adversarial examples without explicit gradient information, suitable for black-box adversarial attacks. These methods optimize the input by querying the target function's output to make the model produce incorrect predictions, thus generating black-box adversarial samples. Suppose our model is f, and the input sample is x. The goal of generating adversarial attack samples is to maximize the error or loss of $f(x)$.

First, we initialize the input sample $x_{adv} = inputs$ and generate a random direction vector $\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, where $\mathcal{N}(\mathbf{0}, \mathbf{I})$ is drawn from a multivariate normal distribution with a mean of $\mathbf{0}$ and a covariance matrix of the identity matrix. The purpose of generating the vector $\boldsymbol{\delta}$ is to explore the solution space around the input sample through random search, thereby indirectly estimating the gradient direction and finding perturbation directions that can effectively generate adversarial samples.

We then compute the loss change for the sample when it is moved in the direction of $\boldsymbol{\delta}$. Let $L(inputs, target)$ denote the model's loss function, where *inputs* is the input sample and *target* is the true label of the sample. The change in loss when the sample is moved in the direction of $\boldsymbol{\delta}$ can be expressed as:

$$\Delta L = L(f(x_{adv} + \alpha\boldsymbol{\delta}), target) - L(f(x_{adv}), target), \tag{51}$$

where α denotes the perturbation strength. After obtaining the loss change in the direction of $\boldsymbol{\delta}$, we estimate the gradient and update the input sample as follows:

$$x_{adv} = x_{adv} + \alpha\boldsymbol{\delta} \cdot sign(\Delta L), \tag{52}$$

To ensure that the generated adversarial samples are perceptually similar to the original samples, we need to keep the perturbation within a certain range to maintain the stealthiness and effectiveness of the attack. Therefore, we apply pruning operations to the generated samples:

$$x_{adv} = clip(x_{adv}, inputs - \epsilon, inputs + \epsilon), \tag{53}$$

where ϵ denotes the perturbation range set for the pruning operation. Table 7 and Table 8 show the robustness test results of SICAF on the ECD and HQF datasets.

Table 7: Experimental results of SICAF on ECD under different noise (lower is better, best in red).

Noise type	**SP**		**Gaussian**		**Black-box attack**		**White-box attack**	
Methods	**[6]**	**SICAF**	**[6]**	**SICAF**	**[6]**	**SICAF**	**[6]**	**SICAF**
boxes	0.97	0.94	0.97	0.94	1.01	1.00	0.99	0.96
0calibration	0.97	0.96	0.98	0.96	1.00	1.00	0.99	0.98
dynamic_6dof	0.94	0.92	0.95	0.92	1.00	0.98	0.97	0.94
office_zigzag	0.97	0.96	0.98	0.97	1.00	1.00	0.99	0.98
poster_6dof	0.97	0.94	0.97	0.94	1.01	0.99	0.99	0.90
shapes_6dof	1.00	0.99	1.01	0.99	1.00	0.99	1.00	0.98
slider_depth	0.91	0.93	0.91	0.93	0.99	0.95	0.92	0.91
Mean	0.96	0.95	0.97	0.95	1.00	0.99	0.98	0.96

Table 8: Experimental results of SICAF on HQF under different noise (lower is better, best in red).

Noise type	**SP**		**Gaussian**		**Black-box attack**		**White-box attack**	
Methods	**[6]**	**SICAF**	**[6]**	**SICAF**	**[6]**	**SICAF**	**[6]**	**SICAF**
bike_bar_hdr	0.97	0.95	0.98	0.95	1.00	0.98	0.98	0.95
boxes	0.94	0.92	0.95	0.93	1.00	0.97	0.97	0.93
desk_6k	0.92	0.90	0.93	0.91	0.99	0.95	0.95	0.92
desk_fast	0.92	0.89	0.93	0.90	0.99	0.95	0.95	0.91
desk_hand_only	1.00	0.98	1.00	0.98	0.99	0.99	0.98	0.97
desk_slow	0.89	0.88	0.91	0.91	0.98	0.95	0.90	0.89
engineering_posters	0.96	0.94	0.97	0.95	0.99	0.95	0.99	0.95
high_texture_plants	0.94	0.92	0.98	0.96	1.00	0.98	0.96	0.92
poster_pillar_1	0.98	0.96	0.98	0.96	0.99	0.97	0.98	0.95
poster_pillar_2	0.97	0.95	0.98	0.95	0.99	0.97	0.98	0.95
reflective_materials	0.97	0.94	0.97	0.95	0.99	0.95	0.98	0.95
slow_and_fast_desk	0.96	0.94	0.97	0.95	0.98	0.95	0.98	0.95
slow_hand	0.97	0.95	0.97	0.95	0.99	0.95	0.99	0.94
still_life	0.94	0.93	0.95	0.93	0.99	0.98	0.93	1.00
Mean	0.95	0.93	0.96	0.94	0.99	0.96	0.97	0.93

A.10 Computational Efficiency Analysis

We evaluate the computational efficiency of SICAF against LIF-EV-FlowNet in terms of inference latency, inference energy per batch, memory usage, and the number of spikes per inference batch. Table 9 summarizes the results.

Table 9: Computational efficiency comparison between LIF-EV-FlowNet and SICAF. Lower values are preferable for all reported metrics.

Model	Latency (ms/batch)	Inference energy per batch (J)	Memory (MiB)	Spikes per inference batch
LF-EV-FlowNet	29.764 ± 7.881	4.354 ± 1.159	241.548 ± 6.885	$197{,}724.67 \pm 769.86$
SICAF	27.164 ± 6.645	3.873 ± 0.436	247.503 ± 6.582	$83{,}978.00 \pm 274.99$

As shown in Table 9, SICAF reduces the mean inference latency from 29.764 to 27.164 ms/batch and the mean inference energy from 4.354 to 3.873 J/batch, corresponding to reductions of approximately 8.7% and 11.0%, respectively. The mean number of spikes per inference batch decreases by approximately 57.5%, indicating sparser spiking activity. Meanwhile, memory usage increases from 241.548 to 247.503 MiB, an increase of approximately 2.5%. These results indicate lower average inference latency and energy consumption with a modest increase in memory usage.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: We have described the proposed method (SICAF) and the scope of experiments in the abstract and introduction.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have outlined the potential limitations in the Appendix A.2.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The theorem and its proof are provided in the Appendix.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We have provided the corresponding experimental details and some parameter settings in the Appendix, along with the pseudocode of the method.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: Our experiments use publicly available benchmark datasets, and the code will be released to ensure that others can reproduce the results.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: We have provided a detailed description of our experimental setup in the Appendix.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: It is worth noting that in the experimental section, we compare our results with the best-reported results from other works.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).

- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: We have specified the GPUs used in the experiments in the Appendix: GeForce RTX 4090 and A6000.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: This work follows the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: This paper focuses on improving the compression process of traditional IB to achieve both compactness and diversity of features, thereby enhancing the generalization and robustness of SNN models in optical flow estimation tasks. In general, this work has no negative societal impacts.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.

- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper proposes a framework for optimizing the SNN model in optical flow estimation tasks. Neither the model nor the standard benchmark data used poses such risk.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We have properly cited the original papers corresponding to the code and datasets used in our work, and have ensured that their licenses and terms of use are respected.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The research does not involve human subjects, therefore IRB approval is not applicable.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The development of the core methods in this paper does not involve large language models.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.