G2N2
Lightweight Event Stream Classification with GRU Graph Neural Networks

T. Mesquida[1], M. Dampfhoffer[1], T. Dalgaty[1], P. Vivet[1], A. Sironi[2], C. Posch[2]

1 CEA LIST, Grenoble, France
2 PROPHESEE, Paris, France
Contact: thomas.mesquida@cea.fr

Event Stream Processing

Event streams are asynchronous in nature, each event containing precise timing information. Legacy processing flows for such data include collapsing back events to frames which degrades inherent data sparsity and timing contained in events.

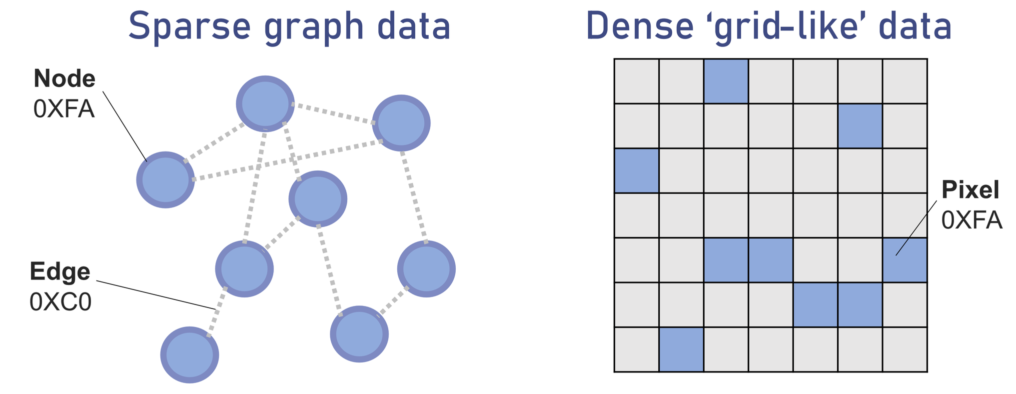


Fig 1. Sparse Graph vs Dense frames.

Graph data structure maintains precise timing, involving data sharing and computation only at nodes and edges. Machine learning techniques were adapted to perform Graph Processing.

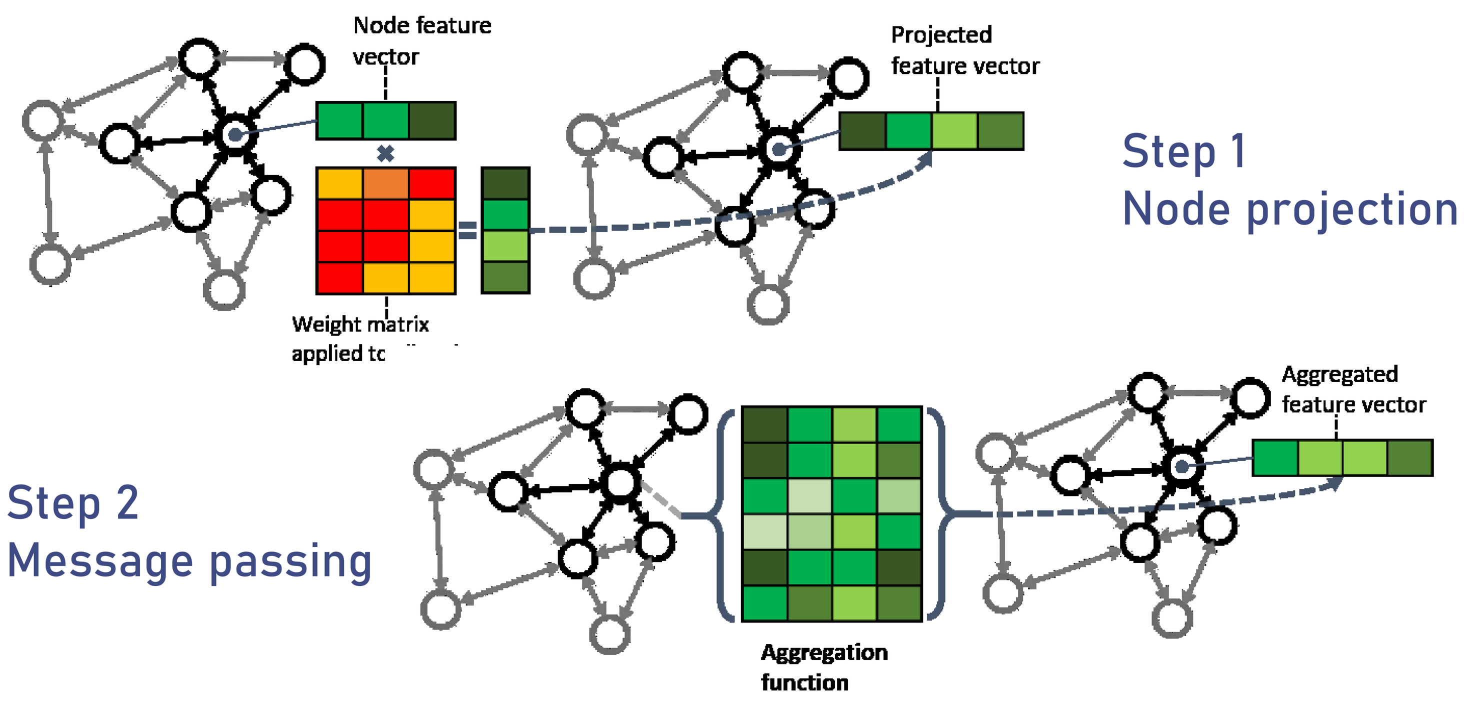
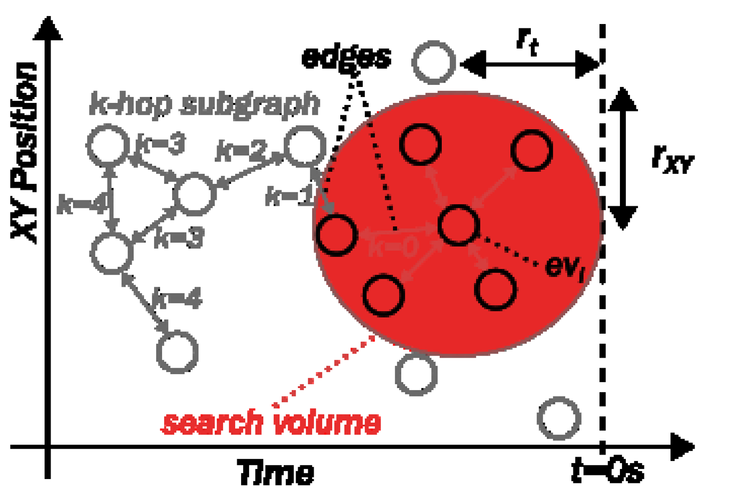


Fig 2. Gconv uses flat kernel to perform node projection and average function for message passing. More complex B-Spline based layer modulates message passing using edge coordinates.

Asynchronous Event Graph Neural Network

Graph building



Algorithm 1 Sparse fully-spherical update

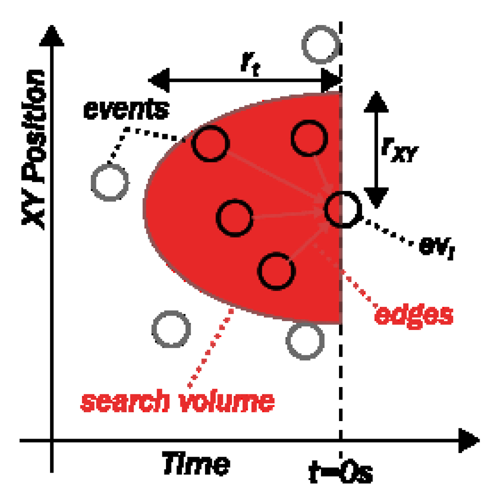
Input : $ev_i = \{x_i, y_i, t_i, p_i\}, G = \{EV, \mathcal{E}\}, r_t, r_{xy}, L$
Output : V_i
$M = EV \cap \mathcal{B}(ev_i, r_{xy}, r_t)$
for $ev_m \in M$ **do**
  $\mathcal{E}_m = knn(ev_m, EV, r_t, r_{xy})$
end for
$G = \pi(G, \mathcal{E}_m, ev_i)$
for l *in range(L)* **do**
  for $ev_j \in hop(M, G, l)$ **do**
    $z_{j,l} = \phi(ev_j, G, l)$
  end for
end for
if $t > t_i + (r_t \times L)$ **then**
  $V_i = \sigma(Z_i)$
end if

For each new event:
i. Update all edges in search volume
ii. Reapply GNN layer k for hop k
iii. Wait r_t * L until final prediction

Fig 3. Asynchronous update of an Event Graph using Fully-spherical update

Applying Event Graph Neural Network in real time requires asynchronous update. Looking for neighbors in a ball around each event in (x,y,t) 3D space is not an efficient search scheme.

Hemi-spherical Update Graph (HUG)



Algorithm 2 Hemi-spherical update

Input : $ev_i = \{x_i, y_i, t_i, p_i\}, EV, r_t, r_{xy}, L$
Output : V_i
$K_i, \mathcal{E}_i = knn(ev_i, EV, r_t, r_{xy})$
$G_i = \{K_i, \mathcal{E}_i\}$
for l *in range(L)* **do**
  $z_{i,l} = \phi'(G_i, l)$
end for
$V_i = \sigma'(Z_i)$

For each new event:
i. Define directed edges forwards to new event
ii. Apply GNN to new event only
iii. Make an immediate prediction

Fig 4. Asynchronous update of an Event Graph using Hemi-spherical update

DVS Lip Classification

We extend the HUG approach tested on Optical Flow [1] to the classification of event stream using the DVS Lip dataset [2].

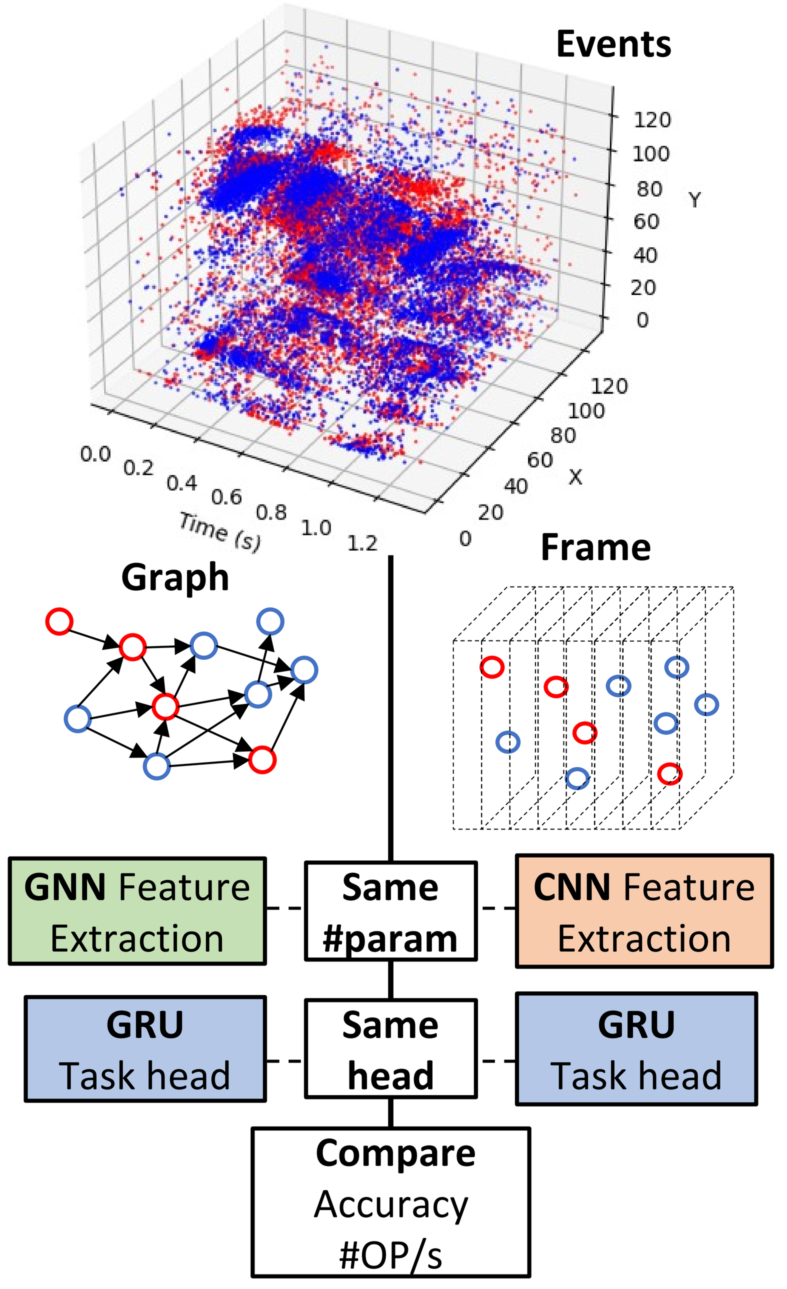


Fig 4. Comparing GNN and CNN feature extraction at iso-parameter for DVS Lip event stream classification

The event stream from each sample is converted to Event Frames and Event Graphs. We design CNN and GNN-based feature extractor with the same number of parameters for fair comparison. Extracted features are then processed with different task heads and performance compared in term of accuracy and real time computational complexity.

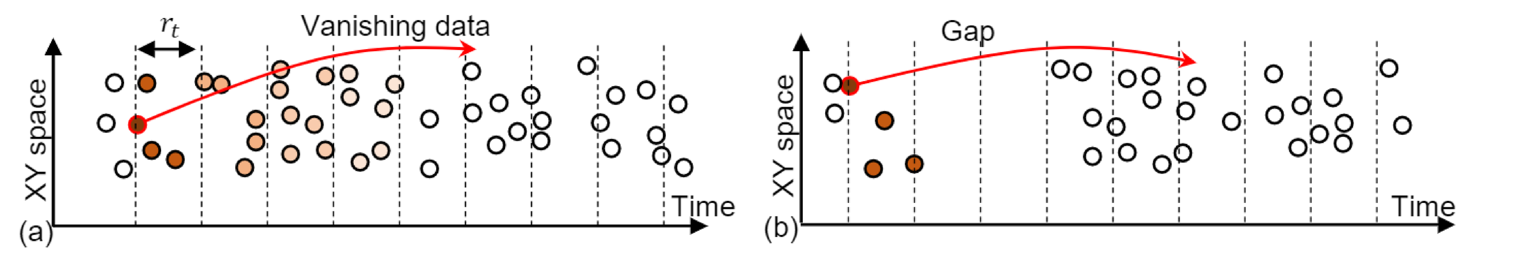


Fig 5. Graph Feature vanishing through (a) inherent decay in time and (b) gaps. Node color represents feature strength in computation

Similar to CNN, EGNN have a finite field of view and thus cannot integrate long term dependancies of variable length. Combining GRU tesk head in **G2N2** enables processing such patterns.

Results

Although classification of event streams relies on the analysis of long term dependancies, enabling future to past data sharing in the graph layer does not improve the accuracy and induces compulsory latency. The HUG approach proves more efficient for classification.

Neighbor Search	Search radii	Search Volume	Latency	Acc
Fully-spherical	(10 pix, 50 ms)	V	250 ms	68.3 %
Fully-spherical	(10 pix, 100 ms)	2V	500 ms	68.0 %
HUG	(10 pix, 100 ms)	V	0 ms	**69.4**%
HUG	(10 pix, 200 ms)	2V	0 ms	68.1 %
HUG	(10 pix, 400 ms)	4V	0 ms	67.3 %

Table 1. Comparing asynchronous graph search methods and geometry

Using the same number of parameters, GNN prove more efficient feature extractor the CNN both in accuracy and operations to compute in real time. Compared to State of the Art approaches, G2N2 is up to x146 lighter while maintaining comparable performances.

	Mode	Rate (Hz)	Acc	Parameters	GOPs/s
CNN-Concat (This work)	Event Frame	75	63.2 %	**418 k**	6.57
CNN-DWPW (This work)	Event Frame	75	66.0 %	**405 k**	8.81
G2N2 (This Work)	Event Graph	75 (GRU)	**69.4** %	**413 k**	**1.68**
RN-Net [31]	Event Frame	33	67.5 %	7.5 M	
[20]	Video Frame	25	65.5 %	11.2 M	
ACTION-Net[30]	Event Frame	25	68.8 %	28.1 M	
MSTP[28]	Event Frame	(25,175)	**72.1** %	60.3 M	

Table 2. Comparison of GNN and CNN feature extractor using the same number of parameters. Comparison with State of the Art approaches, scores taken from [2-3]

[1] T. Dalgaty et al., "HUGNet: Hemi-Spherical Update Graph Neural Network Applied to Low-Latency Event-Based Optical Flow", CVPRW 2023
[2] G. Tan et al., "Multi-grained Spatio-Temporal Features Perceived Network for Event-based Lip-Reading", CVPR 2022
[3] S. Yoo et al., "RN-Net: Reservoir Nodes-Enabled Neuromorphic Vision Sensing Network", CVPRW 2023