PAIR-AWARE SELF-CONTRASTIVE LEARNING IN SPIKING NEURAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Self-Contrastive Forward-Forward (SCFF) learning enables layer-local representation learning from unlabeled data by constructing positive and negative pairs. However, conventional activity-based goodness evaluates the firing strength of a fused pair representation without explicitly assessing the correspondence between its constituent responses. This creates a mismatch between relation-based pair construction and activity-based goodness evaluation in spiking neural networks (SNNs). To address this limitation, we propose Pair-aware SCFF, which incorporates cross-branch spike correspondence into the local goodness function. Specifically, we evaluate observed spike co-activation relative to marginal firing activity and combine the resulting relational signal with evidence from activity-supported pairs. This formulation jointly accounts for neural response strength and cross-branch correspondence while preserving layer-wise local optimization without class-label supervision or network-wide error propagation. Experiments across six static-image and two event-based datasets demonstrate consistent improvements over a matched Spiking SCFF baseline, with accuracy gains of up to 10.71 percentage points. Controlled ablations distinguish the contributions of relation-aware goodness and pair-stream propagation, while pair-level analyses reveal more structured pair relations and improved class-level semantic organization in the learned representations. These results highlight the importance of explicitly evaluating cross-branch correspondence in local self-contrastive SNN learning.

1 INTRODUCTION

Spiking neural networks (SNNs) process information through discrete spike events, providing a computational framework for sparse and temporally structured data. Despite these characteristics, training SNNs typically relies on backpropagation through the full network. The local nature of spiking computation motivates learning rules that enable representations to be optimized without propagating errors throughout the network. Local credit assignment reduces the dependence on network-wide backward error transport and intermediate-state coupling, making the learning process more modular and better aligned with the distributed event-driven computation of SNNs. Learning useful representations without class-label supervision or network-wide error propagation therefore remains an important research direction. Early studies demonstrated that local spike-timing-dependent plasticity can extract visual features from unlabeled inputs (Diehl & Cook, 2015; Kheradpisheh et al., 2018). More recently, self-supervised learning has been explored for spiking representation learning across static-image and event-based tasks (Hagenaars et al., 2021; Bahariasl & Kheradpisheh, 2024; Li et al., 2024; Ororbia, 2024). These developments motivate learning objectives that exploit the structure of unlabeled data while supporting layer-local optimization in SNNs.

Forward-Forward (FF) learning (Hinton, 2022) provides a framework for layer-local optimization by training each layer to assign higher goodness to positive examples than to negative ones, without propagating an error signal through the entire network. FF framework is particularly attractive for SNNs because goodness can be constructed directly from local spiking activity. Recent studies have demonstrated the feasibility of FF learning in fully spiking networks (Terres-Escudero et al., 2025; Ghader et al., 2025). FF therefore allows each spiking layer to form a learning signal from locally available neural responses, avoiding network-wide backward credit assignment while retaining an explicit representation objective. A central challenge, however, is constructing informative positive and negative examples without class labels. Self-Contrastive Forward-Forward (SCFF) (Chen et al.,

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.

2025) addresses this challenge by constructing positive pairs from two observations of the same sample and negative pairs from observations of different samples. The resulting learning signal is derived from relationships between unlabeled observations rather than externally provided class information.

Nevertheless, a distinction remains between how self-contrastive examples are constructed and how their goodness is evaluated. Positive and negative examples are defined by the relationship between two observations, whereas conventional activity-based goodness evaluates the strength of the response in their fused representation. Pair construction thus specifies what the model should distinguish, but the goodness function does not explicitly assess the cross-branch correspondence underlying this distinction. In SNNs, similar firing levels can arise from different patterns of coincident and disjoint spike activation across channels and time steps. Consequently, pairs with different cross-branch activation patterns may receive similar goodness values when their fused firing strengths are comparable (Fig. 1(c)). This motivates a local goodness function that evaluates not only the strength of the neural response but also the correspondence between paired responses.

To this end, we propose Pair-Aware Self-Contrastive Forward-Forward learning for SNNs (Pair-aware SCFF). By retaining the two branch responses before fusion, we explicitly evaluate their observed spike co-activation relative to a reference determined by marginal firing activity. This comparison characterizes cross-branch correspondence beyond that suggested by the individual firing levels. However, a strong relative relation supported by only a few spikes may provide limited evidence for a pair. We therefore combine the relational signal with activity-supported pair evidence, integrating response strength and cross-branch correspondence into a unified local goodness function. The resulting objective enables layer-wise self-contrastive learning without class-label supervision or network-wide error propagation.

We evaluate Pair-aware SCFF on static-image and event-based datasets by comparing it with existing SNN representation learning methods and a matched Spiking SCFF baseline. Controlled ablations examine the contributions of relation-aware goodness and pair-stream propagation, while pair-level analyses investigate how the learned scores organize sample relationships and capture information beyond fused firing activity.

Our contributions are summarized as follows:

- We identify a limitation of activity-based goodness in spiking self-contrastive FF learning: although positive and negative examples are constructed from relationships between observations, their goodness is evaluated without explicitly assessing cross-branch spike correspondence.
- We develop a relation-aware goodness function that evaluates observed spike co-activation relative to marginal firing activity and combines relational information with activity-supported pair evidence, enabling layer-wise local self-contrastive learning in SNNs.
- We demonstrate consistent improvements over a matched Spiking SCFF baseline across static-image and event-based benchmarks. Controlled ablations separate the contributions of goodness formulation and pair-stream propagation, while pair-level analyses characterize how relation-aware learning changes pair organization and frozen representation quality.

2 Related Work

2.1 Unsupervised and Self-Supervised Learning in SNNs

Unsupervised representation learning in spiking neural networks (SNNs) has traditionally been studied through local synaptic plasticity. Competitive learning based on spike-timing-dependent plasticity (STDP) enables visual feature extraction from unlabeled inputs (Diehl & Cook, 2015), while convolutional STDP extends such learning to hierarchical representations (Kheradpisheh et al., 2018). These approaches exploit local spike-driven adaptation, providing an alternative to representation learning based on global error propagation.

More recent studies have introduced self-supervised objectives for learning spiking representations across different data modalities. Temporal consistency has been exploited for self-supervised event-

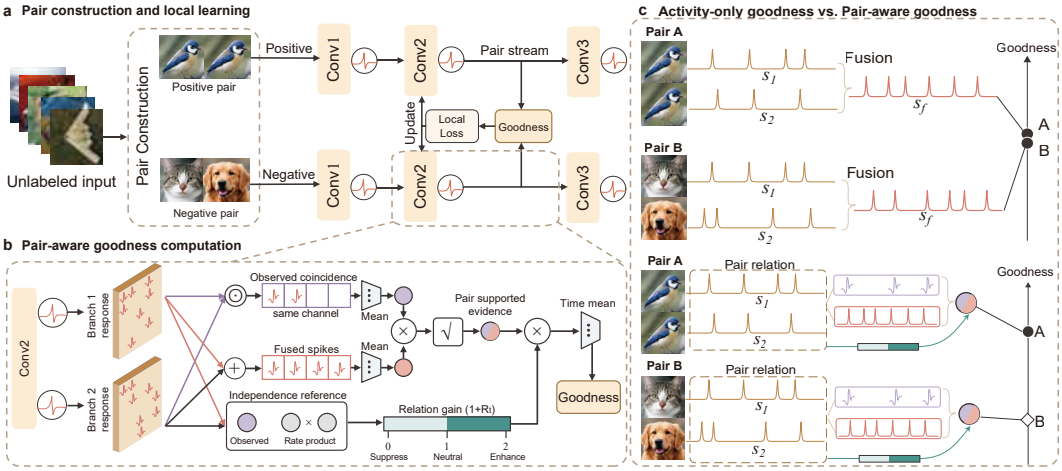

Figure 1: **Overview of Pair-aware SCFF. (a)** Pipeline for Pair-aware SCFF. **(b)** Pair-aware goodness maintains pair relation. **(c)** Activity-only goodness may assign similar scores to pairs despite different cross-branch correspondence.

based optical flow (Hagenaars et al., 2021), while contrastive learning has been adapted to augmented views in SNNs (Bahariasl & Kheradpisheh, 2024). SpikeGCL extends contrastive learning to binary spiking graph representations (Li et al., 2024), and NeuroMoCo introduces temporal-aware momentum contrast for neuromorphic data (Ma et al., 2024). Scalable and few-shot self-supervised SNN pretraining has also been investigated (Zhou & Datta, 2025; Vaillant et al., 2026). These studies demonstrate the potential of unlabeled data for spiking representation learning, although their learning objectives and parameter-update mechanisms differ. In the context of local learning, Contrastive Signal-Dependent Plasticity (CSDP) (Ororbia, 2024) enables self-supervised adaptation for recurrent spiking circuits without conventional network-wide backpropagation.

2.2 Local and Forward Learning

Local learning employs objectives or update rules at individual layers or modules rather than relying exclusively on end-to-end error propagation. Representative approaches include local predictive contrastive learning in CLAPP (Illing et al., 2021), competitive Hebbian learning in Soft-Hebb (Journé et al., 2022), and Dictionary Contrastive Learning (Choi et al., 2024). Among local learning frameworks, Forward-Forward (FF) learning (Hinton, 2022) trains each layer to distinguish positive from negative examples through a scalar goodness function, without propagating an error signal through the entire network.

Subsequent FF studies have explored convolutional architectures, competitive objectives, normalization, and deeper networks (Papachristodoulou et al., 2024; Sun et al., 2025; Scodellaro et al., 2025). FF-style learning has also been adapted to SNNs. FF-SCP (Terres-Escudero et al., 2025) applies forward local learning in fully spiking networks, while FF-SNN (Ghader et al., 2025) investigates layer-wise FF optimization based on rate-coded spike responses.

Two closely related considerations in FF learning are how positive and negative examples are constructed and how their goodness is evaluated. Self-Contrastive Forward-Forward (SCFF) (Chen et al., 2025) addresses the former by constructing positive and negative pairs directly from unlabeled observations. However, evaluating the resulting pairs through fused activity does not explicitly assess the correspondence between their constituent branch responses. Our work focuses on the local goodness used to evaluate these pairs. Pair-aware SCFF explicitly incorporates cross-branch spike correspondence into goodness evaluation, complementing fused firing activity with relational information.

3 METHOD

3.1 CONTRASTIVE FORWARD-FORWARD LEARNING IN SNNS

Forward-Forward (FF) learning provides a layer-local alternative to network-wide backpropagation by replacing global error propagation with layer-wise local optimization. Each layer is trained independently using a goodness function that assesses whether the learned representation satisfies a desired property. Recent contrastive extensions further improve FF by defining goodness in terms of relationships between paired representations rather than in terms of individual samples. Specifically, positive and negative pairs are constructed as

$$p_i^+ = (x_i, \widetilde{x}_i), \qquad p_{ij}^- = (x_i, \widetilde{x}_j), \quad i \neq j, \tag{1}$$

where $\widetilde{x}$ denotes a lightweight support view. Its construction and adaptive strength are detailed in Appendix A.1. Therefore, the effectiveness of contrastive FF depends on whether the local goodness function can capture pair-specific correspondence.

We investigate contrastive FF learning in spiking neural networks (SNNs), where each layer performs local computation using leaky integrate-and-fire (LIF) neurons. For a synaptic input $\mathbf{I}[t]$, the membrane dynamics and spike generation are formulated as

$$\begin{aligned} u[t] &= \lambda u[t-1] + \mathbf{I}[t], \\ s[t] &= H(u[t] - v_{\text{th}}), \end{aligned} \tag{2}$$

where λ denotes the membrane decay factor and $H(\cdot)$ represents the threshold function.

A representative contrastive FF framework is SCFF (Chen et al., 2025), which evaluates paired representations via a fused pathway (Fig. 1(a,c)). Given two paired inputs $\mathbf{x}_1^{(l-1)}[t]$ and $\mathbf{x}_2^{(l-1)}[t]$, their synaptic responses are computed using shared parameters. The two responses are then fused before spike generation:

$$\begin{aligned} \mathbf{I}_1^{(l)}[t] &= W^{(l)} * \mathbf{x}_1^{(l-1)}[t], \\ \mathbf{I}_2^{(l)}[t] &= W^{(l)} * \mathbf{x}_2^{(l-1)}[t], \\ s_f^{(l)}[t] &= \text{LIF}\left(\mathbf{I}_1^{(l)}[t] + \mathbf{I}_2^{(l)}[t]\right). \end{aligned} \tag{3}$$

The corresponding goodness function evaluates the overall firing activity of the fused response:

$$G_{\text{act}}^{(l)} = \frac{1}{TC_l} \sum_{t=1}^{T} \sum_{c=1}^{C_l} s_{f,c}^{(l)}[t]. \tag{4}$$

Although effective, this fusion-based formulation implicitly assumes that the aggregated spike activity is sufficient to characterize the quality of a sample pair. However, contrastive learning is inherently relational: the objective is not only to measure the magnitude of the response but also to determine whether two representations exhibit consistent internal correspondence. The fusion operation removes the identity of individual branches before goodness evaluation, preventing the objective from explicitly modeling cross-branch relationships.

Consequently, two pairs with distinct correspondence patterns may generate similar fused activities and receive indistinguishable goodness scores. This motivates us to develop a relation-aware goodness formulation that preserves branch-wise spike responses and directly models pair-specific correspondences prior to fusion.

3.2 RELATION-AWARE SPIKE CORRESPONDENCE MODELING

The fusion operation used in existing contrastive FF methods removes branch identities before goodness evaluation, preventing the local objective from explicitly modeling the correspondence between paired representations. To preserve such relational information, we maintain separate spike responses for the two branches before fusion (Fig. 1(b)). Specifically, the paired representations are propagated through shared synaptic parameters while keeping independent neuronal states:

$$s_1^{(l)}[t] = \mathrm{LIF}\left(\mathbf{I}_1^{(l)}[t]\right), \qquad s_2^{(l)}[t] = \mathrm{LIF}\left(\mathbf{I}_2^{(l)}[t]\right),$$
$$p_{1,t} = \frac{1}{C}\sum_{c=1}^{C} s_{1,c}[t], \qquad p_{2,t} = \frac{1}{C}\sum_{c=1}^{C} s_{2,c}[t]. \tag{5}$$

Here, s_1 and s_2 denote the branch-wise spike responses, while p_1 and p_2 characterize their marginal firing statistics. Across layers, the two member streams are propagated separately by self-pairing each member and detaching the resulting fused response, preserving pair correspondence while maintaining layer-local optimization (Appendix A.1). Unlike activity-based goodness that only evaluates response magnitude, we explicitly measure whether paired representations exhibit consistent cross-branch correspondence. Since spikes are binary events, channel-wise co-activation provides a direct measurement of shared neural responses. However, raw coincidence can be inflated by high firing propensity in either branch. Therefore, we further separate correspondence induced by marginal activity from pair-specific correspondence.

Specifically, we first compute the observed and activity-explained correspondence, and then derive the normalized relation score:

$$J_t = \frac{1}{C}\sum_{c=1}^{C} s_{1,c}[t]s_{2,c}[t], \qquad M_t = p_{1,t}p_{2,t},$$
$$R_t = \frac{J_t - M_t}{J_t + M_t + \epsilon}. \tag{6}$$

Here, J_t measures the observed cross-branch coincidence, while $M_t = p_{1,t}p_{2,t}$ defines its factorized marginal reference. Thus, $J_t - M_t$ measures the deviation of the observed coincidence from that reference.

The normalized relation score R_t is bounded within $[-1, 1]$. Positive values indicate stronger-than-expected correspondence between the two branches, while negative values indicate reduced correspondence relative to the activity-based reference. For $J_t, M_t > 0$, it is also a bounded monotonic transform of the corresponding joint-to-reference log-ratio (Appendix A).

Although the marginal correction alleviates firing-rate bias, channels that consistently exhibit high activation across samples may still introduce spurious correspondences. To further suppress this channel-wise bias, we optionally apply a lightweight prevalence-based reweighting strategy:

$$\pi_c = \mathrm{sg}\left[\mathbb{E}_{q,b,t,\mathcal{B}}\left(s_{b,c}^{q}[t]\right)\right], \qquad \widetilde{s}_{b,c}[t] = (1 - \pi_c)s_{b,c}[t]. \tag{7}$$

where $q \in \{+, -\}$ denotes positive and negative pairs, $b \in \{1, 2\}$ indexes the two branches, $\mathcal{B}$ represents the minibatch, and $\mathrm{sg}(\cdot)$ denotes stop-gradient operation. This non-parametric correction reduces correspondence dominated by universally active channels while preserving informative cross-branch interactions. This shared minibatch statistic provides a lightweight channel-prevalence calibration rather than an independent estimate of all firing biases.

3.3 Relation-aware Goodness Optimization

The relation score R_t evaluates whether the observed correspondence is specific to the paired representations. However, the strength of the relation alone does not reflect the extent of neural evidence supporting such correspondence, especially in the presence of sparse spike activity. Therefore, the proposed goodness combines the response strength captured by conventional FF objectives with relational consistency between paired representations.

Let F_t denote the overall firing activity of the fused response:

$$F_t = \frac{1}{C}\sum_{c=1}^{C} s_{f,c}[t],$$
$$E_t = \sqrt{(F_t + \epsilon)(J_t + \epsilon)} - \epsilon, \tag{8}$$
$$G_{\mathrm{rel}}^{(l)} = \frac{1}{T}\sum_{t=1}^{T} E_t(1 + R_t).$$

Here, F_t preserves the activity-based evidence underlying the original FF goodness, while R_t introduces relational information beyond marginal firing activity. The evidence term E_t combines response magnitude and observed correspondence through a conjunctive formulation: reliable pair evidence requires both sufficient neural activation and cross-branch agreement. Therefore, sparse accidental coincidence or weak responses cannot dominate the goodness evaluation. The final goodness G_{rel} integrates neural evidence and pair-specific relation, transforming conventional activity-based goodness into a relation-aware objective for contrastive FF learning.

For local optimization, we retain the same positive-negative objective adopted by existing FF-based methods. Specifically, the proposed goodness $G_{\text{rel}}^{(l)}$ is optimized using

$$\begin{aligned}\mathcal{L}^{(l)} =& \mathbb{E}_{+}\left[\frac{1}{\kappa}\,\text{softplus}\left(\kappa\left(\theta_{+}^{(l)} - G_{+}^{(l)}\right)\right)\right] \\ &+ \mathbb{E}_{-}\left[\frac{1}{\kappa}\,\text{softplus}\left(\kappa\left(G_{-}^{(l)} - \theta_{-}^{(l)}\right)\right)\right] + \lambda_l \Omega^{(l)}.\end{aligned} \tag{9}$$

where $\theta_{+}^{(l)}$ and $\theta_{-}^{(l)}$ denote the positive and negative goodness thresholds, respectively, and $\Omega^{(l)}$ represents the regularization term shared with existing FF objectives.

The spike generation process is optimized using surrogate gradients, while the local objective updates only the parameters of the current layer. The representations propagated to subsequent layers are detached, preserving the fully local learning property of FF without requiring network-wide error backpropagation.

4 Experiments

We conduct extensive experiments to evaluate the effectiveness of Pair-aware SCFF in local self-contrastive SNN learning. We first compare the proposed method with existing SNN representation learning approaches, followed by controlled comparisons and ablation studies to assess the contributions of relation-aware goodness and pair-stream propagation. We further investigate the learned pairwise relationships to examine whether relation-aware goodness captures correspondence information beyond that captured by fused firing activity.

4.1 Comparison with Existing SNN Methods

We evaluate Pair-aware SCFF on static-image and event-based benchmarks. The SNN representations are learned through layer-wise local optimization without class-label supervision or network-wide error propagation. After representation learning, the backbone is frozen and evaluated using a linear classifier. Detailed implementation and evaluation settings are provided inAppendix A. Table 1 compares Pair-aware SCFF with representative unsupervised and self-supervised SNN representation learning methods on static-image benchmarks. To account for differences in learning and evaluation protocols, we report the representation-learning supervision, parameter-update mechanism, and downstream evaluation strategy, as well as classification accuracy.

As shown in Table 1, Pair-aware SCFF achieves competitive performance on the evaluated static-image benchmarks using locally learned representations and a frozen linear evaluator. Further comparison results on event-based datasets are in Appendix 6. Since the compared methods differ in network architectures, representation-learning rules, and downstream evaluation protocols, we further assess the proposed method through controlled comparisons under matched experimental settings.

4.2 Controlled Evaluation and Ablation Study

We evaluate Pair-aware SCFF against a matched spiking implementation of SCFF (Chen et al., 2025) across six static-image and two event-based benchmarks. Both methods share the same backbone, neuron model, support-view construction, optimizer, training budget, and downstream evaluator. Designed differs are isolated by the following ablation experiments. We further conduct component-wise ablations to examine the contributions of relation-aware goodness and pair-stream propagation.

As shown in Fig. 2(left), Pair-aware SCFF achieves higher mean accuracy across all eight benchmarks, with gains of 7.82, 8.08, 10.71, and 5.57 percentage points on CIFAR-10, STL-10, CIFAR-

Table 1: Comparison with existing unsupervised and self-supervised SNN representation learning methods on static-image benchmarks. Rep. learn. denotes representation-learning supervision, Rep. update denotes the representation-update mechanism, and Eval. specifies the downstream evaluation protocol.

Method	Rep. learn.	Rep. update	Eval.	Acc. (%)
MNIST				
Querlioz-STDP (Querlioz et al., 2013)	Unsup.	Local	Neuron label	93.50
Diehl-Cook STDP (Diehl & Cook, 2015)	Unsup.	Local	Neuron label	95.00
LC-SNN (Saunders et al., 2019)	Unsup.	Local	2-gram	95.07 ± 0.63
STB-STDP (Dong et al., 2023)	Unsup.	Local	Neuron vote	97.90
Spiking BCPNN (Ravichandran et al., 2023)	Unsup.	Local	Linear	97.70 ± 0.09
SDNN (Kheradpisheh et al., 2018)	Unsup.	Local	Linear SVM	98.40
ReStoCNet (Srinivasan & Roy, 2019)	Unsup.	Local	Sup. MLP	98.54
WTA-STDP (Ferré et al., 2018)	Unsup.	Local	Sup. MLP	98.49
Falez et al. (Falez et al., 2019)	Unsup.	Local	Linear SVM	98.60
VDSP-CSNN (Goupy et al., 2022)	Unsup.	Local	Linear SVM	98.56 ± 0.05
SpikeCNN (Panda & Roy, 2016)	Unsup.	Layer-wise BP	Sup. output	99.08
CSDP, Unsup. (Lg)[‡] (Ororbia, 2024)	SSL	Local	Joint local cls.	95.96 ± 0.05
Contrastive SNN (Bahariasl & Kheradpisheh, 2024)	SSL	Global BP	Full FT (20%)	94.23
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	99.15 ± 0.04
Fashion-MNIST				
FSpiNN (Net1600) (Putra & Shafique, 2020)	Unsup.	Local	Neuron label	68.80
Spiking BCPNN (Ravichandran et al., 2023)	Unsup.	Local	Linear	83.80 ± 0.12
STB-STDP (Dong et al., 2023)	Unsup.	Local	Neuron vote	87.00
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	91.15 ± 0.14
CIFAR-10				
Contrastive SNN (Bahariasl & Kheradpisheh, 2024)	SSL	Global BP	Full FT (20%)	62.24
ReStoCNet (Srinivasan & Roy, 2019)	Unsup.	Local	Sup. MLP	66.23
SpikeCNN (Panda & Roy, 2016)	Unsup.	Layer-wise BP	Sup. output	70.16
WTA-STDP (Ferré et al., 2018)	Unsup.	Local	Sup. MLP	71.20
SpikeCLR (Zhou & Datta, 2026)	SSL	Global BP	Linear	85.60
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	75.25 ± 0.45
STL-10				
WTA-STDP (Ferré et al., 2018)	Unsup.	Local	Sup. MLP	60.10
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	72.76 ± 0.14

Local denotes representation updates without network-wide error propagation. Results follow the evaluation protocols reported in the respective papers. [‡] CSDP, Unsup. learns its hidden representations without class-conditioned signals, while the reported classification accuracy uses additional labeled local classification synapses rather than frozen post-hoc linear evaluation.

Dataset	Spiking SCFF	Pair-aware SCFF	Gain (pp)
MNIST	99.00 ± 0.06	99.15 ± 0.04	+0.15
Fashion-MNIST	89.88 ± 0.09	91.15 ± 0.14	+1.27
CIFAR-10	67.43 ± 0.10	75.25 ± 0.45	+7.82
STL-10	64.68 ± 0.11	72.76 ± 0.14	+8.08
CIFAR-100	39.33 ± 0.42	50.04 ± 0.26	+10.71
Tiny ImageNet	23.46 ± 0.24	29.03 ± 0.49	+5.57
DVS128 Gesture	81.25 ± 1.59	82.63 ± 1.02	+1.38
CIFAR10-DVS	44.86 ± 1.36	52.00 ± 0.95	+7.14

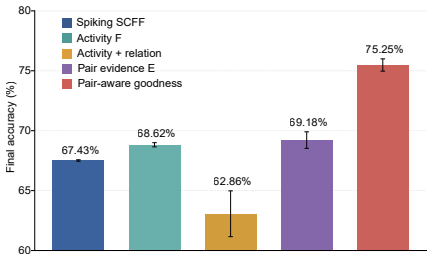

Figure 2: **Performance comparison and component ablation of Pair-aware SCFF. Tab(left)** Controlled comparison with Spiking SCFF across six static-image and two event-based benchmarks, with gains reported in percentage points (pp). **Fig(right)** Component ablation on CIFAR-10, comparing the Spiking SCFF baseline, activity-based scoring, relation-aware scoring, correspondence-aware scoring, and the complete relation-aware goodness formulation.

100, and Tiny ImageNet, respectively. Improvements are also observed on event-based inputs, reaching 7.14 points on CIFAR10-DVS and 1.38 points on DVS128 Gesture.

Figure 2(right) examines the components of the proposed goodness on CIFAR-10. Using activity F alone with Pair Stream propagation achieves 68.62%, while using correspondence $\sqrt{FJ}$ slightly improves performance to 69.18%. In contrast, the relation term $F(1+R)$ alone is insufficient, yielding 62.86%. Combining pair evidence with relation modulation in the full goodness substantially improves accuracy to 75.25%, indicating that relation information is most effective when supported by sufficient pair evidence. Additional robustness ablations are provided in Appendix B.

4.3 Understanding Relation-aware Goodness

We further investigate how relation-aware goodness reorganizes pairwise responses, guides difficult pair comparisons during learning, and affects the semantic structure of the learned representations. All class labels used below are introduced solely for post hoc analysis and are not available during representation learning.

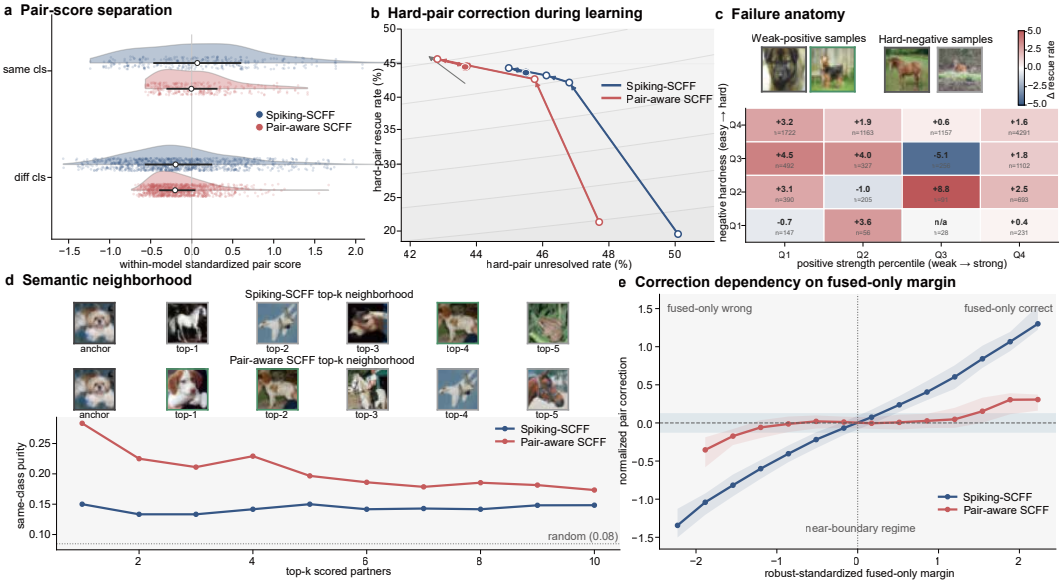

Figure 3: **Understanding relation-aware goodness.** **(a)** Standardized goodness distributions of same- and different-class pairs. **(b)** Learning trajectories of fixed hard-pair comparisons. **(c)** Failure anatomy, rescue-rate difference across fused-only failure regimes defined by positive strength and negative hardness. **(d)** Top-k semantic neighborhoods and same-class purity. **(e)** Pair correction as a function of the fused-only margin.

Figure 3(a) first shows that relation-aware goodness changes the organization of same-class and different-class pair scores rather than acting as a uniform rescaling of fused activity. For the fixed hard subset, movement toward the upper-left corresponds to fewer unresolved comparisons and a higher rescue rate. As shown in Fig. 3(b), Pair-aware SCFF follows a more favorable trajectory during later training than Spiking SCFF.

To understand which failures benefit from this correction, Fig. 3(c) decomposes matched fused-only errors according to positive strength and negative hardness. The accompanying weak-positive and hard-negative examples illustrate two typical cases in which pair relations alter an otherwise incorrect ranking. For the Spiking-SCFF reference, the same relation-aware score is evaluated only post hoc and does not participate in training.

These local corrections also lead to a more semantically coherent pair neighborhood. Figure 3(d) shows higher same-class purity among the top-k partners selected by Pair-aware SCFF, together with representative retrieved pairs. Fig. 3(e) shows that the relation-aware correction depends more weakly on the original fused-only margin, particularly around the ambiguous regime. These results show that relation-aware goodness introduces pair-specific structure beyond a confidence-dependent rescaling of fused activity.

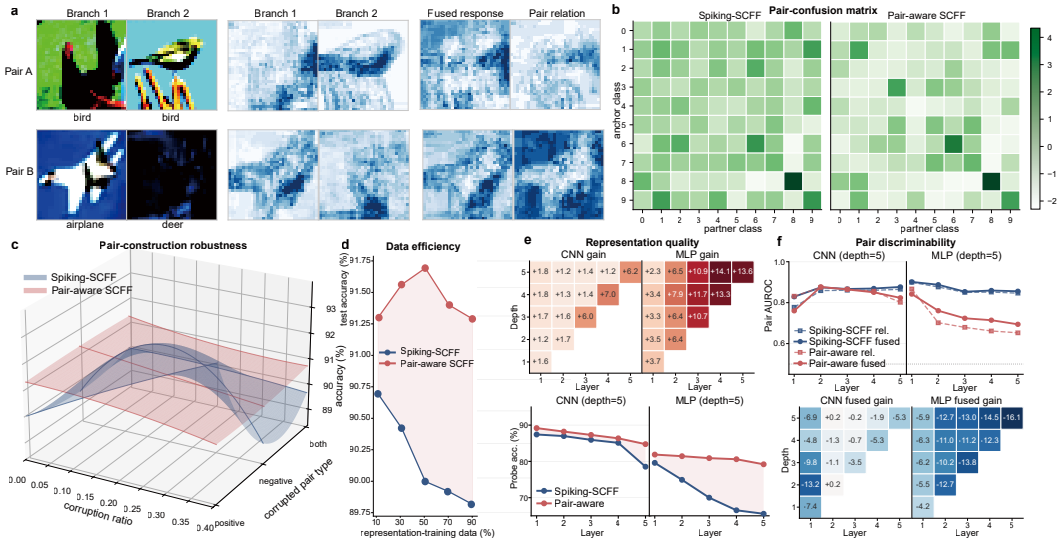

Figure 4: **Analysis of Pair-aware SCFF. (a)** Branch responses, fused responses, and pair relations for representative same- and different-class pairs. **(b)** Pair-confusion matrices across classes. **(c)** Robustness to pair-construction corruption. **(d)** Data efficiency under reduced unlabeled training data. **(e)** Layer-wise frozen-probe accuracy and Pair-aware gains on CNNs and MLPs. **(f)** Corresponding pair discriminability measured by fused-response AUROC.

4.4 Analysis of Pair-Aware Representation Learning

We further analyze how pair-aware learning affects the internal representation and its robustness in Fig. 4. Fig. 4(a) visualizes the branch responses of representative same-class and different-class pairs, showing that branch-wise relation patterns retain pair-dependent information beyond the fused response. Fig. 4(b) presents the pair-confusion matrices of Spiking-SCFF and Pair-aware SCFF, indicating that the learned local goodness better captures the semantic relations within paired samples. This also illustrates that pair-aware learning reorganizes the pair-response structure across semantic classes rather than uniformly increasing the response magnitude. Fig. 4(c) evaluates pair-construction robustness by perturbing positive, negative, and both types of pair construction, where the pair-aware model maintains consistently competitive performance across corruption levels. Fig. 4(d) shows that the advantage also persists when the amount of data used for representation learning is reduced, indicating improved data efficiency under a fixed training schedule. Fig. 4(e) directly evaluates layer-wise representation quality using frozen linear probes. Pair-aware learning consistently improves the probe accuracy across both CNN and MLP architectures, with particularly clear gains in deeper MLP layers. In contrast, Fig. 4(f) shows that these representation gains are not accompanied by a systematic increase in the raw fused-pair discriminability. Fig. 4(e) and (f) show that semantic separability can be decoupled from pair discriminability. A mechanistic analysis of this distinction is provided in Appendix C.

5 Conclusion

In this work, we revisit the design of local goodness in self-contrastive Forward-Forward learning for spiking neural networks. We identify a limitation of conventional activity-based goodness: aggregating paired responses before evaluation obscures their cross-branch correspondence, leaving relational information introduced by pair construction underutilized. To address this issue, we propose Pair-aware SCFF, which explicitly models spike correspondence relative to marginal firing activity and integrates relational information with pair-supported neural evidence to construct a relation-aware goodness function. Across static-image and event-based benchmarks, the method consistently improves a matched Spiking SCFF baseline. Controlled ablations support the complementary roles of pair evidence and relation-aware modulation, and that improved frozen representation quality does not require stronger raw fused-pair separability. More broadly, these results suggest that matching the structure of the self-supervisory signal to that of the local objective is an important design principle for learning local SNN representations.

AI USE STATEMENT

Generative AI tools were used solely for language translation and final-stage polishing to improve the clarity, fluency, and readability of the manuscript. They were not used for research ideation. methodology design, experimental design, data analysis, result interpretation, or the formulation of scientific claims. The authors take full responsibility for the content of the manuscript.

REPRODUCIBILITY STATEMENT

Detailed experimental configurations, including implementation and training settings, are provided in the Appendix. The complete experimental code necessary to reproduce the results is included in the supplementary material submitted with this paper.

REFERENCES

Yeganeh Bahariasl and Saeed Reza Kheradpisheh. Self-supervised contrastive learning in spiking neural networks. In *2024 13th Iranian/3rd International Machine Vision and Image Processing Conference (MVIP)*, pp. 1–5, 2024. doi: 10.1109/MVIP62238.2024.10491173.

Xing Chen, Dongshu Liu, J'er'emie Laydevant, and Julie Grollier. Self-contrastive forward-forward algorithm. *Nature Communications*, 16(1):5978, 2025.

Suhwan Choi, Myeongho Jeon, Yeonjung Hwang, Jeonglyul Oh, Sungjun Lim, Joonseok Lee, and Myungjoo Kang. Dictionary contrastive learning for efficient local supervision without auxiliary networks. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=Gg7cXo3S8l.

Peter U. Diehl and Matthew Cook. Unsupervised learning of digit recognition using spike-timing-dependent plasticity. *Frontiers in Computational Neuroscience*, Volume 9 - 2015, 2015. ISSN 1662-5188. doi: 10.3389/fncom.2015.00099. URL https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2015.00099.

Yiting Dong, Dongcheng Zhao, Yang Li, and Yi Zeng. An unsupervised stdp-based spiking neural network inspired by biologically plausible learning rules and connections. *Neural Networks*, 165:799–808, 2023. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2023.06.019. URL https://www.sciencedirect.com/science/article/pii/S0893608023003301.

Pierre Falez, Pierre Tirilly, Ioan Marius Bilasco, Philippe Devienne, and Pierre Boulet. Multi-layered spiking neural network with target timestamp threshold adaptation and stdp. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2019. doi: 10.1109/IJCNN.2019.8852346.

Paul Ferré, Franck Mamalet, and Simon J. Thorpe. Unsupervised feature learning with winner-takes-all based stdp. *Frontiers in Computational Neuroscience*, Volume 12 - 2018, 2018. ISSN 1662-5188. doi: 10.3389/fncom.2018.00024. URL https://www.frontiersin.org/journals/computational-neuroscience/articles/10.3389/fncom.2018.00024.

Mohammadnavid Ghader, Saeed Reza Kheradpisheh, Bahar Farahani, and Mahmood Fazlali. Backpropagation-free spiking neural networks with the forward–forward algorithm. *Scientific Reports*, 16, 2025. URL https://api.semanticscholar.org/CorpusID:276725450.

Gaspard Goupy, Alexandre Juneau-Fecteau, Nikhil Garg, Ismael Balafrej, Fabien Alibart, Luc G. Fréchette, Dominique Drouin, and Yann Beilliard. Unsupervised and efficient learning in sparsely activated convolutional spiking neural networks enabled by voltage-dependent synaptic plasticity. *Neuromorphic Computing and Engineering*, 3, 2022. URL https://api.semanticscholar.org/CorpusID:255037762.

Jesse Hagenaars, Federico Paredes-Vallés, and Guido de Croon. Self-supervised learning of event-based optical flow with spiking neural networks. *Advances in Neural Information Processing Systems*, 34, 2021.

Geoffrey E. Hinton. The forward-forward algorithm: Some preliminary investigations. *ArXiv*, abs/2212.13345, 2022. URL https://api.semanticscholar.org/CorpusID:254537921.

Bernd Illing, Jean Robin Ventura, Guillaume Bellec, and Wulfram Gerstner. Local plasticity rules can learn deep representations using self-supervised contrastive predictions. In A. Beygelzimer, Y. Dauphin, P. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, 2021. URL https://openreview.net/forum?id=Yu8Q6341U7W.

Adrien Journé, Hector Garcia Rodriguez, Qinghai Guo, and Timoleon Moraitis. Hebbian deep learning without feedback. *ArXiv*, abs/2209.11883, 2022. URL https://api.semanticscholar.org/CorpusID:252531820.

Saeed Reza Kheradpisheh, Mohammad Ganjtabesh, Simon J. Thorpe, and Timothée Masquelier. Stdp-based spiking deep convolutional neural networks for object recognition. *Neural Networks*, 99:56–67, 2018. ISSN 0893-6080. doi: https://doi.org/10.1016/j.neunet.2017.12.005. URL https://www.sciencedirect.com/science/article/pii/S0893608017302903.

Jintang Li, Huizhe Zhang, Ruofan Wu, Zulun Zhu, Baokun Wang, Changhua Meng, Zibin Zheng, and Liang Chen. A graph is worth 1-bit spikes: When graph contrastive learning meets spiking neural networks. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=LnLySuf1vp.

Yuhang Li, Youngeun Kim, Hyoungseob Park, Tamar Geller, and Priyadarshini Panda. Neuromorphic data augmentation fornbsp;training spiking neural networks. Berlin, Heidelberg, 2022. Springer-Verlag. ISBN 978-3-031-20070-0. doi: 10.1007/978-3-031-20071-7_37. URL https://doi.org/10.1007/978-3-031-20071-7_37.

Yuqi Ma, Huamin Wang, Hangchi Shen, Xuemei Chen, Shukai Duan, and Shiping Wen. Neuromoco: a neuromorphic momentum contrast learning method for spiking neural networks: Neuromoco: a neuromorphic momentum contrast learning... *Applied Intelligence*, 55(2), December 2024. ISSN 0924-669X. doi: 10.1007/s10489-024-05982-1. URL https://doi.org/10.1007/s10489-024-05982-1.

Alexander G. Ororbia. Contrastive signal–dependent plasticity: Self-supervised learning in spiking neural circuits. *Science Advances*, 10(43):eadn6076, 2024. doi: 10.1126/sciadv.adn6076. URL https://www.science.org/doi/abs/10.1126/sciadv.adn6076.

Priyadarshini Panda and Kaushik Roy. Unsupervised regenerative learning of hierarchical features in spiking deep networks for object recognition. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 299–306, 2016. doi: 10.1109/IJCNN.2016.7727212.

Andreas Papachristodoulou, Christos Kyrkou, Stelios Timotheou, and Theocharis Theocharides. Convolutional channel-wise competitive learning for the forward-forward algorithm. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(13):14536–14544, Mar. 2024. doi: 10.1609/aaai.v38i13.29369. URL https://ojs.aaai.org/index.php/AAAI/article/view/29369.

Rachmad Vidya Wicaksana Putra and Muhammad Shafique. Fspinn: An optimization framework for memory-efficient and energy-efficient spiking neural networks. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 39(11):3601–3613, 2020. doi: 10.1109/TCAD.2020.3013049.

Damien Querlioz, Olivier Bichler, Philippe Dollfus, and Christian Gamrat. Immunity to device variations in a spiking neural network with memristive nanodevices. *IEEE Transactions on Nanotechnology*, 12(3):288–295, 2013. doi: 10.1109/TNANO.2013.2250995.

Naresh Balaji Ravichandran, Anders Lansner, and Pawel Herman. Spiking neural networks with hebbian plasticity for unsupervised representation learning. In *ESANN 2023 Proceedings - 31st European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, pp. 611–616. Universite Catholique de Louvain, 2023. doi: 10.14428/esann/2023.ES2023-169.

Ali Safa, Ilja Ocket, André Bourdoux, Hichem Sahli, Francky Catthoor, and Georges Gielen. A new look at spike-timing-dependent plasticity networks for spatio-temporal feature learning, 2022. URL https://arxiv.org/abs/2111.00791.

Daniel J. Saunders, Devdhar Patel, Hananel Hazan, Hava T. Siegelmann, and Robert Thijs Kozma. Locally connected spiking neural networks for unsupervised feature learning. *Neural networks : the official journal of the International Neural Network Society*, 119:332–340, 2019. URL https://api.semanticscholar.org/CorpusID:118646104.

Riccardo Scodellaro, Ajinkya Kulkarni, Frauke Alves, and Matthias Schröter. Training convolutional neural networks with the forward–forward algorithm. *Scientific Reports*, 15(1), November 2025. ISSN 2045-2322. doi: 10.1038/s41598-025-26235-2. URL http://dx.doi.org/10.1038/s41598-025-26235-2.

Raghav Singhal, Jan Finkbeiner, and Emre Neftci. Self-supervised pre-training of spiking neural networks by contrasting events and frames. In *UniReps: 2nd Edition of the Workshop on Unifying Representations in Neural Models*, 2024. URL https://openreview.net/forum?id=DNopfn4hZf.

Gopalakrishnan Srinivasan and Kaushik Roy. Restocnet: Residual stochastic binary convolutional spiking neural network for memory-efficient neuromorphic computing. *Frontiers in Neuroscience*, Volume 13 - 2019, 2019. ISSN 1662-453X. doi: 10.3389/fnins.2019.00189. URL https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2019.00189.

Liang Sun, Yang Zhang, Weizhao He, Jiajun Wen, Linlin Shen, and Weicheng Xie. Deeperforward: Enhanced forward-forward training for deeper and better performance. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=kOYnXVQCtA.

Erik B. Terres-Escudero, Javier Del Ser, Aitor Martínez-Seras, and Pablo Garcia-Bringas. Forward-forward learning achieves highly selective latent representations for out-of-distribution detection in fully spiking neural networks, 2025. URL https://arxiv.org/abs/2407.14097.

Maxime Vaillant, Axel Carlier, Lai Xing Ng, Christophe Hurter, and Benoit R. Cottereau. Spikeclr: Contrastive self-supervised learning for few-shot event-based vision using spiking neural networks, 2026. URL https://arxiv.org/abs/2603.16338.

Cheng-Wei Zhou and Gourav Datta. Learning scalable temporal representations in spiking neural networks without labels. *ArXiv*, abs/2511.18542, 2025. URL https://api.semanticscholar.org/CorpusID:283245102.

Chengwei Zhou and Gourav Datta. SpikeCLR: Self-supervised contrastive learning for visual representations with spiking neural networks. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=n3Rj1wRtlE.

Algorithm 1 Training procedure of Pair-aware SCFF

Require: Anchor batch X, support batch $\widetilde{X}$, layers $\{f_l\}_{l=1}^{L}$
Sample a non-zero minibatch permutation ν
$Z_1^{(1)} \leftarrow X, \quad Z_2^{(1)} \leftarrow \widetilde{X}$
for $l = 1, \ldots, L$ **do**
  $P_l^{+} \leftarrow (Z_1^{(l)}, Z_2^{(l)})$
  $P_l^{-} \leftarrow (Z_1^{(l)}, Z_2^{(l)}[\nu])$
  $(S_f^{\pm}, S_1^{\pm}, S_2^{\pm}) \leftarrow \textsc{PairForward}(f_l, P_l^{\pm})$
  $\pi^{(l)} \leftarrow \mathrm{sg}\big(\textsc{Prevalence}(S_1^{\pm}, S_2^{\pm})\big)$
  $G_l^{\pm} \leftarrow \textsc{PairGoodness}(S_f^{\pm}, S_1^{\pm}, S_2^{\pm}, \pi^{(l)})$
  $\mathcal{L}^{(l)} \leftarrow \textsc{LocalLoss}(G_l^{+}, G_l^{-})$
  if $l < L$ **then**
    $\widehat{Z}_{1,2}^{(l+1)} \leftarrow \textsc{SelfPairForward}(f_l, Z_{1,2}^{(l)})$
  end if
  Update θ_l using $\nabla_{\theta_l} \mathcal{L}^{(l)}$
  if $l < L$ **then**
    $Z_{1,2}^{(l+1)} \leftarrow \mathrm{sg}(\widehat{Z}_{1,2}^{(l+1)})$
    if event-based input **then**
      $Z_{1,2}^{(l+1)} \leftarrow \textsc{TemporalOR}(Z_{1,2}^{(l+1)})$
    end if
  end if
end for

A Experiment Details

This section summarizes the implementation and evaluation protocol used throughout the paper. Pair-aware SCFF and the matched Spiking SCFF baseline use the same SNN backbone, neuron dynamics, support-view construction, optimizer family, layer-wise training budget, and frozen linear evaluator. Their method-specific differences are the propagated representation stream and the local goodness defined in the main text. Unless otherwise stated, multi-seed results use $\{730, 1234, 3407\}$, single-model analyses use seed 1234, and the frozen readout uses initialization seed 42.

A.1 Pair-aware SCFF training

Here, PairForward returns the fused and two branch-wise spike responses using shared synaptic parameters and independent neuronal states. PairGoodness denotes the pair-aware goodness, while SelfPairForward generates the representation propagated to the next layer by self-pairing an individual member. All propagated representations are detached before entering the next layer. For event-based inputs, temporal coarsening is applied only to the propagated stream.

The same negative mapping is retained across depth, so the two member trajectories remain aligned with the original pair construction. No gradient crosses a layer boundary.

Pair construction. The positive pair consists of an anchor and a lightly perturbed support view of the same sample. For static inputs,

$$\widetilde{x} = a \odot x + n, \qquad a_c \sim \mathcal{U}(1 - 0.25\rho, 1 + 0.25\rho), \qquad \sigma_n = 0.08\rho\,\sigma(x), \tag{10}$$

whereas event inputs use independent event dropout $p_{\mathrm{drop}} = 1 - (1 - 0.1)^{\rho}$. The support strength starts at $\rho = 1$ and is adjusted once per epoch based on the observed positive/negative branch-overlap hardness. One negative partner is sampled for each anchor; class labels are never used to construct training pairs.

Stream recurrence and branch threshold. At layer l, each member stream has shape $[T, B, C_{l-1}, H_{l-1}, W_{l-1}]$, and pair concatenation produces $[T, B, 2C_{l-1}, H_{l-1}, W_{l-1}]$. The fused and two branch responses have shape $[T, B, C_l, H_l, W_l]$. The explicit recurrence used by Pair-aware SCFF is

$$z_b^{(l+1)} = \mathrm{sg}\left[P_l\left(S_{f,l}(z_b^{(l)}, z_b^{(l)})\right)\right], \qquad b \in \{1, 2\}. \tag{11}$$

Table 2: Dataset-specific settings used in the main experiments. B is the representation-learning batch size.

Dataset	T	B	Depth	Channels	Epochs/layer
MNIST	4	100	3	96–384–1536	[6,4,21]
Fashion-MNIST	4	100	3	96–384–1536	[6,4,21]
CIFAR-10	4	100	3	96–384–1536	[6,6,13]
CIFAR-100	4	100	3	96–384–1536	[6,6,13]
STL-10	4	64	4	96–384–1536–6144	[4,5,12,12]
Tiny ImageNet	4	100	5	64–192–384–256–256	[6,15,17,19,25]
CIFAR10-DVS	16	8	4	128–128–128–128	[64,64,64,64]
DVS128 Gesture	16	16	5	128–128–128–128–128	[64,64,64,64,64]

A single carried stream can equivalently be evaluated at

$$v_{\mathrm{br}} = v_{\mathrm{reset}} + \frac{1}{2}\left(v_{\mathrm{th}} - v_{\mathrm{reset}}\right). \tag{12}$$

For an exact self-pair (z, z), the fused current is $2I(z)$; with identical LIF parameters, reset-state initialization, and subtractive reset, $I(z)$ at v_{br} is spike-equivalent to $2I(z)$ at v_{th}. This equivalence applies to self-pairing or single-stream transfer, but not to an arbitrary two-member pair.

Relation implementation. Channel reduction is performed before temporal reduction: $F_t, J_t, M_t, R_t, E_t \in \mathbb{R}^{T\times B\times H_l\times W_l}$, followed by $G^{(l)} \in \mathbb{R}^{B\times H_l\times W_l}$. Raw coincidence forms the evidence,

$$F_t = \frac{1}{C}\sum_c s_{f,c}[t], \qquad J_t = \frac{1}{C}\sum_c s_{1,c}[t]s_{2,c}[t], \qquad E_t = \sqrt{(F_t+\epsilon)(J_t+\epsilon)} - \epsilon. \tag{13}$$

The relation uses prevalence-weighted branches $\widetilde{s}_{b,c} = (1-\pi_c)s_{b,c}$, where π_c is estimated over positive and negative pairs and detached before use. With weighted joint and marginal-reference terms $\widetilde{J}_t$ and $\widetilde{M}_t$,

$$R_t = \frac{\widetilde{J}_t - \widetilde{M}_t}{\widetilde{J}_t + \widetilde{M}_t + \epsilon}, \qquad G_{\mathrm{pair}} = \frac{1}{T}\sum_t E_t(1+R_t). \tag{14}$$

For positive J_t, M_t and ignoring ϵ,

$$\frac{J_t - M_t}{J_t + M_t} = \tanh\left[\frac{1}{2}\log\left(\frac{J_t}{M_t}\right)\right], \tag{15}$$

so the bounded relation is a monotonic transform of the corresponding joint-to-reference log-ratio.

Event-stream propagation. For CIFAR10-DVS and DVS128 Gesture, only the downstream Pair Stream is temporally coarsened:

$$\widetilde{s}[0] = s[0], \qquad \widetilde{s}[t] = \max\{s[t], s[t-1]\}, \quad t > 0. \tag{16}$$

Current-layer goodness and readout features are computed from the uncoarsened response.

A.2 Datasets and optimization

Static inputs are standardized per sample and repeated for $T = 4$ simulation steps; no Poisson or Bernoulli spike encoder is used. CIFAR10-DVS and DVS128 Gesture are integrated into $T = 16$ frames with SpikingJelly using `split_by=number`. MNIST and Fashion-MNIST use their standard train/test partitions; CIFAR-10 and CIFAR-100 use their standard splits with random horizontal flipping during representation learning; STL-10 uses the official unlabeled split for representation learning and the labeled training split for the frozen readout. CIFAR10-DVS uses the deterministic class-wise $90/10$ split generated by the SpikingJelly utility, and DVS128 Gesture uses its official train/test split.

The LIF parameters are $\tau = 2$, $v_{\mathrm{th}} = 1$, $v_{\mathrm{reset}} = 0$, and fast-sigmoid surrogate parameter $\beta = 4$. Each layer has an independent AdamW optimizer. For static datasets, the learning-rate profile

is $[0.02, 0.001, 0.0004, \ldots]$, weight decay is $[10^{-4}, 3 \times 10^{-4}, 10^{-4}, \ldots]$, and scheduler decay is $[0.99, 0.90, 0.99, \ldots]$. Event datasets use learning rate 10^{-3}, weight decay 10^{-4}, and decay 0.99 at every layer. Schedulers are stepped every 500 optimization minibatches. The default local thresholds are $\theta_+ = [0.12, 0.10, 0.08, \ldots]$ and $\theta_- = [0.04, 0.035, 0.03, \ldots]$. All remaining layer, pooling, and dataset parameters are provided in the released configuration files.

A.3 Frozen evaluation and Tiny ImageNet staging

After representation learning, the SNN backbone is frozen. Selected layer responses are averaged over time, spatially average-pooled, standardized per sample, flattened, and concatenated. The evaluator is a dropout layer followed by one linear layer, trained for 50 epochs with Adam, a learning rate of 10^{-3}, zero weight decay, and cross-entropy loss. Dropout is 0.2 by default, 0.4 for STL-10, and 0.3 for Tiny ImageNet. The main paper reports final-epoch readout accuracy.

Tiny ImageNet follows the two-stage procedure used by the released SCFF implementation (Chen et al., 2025): the first two layers are trained in Stage 1, the selected Stage 1 weights are frozen, and the remaining three layers are trained in Stage 2. For the matched SNN reproduction, both Spiking SCFF and Pair-aware SCFF use the same Stage 1 selector: a one-epoch frozen linear probe ranks checkpoints on the Tiny ImageNet evaluation loader, and the highest-scoring checkpoint is transferred to Stage 2. The official validation set is the evaluation split used for this stage handoff and for final reporting. The reported final result is nevertheless the final epoch of the fixed 50-epoch downstream readout, rather than the maximum readout accuracy.

A.4 Quantities used in the main-paper diagnostics

Only quantities not defined explicitly in the main text are collected here. Class labels are used only for post-hoc grouping, retrieval evaluation, AUROC computation, and frozen linear probes.

For Fig. 3(a), pair scores are standardized within each model as

$$\widehat{G}_{ij} = \frac{G_{ij} - \mu_G}{\sigma_G + \epsilon}. \tag{17}$$

For a same-class partner j and a different-class partner k of anchor i, the fused-only and relation-aware margins are

$$m_0(i,j,k) = G_0(i,j) - G_0(i,k), \qquad m_{\mathrm{P}}(i,j,k) = G_{\mathrm{P}}(i,j) - G_{\mathrm{P}}(i,k), \qquad \Delta m = m_{\mathrm{P}} - m_0. \tag{18}$$

The hard subset in Fig. 3(b) is the lowest quartile of $|m_0|$ at the first recorded epoch and is then held fixed. Its unresolved and rescue rates are

$$U^{(t)} = \Pr(m_{\mathrm{P}}^{(t)} \le 0 \mid r \in \mathcal{H}), \qquad R^{(t)} = \Pr(m_{\mathrm{P}}^{(t)} > 0 \mid r \in \mathcal{H},\ m_0^{(t_0)} \le 0). \tag{19}$$

Figure 3(c) partitions shared fused-only failures into quartiles of anchor-relative positive strength and negative hardness and reports the difference in rescue rate between the two matched representations.

For Fig. 3(d), semantic neighborhood purity is

$$P_k = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{k} \sum_{j \in \mathcal{N}_k(i)} \mathbb{I}[y_j = y_i], \qquad k = 1, \ldots, 10. \tag{20}$$

For Fig. 3(e), the fused-only margin and correction are placed on a common robust scale

$$s_m = \max\{Q_{0.75}(m_0) - Q_{0.25}(m_0), \mathrm{Std}(m_0), \epsilon\}, \qquad \widetilde{m}_0 = \frac{m_0}{s_m}, \qquad \widetilde{\Delta m} = \frac{m_{\mathrm{P}} - m_0}{s_m}. \tag{21}$$

For Fig. 4(b), the class-pair score matrix is

$$C_{uv} = \frac{1}{|\mathcal{P}_{uv}|} \sum_{(i,j) \in \mathcal{P}_{uv}} G_{ij}, \qquad \mathcal{P}_{uv} = \{(i,j) : y_i = u,\ y_j = v\}, \tag{22}$$

with self-pairs excluded and each matrix standardized independently before visualization. Figure 4(c) corrupts positive and/or negative pair construction with $\rho \in \{0, 0.1, 0.2, 0.3, 0.4\}$.

Table 3: Alternative relation formulations on CIFAR-10. All variants use the same Pair Stream and pair-supported evidence; values are final-epoch accuracy over three seeds.

Relation	Prevalence	Statistic	Final Acc. (%)
Full Pair Relation	Yes	$(\widetilde{J}-\widetilde{M})/(\widetilde{J}+\widetilde{M}+\epsilon)$	$\mathbf{75.25 \pm 0.45}$
No prevalence	No	$(J-M)/(J+M+\epsilon)$	74.75 ± 0.20
Centered Dice	Yes	$2D-1$	75.04 ± 0.20
Normalized covariance	Yes	R^{Cov}	73.16 ± 0.21

Figure 4(d) trains the native three-layer Fashion-MNIST backbone using

$$r_{\mathrm{data}} \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$$

of the unlabeled representation-training set. For each fraction, the labeled linear evaluation split, network architecture, layer-wise epoch schedule, and downstream readout protocol are held fixed.

Because the number of epochs is unchanged, reducing r_{data} simultaneously reduces both the amount of unlabeled data observed during representation learning and the total number of local optimization updates. The low-data settings therefore constitute a strictly reduced representation-learning budget rather than compensating for fewer samples with additional training. Pair-aware SCFF maintains a clear advantage over Spiking SCFF across these fractions, supporting improved data efficiency under the common training schedule.

The curves are not expected to vary strictly monotonically with r_{data}, since each fraction changes the sampled training subset and stochastic optimization trajectory. We therefore focus on the consistent relative advantage across data regimes rather than on the monotonicity of absolute accuracy.

For Fig. 4(e), the layer-wise probe gain is

$$\Delta Q_{D,l} = 100\left(Q_{D,l}^{\mathrm{Pair}} - Q_{D,l}^{\mathrm{SCFF}}\right) \quad \text{percentage points.} \tag{23}$$

For Fig. 4(f), same/different-class discriminability of a scalar score S is

$$A(S) = \Pr(S^+ > S^-) + \frac{1}{2}\Pr(S^+ = S^-), \tag{24}$$

and the heatmap reports

$$\Delta A_{D,l}^{\mathrm{fused}} = 100\left[A_{\mathrm{Pair}}^{\mathrm{fused}}(D,l) - A_{\mathrm{SCFF}}^{\mathrm{fused}}(D,l)\right]. \tag{25}$$

B Robustness Analysis and Further Results

B.1 Alternative relation formulations

We keep Pair Stream propagation, branch thresholds, pair-supported evidence, optimizer, and training budget fixed, and replace only the relation term. Let $a_c = \widetilde{s}_{1,c}$ and $b_c = \widetilde{s}_{2,c}$ denote prevalence-weighted branch responses. The centered Dice alternative is

$$D_t = \frac{2\mathbb{E}_c[a_c b_c]}{\mathbb{E}_c[a_c^2] + \mathbb{E}_c[b_c^2] + \epsilon}, \qquad R_t^{\mathrm{Dice}} = 2D_t - 1, \tag{26}$$

and normalized covariance is

$$R_t^{\mathrm{Cov}} = \mathrm{clip}\left[\frac{\mathbb{E}_c[(a_c-\bar{a})(b_c-\bar{b})]}{\sqrt{\mathbb{E}_c[(a_c-\bar{a})^2]\mathbb{E}_c[(b_c-\bar{b})^2] + \epsilon}}, -1, 1\right]. \tag{27}$$

Each alternative uses the same $E_t(1+R_t)$ mapping.

Centered Dice is a competitive alternative, while the full formulation gives the highest mean accuracy in this matched comparison. Together with the no-prevalence variant, the results show that the main gain comes from retaining cross-branch correspondence, with marginal calibration providing additional refinement.

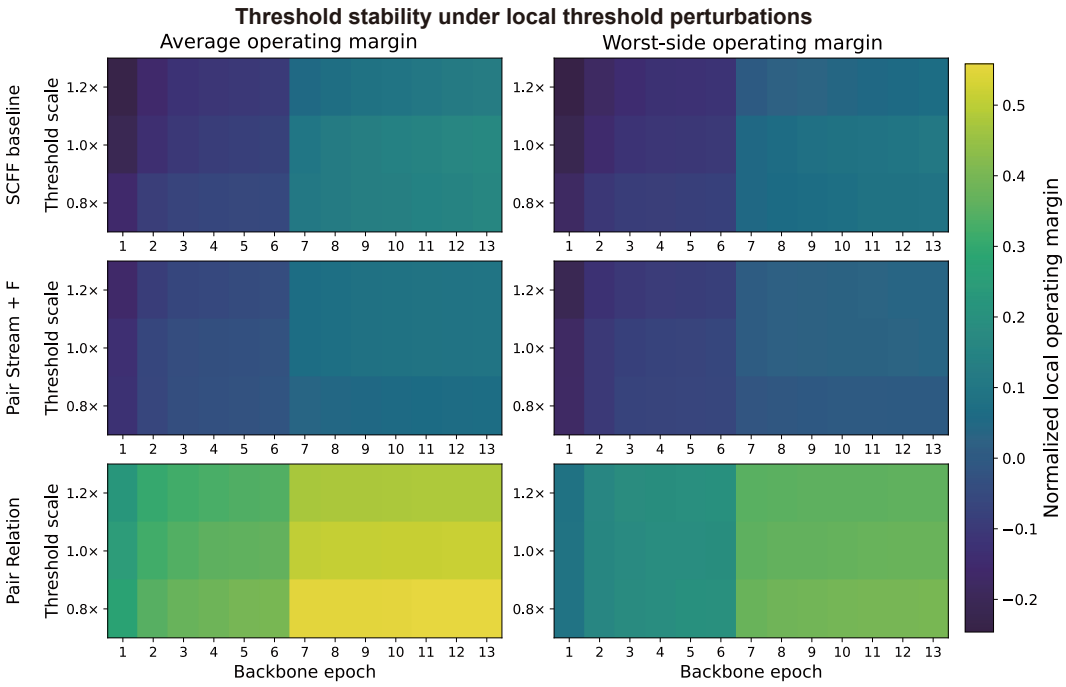

Figure 5: **Robustness to local-threshold scaling on CIFAR-10.** Average operating margin (left) and worst-side margin (right) under $0.8\times$, $1.0\times$, and $1.2\times$ threshold scaling. Pair Relation remains in a positive operating regime across the tested scales.

Table 4: Temporal-OR ablation on CIFAR10-DVS. Values are final accuracies from the corresponding diagnostic runs; the three-seed comparison in the main paper remains the primary benchmark.

Method	**Temporal OR**	**Final Acc. (%)**
Spiking SCFF	Off	45.80
Pair-aware SCFF	Off	51.90
Pair-aware SCFF	On	52.90

B.2 Threshold robustness

We characterize the operating range of the local objective by jointly scaling the positive and negative thresholds by $\alpha \in \{0.8, 1.0, 1.2\}$ while keeping the remaining training configuration fixed. For each active layer,

$$M_{\text{avg}} = \frac{1}{2}\left[\frac{G_+ - \theta_+}{\theta_+ - \theta_-} + \frac{\theta_- - G_-}{\theta_+ - \theta_-}\right], \qquad M_{\text{worst}} = \min\left\{\frac{G_+ - \theta_+}{\theta_+ - \theta_-}, \frac{\theta_- - G_-}{\theta_+ - \theta_-}\right\}. \tag{28}$$

The relation-aware objective retains a broader positive–negative operating region across the tested scales, while the activity-based objectives exhibit smaller margins. The behavior is therefore stable under moderate perturbations of the local thresholds.

B.3 Temporal adaptation on event streams

Temporal OR complements the Pair-aware goodness by stabilizing the representation passed between event-driven layers. The recorded CIFAR10-DVS runs are:

Without temporal OR, Pair-aware SCFF improves the corresponding Spiking SCFF run by 6.10 percentage points; temporal OR adds a further 1.00 point. The event-specific stream adaptation is therefore complementary to the Pair-aware local objective.

Table 5: Frozen linear-readout accuracy (%). The main paper uses final-epoch results. Values are mean ± standard deviation over three representation-learning seeds.

Dataset	Pair-aware SCFF		Spiking SCFF	
	Best	Final	Best	Final
MNIST	99.20 ± 0.09	99.15 ± 0.04	99.03 ± 0.07	99.00 ± 0.06
Fashion-MNIST	91.25 ± 0.12	91.15 ± 0.14	89.94 ± 0.03	89.88 ± 0.09
CIFAR-10	75.48 ± 0.51	75.25 ± 0.45	67.51 ± 0.06	67.43 ± 0.10
CIFAR-100	50.16 ± 0.31	50.04 ± 0.26	39.62 ± 0.33	39.33 ± 0.42
STL-10	72.87 ± 0.08	72.76 ± 0.14	65.01 ± 0.12	64.68 ± 0.11
Tiny ImageNet	29.15 ± 0.38	29.03 ± 0.49	23.66 ± 0.37	23.46 ± 0.24
CIFAR10-DVS	52.33 ± 0.35	52.00 ± 0.95	45.63 ± 1.12	44.86 ± 1.36
DVS128 Gesture	83.34 ± 1.59	82.63 ± 1.02	82.48 ± 1.40	81.25 ± 1.59

Table 6: Further comparison with representative unsupervised and self-supervised SNN representation learning on event-based benchmarks. Evaluation protocols are shown because frozen-feature, fine-tuning, and transfer settings are not directly matched.

Method	Rep. learn.	Rep. update	Eval.	Acc. (%)
CIFAR10-DVS				
SNN-STDP (Safa et al.) (Safa et al., 2022)	Unsup.	Local	Desc.+Linear SVM	73.98 ± 0.80
Event-Frame SSL (Singhal et al., 2024)	SSL	Global BP	Full FT	64.10
SpikeCLR (Zhou & Datta, 2026)	SSL	Global BP	Linear	63.90
SpikeCLR (Vaillant et al.) (Vaillant et al., 2026)	SSL	Global BP	Full FT	79.50
NDA+SimSiam† (Li et al., 2022)	SSL-transfer	Global BP	Full FT	82.10
NeuroMoCo (Spikformer-2-256) (Ma et al., 2024)	SSL	Global BP	Full FT	83.60
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	52.00 ± 0.95
DVS128 Gesture				
SNN-STDP (Safa et al.) (Safa et al., 2022)	Unsup.	Local	Desc.+Linear SVM	92.50
NeuroMoCo (Spikformer-2-256) (Ma et al., 2024)	SSL	Global BP	Full FT	98.62
Pair-aware SCFF (ours)	**SSL**	**Local**	**Linear**	82.63 ± 1.02

† NDA+SimSiam is self-supervisedly pre-trained on N-Caltech101 and then fully fine-tuned on CIFAR10-DVS; the reported 300-epoch fine-tuning result is 82.10%.

B.4 Final-epoch reporting and further benchmark comparison

For completeness, Table 5 reports both the maximum and final-epoch accuracy observed during the fixed readout schedule. All main-paper matched comparisons use the final-epoch column.

Computational profile. Pair-aware goodness introduces no learnable parameters. Training retains the two branch spike responses in addition to the fused response and adds channel-wise products, reductions, and the bounded relation gain. The relation statistic is not required for single-sample inference by the frozen downstream classifier.

C Mechanistic Analysis

The main analysis in Fig. 4(e)-(f) shows that improvements in layer-wise frozen linear probes do not require a corresponding increase in same/different-class AUROC of the raw fused score. This section formalizes why the two quantities need not be optimized together and uses two controlled analyses to characterize the resulting representation geometry.

C.1 Pair objective versus semantic representation geometry

Let $q \in \{+, -\}$ denote the self-contrastive pair label used for local learning: $q = +$ for an anchor and its support view, and $q = -$ for two different sampled instances. At layer l, write the local

objective as

$$\mathcal{J}_l = \mathbb{E}_{p^+}\left[\phi_+\big(G_l(p^+)\big)\right] + \mathbb{E}_{p^-}\left[\phi_-\big(G_l(p^-)\big)\right] + \lambda_l \Omega_l, \tag{29}$$

with

$$\phi_+(g) = \mathrm{softplus}(\kappa(\theta_+ - g))\,, \qquad \phi_-(g) = \mathrm{softplus}(\kappa(g - \theta_-))\,. \tag{30}$$

This objective distinguishes the *construction relation* of a pair; it does not directly optimize class labels.

This distinction is explicit for negative sampling. Let the class prior be $\pi_k = \Pr(y = k)$. If the negative partner is sampled independently of class, then, up to the negligible exclusion of the identical instance,

$$\Pr(y_j = y_i \mid q = -) = \sum_k \pi_k^2. \tag{31}$$

For a balanced K-class dataset, this is $1/K$. Hence, a non-zero subset of self-contrastive negative pairs is semantically of the same class. Conversely, the positive set contains another view of the same instance rather than all members of the same class. The self-contrastive pair objective is therefore a proxy objective for semantic organization, not a class-contrastive objective.

To make the distinction measurable, let $z_l(x)$ denote the frozen self-paired representation at layer l. Its downstream linear quality can be written as

$$\mathcal{R}^{(l)}_{\mathrm{lin}} = \min_{W,b} \mathbb{E}_{(x,y)}\left[\ell_{\mathrm{cls}}\left(W z_l(x) + b, y\right)\right], \tag{32}$$

whereas the raw fused-pair statistic in Fig. 4(f) is summarized by

$$A^{(l)}_{\mathrm{fused}} = \Pr\left(S^{(l)}_f(x_i, x_j) > S^{(l)}_f(x_i, x_k)\right) + \frac{1}{2}\Pr(\mathrm{tie}), \tag{33}$$

For $y_j = y_i$ and $y_k \neq y_i$. The first quantity depends on the geometry of a high-dimensional unary representation; the second depends on the ordering induced by one scalar pair statistic. Without the additional assumption that this scalar statistic is sufficient for the downstream label, there is no order-preserving implication between $A^{(l)}_{\mathrm{fused}}$ and $\mathcal{R}^{(l)}_{\mathrm{lin}}$. This explains why layer-wise probe quality and raw fused-pair AUROC can evolve differently.

C.2 How relation-aware goodness changes the local update

The distinction is also visible directly in the direction of optimization. For the positive and negative softplus terms,

$$\nabla_{\theta_l}\mathcal{J}_l = -\mathbb{E}_{p^+}\left[w_+(G_+)\nabla_{\theta_l}G_+\right] + \mathbb{E}_{p^-}\left[w_-(G_-)\nabla_{\theta_l}G_-\right] + \lambda_l \nabla_{\theta_l}\Omega_l, \tag{34}$$

where

$$w_+(G) = \kappa\sigma(\kappa(\theta_+ - G)) > 0, \qquad w_-(G) = \kappa\sigma(\kappa(G - \theta_-)) > 0.$$

For Pair-aware goodness,

$$G_{\mathrm{pair}} = \frac{1}{T}\sum_t E_t(1 + R_t), \tag{35}$$

and therefore

$$\nabla_{\theta_l} G_{\mathrm{pair}} = \frac{1}{T}\sum_t \left[(1 + R_t)\nabla_{\theta_l}E_t + E_t\nabla_{\theta_l}R_t\right]. \tag{36}$$

The first term reweights activity-supported evidence, while the second introduces a gradient component that depends on how the two branches correspond. Unless ∇E_t and ∇R_t are everywhere collinear, the relation term changes the update direction, not merely the numerical scale of fused activity. Consequently, the local objective can reorganize the feature geometry without requiring a monotonic increase in raw fused-score separation. This is consistent with the boundary-focused correction and layer-wise probe behavior observed in the main analysis.

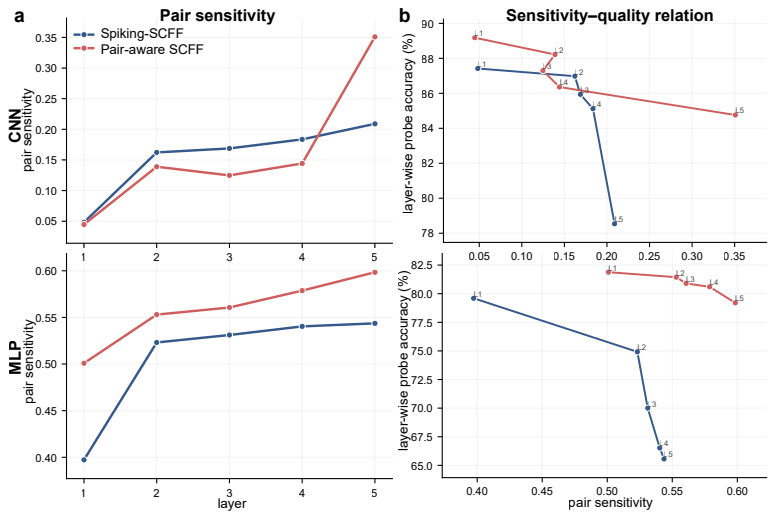

Figure 6: **Pair sensitivity and frozen-feature quality on Fashion-MNIST.** The top and bottom rows show the five-layer CNN and MLP backbones. **(a)** Layer-wise pair sensitivity relative to the self-pair response. **(b)** Pair sensitivity versus frozen linear-probe accuracy. Pair-aware representations improve the probe across different sensitivity regimes.

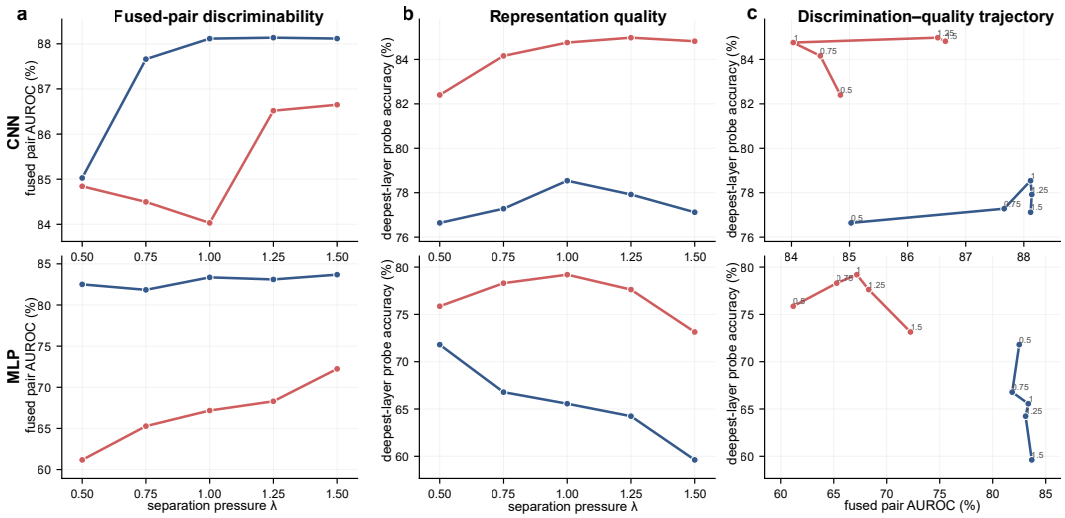

Figure 7: **Controlled separation-pressure sweep on Fashion-MNIST.** The top and bottom rows show the five-layer CNN and MLP backbones. **(a)** Deepest-layer same/different-class AUROC of the raw fused response. **(b)** Corresponding frozen linear-probe accuracy. **(c)** Direct AUROC–probe trajectory as λ varies.

C.3 Pair sensitivity and frozen-feature quality

To separate partner dependence from downstream feature quality, define the channel-wise fused response

$$\mathbf{z}_{i,j}^{(l)} = \frac{1}{THW} \sum_{t,h,w} \mathbf{s}_f^{(l)}(x_i, x_j)[t, :, h, w], \tag{37}$$

and its deviation from the self-pair reference

$$D_{\text{pair}}^{(l)} = \mathbb{E}_{(i,j)} \left[1 - \cos\left(\mathbf{z}_{i,j}^{(l)}, \mathbf{z}_{i,i}^{(l)} \right) \right]. \tag{38}$$

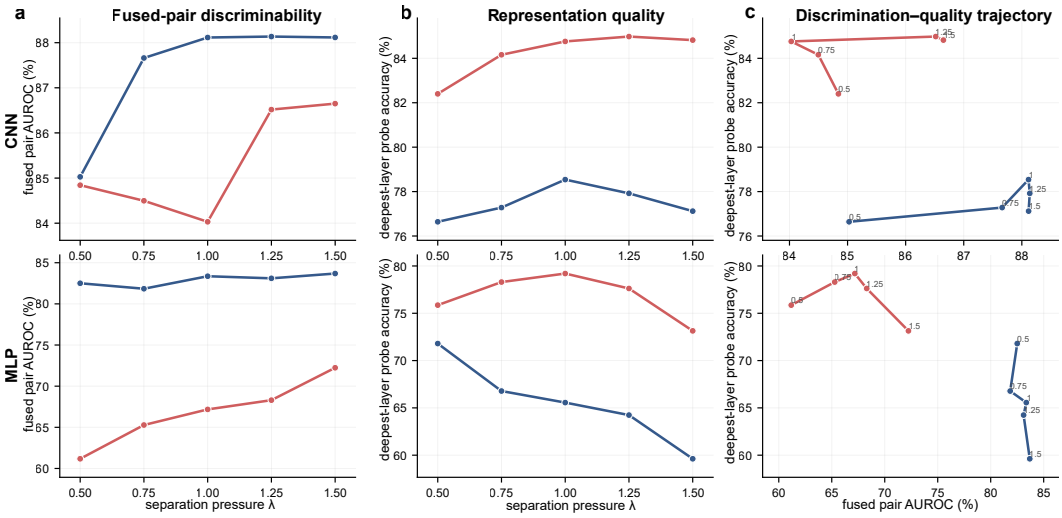

Figure 8: **Controlled separation-pressure sweep on Fashion-MNIST.** The top and bottom rows show the five-layer CNN and MLP backbones. **(a)** Deepest-layer same/different-class AUROC of the raw fused response. **(b)** Corresponding frozen linear-probe accuracy. **(c)** Direct AUROC–probe trajectory as λ varies.

The two architectures exhibit different sensitivity profiles, while Pair-aware SCFF maintains higher probe quality across these regimes. Pair sensitivity, therefore, describes how strongly the fused response depends on its partner, but it is not, by itself, a measure of semantic representation quality.

C.4 Controlled separation pressure

We next change the explicit positive–negative separation pressure while preserving the center of each threshold band. For layer l,

$$c_l = \frac{\theta_{+,l} + \theta_{-,l}}{2}, \qquad b_l = \theta_{+,l} - \theta_{-,l}, \tag{39}$$

and

$$\theta^{(\lambda)}_{+,l} = c_l + \frac{\lambda b_l}{2}, \qquad \theta^{(\lambda)}_{-,l} = c_l - \frac{\lambda b_l}{2}, \qquad \lambda \in \{0.5, 0.75, 1.0, 1.25, 1.5\}. \tag{40}$$

Increasing separation pressure alters raw fused-pair discrimination, but the probe follows a different trajectory: the settings that maximize fused AUROC need not maximize frozen-feature quality. Together with Fig. 4(e)–(f), this supports a simple optimization interpretation. Local self-contrastive learning benefits from resolving informative pair relations, but maximizing a scalar pair-separation statistic is not, by itself, the downstream objective.

The resulting design insight is that the local goodness should match the relational structure of the self-supervisory signal while preserving the degrees of freedom required by the unary representation. Pair-aware SCFF addresses the first part by retaining pre-fusion correspondence in the local update. The remaining challenge for local learning is to control how instance-level pair objectives shape semantic geometry across depth without introducing labels or network-wide credit assignment.