UNIVERSITÄT ZU LÜBECK

BASLER

Artificial Intelligence for Processing of Event-Based Vision Sensor Data

Steffen Hagedorn | 29.09.2022

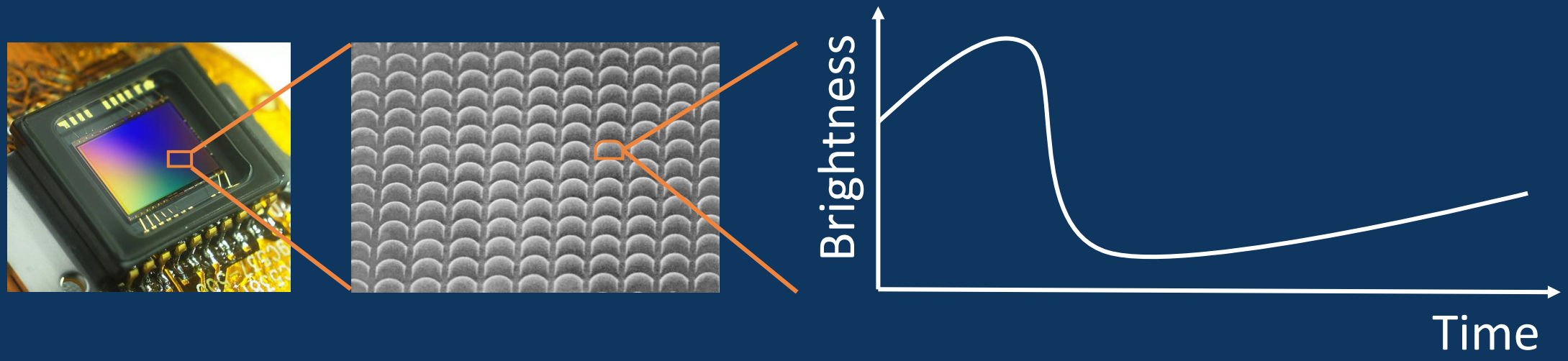
Steffen Hagedorn

B.Sc. **Medizinische Ingenieurwissenschaften**
Lübeck

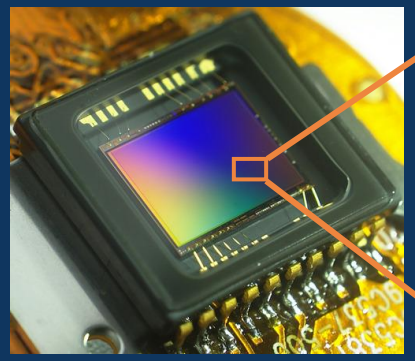
M.Sc. **Robotics and Autonomous Systems**
Lübeck, Oslo

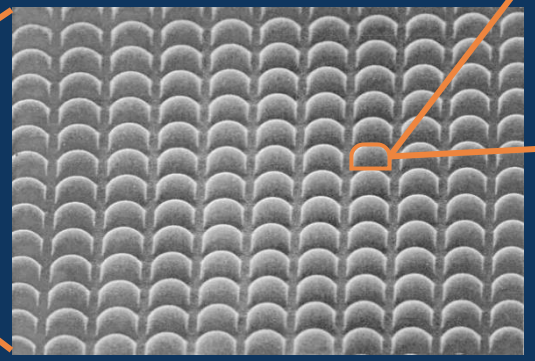
Ph.D. **Perception for Autonomous Driving**
Stuttgart

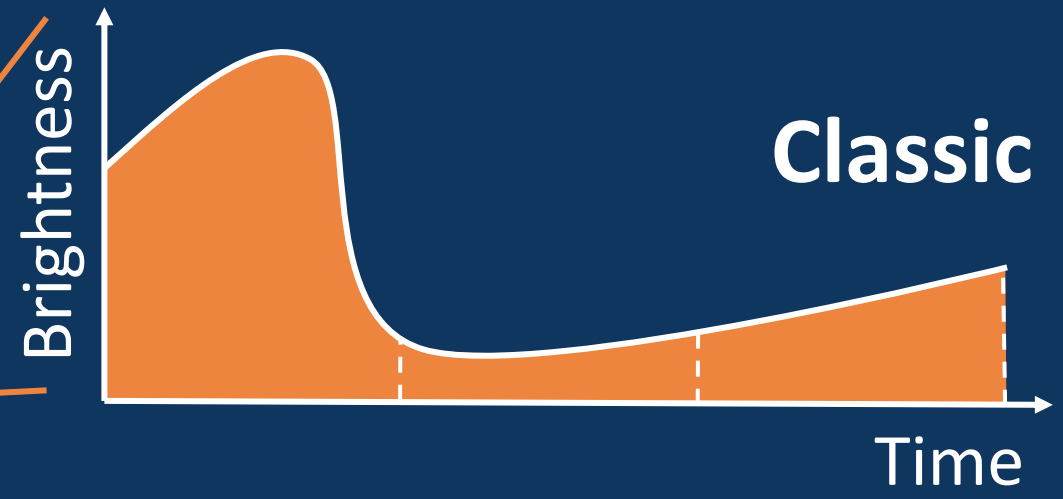
What is Event-Based Vision?



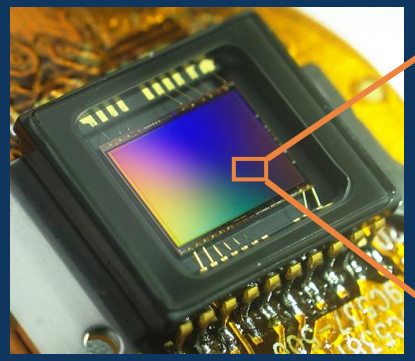
What is Event-Based Vision?

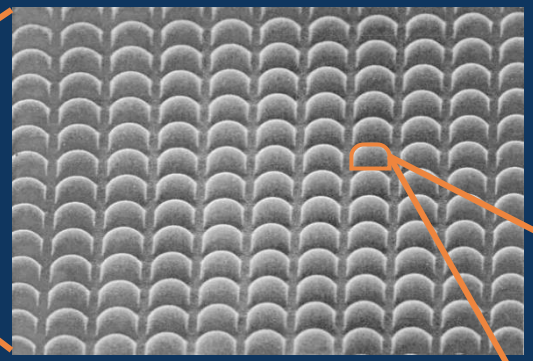


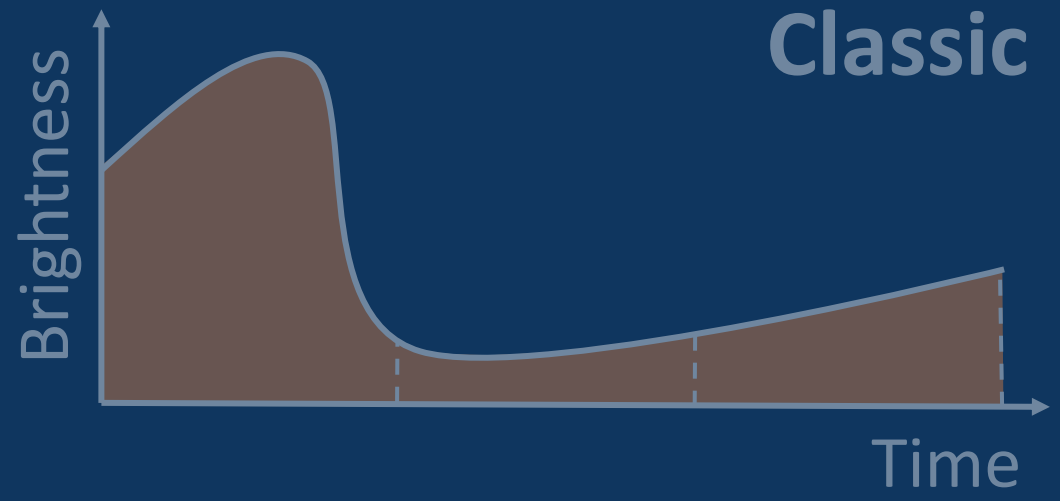


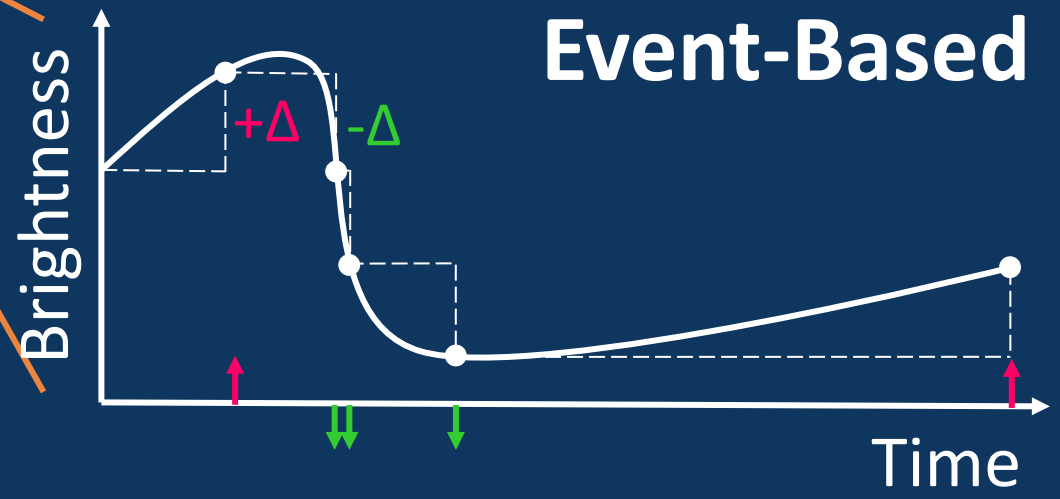


What is Event-Based Vision?







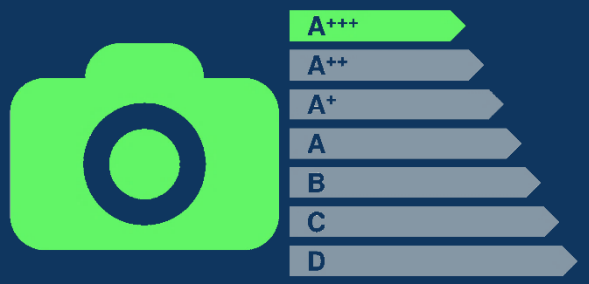


[1]

Advantages

Low

Data Output



Energy Use

High

Dynamic Range

Time Resolution

Problem

State of the Art

Concept

Results

Insights

Problem

State of the Art

Concept

Results

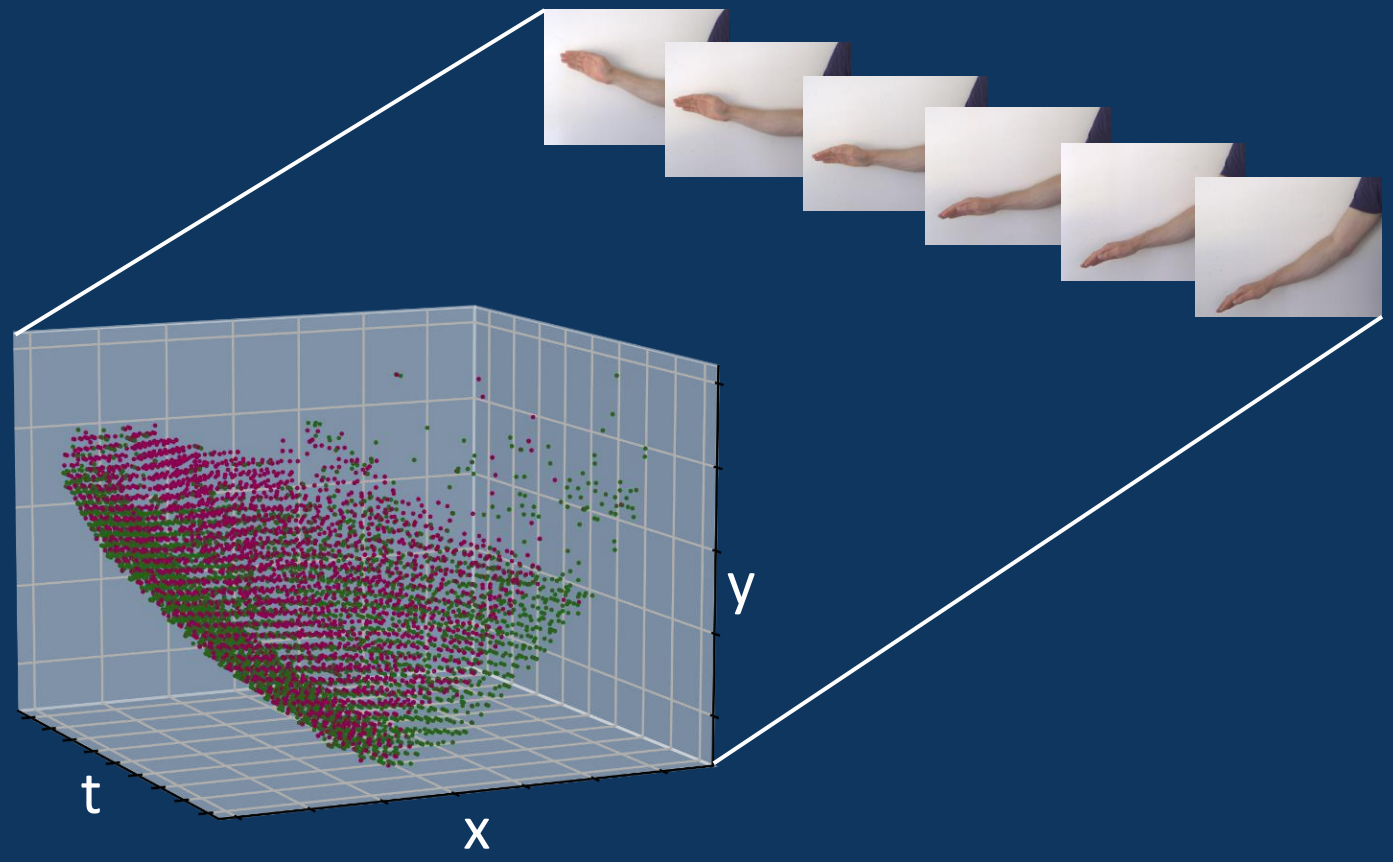
Insights

Event Data ≠ Image Data

- Classic image processing not applicable
- ~~Huge effort to develop new methods~~

→ Use AI!

- **Existing solutions don‘t use properties of event data**



Research Questions

- How does **current AI** process event data?
- **Requirements** for processing that suits event data?
- **Develop model** according to requirements

Problem

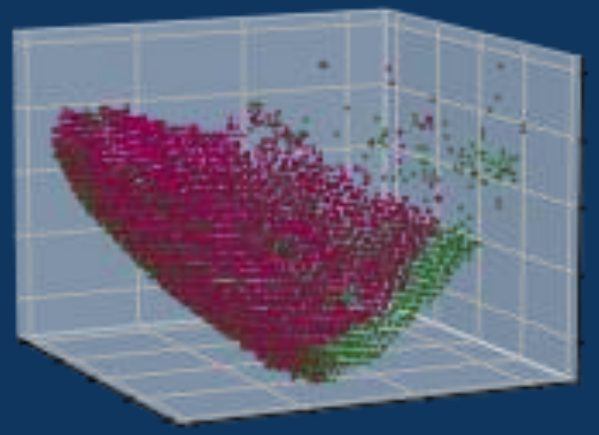
State of the Art

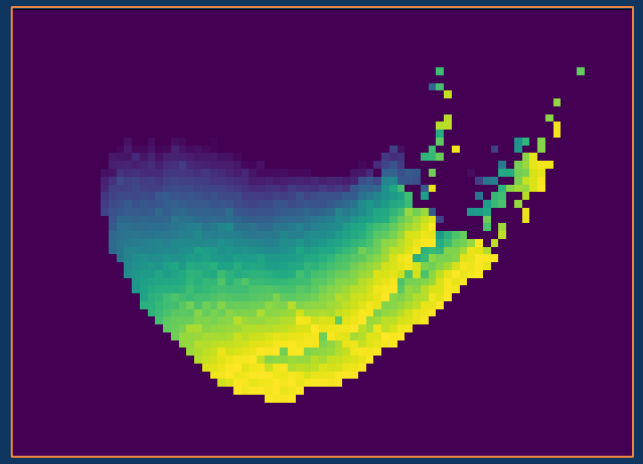
Concept

Results

Insights

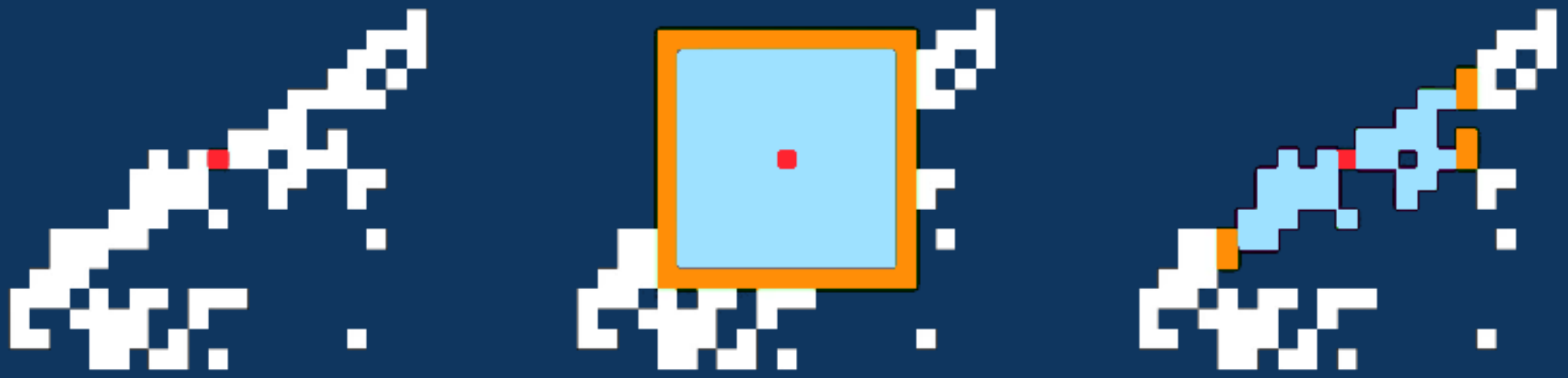
CNN on 2D Representation [3]





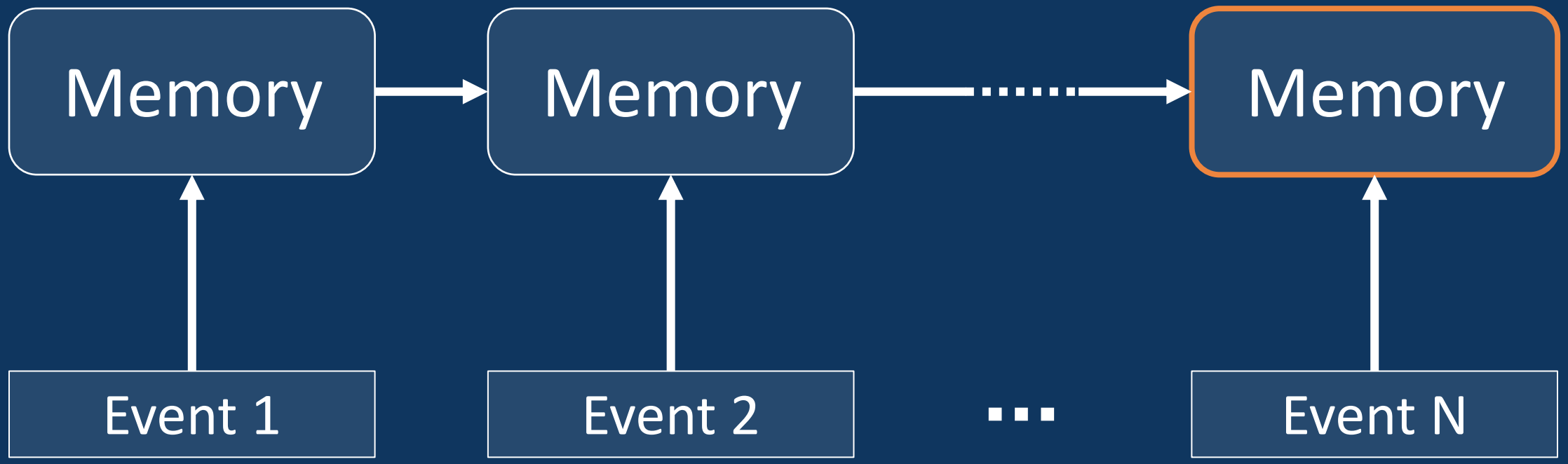
CNN

Sparse CNN

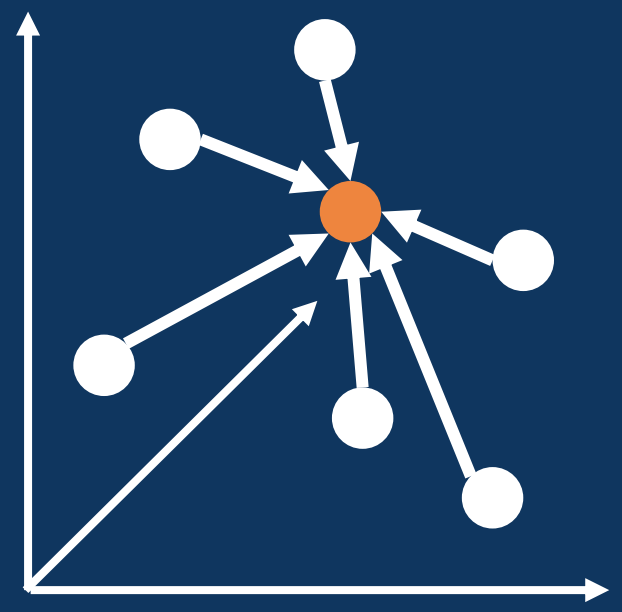


New Event | Conv | Sparse Conv [4]

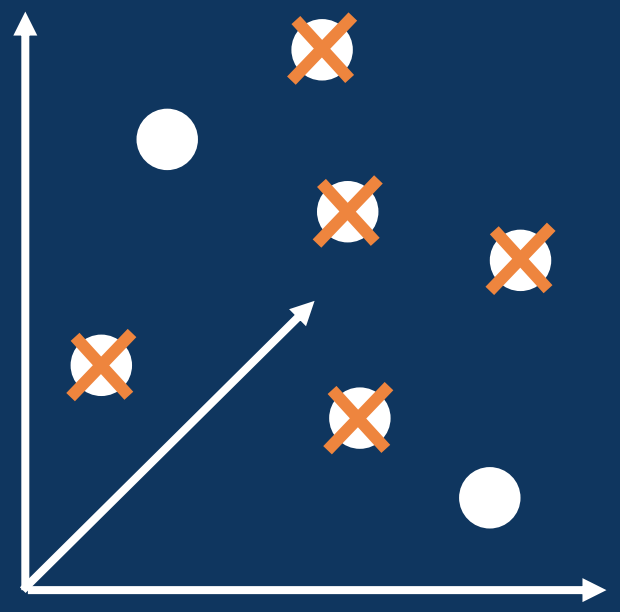
Sequence Processing



3D Processing [6]



Aggregate Information



Reduce

Research Questions

- How does **current AI** process event data?
  - CNN on 2D Representation
  - Sparse CNN
  - Sequence Processing
  - 3D Processing
- **Requirements** for processing that suits event data?
- **Develop model** according to requirements

Problem

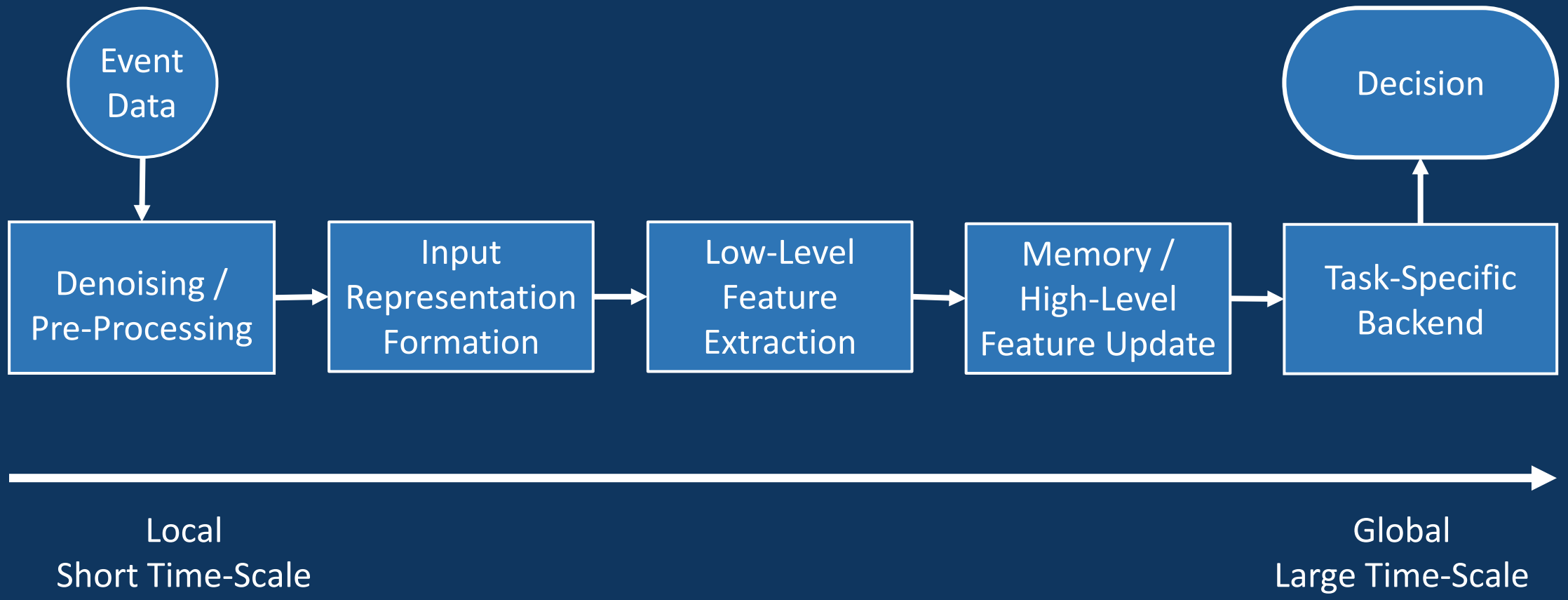
State of the Art

Concept

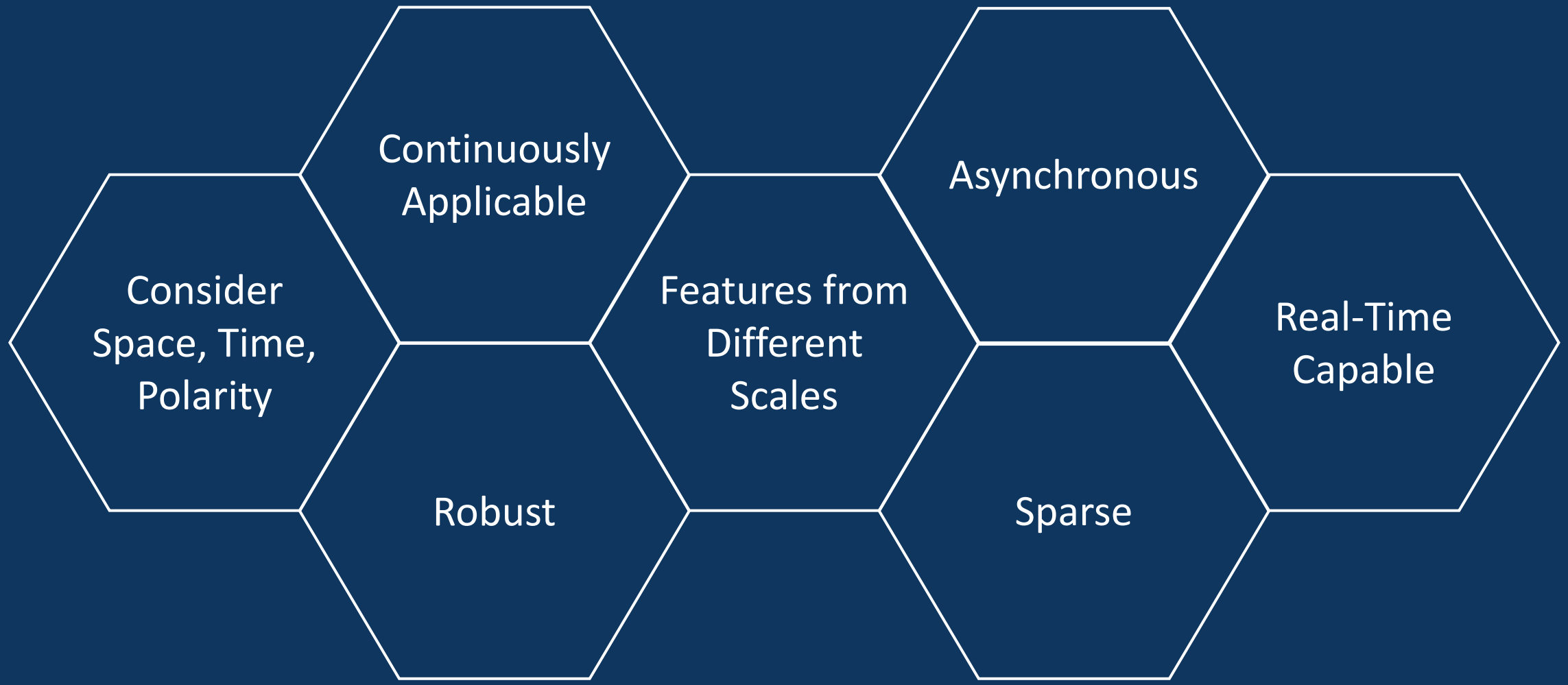
Results

Insights

Processing Model

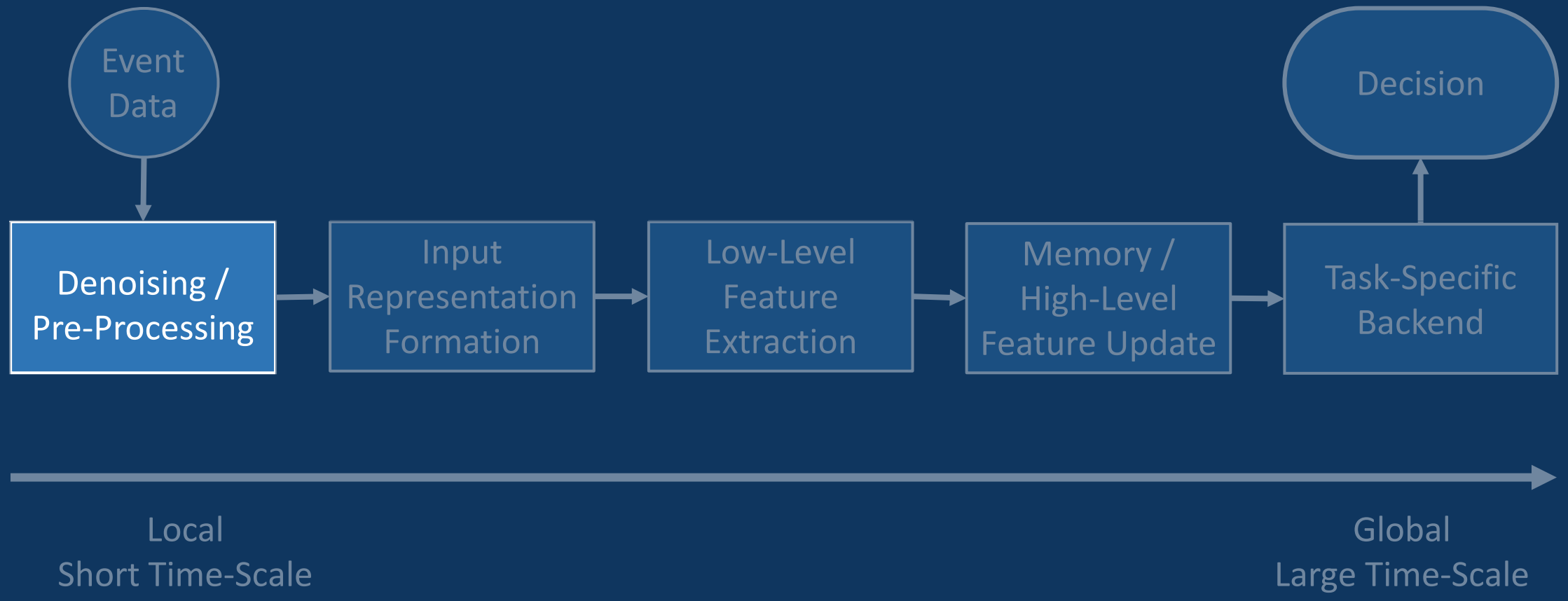


Requirements

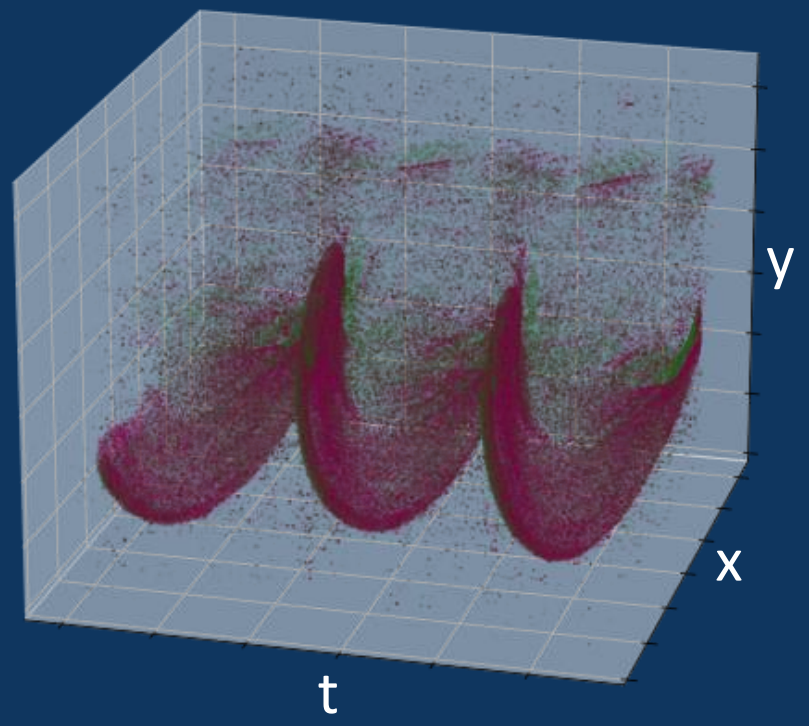


Research Questions

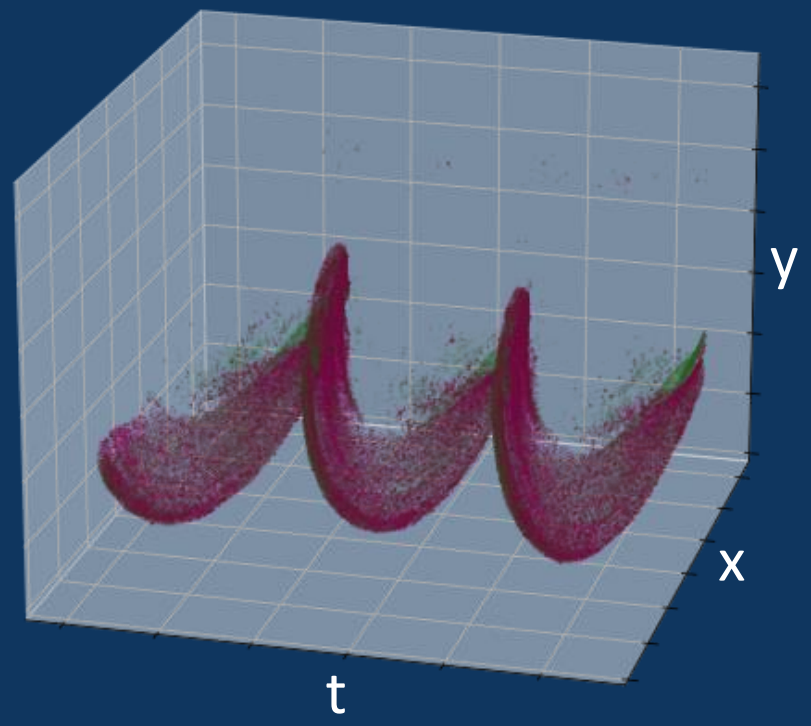
- How does **current AI** process event data?
- **Requirements** for processing that suits event data?
- **Develop model** according to requirements

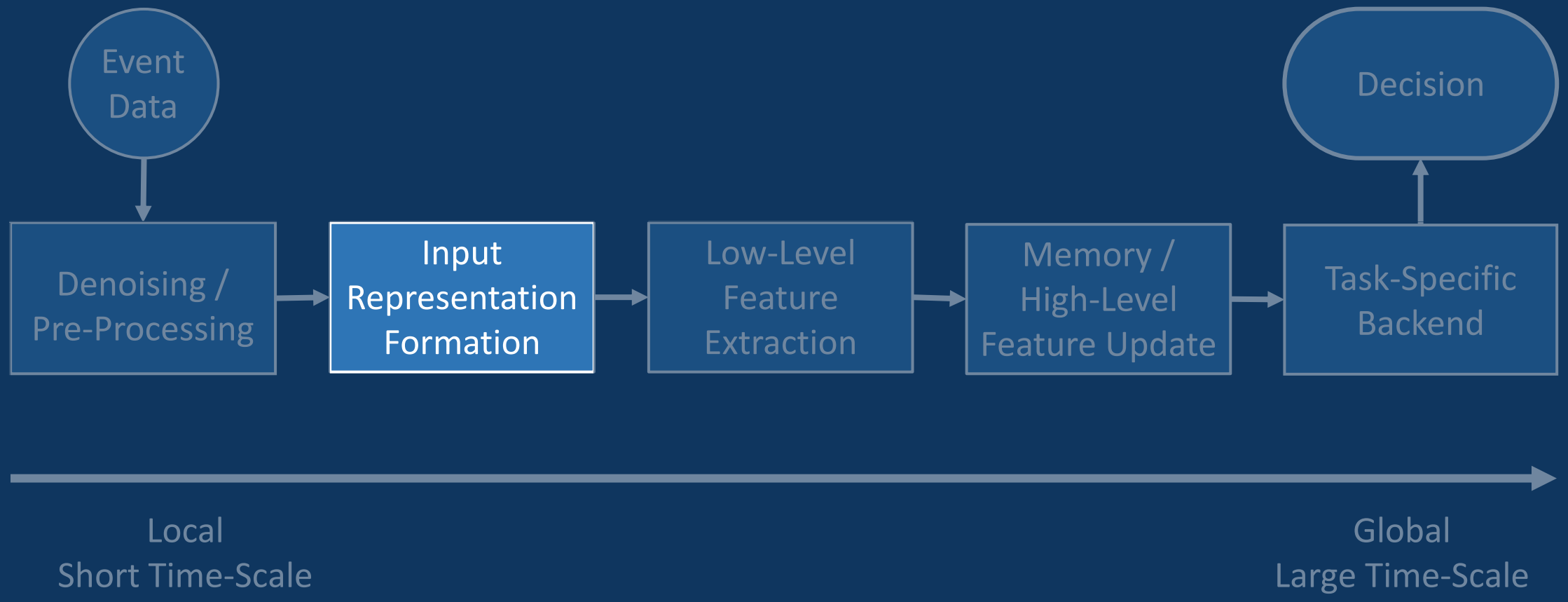

Event Data
Decision
Denoising / Pre-Processing
Input Representation Formation
Low-Level Feature Extraction
Memory / High-Level Feature Update
Task-Specific Backend
Local
Short Time-Scale
Global
Large Time-Scale

Denoising

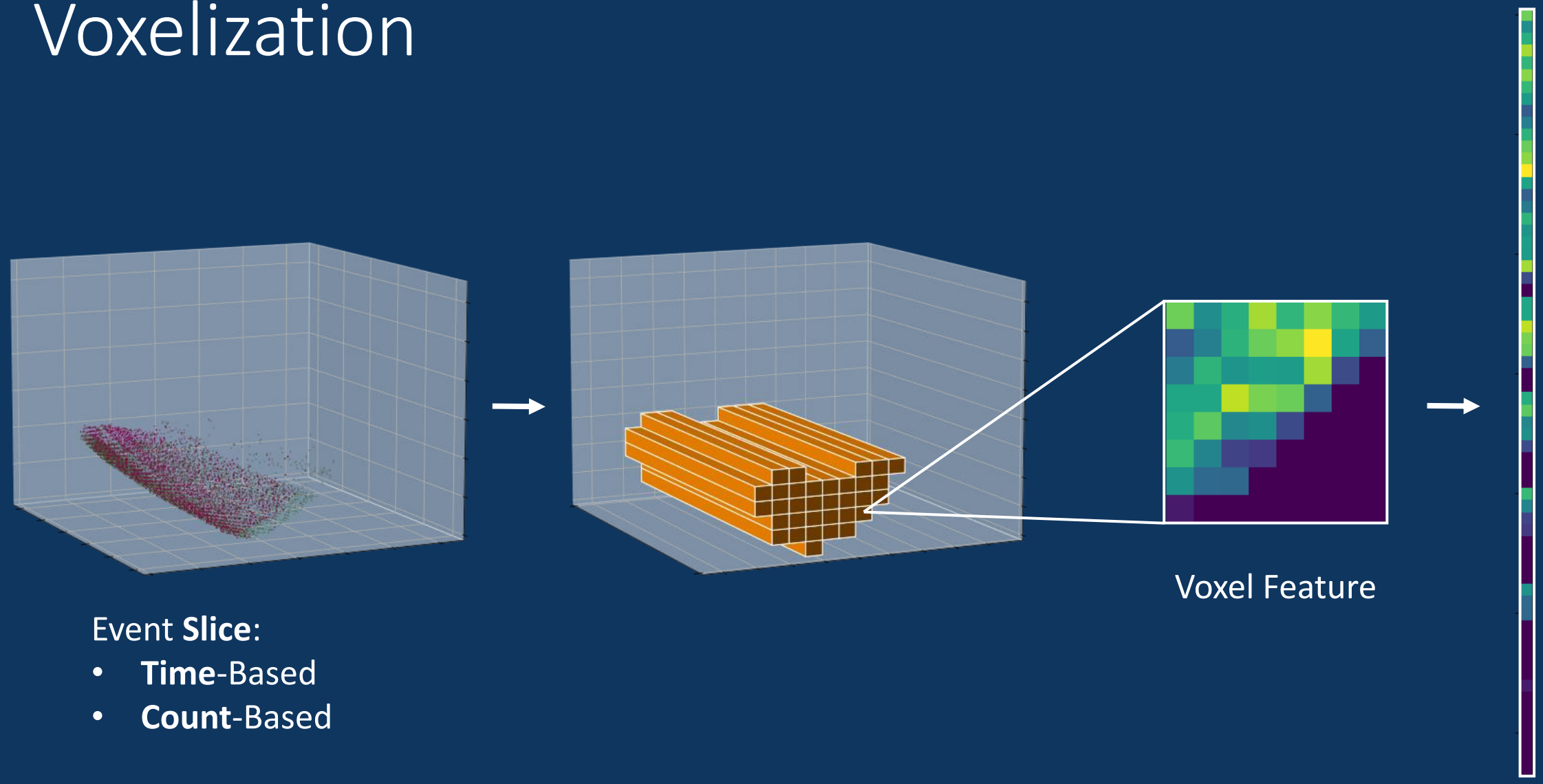


Spatio-Temporal Correlation Filter (STCF) [7]




Event Data
Decision
Denoising / Pre-Processing
Input Representation Formation
Low-Level Feature Extraction
Memory / High-Level Feature Update
Task-Specific Backend
Local
Short Time-Scale
Global
Large Time-Scale

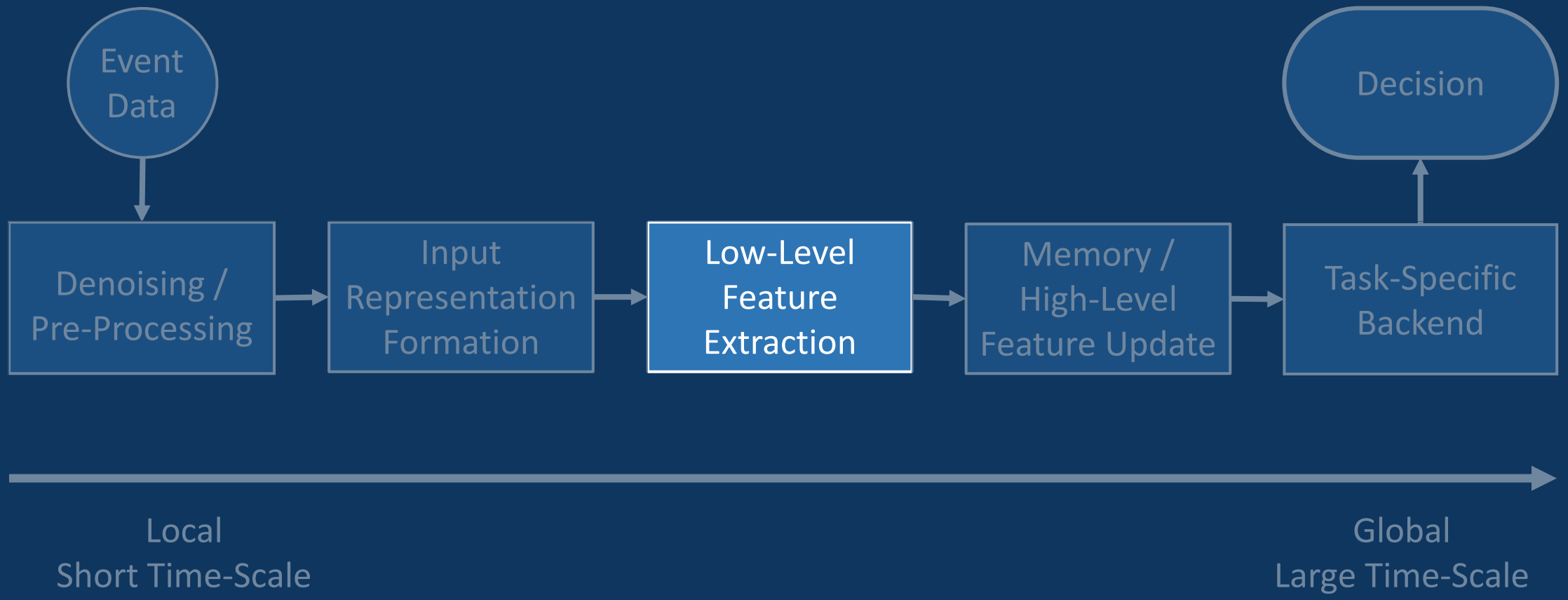
Voxelization



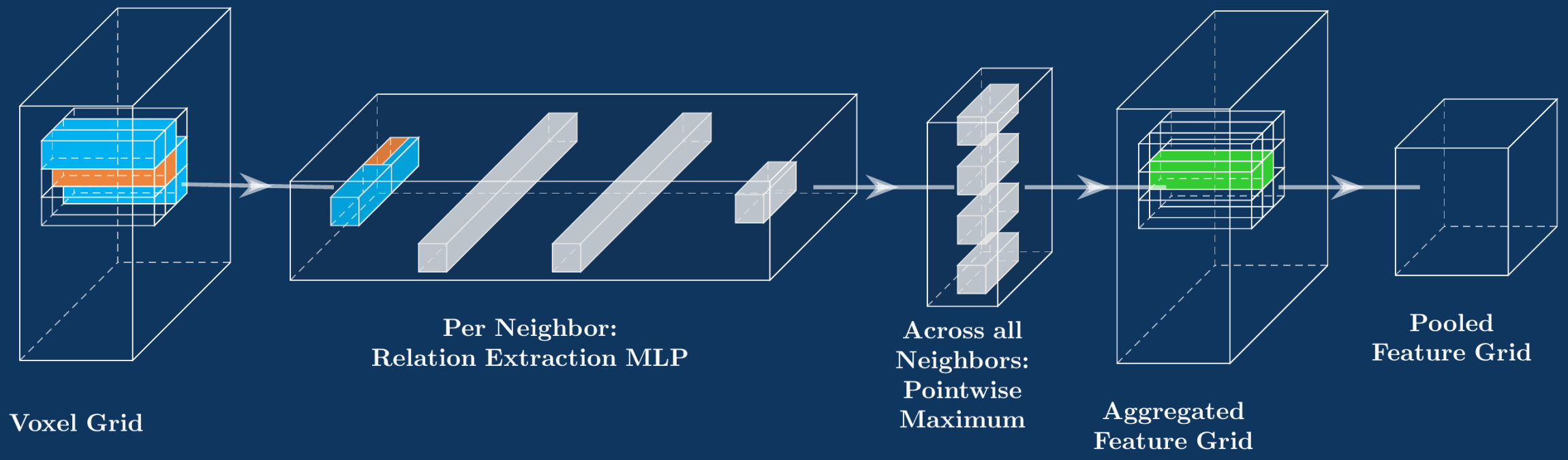
Event **Slice**:

- **Time**-Based
- **Count**-Based

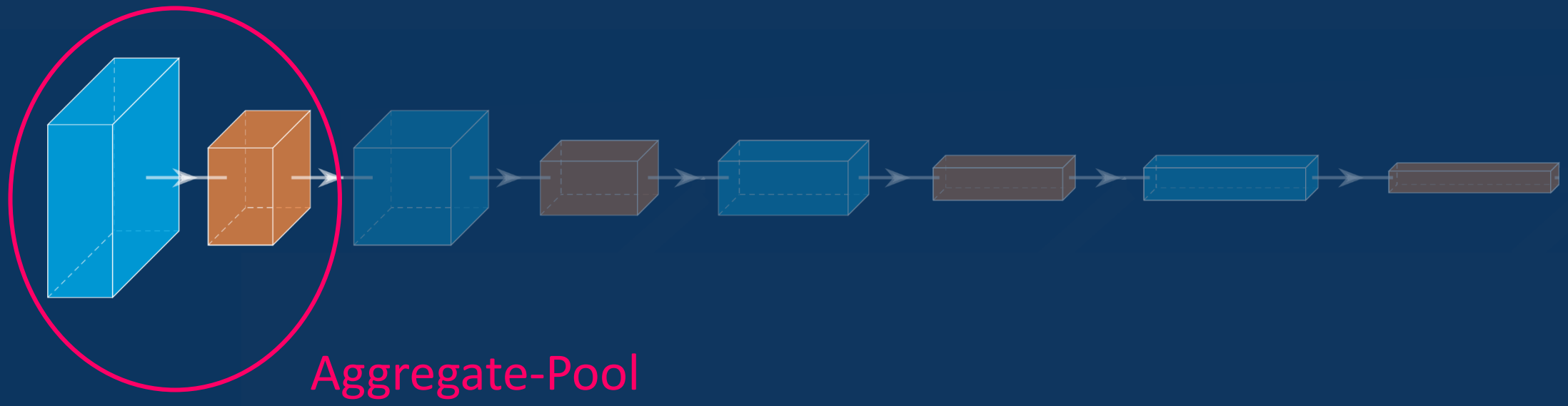
[8]

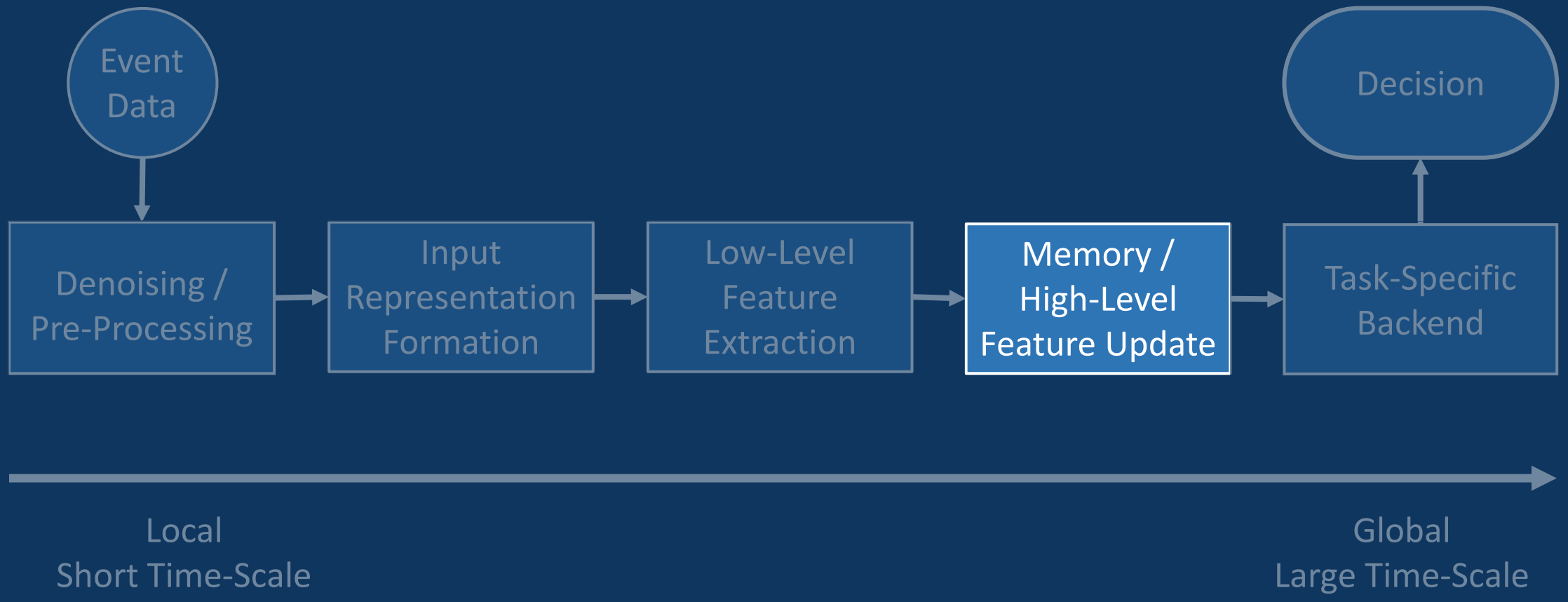

Event Data
Decision
Denoising / Pre-Processing
Input Representation Formation
Low-Level Feature Extraction
Memory / High-Level Feature Update
Task-Specific Backend
Local
Short Time-Scale
Global
Large Time-Scale

Aggregate-Pool

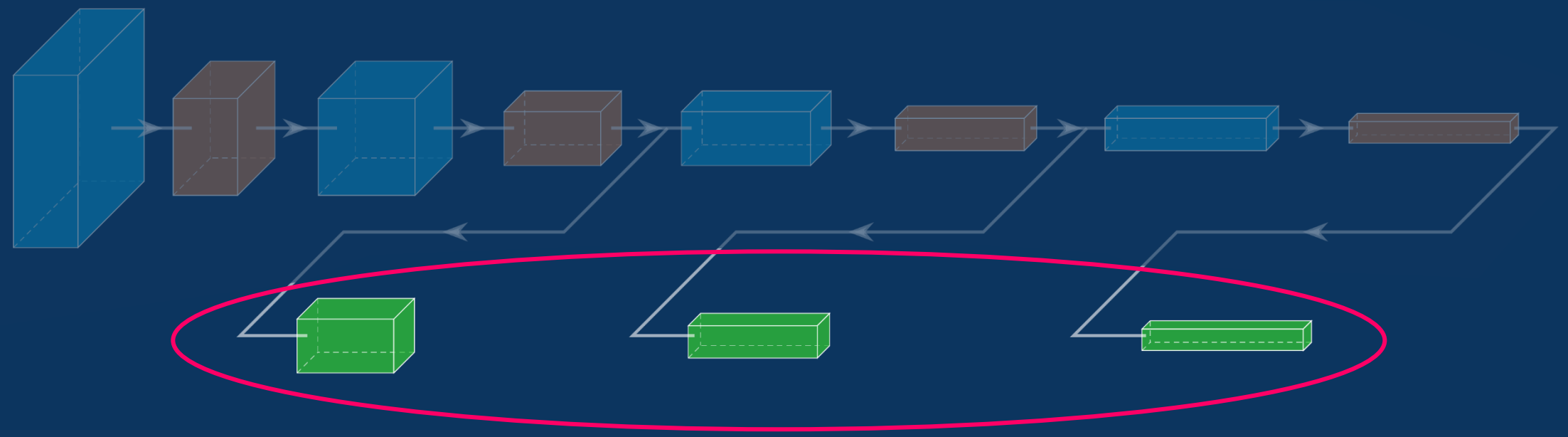


Aggregate-Pool

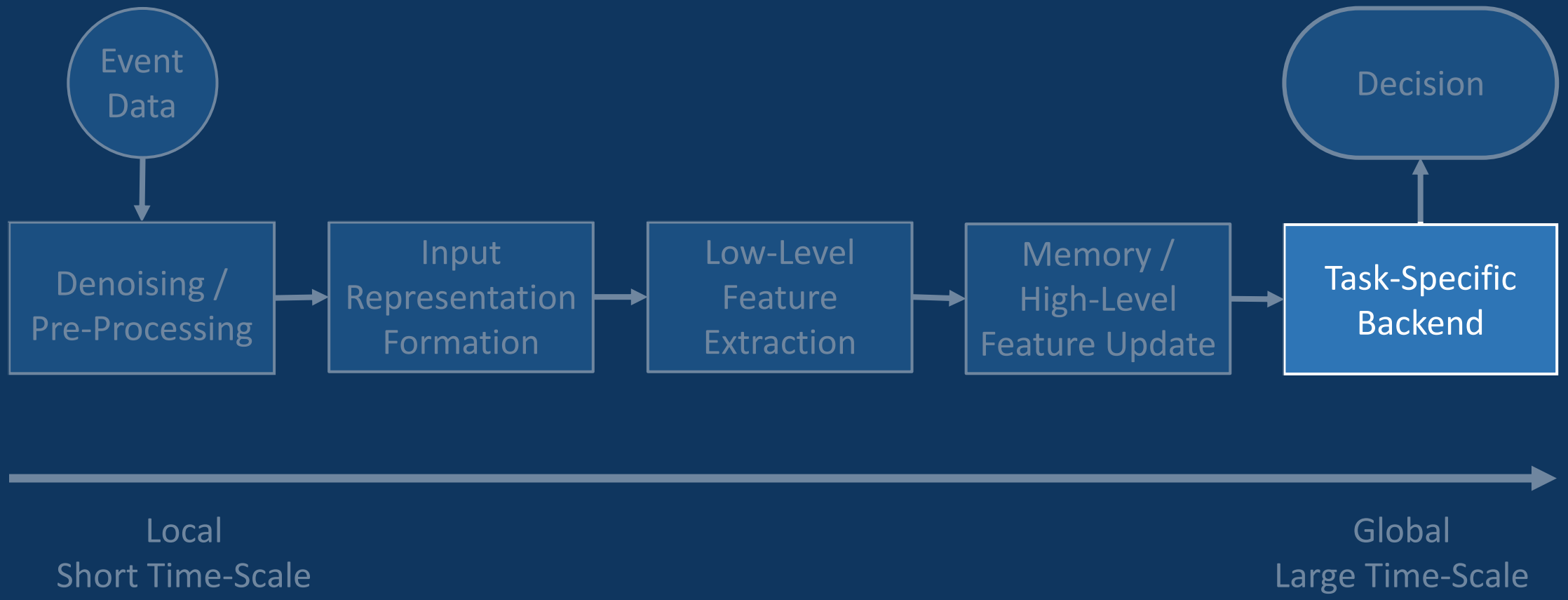



Event Data
Decision
Denoising / Pre-Processing
Input Representation Formation
Low-Level Feature Extraction
Memory / High-Level Feature Update
Task-Specific Backend
Local
Short Time-Scale
Global
Large Time-Scale

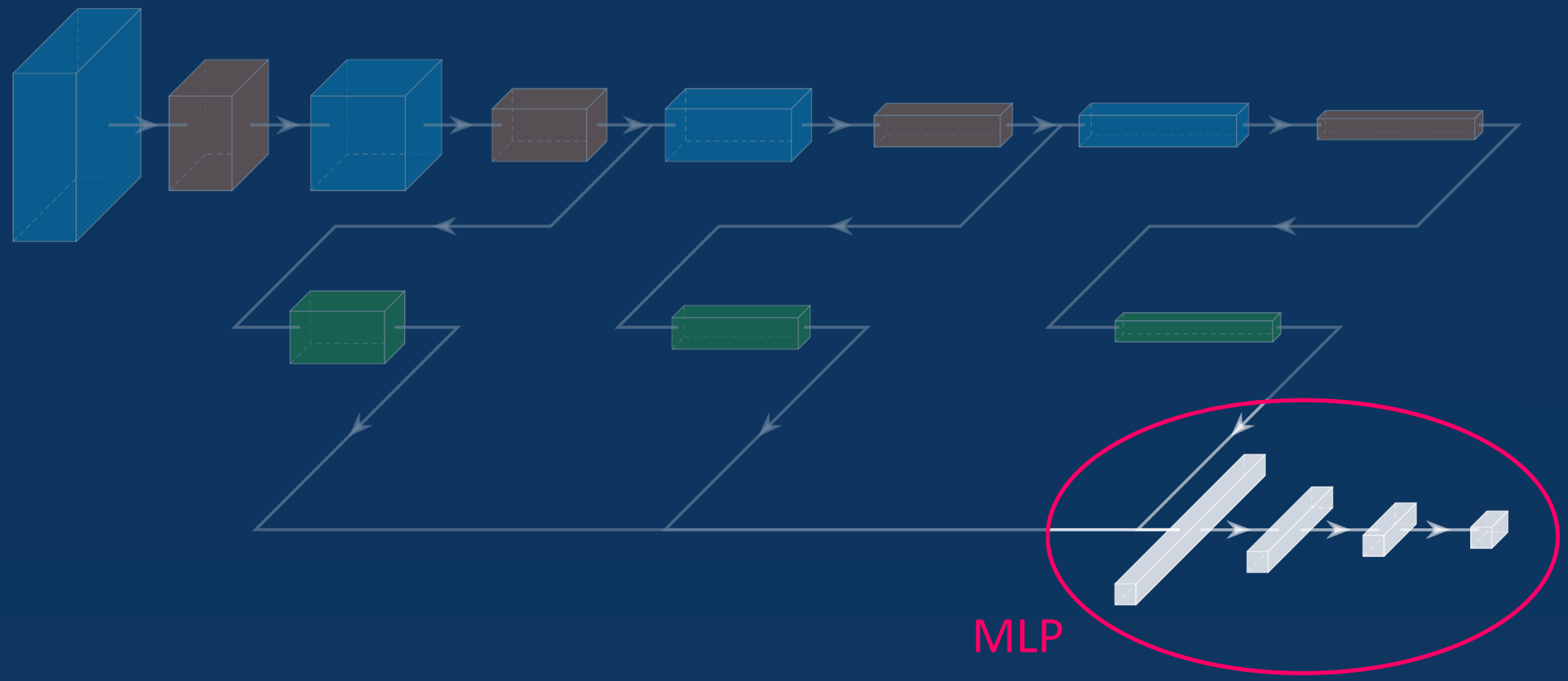
High-Level Memory



LSTM [9]

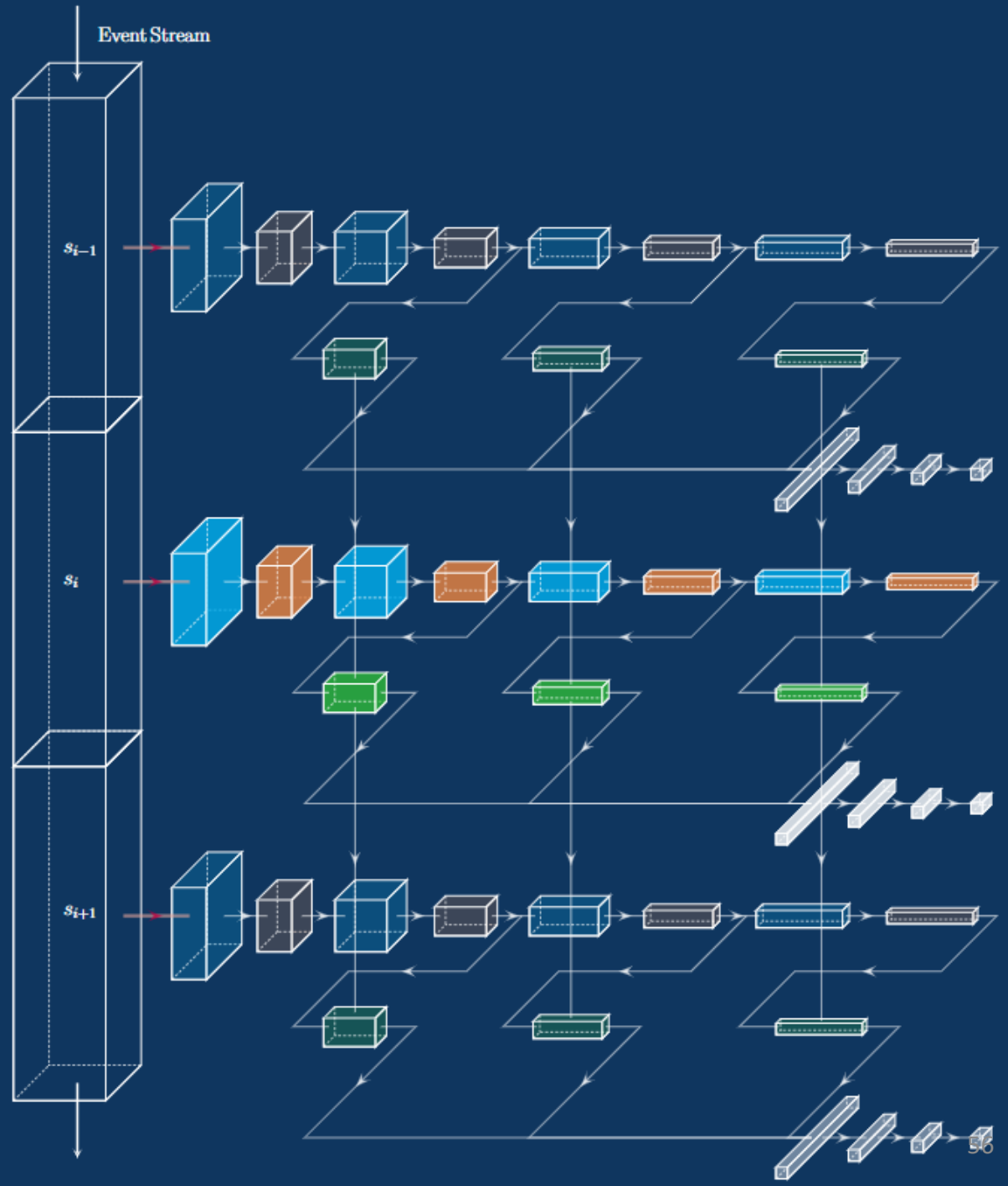

Event Data
Decision
Denoising / Pre-Processing
Input Representation Formation
Low-Level Feature Extraction
Memory / High-Level Feature Update
Task-Specific Backend
Local
Short Time-Scale
Global
Large Time-Scale

Backend

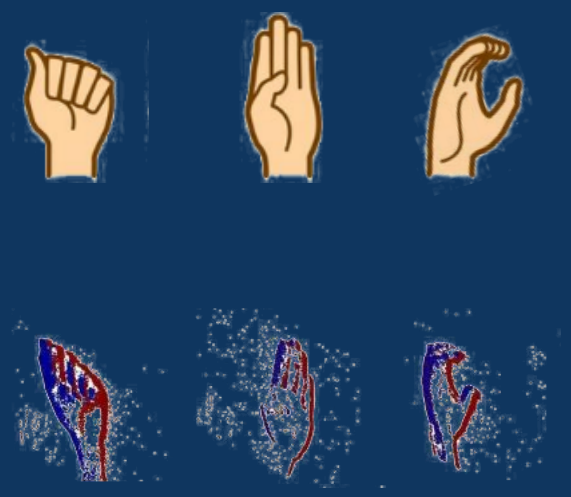


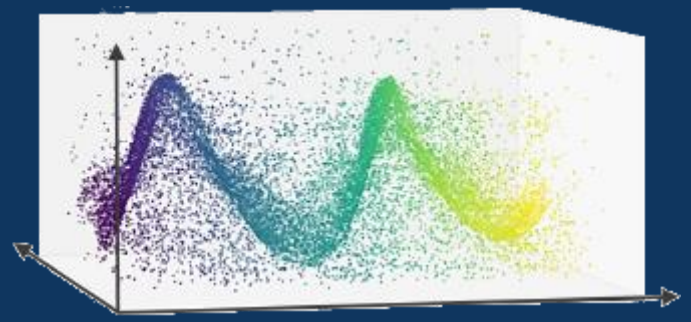
HAREM Net

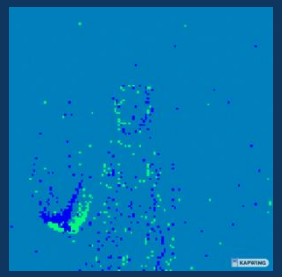
Hierarchically **A**ggregated **R**ecursive **E**vent **M**emory **Net**work

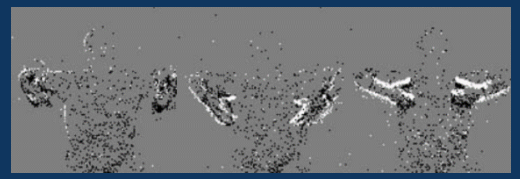


Gesture Classification









ASL DVS
Alphabet

DVS 128 Gesture
Smart Device Control

SL Animals DVS
Animal Gestures

[12]

Easy → Hard

Problem

State of the Art

Concept

Results

Insights

Accuracy

Model	Accuracy [%]	
	ASL DVS	
SLAYER		
STBP		
GCNN	90.1	
EventNet	83.33	
PointNet++		
Timestamp CNN		
DGCNN		
Matrix LSTM	99.73	21.43 M parameters
EV-VGCNN	98.3	
HAREM Net	**98.1**	8.8 M parameters

Accuracy

Model	Accuracy [%]	
	ASL DVS	
SLAYER		
STBP		
GCNN	90.1	
EventNet	83.33	
PointNet++		
Timestamp CNN		
DGCNN		
Matrix LSTM	99.73	
EV-VGCNN	98.3	not continuously applicable
HAREM Net	**98.1**	continuously applicable

Accuracy

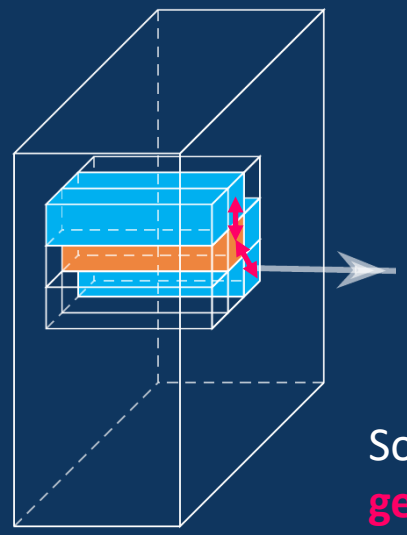
Model	Accuracy [%]	
	ASL DVS	DVS 128 Gesture
SLAYER		93.64
STBP		
GCNN	90.1	
EventNet	83.33	
PointNet++		97.08
Timestamp CNN		97.35
DGCNN		98.56
Matrix LSTM	99.73	
EV-VGCNN	98.3	
HAREM Net	**98.1**	**88.5**

Δ = -10 %

Accuracy

Model	Accuracy [%]	
	ASL DVS	DVS 128 Gesture
SLAYER		93.64
STBP		
GCNN	90.1	
EventNet	83.33	
PointNet++		97.08
Timestamp CNN		97.35
DGCNN		98.56
Matrix LSTM	99.73	
EV-VGCNN	98.3	
HAREM Net	**98.1**	**88.5**

Δ = -10 %



Solution: Consider **geometric relation** in aggregation

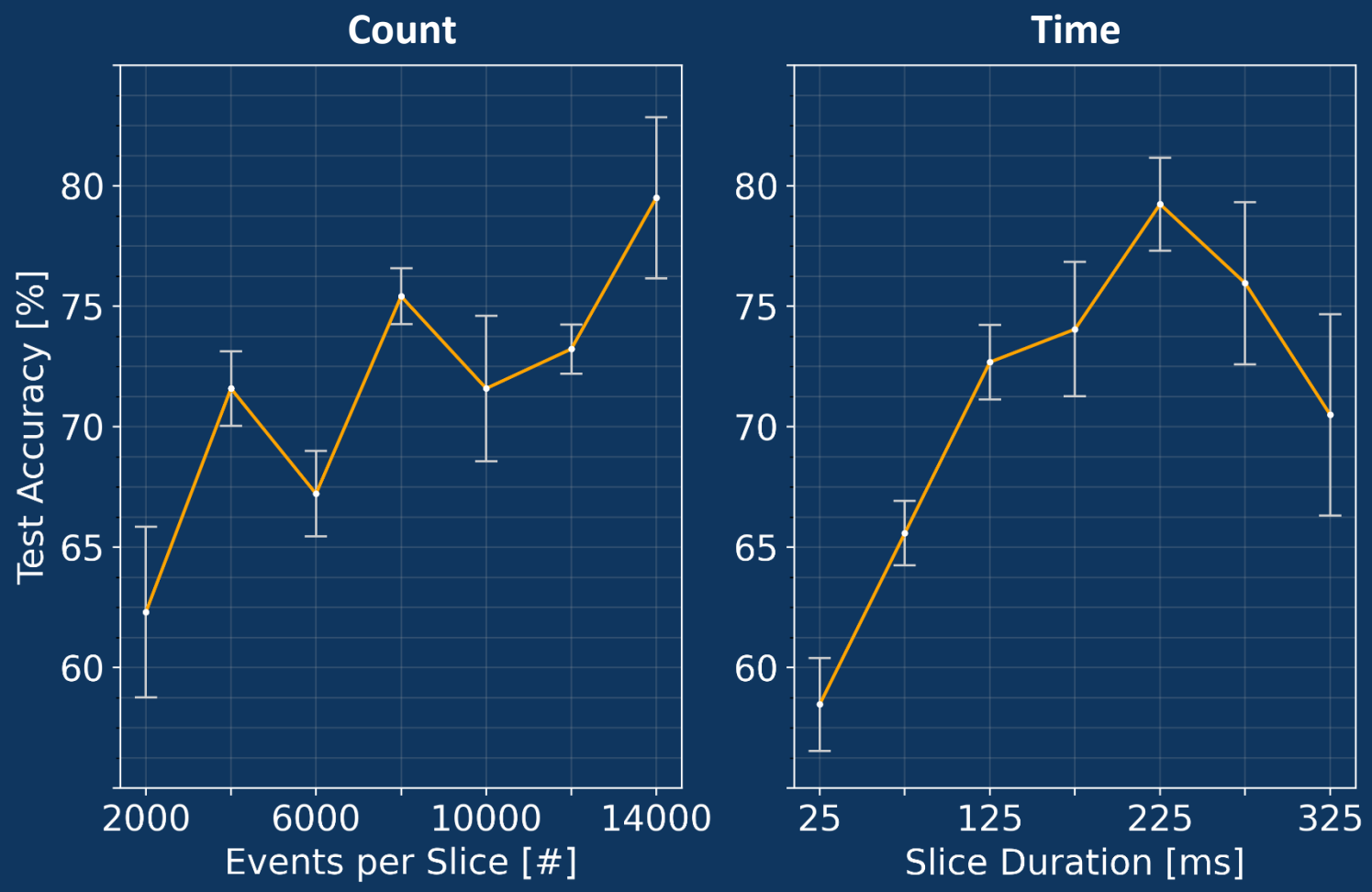
Accuracy

Model	Accuracy [%]		
	ASL DVS	DVS 128 Gesture	SL Animals DVS
SLAYER		93.64	60.9
STBP			56.2
GCNN	90.1		
EventNet	83.33		
PointNet++		97.08	
Timestamp CNN		97.35	
DGCNN		98.56	
Matrix LSTM	99.73		
EV-VGCNN	98.3		
HAREM Net	**98.1**	**88.5**	**73.2**

> State of the Art

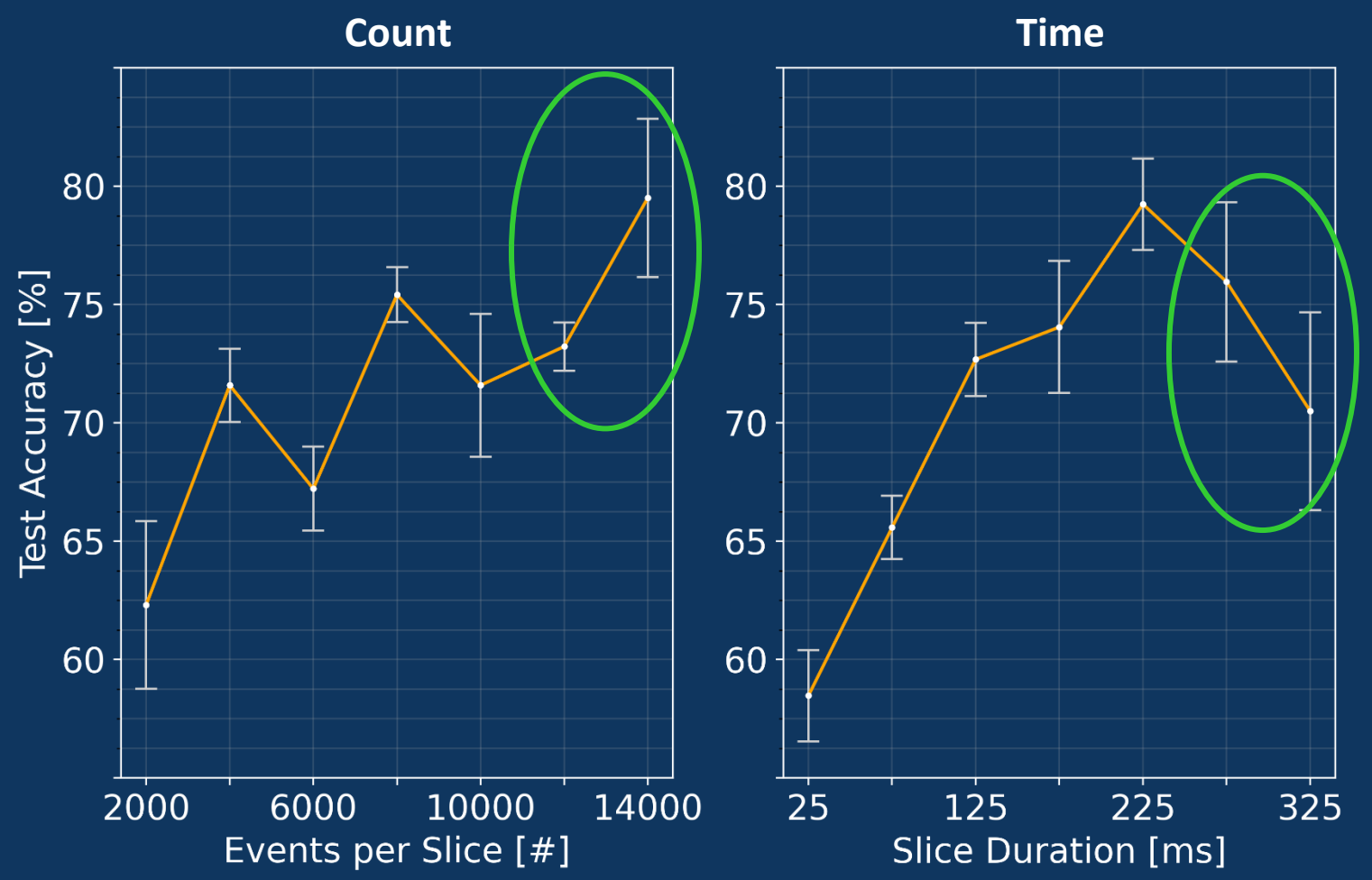
- Robust against noise
- Good on long sequences

Slicing



1000 Events ≈ 24 ms

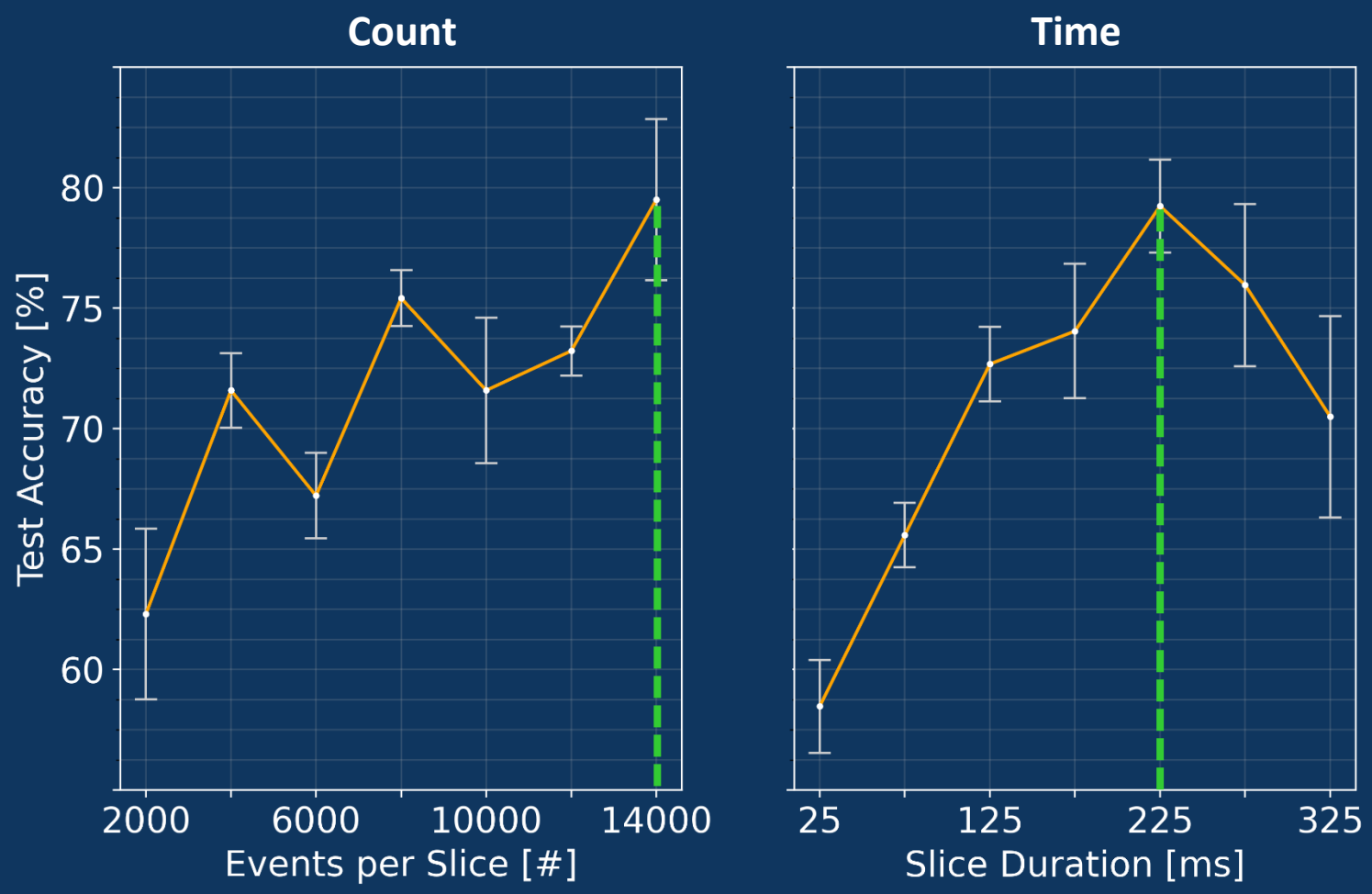
Slicing



1000 Events ≈ 24 ms

- Count > time for long slices

Slicing

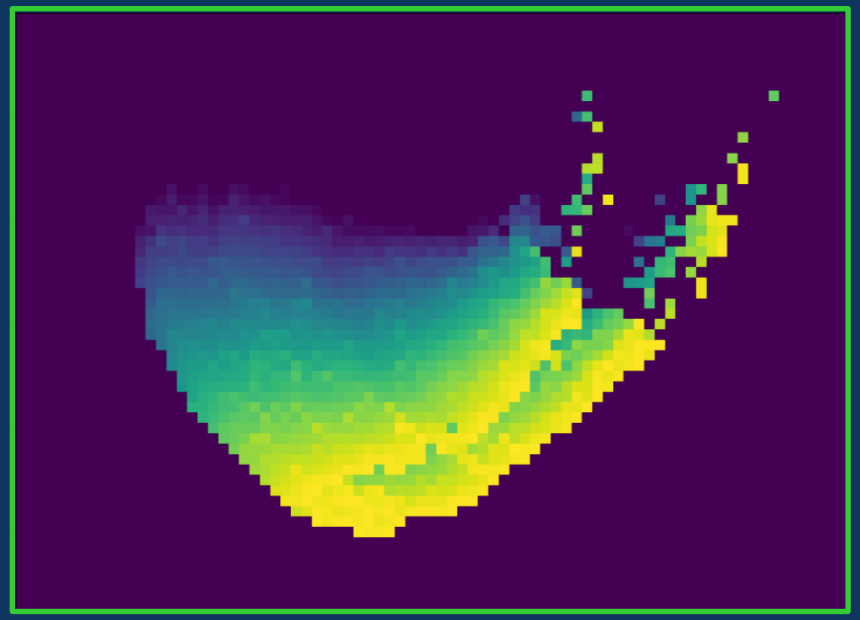


1000 Events ≈ 24 ms

- Count > time for long slices
- Time: Similar peak accuracy with shorter slices

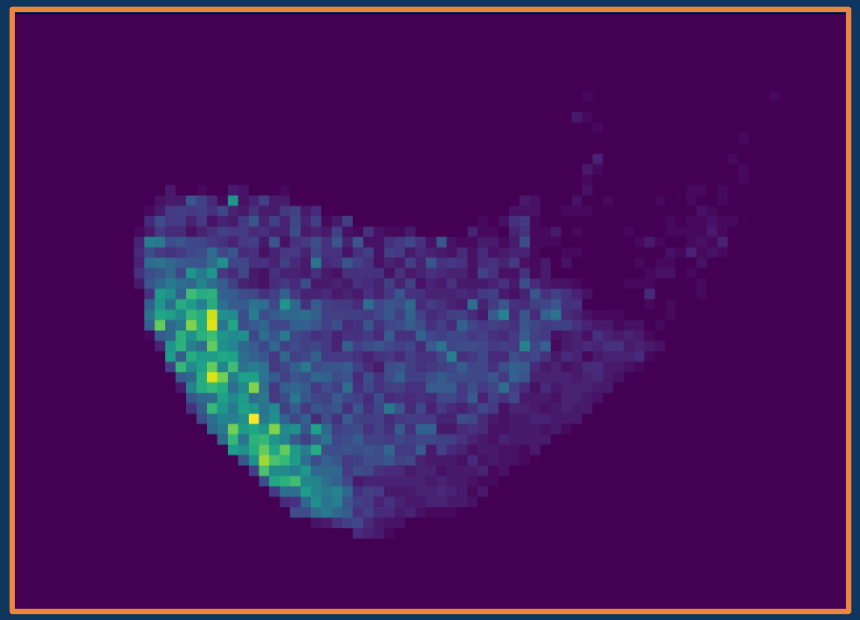
Voxel Feature

Timestamp Image



Per pixel: Time of last event

Event Image



Per pixel: Number of events in slice

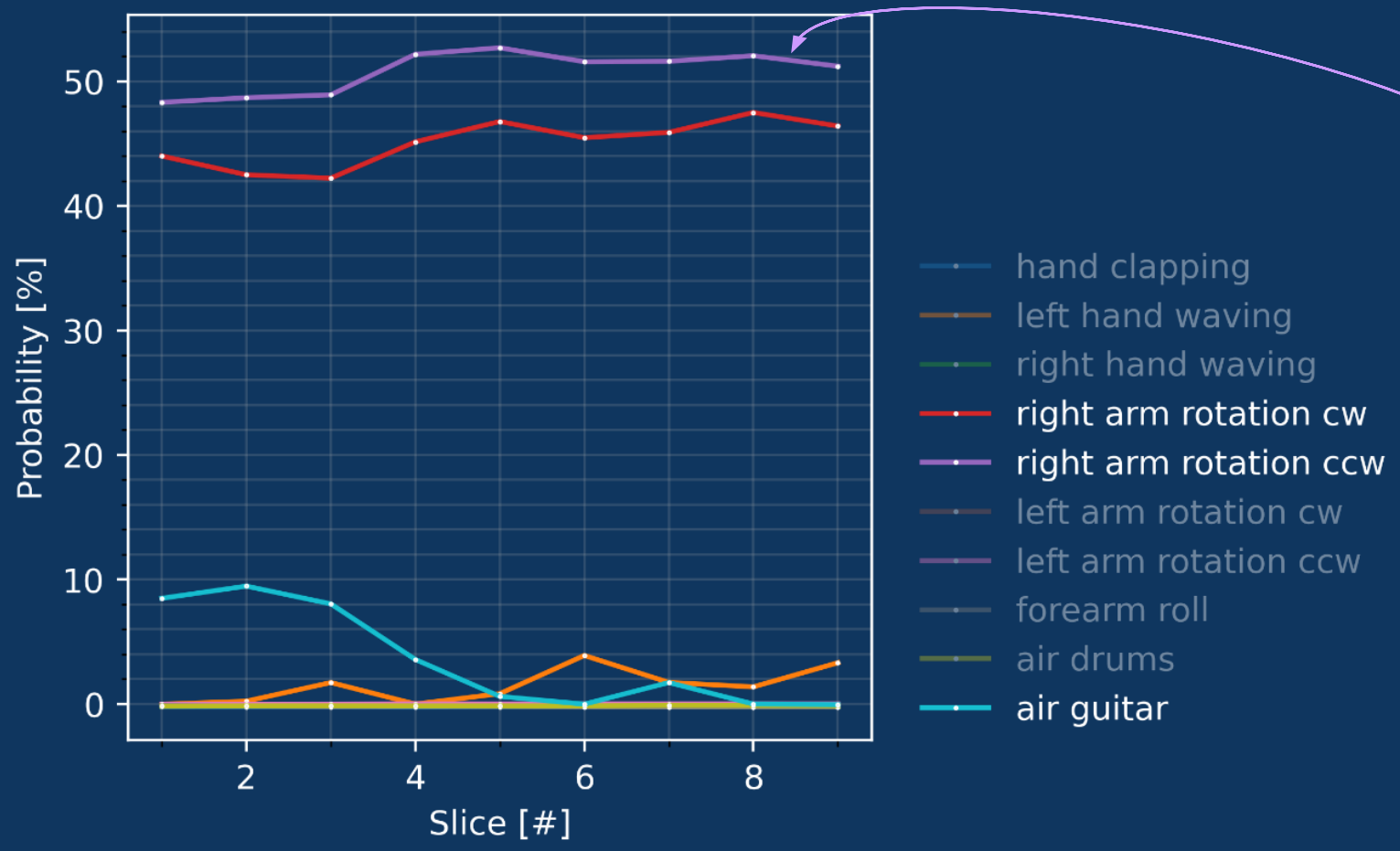
55 % Accuracy

Slice level:

Time information < Statistics

77 % Accuracy

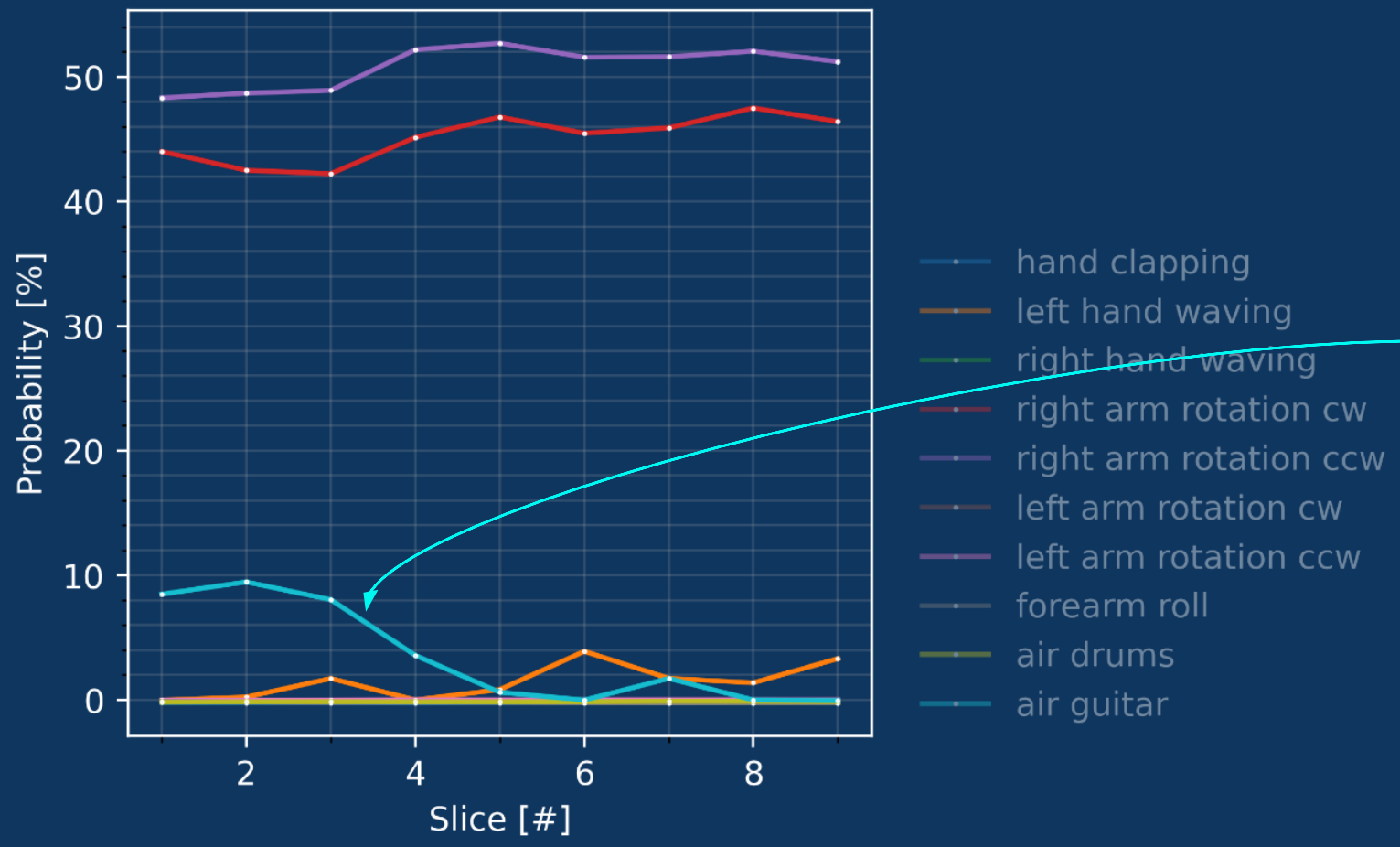
Confidence



- High-level information aggregation works:

- Correct class:
  Increases & scores high

Confidence

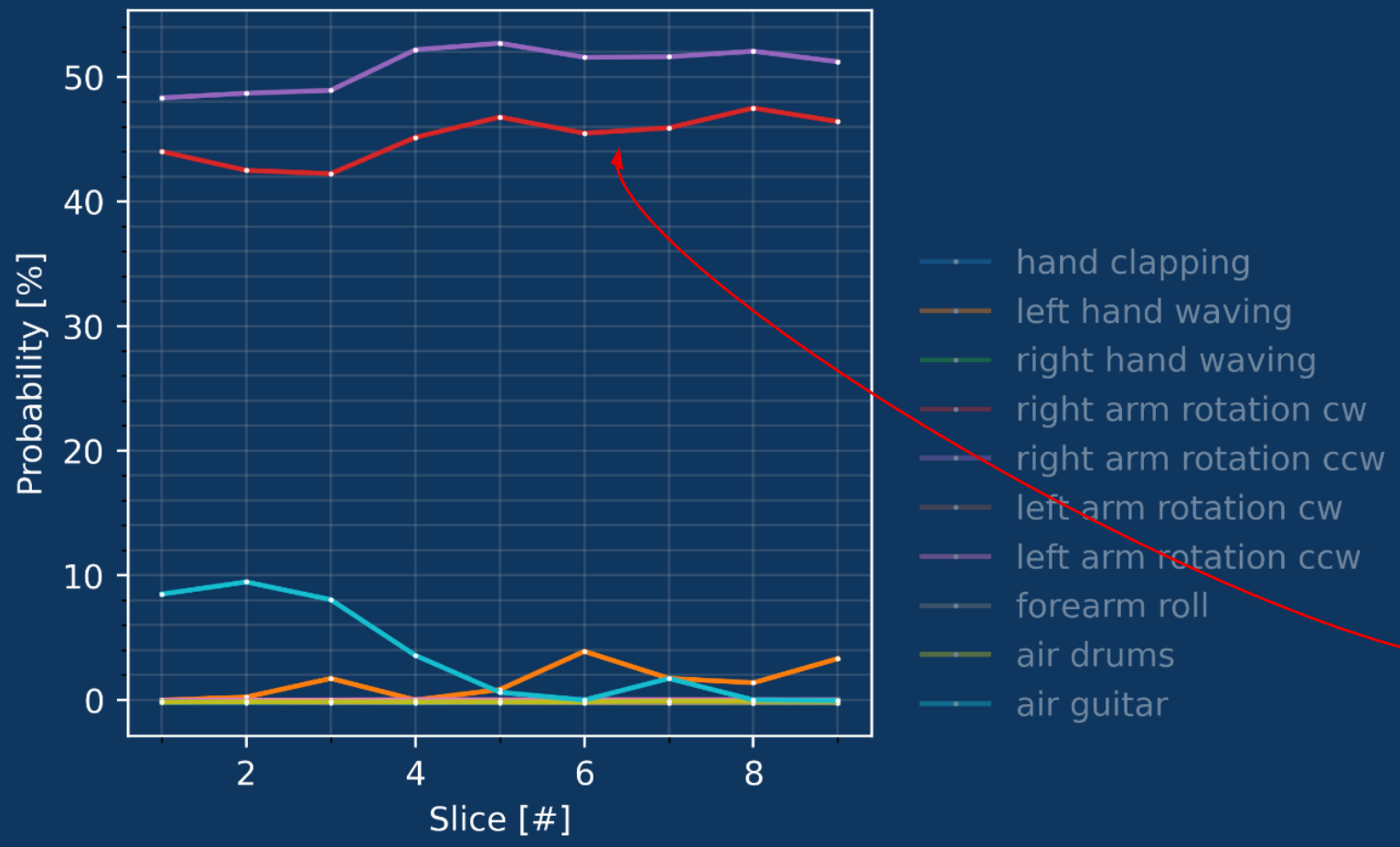


- High-level information aggregation works:

  - Correct class:
    Increases & scores high

  - Incorrect classes:
    Decrease & score low

Confidence



- High-level information aggregation works:

  - Correct class:
    Increases & scores high

  - Incorrect classes:
    Decrease & score low

  - Similar class:
    Closer probability

Problem

State of the Art

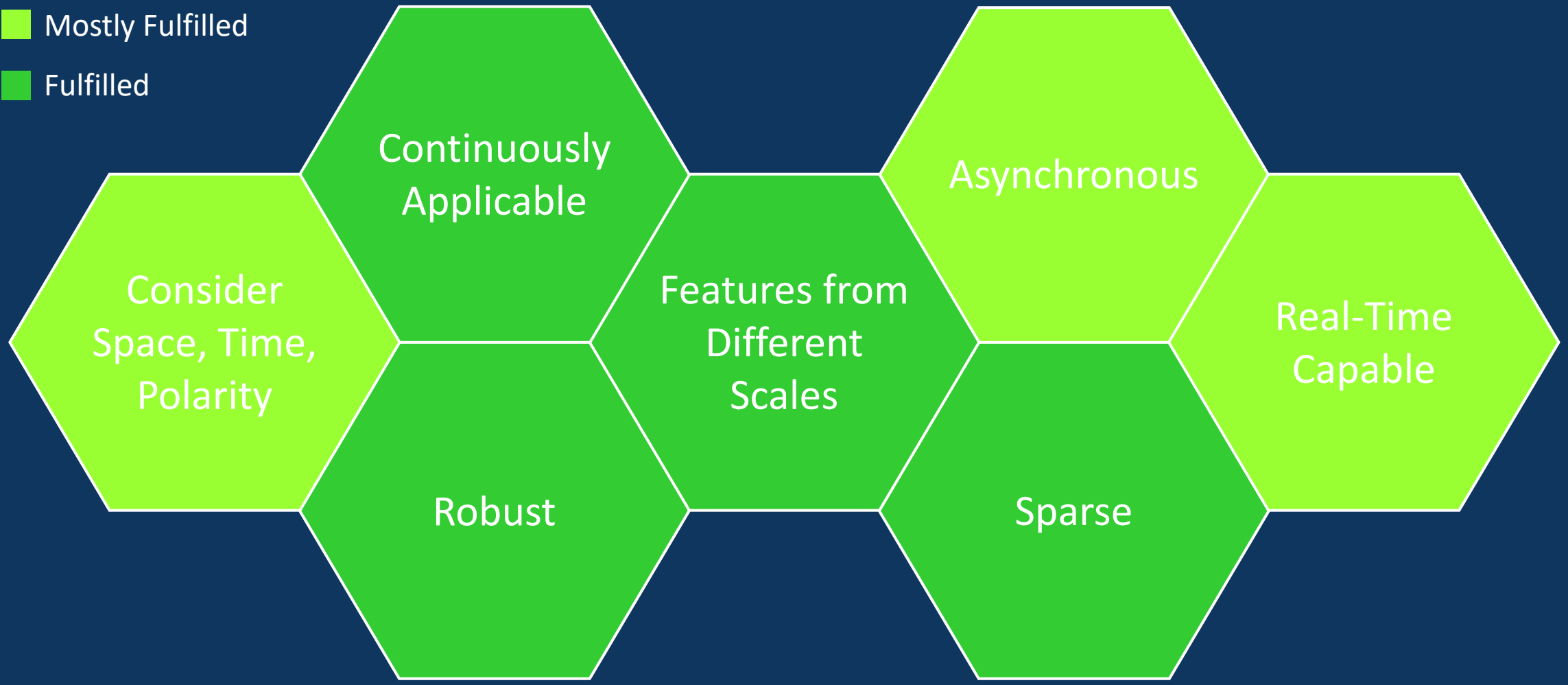
Concept

Results

Insights

Requirement Check

- Mostly Fulfilled
- Fulfilled



Research Questions

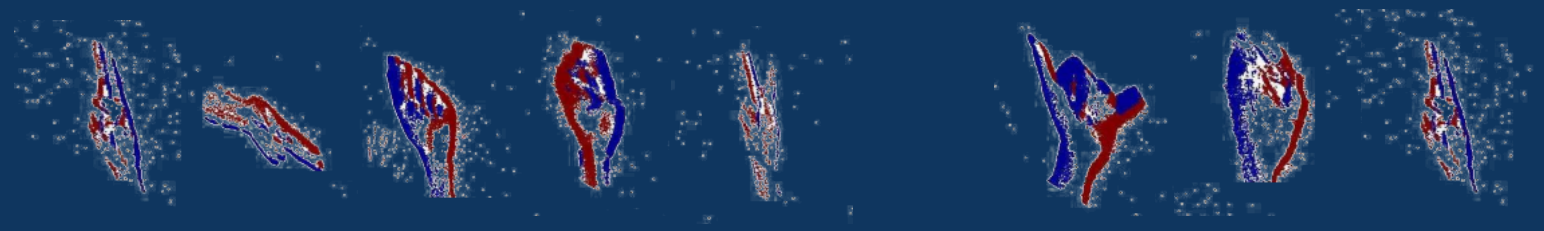
- How does **current AI** process event data?
- **Requirements** for processing that suits event data?
- **Develop model** according to requirements

Take Home

- AI approaches for event processing
  - CNN on 2D Representation
  - Sparse CNN
  - Sequence Processing
  - 3D Processing
- Processing model
  - Pre-Processing
  - Representation
  - Low-level features
  - High-level features
  - Backend
- Requirements ≠ conventional image processing e.g. sparse & asynchronous
- HAREM Net: Universal event processing network

Future Work

- Voxel Feature: Exploit polarity
- Aggregation: Include geometric information



HAREM
Net

Thank you

Sources

[1] Delbruck, Tobias
Frame-free dynamic digital vision
2008
Proceedings of International Symposium on Secure-Life Electronics, Advanced Electronics for Quality Life and Society, Univ. of Tokyo, Mar. 6-7, 2008
p. 21-26

[2] Gallego, Guillermo / Delbruck, Tobi / Orchard, Garrick / Bartolozzi, Chiara / Taba, Brian / Censi, Andrea / Leutenegger, Stefan / Davison, Andrew J. / Conradt, Jörg / Daniilidis, Kostas / others
Event-based vision: A survey
2020
IEEE transactions on pattern analysis and machine intelligence , Vol. 44, No. 1
p. 154-180

[3] Maqueda, Ana I. / Loquercio, Antonio / Gallego, Guillermo / Garc\ia, Narciso / Scaramuzza, Davide
Event-based vision meets deep learning on steering prediction for self-driving cars
2018
Proceedings of the IEEE conference on computer vision and pattern recognition
p. 5419-5427

[4] Messikommer, Nico / Gehrig, Daniel / Loquercio, Antonio / Scaramuzza, Davide
Event-based asynchronous sparse convolutional networks
2020
European Conference on Computer Vision
p. 415-431

[5] Sekikawa, Yusuke / Hara, Kosuke / Saito, Hideo
Eventnet: Asynchronous recursive event processing
2019
Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition
p. 3887-3896

[6] Wang, Yue / Sun, Yongbin / Liu, Ziwei / Sarma, Sanjay E. / Bronstein, Michael M. / Solomon, Justin M.
Dynamic graph cnn for learning on point clouds
2019
Acm Transactions On Graphics (tog) , Vol. 38, No. 5
ACM New York, NY, USA
p. 1-12

[7] Guo, Shasha / Delbruck, Tobi
Low Cost and Latency Event Camera Background Activity Denoising
2022
IEEE Transactions on Pattern Analysis and Machine Intelligence

[8] Deng, Yongjian / Chen, Hao / Chen, Huiying / Li, Youfu
EV-VGCNN: A voxel graph CNN for event-based object classification
2021
arXiv preprint arXiv:2106.00216

[9] Hochreiter, Sepp / Schmidhuber, Jürgen
Long short-term memory
1997
Neural computation , Vol. 9, No. 8 , MIT Press
p. 1735-1780

[10] Bi, Yin / Chadha, Aaron / Abbas, Alhabib / Bourtsoulatze, Eirina / Andreopoulos, Yiannis
Graph-based object classification for neuromorphic vision sensing
2019
Proceedings of the IEEE/CVF International Conference on Computer Vision
p. 491-501

[11] Amir, Arnon / Taba, Brian / Berg, David / Melano, Timothy / McKinstry, Jeffrey / Di Nolfo, Carmelo / Nayak, Tapan / Andreopoulos, Alexander / Garreau, Guillaume / Mendoza, Marcela / others
A low power, fully event-based gesture recognition system
2017
Proceedings of the IEEE conference on computer vision and pattern recognition
p. 7243-7252

[12] Vasudevan, Ajay / Negri, Pablo / Linares-Barranco, Bernabe / Serrano-Gotarredona, Teresa
Introduction and analysis of an event-based sign language dataset
2020
2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)
p. 675-682

Sources

[13] Shrestha, Sumit B. / Orchard, Garrick
Slayer: Spike layer error reassignment in time
2018
Advances in neural information processing systems , Vol. 31

[14] Wu, Yujie / Deng, Lei / Li, Guoqi / Zhu, Jun / Shi, Luping
Spatio-temporal backpropagation for training high-performance spiking neural networks
2018
Frontiers in neuroscience , Vol. 12, Frontiers Media SA
p. 331

[15] Bi, Yin / Chadha, Aaron / Abbas, Alhabib / Bourtsoulatze, Eirina / Andreopoulos, Yiannis
Graph-based object classification for neuromorphic vision sensing
2019
Proceedings of the IEEE/CVF International Conference on Computer Vision
p. 491-501

[16] Wang, Qinyi / Zhang, Yexin / Yuan, Junsong / Lu, Yilong
Space-time event clouds for gesture recognition: From RGB cameras to event cameras
2019
2019 IEEE Winter Conference on Applications of Computer Vision (WACV)
p. 1826-1835

[17] Huang, Chaoxing
Event-based action recognition using timestamp image encoding network
2020
arXiv preprint arXiv:2009.13049

[18] Cannici, Marco / Ciccone, Marco / Romanoni, Andrea / Matteucci, Matteo
A differentiable recurrent surface for asynchronous event-based data
2020
European Conference on Computer Vision
p. 136-152