CROSS-SPACE ATTENTIVE FUSION OF HYPERBOLIC AND EUCLIDEAN GRAPH CONVOLUTIONAL NETWORK FOR EVENT-BASED OBJECT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Event streams, which encode brightness variations asynchronously and sparsely, provide unique benefits for event-based visual perception and neural representation learning, making them particularly suitable for tasks that require fine-grained temporal modeling, such as object detection. While traditional Graph Neural Networks (GNNs) based on neighborhood aggregation and message passing excel at characterizing local structures, they suffer from a geometric mismatch with the inherent physical-geometric hierarchy of event streams, which limits their ability to capture global semantic dependencies and spatio-temporal relationships. To this end, we propose an Event-based Cross-space Attentive Fusion Graph Convolutional Network (ECAF-GCN) that leverages a dual geometry-aware architecture in Euclidean and hyperbolic spaces, respectively enabling targeted enhancement of structural and hierarchical features. To dynamically balance and fuse information across geometries, we introduce cross-space fusion to effectively and efficiently alleviate geometric mismatch and enhance hierarchical modeling for hybrid event perception. Extensive experiments validate the effectiveness of our approach. Our code will be released for public use.

1 INTRODUCTION

Unlike conventional cameras that sample scenes at fixed frame rates, event cameras asynchronously and sparsely record per-pixel logarithmic brightness changes, generating spatio-temporal event streams with high temporal resolution Lu et al. (2025); Dong et al. (2026). Compared with frame-based appearance representations, event streams more directly encode scene dynamics and preserve informative temporal cues under rapid motion and challenging illumination Ma et al. (2026); Yu et al. (2026). This event-driven signal provides a natural basis for fine-grained spatio-temporal representation learning and has been increasingly explored in event-based vision tasks.

With the widespread use of GNNs, Euclidean GNNs have gradually become a promising approach for processing event streams. As described in Liang et al. (2023), traditional Euclidean GNNs excel at extracting local structural information and short-range dependencies through the connections between nodes, providing an effective tool for modeling local patterns in event streams. Building on this, retinal events are processed into spatio-temporal graph structures, where Euclidean GNNs capture local spatial relationships and short-range temporal features within the event stream, offering effective support for routine event detection tasks.

However, Euclidean geometry Li et al. (2023) alone cannot accommodate the physical-geometric hierarchy inherent in event streams. First, event streams exhibit multi-scale dynamic evolution, ranging from pixel-level brightness changes and local edge responses to object-level motion patterns. Meanwhile, different motion regions and event clusters may form non-uniform and approximately tree-like structural relations, resulting in a hierarchical organization that progressively evolves from local details to higher-level structures across different scales. In addition, Euclidean space is geometrically flat, with volume growing polynomially with radius. When the underlying data exhibit pronounced hierarchical organization or rapidly expanding structures, Euclidean representations may incur larger embedding distortion, leading to a geometric mismatch. Therefore, local Euclidean

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.

neighborhood aggregation alone may be insufficient to simultaneously preserve fine-grained local information and model long-range hierarchical relations.

To alleviate this geometric limitation, we propose an Event-based Cross-space Attentive Fusion Graph Convolutional Network (ECAF-GCN), which combines Euclidean and hyperbolic representations to jointly model local structural information and hierarchical long-range relations. The Euclidean space excels at capturing local details and enhancing structural features but faces the inevitable intermixing of nodes from different classes in the feature space. In contrast, the hyperbolic space, with its negative curvature space, naturally fits tree-like and exponential expansion structures Si et al. (2026), making it effective at representing global hierarchies and long-range dependencies, thus enhancing hierarchical features. However, due to curvature stretching, it may cause fragmentation of intra-class structures, leading to a loss of local feature precision. Therefore, we present a cross-space fusion module that introduces a cross-space attention mechanism to adaptively align and dynamically fuse the two geometric representations. This fusion module optimizes their synergy, thereby producing a mixed geometric representation that combines global discriminability with local robustness. Specifically, our contributions are as follows:

- **Hyperbolic Learning for Retinal Events:** We show that event streams contain not only local spatio-temporal neighborhood relationships but also hierarchical topological structures induced by asynchronous triggering mechanisms. Moreover, we introduce hyperbolic space into event representation learning, providing a new geometric perspective for modeling the hierarchical structure of event data and further revealing the potential of cross-space modeling for event streams.
- **Dual-geometry Perception Architecture:** We propose a dual geometry-aware architecture, in which Euclidean geometry is employed to capture fine-grained information such as local boundaries, while hyperbolic geometry is utilized to characterize the global consistency of holistic objects, thereby solving the limitation of a single Euclidean space in simultaneously modeling local structural detail and global hierarchical relations.
- **Cross-space Fusion:** We present a cross-space fusion with a hybrid attention mechanism for adaptive alignment and dynamic fusion between Euclidean and hyperbolic representations, optimizing their synergy to produce a mixed geometric representation that combines local robustness with global discriminability.

2 Related Works

2.1 Event-based Object Detection

In event-based object detection Chakravarthi et al. (2024); Zhu et al. (2025), CNN-based methods remain the mainstream in the early stages. The work by Cannici et al. Cannici et al. (2019) is the first to systematically explore how CNNs can be applied to asynchronous event data, which lays the foundation for subsequent research. The study by Messikommer et al. Messikommer et al. (2020) introduces sparse convolution, which performs convolution only at the active pixel locations where events occur. Although the convolution becomes sparser, it essentially still relies on local window aggregation. These methods typically convert asynchronous event streams into frame-like representations and then process them with mature CNN models Smagulova et al. (2025), thereby sacrificing the inherent spatio-temporal and asynchronous characteristics of event data. Spiking Neural Networks (SNNs), as neuromorphic computing models, naturally align with event cameras because of their spiking response mechanism Voudaskas et al. (2025); Shen et al. (2025); Ren et al. (2024). The framework proposed by Nagaraj et al. Nagaraj et al. (2023) achieves low-latency detection through spatio-temporal spike encoding. However, deep SNNs suffer from the vanishing gradient problem and perform poorly in modeling multi-scale event streams, often resulting in limited detection performance. Moreover, selective state space models, exemplified by Mamba, have recently emerged as an efficient alternative to attention for long-range sequence modeling and show promise in event-based tasks Yang et al. (2025). While these methods effectively capture long-term dependencies, their primary role remains the enhancement of temporal or state-space modeling.

By constructing events into spatio-temporal graphs Dampfhoffer et al. (2025); Goyal et al. (2025), GNN-based methods provide a more natural and lossless representation of event streams and thus become one of the mainstream paradigms. They effectively integrate local neighborhood informa-

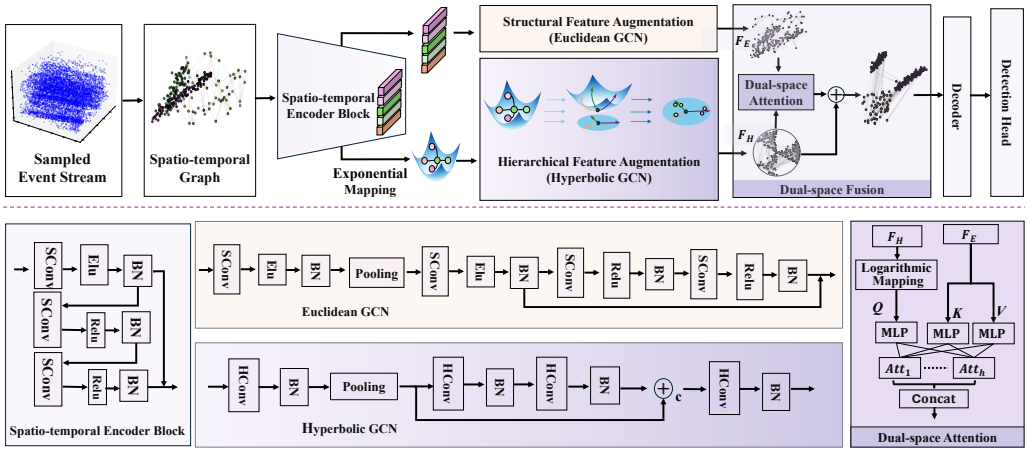


Figure 1: Framework of the proposed ECAF-GCN approach. It includes a dual geometry-aware architecture, where the Euclidean branch captures local structural information, while the hyperbolic branch uses the negative curvature of hyperbolic geometry to encode global hierarchical structures and long-range relationships. The two branches are fused through a cross-space multi-head attention, which dynamically integrates features from both spaces.

tion among nodes. For densely connected clusters in the graph, Graph Convolutional Networks (GCNs) learn smooth representations well and are particularly effective at capturing local similarities, demonstrating clear advantages in object detection Wu et al. (2024); Schaefer et al. (2022). Notably, the Asynchronous Event-based Graph Neural Network (AEGNN) proposed by Schaefer et al. Schaefer et al. (2022) dynamically updates the graph structure upon event arrival through sparse activations and asynchronous message passing, achieving groundbreaking progress.

However, existing GNN-based methods mainly rely on Euclidean representations, while the hierarchical structures and long-range dependencies in event graphs may not be well captured by flat Euclidean geometry. This geometric mismatch can limit the modeling of high-level structural relations and global spatio-temporal dependencies Grover et al. (2025). Thus, we introduce hyperbolic geometry to complement Euclidean GNNs, with the aim of overcoming the geometric constraints of the single Euclidean space.

3 METHODOLOGY

Overview. Our ECAF-GCN constructs a cross-space attentive fusion framework for event detection. As shown in Fig.1, we propose a dual geometry-aware architecture, where the Euclidean branch enhances local structural features from the event stream, while the hyperbolic branch uses the negative curvature of hyperbolic geometry to encode global hierarchical structures and long-range relationships, thus achieving augmentation of global hierarchical features. The two branches are fused through a cross-space attention module, which dynamically integrates features from both spaces, producing a mixed geometric representation that combines local robustness with global discriminability.

3.1 DUAL GEOMETRY-AWARE ARCHITECTURE

An event $e = (x, y, t, p)$, where $p \in \{-1, +1\}$ denotes the polarity. A set of event streams within a time window ΔT is represented as an ordered tuple list $\varepsilon = \{e_i\}_{i=1}^{M}$, where M is the number of events Liu et al. (2024); Zhou et al. (2026). After performing rapid uniform sampling, we construct a spatio-temporal event graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertices $\mathcal{V}$ and edges $\mathcal{E}$. For any pair of nodes i and j, an edge e_{ij} is established between them if their spatio-temporal distance does not exceed R. Once the spatio-temporal graph is constructed, it can be input into the graph neural network. Before entering the dual-branch network, the spatio-temporal graph is fed into a shared spatio-temporal encoder block to extract common features. While maintaining homology, these features provide high-quality input for exponential mapping, guiding the features to be effectively organized into a

hierarchical distribution from the center to the edges in the hyperbolic space. In this way, the shared encoding block provides a common reference for the subsequent dual-geometric space analysis, ensuring that the local structural information from the Euclidean branch and the global hierarchical information from the hyperbolic branch achieve effective cross-space fusion.

Euclidean branch. Between the two independent yet complementary branches, the Euclidean GCN branch is primarily designed to capture fine-grained local structural details and is particularly effective at modeling short-range spatio-temporal dependencies. And we mainly employ graph convolution layers based on B-splines Bekkers (2019), which are piece-wise polynomial functions characterized by local support and defined in the Euclidean space (R^d), and are widely used for extracting spatio-temporal neighborhood features. Its basic form can be written as:

$$h_i^{(l+1)} = \sum_{j \in \mathcal{N}(i)} g_{\Theta}^{(l)}(\mathbf{u}_{ij})\, h_j^{(l)}, \tag{1}$$

where $\mathcal{N}(i)$ denotes the local neighborhood of node i, $\mathbf{h}_j^{(l)}$ is the feature representation of node j at the l-th layer, and $g_{\Theta}^{(l)}(\mathbf{u}_{ij})$ denotes the B-spline kernel defined on the edge attribute $\mathbf{u}_{ij}$, whose output determines the weighted contribution of neighboring nodes to the current node. This formulation indicates that the Euclidean branch does not rely solely on node-wise features, but instead performs position-sensitive aggregation over local neighborhoods by jointly considering node features, graph connectivity, and relative spatio-temporal offsets. Object boundaries and local structures typically manifest as event responses that are temporally continuous and spatially coherent within short time intervals, while background noise tends to appear sparse, fragmented, and inconsistent. Specifically, spatial information determines local neighborhood connectivity during graph construction and is further encoded in the edge attribute to modulate the convolutional kernel response, enabling the network to distinguish local structural patterns under different relative spatial configurations. Temporal information is not only embedded in node features but also reflected by relative temporal offsets, which further characterize the sequential continuity among neighboring events. In addition, polarity encodes local brightness change directions, which is helpful for identifying the consistency of edges, contours, and local motion responses.

Through hierarchical propagation and aggregation with multiple spline convolution layers, the Euclidean branch progressively enhances local boundaries, shapes, spatio-temporal consistency, and polarity variation patterns, thereby producing a local geometric representation that is sensitive to object localization. Nevertheless, since the Euclidean branch mainly relies on local neighborhood aggregation, it remains limited in modeling long-range semantic dependencies and global hierarchical structures, which in turn motivates the introduction of the hyperbolic branch to complement global consistency modeling.

Hyperbolic branch. We leverage the Poincaré ball model Chami et al. (2019) to define a hyperbolic space of n dimensions $\mathbb{D}_c^n$, where c is the curvature. Unlike Euclidean space, hyperbolic space is a non-Euclidean manifold where standard linear operations cannot be directly applied. Therefore, hyperbolic neural operations are typically performed by mapping features between the manifold and its tangent space through exponential and logarithmic maps. The detailed definitions of the hyperbolic operations, including hyperbolic distance, Möbius addition, and Möbius matrix-vector multiplication, are provided in the Appendix.

In this branch, we address the fact that conventional Hyperbolic Graph Convolutions (HGCN) are not developed for asynchronous spatio-temporal event graphs, whose sparse, causally evolving, and hierarchically organized structures differ fundamentally from those of generic graphs Chami et al. (2019). To better accommodate these characteristics, we construct a manifold representation tailored to spatio-temporal event graphs. Unlike the Euclidean branch, which mainly emphasizes local continuity and fine-grained neighborhood structures, the hyperbolic branch is designed to capture whether event nodes can be organized into a semantically coherent object structure at a higher level, thereby alleviating the limitations of a single Euclidean space in modeling non-flat hierarchical patterns. And it is worth noting that the hyperbolic branch does not directly re-read the original event points. Instead, it further reorganizes the intermediate features extracted by the shared encoder and the shallow Euclidean convolutions. Given Euclidean features $\mathcal{F}_i^E \in \mathbb{R}^n$ of event i, we embed them into hyperbolic space via exponential mapping $\mathcal{F}_i^H = \exp_{\mathbf{o}}^{\mathrm{c}}\left(\mathcal{F}_i^E\right)$ with the curvature c:

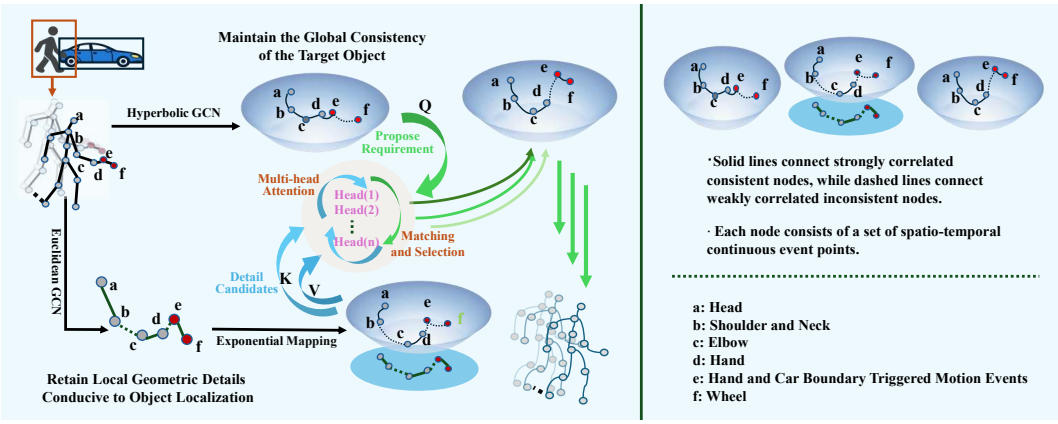


Figure 2: Overview of our cross-space attentive fusion. The hyperbolic space is more effective at preserving global structural consistency and provides stronger modeling capacity for the hierarchical organization of objects such as humans and vehicles, but is less sensitive to fine-grained local details. In contrast, the Euclidean space better describes local geometric details and effectively separates components such as the head, elbow, and hand, yet it is weaker in modeling long-range semantics and global hierarchy. Therefore, the multi-head self-attention fusion enables globally guided selective integration, injecting relevant local details while maintaining global structural stability.

$$\exp_{\mathbf{o}}^{\mathrm{c}}\left(\mathcal{F}_i^E\right)=\mathbf{o}\oplus_c\left(\tanh\left(\sqrt{\mathbf{c}}\frac{\left\|\mathcal{F}_i^E\right\|}{2}\right)\frac{\mathcal{F}_i^E}{\sqrt{\mathbf{c}}\left\|\mathcal{F}_i^E\right\|}\right), \tag{2}$$

where $\mathbf{o}$ denotes the origin in the Poincaré ball model and $\oplus_c$ is the Möbius addition Ganea et al. (2018). This process implies that the input to the hyperbolic branch is an intermediate node representation that already carries certain local spatio-temporal semantics. Through this mapping, features can be organized in the Poincaré ball as a hierarchical distribution that progressively expands from the center toward the boundary, thereby providing a suitable geometric substrate for subsequent global structure modeling. Let the input feature at layer l be $\mathcal{F}_i^{H_l}\in\mathbb{D}_c^n$, the message passing process of one hyperbolic GCN layer can be divided into the following three steps:

$$\widetilde{\mathcal{F}}_i^{H_l}=\Gamma\otimes_c\mathcal{F}_i^{H_l}\oplus_c b, \tag{3}$$

$$\hat{\mathcal{F}}_i^{H_l}=\exp_{\mathbf{o}}^c\left(\sum_{j\in N(i)}A_{i,j}\log_{\mathbf{o}}^c\left(\widetilde{\mathcal{F}}_j^{H_l}\right)\right), \tag{4}$$

$$\mathcal{F}_i^{H_{l+1}}=\exp_{\mathbf{o}}^c\varphi\left(\log_{\mathbf{o}}^c\left(\hat{\mathcal{F}}_i^{H_l}\right)\right). \tag{5}$$

In the above three equations, $\Gamma\in\mathbb{R}^{d'\times d}$ is the weight matrix, $b\in\mathbf{T}_{\mathbf{o}}\mathbb{D}_{\mathrm{c}}^{\mathrm{n}}$ is the bias vector defined on the tangent space at the origin, $\otimes_c$ denotes Möbius matrix vector multiplication. Furthermore, $\log_{\mathbf{o}}^{\mathrm{c}}(\cdot)$ represents the logarithmic map, which is the inverse process of the exponential map. And $A_{i,j}$ is the adjacency matrix, $\varphi(\cdot)$ denotes the activation function. In essence, this process first maps hyperbolic features back to the tangent space for linear transformation and neighborhood aggregation, and then projects them back into the hyperbolic space, thereby modeling hierarchical and complex relationships among nodes while preserving the geometric constraints of hyperbolic space.

Although the hyperbolic branch is also built upon graph topology and edge connectivity, its focus is no longer limited to local spatio-temporal proximity within neighborhoods, but instead shifts toward organizing information at a higher semantic level. It aggregates scattered local substructures into coherent object-level representations, captures longer-range temporal associations, and exploits polarity-induced event patterns to support global semantics. In this way, it performs hierarchy-oriented semantic reorganization, maintaining stable target representations under fragmented responses, dispersed structures, and noisy events. Thus, the hyperbolic branch complements the Euclidean branch by modeling hierarchical relations and long-range dependencies that are difficult to

capture through local aggregation alone. For spatially separated regions belonging to the same object, it forms a more compact and hierarchical representation, yielding stronger global consistency and clearer cross-region dependencies. This provides discriminative high-level structural priors for subsequent cross-space fusion.

3.2 Cross-space Attentive Fusion

Assume that the node features $\mathcal{F}^E$ are obtained from the Euclidean branch and the corresponding node features $\mathcal{F}^H$ are obtained from the hyperbolic branch. Since hyperbolic features cannot directly interact with Euclidean features through standard linear operations, we first map the hyperbolic features to the tangent space at the origin $\mathbf{o}$ using the logarithmic map: $\mathcal{F}^{H\to T} = \log_{\mathbf{o}}^{c}\left(\mathcal{F}^{H}\right)$. The tangent space is locally Euclidean, allowing the mapped hyperbolic features to interact with the Euclidean features $\mathcal{F}^E$ in a common representation space. For the i-th attention head, the mapped hyperbolic features are used as the Query, while the Euclidean features serve as the Key and Value:

$$Q = W_i^Q \mathcal{F}^{H\to T}, \qquad K = W_i^K \mathcal{F}^E, \qquad V = W_i^V \mathcal{F}^E. \tag{6}$$

The attention output of the i-th head is then computed as

$$Att_i = \mathrm{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right) V, \tag{7}$$

where W_i^Q, W_i^K, and W_i^V denote the linear projection matrices for the Query, Key, and Value, respectively, and d_k is the feature dimension of each attention head. In this way, the hierarchical information encoded by the hyperbolic branch acts as a structural query to selectively retrieve fine-grained local evidence from the Euclidean branch, enabling effective interaction and fusion between the two geometric representations within the tangent space.

The module does not simply concatenate Euclidean and hyperbolic features; instead, it performs *selective injection of local details guided by global hierarchical semantics* in a common tangent space. Specifically, the hyperbolic feature $\mathcal{F}^H$ encodes target-level hierarchical relations, long-range dependencies, and global consistency, and is therefore used as the Query Q to represent high-level structural requirements. In contrast, the Euclidean feature $\mathcal{F}^E$ serves as the Key K and Value V, providing fine-grained cues such as local boundaries, shapes, and short-range spatio-temporal continuity. Thus, the cross-space attention does not perform equal-weight fusion, but selectively retrieves local Euclidean evidence that is most relevant to the structural requirements encoded by the hyperbolic branch.

After that, the attention outputs of h heads are concatenated, and a linear transformation is applied to obtain the final output:

$$Att_{\mathrm{out}} = W^o \mathrm{Concat}(Att_1, \ldots, Att_h), \tag{8}$$

where W^o denotes the output projection matrix. The multi-head attention mechanism enables the model to capture diverse cross-space correspondences across subspaces. Different attention heads can focus on boundary details, local topological changes, short-range motion-triggered patterns, or matching relationships among different regions within the target, thereby enhancing the richness and robustness of the fused representation and effectively complementing, correcting, and refining the global hierarchical representation in the hyperbolic branch.

As illustrated in Fig.2, the hyperbolic and Euclidean branches provide complementary event-graph representations. The hyperbolic branch better preserves global object consistency and hierarchical organization, encouraging semantically related nodes from the same entity, such as humans or vehicles, to form coherent representations. However, this global modeling is less sensitive to fine-grained local details, which may cause ambiguous boundary events, e.g., between a hand and a vehicle, to be incorrectly grouped. In contrast, the Euclidean branch is better at capturing local geometric details, distinguishing object parts such as the head, elbow, and hand, and providing localization-sensitive structural cues. However, due to its reliance on local connectivity, it is weaker in modeling long-range semantic dependencies and global hierarchical structures, which may cause spatially distant parts of the same object, such as the head and hand regions shown in the figure, to appear weakly correlated. This complementarity directly motivates our dual-geometry encoding and cross-space fusion.

Table 1: Comparison of detection performance on the N-Caltech101 and Gen1 datasets.

Method	Backbone	N-Caltech101		Gen1		Param(M)
		mAP	MFLOPS/ev	mAP	MFLOPS/ev	
SFOD Fan et al. (2024)	SNNs	-	-	0.321	-	11.9
EAS-SNN(M) Wang et al. (2024)	SNNs	0.664	-	0.409	-	25.3
Events+YOLOv3 Jiang et al. (2019)	CNNs	-	-	0.312	65558	>60
YOLE Cannici et al. (2019)	CNNs	0.398	3682	-	-	-
Inception+SSD Iacono et al. (2018)	CNNs	-	-	0.301	27183	>60
Standard-Cov Messikommer et al. (2020)	CNNs	0.623	1584	0.145	2098	-
Farse-CNN Santambrogio et al. (2024)	CNNs	-	-	0.300	-	0.137
AsyNet Messikommer et al. (2020)	CNNs	0.643	200	0.129	205	11.4
RED Perot et al. (2020)	CNNs	-	-	0.400	4712	24.1
DAGR_N Gehrig & Scaramuzza (2024)	GNNs	0.629	2.28	0.263	1.36	2.7
DAGR Gehrig & Scaramuzza (2024)	GNNs	0.702	6.85	0.304	4.58	2.7
AEGNN Schaefer et al. (2022)	GNNs	0.595	7.41	0.163	5.26	20.4
EDGCN Deng et al. (2024)	GNNs	0.657	-	-	-	0.77
NVS-S Li et al. (2021)	GNNs	0.346	7.8	0.086	7.8	0.9
Ours	GNNs	**0.706**	24.7	**0.435**	22.6	28.5

In summary, this module mitigates the limitations of using either branch alone. The Euclidean branch preserves local boundaries and fine-grained details, but may lose global object consistency in sparse, fragmented, or noisy event graphs. Conversely, the hyperbolic branch better models global hierarchies, yet remains less precise in capturing local boundaries and localization cues. Our cross-space multi-head attention avoids simple feature stacking and instead performs semantically guided selective fusion, injecting relevant local details while preserving global structural stability.

4 EXPERIMENTS

Datasets. We conduct systematic experiments on two challenging benchmark datasets. The N-Caltech101 dataset Orchard et al. (2015) is a converted event stream derived from the Caltech101 image dataset, with each event stream lasting 300 milliseconds. While every sample contains only one bounding box, it covers 101 object categories. The Gen1 dataset De Tournemire et al. (2020) is a large real-world dataset consisting of 39 hours of open road and various driving scenarios, with 228,123 bounding boxes for cars and 27,658 for pedestrians.

Implementation details. The curvature of each HGCN layer is treated as a learnable parameter, initialized with 0.4, while the dropout rate is set to 0.1. Our model employs the YOLOX detection head for bounding box regression and category prediction. And we use AdamW with batch size 32 and an initial learning rate of 10^{-3}. All experiments are conducted on a workstation equipped with an NVIDIA GeForce RTX 4090 GPU.

4.1 EXPERIMENTAL RESULTS

We employ the mean average precision (mAP) score to evaluate the detection performance of our model. And we conduct comparisons with several competitive SOTA models that represent leading approaches in event-based object detection. As shown in Tab. 1, our approach achieves 0.706 mAP on N-Caltech101 and 0.435 mAP on Gen1, outperforming the compared methods. These results highlight the efficacy of our approach, particularly with the integration of Hyperbolic GCN and cross-space attention fusion modules, which enable the model to capture long-range dependencies and improve feature fusion for enhanced detection performance. Unlike Mamba-based methods, which primarily improve representation learning from the perspective of temporal modeling, our work focuses on the geometric mismatch inherent in graph representations of event streams. Since the two approaches target fundamentally different bottlenecks, they are better viewed as complementary approaches rather than direct alternatives. In addition, from the perspective of model complexity, the Euclidean–hyperbolic dual-branch architecture and cross-space fusion mechanism increase the parameter count from 20.4M in AEGNN to 28.5M, while the computational cost on N-

Table 2: Component ablation on N-Caltech101 and Gen1.

Hyperbolic GCN	Attentive Fusion	N-Caltech101	Gen1
×	×	0.612	0.245
✓	×	0.644	0.328
✓	✓	**0.706**	**0.435**

Table 3: Comparison of different geometric branch configurations on N-Caltech101.

Branch configuration	Parameters	mAP
Euclidean + Euclidean	55.8M	0.593
Hyperbolic + Hyperbolic	1.8M	0.683
Euclidean + Hyperbolic	**28.5M**	**0.706**

Caltech101 and Gen1 rises from 7.41/5.26 to 24.7/22.6 MFLOPS/ev, respectively. This additional overhead mainly stems from dual-geometry modeling and cross-space fusion and can be regarded as a reasonable cost for the performance gain. Despite the increased complexity, ECAF-GCN remains less computationally expensive than most CNN-based methods and achieves the best detection performance in the table, demonstrating a favorable accuracy–efficiency trade-off.

4.2 Ablation

To evaluate the key components of ECAF-GCN, we conduct ablation studies under a unified setting, with the same data split, training protocol, and detection head for all variants.

Component ablation. As shown in Table 2, introducing the Hyperbolic GCN improves mAP from 0.612/0.245 to 0.644/0.328 on N-Caltech101/Gen1, indicating that hyperbolic representations complement the Euclidean branch with hierarchical and long-range structural information. With cross-space attentive fusion, performance further increases to 0.706/0.435, validating the benefit of explicit interaction between heterogeneous geometric representations.

Effect of complementary geometries. We further investigate whether the performance gain mainly originates from the dual-branch architecture or increased model capacity. To this end, we replace the heterogeneous Euclidean–hyperbolic design with homogeneous Euclidean–Euclidean and Hyperbolic–Hyperbolic variants while keeping the remaining experimental settings unchanged. As shown in Table 3, the Euclidean–Euclidean variant contains 55.8M parameters, substantially more than our 28.5M model, but achieves only 0.593 mAP. In contrast, the Hyperbolic–Hyperbolic variant contains only 1.8M parameters and achieves 0.683 mAP. The proposed heterogeneous Euclidean–Hyperbolic architecture further improves the performance to 0.706 mAP. The non-monotonic relationship between parameter count and accuracy indicates that the gain cannot be explained simply by additional model capacity. Instead, the result supports the complementarity between local Euclidean representations and hierarchical hyperbolic representations.

Cross-space fusion design. We further examine whether the direction of cross-space interaction is arbitrary. In the proposed design, the hyperbolic features are mapped back to the tangent space and used as the Query, while the Euclidean features serve as the Key and Value. Reversing this interaction direction reduces the mAP from 0.706 to 0.648, indicating the importance of this asymmetric role assignment. Specifically, the hyperbolic branch provides high-level structural requirements, whereas the Euclidean branch supplies fine-grained local evidence for selective retrieval. We further compare our fusion strategy with several generic alternatives. Table 4 shows that simple concatenation, conventional cross-attention, joint self-attention, and Transformer-style fusion all perform worse than the proposed cross-space attentive fusion. ConcatMLPFusion only performs point-wise feature mixing and therefore cannot adaptively select information according to the structural state of each node. SimpleCrossAttention introduces directional query–key–value interaction, but does not explicitly preserve the distinct geometric roles of the two representations. JointSelfAttentionFusion and TransformerFusion permit more general interactions among all Euclidean and hyperbolic tokens, but treat them more symmetrically. In contrast, our method explicitly assigns the hyperbolic representation to structural querying and the Euclidean representation to local evidence provision.

Table 4: Comparison of fusion strategies on N-Caltech101.

Fusion method	mAP
SimpleCrossAttention	0.646
ConcatMLPFusion	0.662
JointSelfAttentionFusion	0.663
TransformerFusion	0.666
Proposed cross-space attentive fusion	**0.706**

Table 5: Gromov hyperbolicity on N-Caltech101.

Metric	Event graph	Random graph
Estimated $\hat{\delta}$	21.54 ± 10.30	1.68 ± 0.24
Median $\hat{\delta}$	21.25	1.50
Estimated diameter $\hat{D}$	157.74 ± 49.78	5.01 ± 0.08
Normalized $\hat{\delta}/\hat{D}$	$\mathbf{0.151 \pm 0.088}$	$\mathbf{0.336 \pm 0.048}$
Largest connected-component ratio	0.779 ± 0.130	1.000

The results suggest that the improvement does not simply come from using a more expressive attention module, but from the asymmetric cross-space inductive bias.

Demonstration of hierarchical graph structure. To examine whether the constructed event graphs exhibit a metric structure suitable for hierarchical representation learning, we analyze their Gromov hyperbolicity on N-Caltech101, with the full definition provided in Appendix D. As summarized in Table 5, the raw hyperbolicity estimate $\hat{\delta}$ of the event graphs is larger than that of randomized graphs. However, raw $\hat{\delta}$ is strongly affected by graph scale: the estimated geodesic diameter of the event graphs is 157.74 ± 49.78, compared with only 5.01 ± 0.08 for the randomized graphs. After normalization by the estimated graph diameter, the event graphs achieve $\hat{\delta}/\hat{D} = 0.151 \pm 0.088$, substantially lower than 0.336 ± 0.048 for the randomized controls. A paired Wilcoxon test further confirms a significant difference between the event graphs and randomized graphs. The test gives $p = 1.78 \times 10^{-15}$ for the raw $\hat{\delta}$ and $p = 5.33 \times 10^{-15}$ for the normalized $\hat{\delta}/\hat{D}$. Since the normalized metric removes the dominant effect of graph scale, the latter result provides stronger evidence that the event graphs exhibit a more tree-like metric organization relative to their overall geodesic scale. This observation provides quantitative structural motivation for introducing hyperbolic representation learning, while it does not imply that the input event graph itself forms an explicit tree.

Additional analyses. Additional analyses of training convergence, initial learnable curvature, graph-construction hyperparameters, and qualitative detection results are provided in the Appendix. Specifically, we report the epoch-wise validation behavior, curvature-initialization sensitivity, the effects of the graph radius R and maximum neighborhood size $d_{\max}$, as well as representative detection visualizations on Gen1 compared with AEGNN.

5 Conclusion

In this work, we propose ECAF-GCN, a novel event-based object detection framework that unifies Euclidean and hyperbolic geometries within a cross-space parallel architecture. By jointly modeling fine-grained local structures and high-level hierarchical organization, the proposed method alleviates the limitations of conventional single-space graph representations for asynchronous event streams. In addition, the cross-space attention mechanism enables adaptive interaction and fusion between heterogeneous geometric representations, resulting in more expressive event features. Extensive experiments on N-Caltech101 and Gen1 demonstrate that ECAF-GCN outperforms comparable SOTA methods, confirming the effectiveness of both the dual-space design and the attentive fusion mechanism. Beyond the performance gains, our results indicate that multi-space interaction provides a new way to explore the geometry-aware representation learning for event streams, opening a promising path for future event-based perception models.

AI USE STATEMENT

In this work, we used generative AI tools to assist with language editing, sentence refinement, and improving the readability and conciseness of parts of the manuscript. Generative AI was not used to generate experimental results, formulate mathematical claims, design the core methodology, or interpret the experimental findings. All AI-assisted text was carefully reviewed and revised by the authors to ensure technical accuracy, consistency with the conducted research, and appropriate attribution to the relevant literature. The authors take full responsibility for the final content of this work, including all text, claims, experimental results, and conclusions.

REFERENCES

Erik J Bekkers. B-spline cnns on lie groups. In *International Conference on Learning Representations*, 2019.

Marco Cannici, Marco Ciccone, Andrea Romanoni, and Matteo Matteucci. Asynchronous convolutional networks for object detection in neuromorphic cameras. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019.

Bharatesh Chakravarthi, Aayush Atul Verma, Kostas Daniilidis, Cornelia Fermuller, and Yezhou Yang. Recent event camera innovations: A survey. In *the European conference on computer vision*, pp. 342–376, 2024.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. *Advances in Neural Information Processing Systems*, 2019.

Manon Dampfhoffer, Thomas Mesquida, Damien Joubert, Thomas Dalgaty, Pascal Vivet, and Christoph Posch. Graph neural network combining event stream and periodic aggregation for low-latency event-based vision. In *the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6909–6918, 2025.

Pierre De Tournemire, Davide Nitti, Etienne Perot, Davide Migliore, and Amos Sironi. A large scale event-based detection dataset for automotive. *arXiv preprint arXiv:2001.08499*, 2020. URL `https://api.semanticscholar.org/CorpusID:210860813`.

Yongjian Deng, Hao Chen, and Youfu Li. A dynamic gcn with cross-representation distillation for event-based learning. In *AAAI Conference on Artificial Intelligence*, pp. 1492–1500, 2024.

Wen Dong, Haiyang Mei, Yinglian Ji, Yutong Jiang, Ziqi Wei, Shengfeng He, and Xin Yang. You only look intensity once: Event-driven long-term high-speed object detection. *International Journal of Computer Vision*, pp. 149, 2026.

Yimeng Fan, Wei Zhang, Changsong Liu, Mingyang Li, and Wenrui Lu. Sfod: Spiking fusion object detector. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 17191–17200, 2024.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in Neural Information Processing Systems*, 2018.

Daniel Gehrig and Davide Scaramuzza. Low-latency automotive vision with event cameras. *Nature*, pp. 1034–1040, 2024.

Gaurvi Goyal, Pham Cong Thuong, Arren Glover, Masayoshi Mizuno, and Chiara Bartolozzi. Graphenet: Event-driven human pose estimation with a graph neural network. In *the IEEE International Conference on Computer Vision*, pp. 4665–4674, 2025.

Karish Grover, Haiyang Yu, Qi Zhu, Han Xie, Vassilis Ioannidis, Christos Faloutsos, et al. Spectro-riemannian graph neural networks. In *International Conference on Learning Representations*, volume 2025, pp. 57833–57859, 2025.

Massimiliano Iacono, Stefan Weber, Arren Glover, and Chiara Bartolozzi. Towards event-driven object detection with off-the-shelf deep learning. In *the IEEE International Conference on Intelligent Robots and Systems*, pp. 1–9, 2018.

Zhuangyi Jiang, Pengfei Xia, Kai Huang, Walter Stechele, Guang Chen, Zhenshan Bing, and Alois Knoll. Mixed frame-/event-driven fast pedestrian detection. In *the International Conference on Robotics and Automation*, pp. 8332–8338, 2019.

Lingxiao Li, Yi Zhang, and Shuhui Wang. The euclidean space is evil: Hyperbolic attribute editing for few-shot image generation. In *the IEEE International Conference on Computer Vision*, pp. 22714–22724, 2023.

Yijin Li, Han Zhou, Bangbang Yang, Ye Zhang, Zhaopeng Cui, Hujun Bao, and Guofeng Zhang. Graph-based asynchronous event processing for rapid object recognition. In *the IEEE International Conference on Computer Vision*, pp. 934–943, 2021.

Jiye Liang, Zijin Du, Jianqing Liang, Kaixuan Yao, and Feilong Cao. Long and short-range dependency graph structure learning framework on point cloud. *the IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 14975–14989, 2023.

Zibin Liu, Banglei Guan, Yang Shang, Shunkun Liang, Zhenbao Yu, and Qifeng Yu. Optical flow-guided 6dof object pose tracking with an event camera. In *the ACM International Conference on Multimedia*, pp. 6501–6509, 2024.

Yunfan Lu, Yanlin Qian, Ziyang Rao, Junren Xiao, Liming Chen, and Hui Xiong. Rgb-event isp: The dataset and benchmark. In *International Conference on Learning Representations*, volume 2025, pp. 16556–16570, 2025.

Wenzong Ma, Zhuoxiao Li, Jinjing Zhu, Tongyan Hua, Kanghao Chen, Zidong Cao, Da Yang, Peilun Shi, Yibo Zhou, Wufan Zhao, et al. Skyevents: A large-scale event-enhanced uav dataset for robust 3d scene reconstruction. In *International Conference on Learning Representations*, 2026.

Nico Messikommer, Daniel Gehrig, Antonio Loquercio, and Davide Scaramuzza. Event-based asynchronous sparse convolutional networks. In *the European Conference on Computer Vision*, pp. 415–431, 2020.

Manish Nagaraj, Chamika Mihiranga Liyanagedera, and Kaushik Roy. Dotie-detecting objects through temporal isolation of events using a spiking architecture. In *IEEE International Conference on Robotics and Automation*, pp. 4858–4864, 2023.

Garrick Orchard, Ajinkya Jayawant, Gregory K Cohen, and Nitish Thakor. Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in Neuroscience*, 9:437, 2015. URL https://www.frontiersin.org/journals/neuroscience/articles/10.3389/fnins.2015.00437.

Etienne Perot, Pierre De Tournemire, Davide Nitti, Jonathan Masci, and Amos Sironi. Learning to detect objects with a 1 megapixel event camera. *Advances in Neural Information Processing Systems*, pp. 16639–16652, 2020.

Hongwei Ren, Yue Zhou, Xiaopeng Lin, Yulong Huang, Haotian Fu, Jie Song, and Bojun Cheng. Spikepoint: An efficient point-based spiking neural network for event cameras action recognition. In *International Conference on Learning Representations*, volume 2024, pp. 27827–27846, 2024.

Riccardo Santambrogio, Marco Cannici, and Matteo Matteucci. Farse-cnn: Fully asynchronous, recurrent and sparse event-based cnn. In *the European Conference on Computer Vision*, pp. 1–18, 2024.

Simon Schaefer, Daniel Gehrig, and Davide Scaramuzza. Aegnn: Asynchronous event-based graph neural networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 12371–12381, 2022.

Jiangrong Shen, Qi Xu, Gang Pan, and Badong Chen. Improving the sparse structure learning of spiking neural networks from the view of compression efficiency. In *International Conference on Learning Representations*, volume 2025, pp. 92834–92846, 2025.

Leping Si, Meimei Yang, Hui Xue, Shipeng Zhu, and Pengfei Fang. Adaptive hyperbolic kernels: Modulated embedding in de branges-rovnyak spaces. In *the AAAI Conference on Artificial Intelligence*, volume 40, pp. 25437–25445, 2026.

Kamilya Smagulova, Ahmed Elsheikh, Diego A Silva, Mohammed E Fouda, and Ahmed M Eltawil. Efficient and real-time perception: a survey on end-to-end event-based object detection in autonomous driving. *Frontiers in Robotics and AI*, pp. 1674421, 2025.

Michael Voudaskas, Jack Iain MacLean, Neale AW Dutton, Brian D Stewart, and Istvan Gyongy. Spiking neural networks in imaging: A review and case study. *Sensors*, pp. 6747, 2025.

Ziming Wang, Ziling Wang, Huaning Li, Lang Qin, Runhao Jiang, De Ma, and Huajin Tang. Eassnn: End-to-end adaptive sampling and representation for event-based detection with recurrent spiking neural networks. In *the European Conference on Computer Vision*, pp. 310–328, 2024.

Sheng Wu, Hang Sheng, Hui Feng, and Bo Hu. Egsst: Event-based graph spatiotemporal sensitive transformer for object detection. *Advances in Neural Information Processing Systems*, 37: 120526–120548, 2024.

Nan Yang, Yang Wang, Zhanwen Liu, Meng Li, Yisheng An, and Xiangmo Zhao. Smamba: Sparse mamba for event-based object detection. In *AAAI Conference on Artificial Intelligence*, pp. 9229–9237, 2025.

Wangbo Yu, Chaoran Feng, Jianing Li, Aofan Zhang, Zhenyu Tang, Mingyi Guo, Wei Zhang, Zhengyu Ma, Li Yuan, and Yonghong Tian. Ea3d: Event-augmented 3d diffusion for generalizable novel view synthesis. In *The Fourteenth International Conference on Learning Representations*, 2026.

Yun Zhou, Hongfu Yin, Chunyu Tan, Qiaoyun Wu, Changyin Sun, and Richang Hong. An efficient hybrid cascade tracker with spiking neural networks for event domain tracking. *ACM Transactions on Multimedia Computing, Communications and Applications*, 2026.

Lin Zhu, Xiao Wang, Lizhi Wang, Hua Huang, et al. Rethinking scale-aware temporal encoding for event-based object detection. In *Advances in Neural Information Processing Systems*, 2025.

A Training Convergence and Curvature Sensitivity

The main paper primarily focuses on the architectural and geometric ablation studies that are most directly related to the proposed contributions. For completeness, we additionally report the validation mAP across training epochs and the model's sensitivity to the initialization of the learnable curvature. As shown in Fig. 3, the validation mAP increases rapidly during early training and becomes stable in later epochs. The best validation result is obtained at epoch 19 with a Val/mAP of 0.7063. We further initialize the learnable curvature with $c \in \{0.2, 0.4, 0.6, 0.8\}$, resulting in mAP values of 0.705, 0.706, 0.702, and 0.703, respectively. The variation is only 0.004, indicating low sensitivity to curvature initialization.

B Event-Graph Construction and Density Sensitivity

For event e_i, we define the normalized spatio-temporal coordinate as

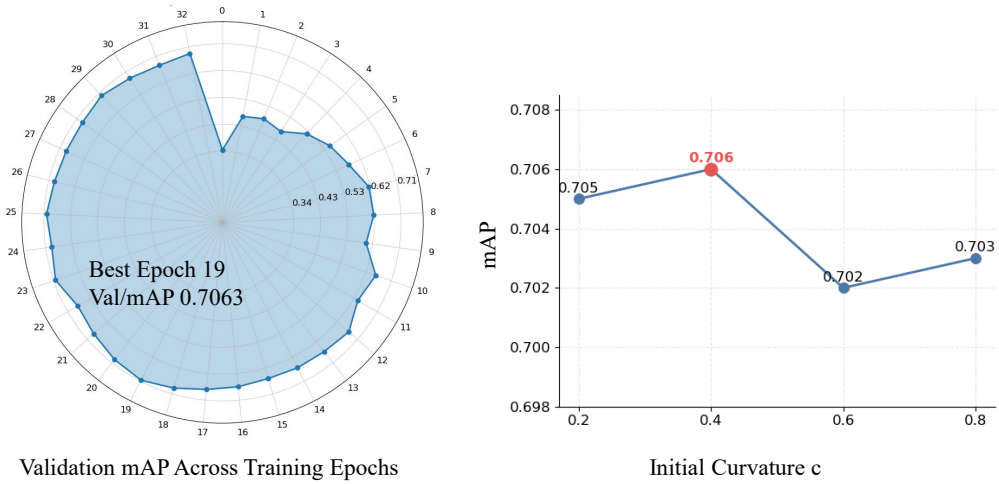
$$\mathbf{q}_i = [x_i, y_i, \tilde{t}_i], \tag{9}$$

where x_i and y_i are the spatial coordinates and $\tilde{t}_i$ is the normalized timestamp. The spatio-temporal distance between nodes v_i and v_j is

$$d_{ij} = \|\mathbf{q}_i - \mathbf{q}_j\|_2 = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (\tilde{t}_i - \tilde{t}_j)^2}. \tag{10}$$

The radius-based candidate neighborhood is

$$\mathcal{N}_R(i) = \{j \mid j \neq i,\ d_{ij} \leq R\}. \tag{11}$$



Validation mAP Across Training Epochs

Initial Curvature c

Figure 3: The validation mAP improves rapidly during training and reaches its best value of 0.7063 at epoch 19, while different initial curvature values lead to only minor performance variations, indicating low sensitivity to curvature initialization.

To control graph density and computational cost, the actual neighborhood is further constrained by the maximum number of neighbors $d_{\max}$:

$$\mathcal{N}(i) \subseteq \mathcal{N}_R(i), \qquad |\mathcal{N}(i)| \leq \min(|\mathcal{N}_R(i)|, d_{\max}). \tag{12}$$

Thus, R determines the spatio-temporal support of each neighborhood, whereas $d_{\max}$ limits the maximum node degree and the upper bound of graph complexity.

We first fix $R = 3$ and vary $d_{\max}$. Tab. 6 shows that a very small neighbor limit removes useful local spatio-temporal connections, whereas the gains progressively diminish as the graph becomes denser while computational cost continues to increase.

Table 6: Sensitivity to the maximum number of neighbors on N-Caltech101 ($R = 3$).

$d_{\max}$	8	16	20	24
mAP	0.690	0.706	0.710	0.713

We then fix $d_{\max} = 16$ and vary R. As shown in Tab. 7, a larger radius can introduce events that are spatially or temporally less relevant to the target node. These additional edges increase graph density but may weaken local structural consistency and introduce noisy message propagation.

Table 7: Sensitivity to graph radius on N-Caltech101 ($d_{\max} = 16$).

Radius R	2	3	4	5
mAP	0.708	0.706	0.701	0.688

These results show that graph quality cannot be characterized by edge density alone. An excessively sparse graph may fragment meaningful motion patterns, while an excessively dense graph may connect weakly related events, reduce local discriminability, and increase the risk of feature over-smoothing. The tested configuration with the highest mAP should be interpreted as an empirical optimum within the evaluated range rather than a theoretically universal graph construction.

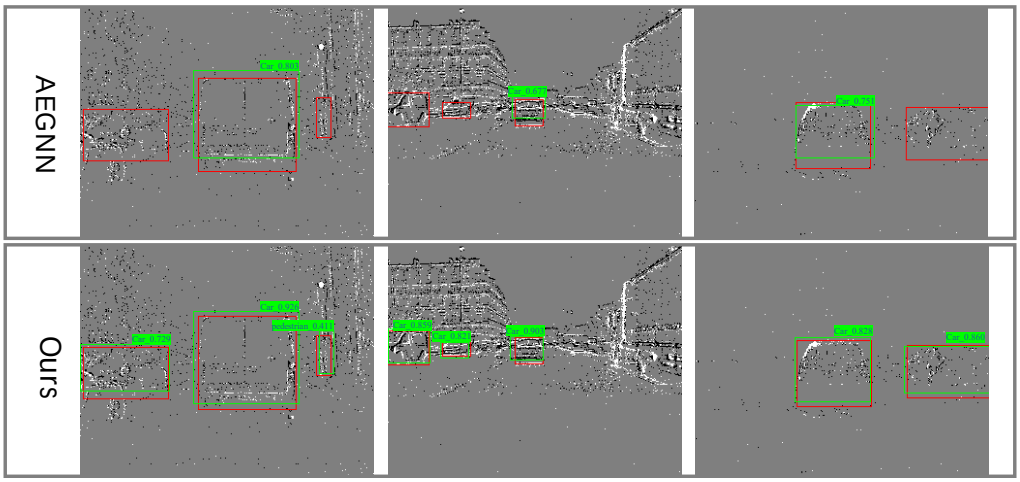


Figure 4: Detection visualization on the Gen1 dataset. The red boxes denote the ground-truth annotations, while the green boxes represent our detected bounding boxes. And the text in the top-right corner of the bounding boxes indicates the detected class and confidence score.

C Qualitative Analysis on the Gen1 Dataset

To visually demonstrate the effectiveness of our approach in object detection, we conducted a qualitative analysis on the Gen1 dataset. As shown in Fig.4, compared with AEGNN, a representative GNN-based method, our approach achieves better overlap between predicted bounding boxes and ground-truth annotations. Meanwhile, AEGNN suffers from more noticeable false negatives, whereas our method effectively reduces missed detections, leading to improved overall detection accuracy.

D Gromov Hyperbolicity Definition and Estimation

To quantify whether the event graph exhibits a tree-like metric organization, we use the four-point Gromov hyperbolicity condition. For any four graph nodes a, b, c, d, define

$$S_1 = d(a,b) + d(c,d), \qquad S_2 = d(a,c) + d(b,d), \qquad S_3 = d(a,d) + d(b,c). \tag{13}$$

After sorting the three sums such that

$$S_{(1)} \le S_{(2)} \le S_{(3)}, \tag{14}$$

the four-point hyperbolicity is

$$\delta(a,b,c,d) = \frac{S_{(3)} - S_{(2)}}{2}. \tag{15}$$

A smaller value indicates that the graph metric is closer to a tree metric. Because the raw hyperbolicity constant depends on graph scale, we additionally report the diameter-normalized measure $\hat{\delta}/\hat{D}$, where $\hat{D}$ is the estimated geodesic diameter. This normalization is important here because the event graphs have much larger diameters than the randomized controls. Statistical significance is assessed using paired Wilcoxon tests; The corresponding quantitative results are reported in Table 5 in the main paper.

Scope of the Hierarchical Interpretation The graph edges are constructed from spatial and normalized temporal distances and therefore do not explicitly encode temporal direction, physical connectivity, force transmission, or object identity. Accordingly, ECAF-GCN learns hierarchical representations of spatio-temporal correlations rather than explicit directed causal dependencies. The hyperbolic branch also does not require the input graph to be an explicit tree. Through repeated neighborhood aggregation in an L-layer graph network, each node representation depends on its L-hop receptive field, allowing local event neighborhoods to be progressively organized into motion fragments, object parts, and higher-level event structures.

E HYPERBOLIC GEOMETRY PRELIMINARIES

Hyperbolic Space and Tangent Space. The hyperbolic space is not a vector space, prohibiting direct linear operations. To preserve its global hyperbolic geometry while leveraging the linear convenience of Euclidean space locally, we need the tangent space to unfold the manifold at a point into a local Euclidean space. For a point $\mathbf{x} \in \mathbb{D}_c^n$, the tangent space $\mathbb{T}_{\mathbf{x}}\mathbb{D}_c^n$ describes the set of tangent vectors through it. A tangent vector $\mathbf{u} \in \mathbb{T}_{\mathbf{x}}\mathbb{D}_c^n$ must satisfy the orthogonality condition: $\langle \mathbf{u}, \mathbf{x} \rangle_{\mathbb{D}_c^n} = 0$, indicating that $\mathbf{u}$ is orthogonal to the radial direction from the origin to $\mathbf{x}$.

Hyperbolic Distance. The hyperbolic distance is the *shortest path* metric in negatively curved spaces. By embedding data into the Poincaré disk and using the hyperbolic distance, we can capture the complex relationships among data points with lower dimensions and higher efficiency. For two points $\mathbf{x}, \mathbf{y} \in \mathbb{D}_c^n$ the hyperbolic distance is defined as:

$$\mathbf{d}_{\mathbb{D}_c^n}(\mathbf{x}, \mathbf{y}) = \frac{1}{\sqrt{c}} \operatorname{arcosh}\left(1 + \frac{2c \left\|\mathbf{x} - \mathbf{y}\right\|^2}{\left(1 - c \left\|\mathbf{x}\right\|^2\right)\left(1 - c \left\|\mathbf{y}\right\|^2\right)}\right), \tag{16}$$

Möbius Addition and Möbius Matrix-Vector Multiplication. Möbius addition and matrix-vector multiplication form the foundation for building hyperbolic neural networks. The operations of Möbius addition and matrix-vector multiplication in hyperbolic space are defined as follows:

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{\left(1 + 2c \left\langle \mathbf{x}, \mathbf{y} \right\rangle + c \left\|\mathbf{y}\right\|^2\right)\mathbf{x} + \left(1 - c \left\|\mathbf{x}\right\|^2\right)\mathbf{y}}{1 + 2c \left\langle \mathbf{x}, \mathbf{y} \right\rangle + c^2 \left\|\mathbf{x}\right\|^2 \left\|\mathbf{y}\right\|^2}, \tag{17}$$

$$\mathbf{M} \otimes_c \mathbf{x} = \exp_{\mathbf{o}}^{\mathrm{c}}\left(\mathbf{M} \log_{\mathbf{o}}^{\mathrm{c}}\left(\mathbf{x}\right)\right), \tag{18}$$

where $\mathbf{M}$ is a learnable weight matrix, $\exp_{\mathbf{o}}^{\mathrm{c}}$ and $\log_{\mathbf{o}}^{\mathrm{c}}$ which is described in detail in the Methodology of Section 3, denote the exponential and logarithmic maps anchored at the origin respectively. And here, $(\langle x, y \rangle = |x|\,|y| \cos(\measuredangle(x, y)))$, where $(\cos(\measuredangle(x, y)))$ represents the cosine of the angle between the vectors x and y.