UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN

Physical Data Curation and Augmentation

Tarek Abdelzaher, *University of Illinois at Urbana Champaign*

Reminders and Announcements

Elevator Talks on March 12

- Plan for a 5-minute talk that answers the four key Heilmeier questions about your project (see https://en.wikipedia.org/wiki/George_H._Heilmeier), namely:
  - *What are you trying to do?* Articulate your objectives using no jargon.
  - *How is it done today*, and what are the limits of current practice?
  - *What's new* in your approach and *why* do you think it will be *successful?*
  - *Who cares?* If you're successful, what difference will it make?
- You are encouraged to use visuals (it's a short talk) but plan them well. The visuals should not be pure "eye candy". They should serve as a vehicle to convey information more efficiently.
- Add a slide on the current *status and timeline*.

A Comment on Debates

Debate	G1	G2	G3	G4	G5	G6	G7	G8
D1	Contrastive	MAE	MAE	MAE				
D2						Small	Large	Small
D3	Time		Freq				Time	
D4		MAE		MAE		MAE		MAE
D5	Fixed		Variable				Variable	
D6		S3		S3		S3		S3

Agreement

Dissenting opinion

A Comment on Debates

Departing from consensus opinion (with good reasons) is highly correlated with innovation and success in research environments!

Debate	G1	G2	G3	G4	G5	G6	G7	G8
D1	Contrastive	MAE	MAE	MAE				
D2						Small	Large	Small
D3	Time		Freq				Time	
D4		MAE		MAE		MAE		MAE
D5	Fixed		Variable				Variable	
D6		S3		S3		S3		S3

Agreement

Dissenting opinion

Group 1

Soham Kaje Mathew Pan Davis Zhang

A Comment on Debates

Departing from consensus opinion (with good reasons) is highly correlated with innovation and success in research environments!

Debate	G1	G2	G3	G4	G5	G6	G7	G8
D1	Contrastive	MAE	MAE	MAE				
D2						Small	Large	Small
D3	Time		Freq *				Time	
D4		MAE		MAE		MAE		MAE
D5	Fixed		Variable				Variable	
D6		S3		S3		S3		S3

Agreement

Dissenting opinion (a "*" denotes that dissenting opinion won!)

What We Covered: The Data Bottleneck in Self-Supervised Learning for IoT

CS 537 (AIoT) — Home | Syllabus | Schedule | Homework | Piazza | Canvas

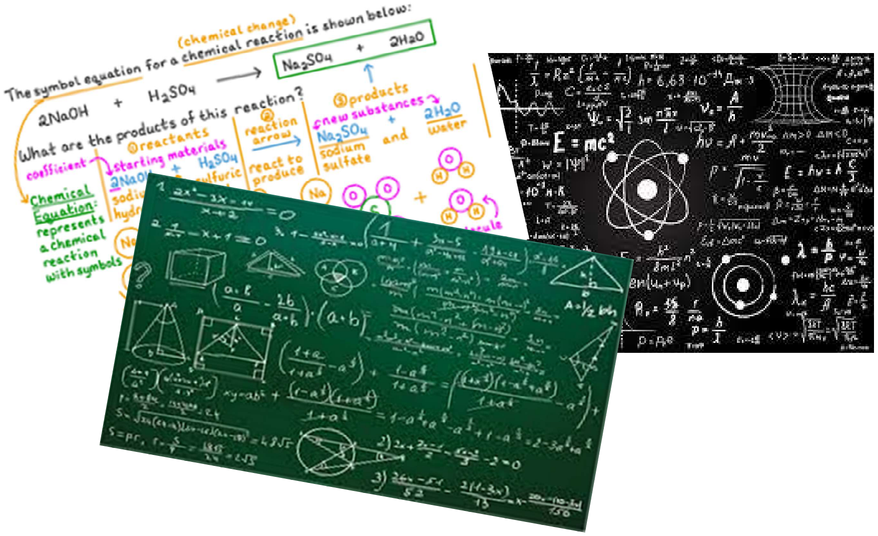
The Data Bottleneck: Self-Supervised Data-Efficient Learning for IoT	2/3	Class Project Ideas Introduction.	Projects	Slide set	
	2/5	Fundamentals of Self-Supervised Learning: Tokenization, Pre-training, Fine-tuning, Backbone Architectures (e.g., auto-encoders, transformers, etc), and Issues with Scaling Laws for IoT Applications	Self-Supervised Learning (SSL)	Slide set	
	2/10	Self-supervised Learning Architectures for Time-Series Data: RNNs, LSTMs, and State Space Models	SSL Models for Time-Series Data	1. Schmidt, Robin M. "Recurrent neural networks (rnns): A gentle introduction and overview." arXiv preprint arXiv:1912.05911 (2019). 2. Kexin Zhang, Qingsong Wen, Chaoli Zhang, Rongyao Cai, Ming Jin, Yong Liu, James Y. Zhang et al. "Self-supervised learning for time series analysis: Taxonomy, progress, and prospects." IEEE transactions on pattern analysis and machine intelligence 46, no. 10 (2024): 6775-6794. 3. Albert Gu, Karan Goel, and Christopher Ré. "Efficiently modeling long sequences with structured state spaces." arXiv preprint arXiv:2111.00396 (2021). 4. Albert Gu, and Tri Dao. "Mamba: Linear-time sequence modeling with selective state spaces." In First conference on language modeling. 2024.	Note: Project title and abstract due
	2/12	Representation Learning from Multimodal Sensor Data	Multimodal Intro HW1 Out	1. Chao Zhang, Zichao Yang, Xiaodong He, and Li Deng. "Multimodal intelligence: Representation learning, information fusion, and applications." Journal of Selected Topics in Signal Processing, 2020. 2. Dave Vedant, Fotios Lygerakis, and Elmar Rueckert. "Multimodal visual-tactile representation learning through self-supervised contrastive pre-training." In 2024 IEEE ICRA, pp. 8013-8020. IEEE, 2024. 3. Chen Sun, Austin Myers, Carl Vondrick, Kevin Murphy, and Cordelia Schmid. "Videobert: A joint model for video and language representation learning." In Proceedings of the IEEE/CVF international conference on computer vision, pp. 7464-7473. 2019.	
	2/17	Representation Learning from Multimodal Sensor Data (Student Led)	G4 Multimodal Papers	1. Shengzhong Liu, Tomoyoshi Kimura, Dongxin Liu, Ruijie Wang, Jinyang Li, Suhas Diggavi, Mani Srivastava, and Tarek Abdelzaher. "Focal: Contrastive learning for multimodal time-series sensing signals in factorized orthogonal latent space." Advances in Neural Information Processing Systems 36 (2023): 47309-47338. 2. Chen, Yatong, Chenzhi Hu, Tomoyoshi Kimura, Qinya Li, Shengzhong Liu, Fan Wu, and Guihai Chen. "SemiCMT: Contrastive cross-modal knowledge transfer for iot sensing with semi-paired multi-modal signals." Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies 8, no. 4 (2024): 1-30. 3. Xiaomin Ouyang, Jason Wu, Tomoyoshi Kimura, Yihan Lin, Gunjan Verma, Tarek Abdelzaher, and Mani Srivastava. "MMbind: Unleashing the potential of distributed and heterogeneous data for multimodal learning in iot." In Proceedings of the 23rd ACM Conference on Embedded Networked Sensor Systems, pp. 491-503. 2025. 4. Tomoyoshi Kimura, Xinlin Li, Osama Hanna, Yatong Chen, Yizhuo Chen, Denizhan Kara, Tianshi Wang et al. "InfoMAE: Pair-efficient cross-modal alignment for multimodal time-series sensing signals." In Proceedings of the ACM on Web Conference 2025, pp. 3084-3095. 2025. 5. Li, Zechen, Shohreh Deldari, Linyao Chen, Hao Xue, and Flora D. Salim. "SensorLLM: Aligning large language models with motion sensors for human activity recognition." In Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing, pp. 354-379. 2025. 6. Yuwei Zhang, Kumar Ayush, Siyuan Qiao, A. Ali Heydari, Girish Narayanswamy, Maxwell A. Xu, Ahmed A. Metwally et al. "SensorLM: Learning the Language of Wearable Sensors." arXiv preprint arXiv:2506.09108 (2025).	Debate #1 (20 min) Student led talk (45 min + 10 min Q&A) See note @48 for debate concluding remarks on Piazza.
	2/19	Self-supervised Learning from Frequency Domain Data	Frequency Domain Intro HW2 Out	Slide set	Debate #2 (20 min)
	2/24	Self-supervised Learning from Frequency Domain Data (Student Led)	G3 Frequency Domain Papers	1. Shuochao Yao, Ailing Piao, Wenjun Jiang, Yiran Zhao, Huajie Shao, Shengzhong Liu, Dongxin Liu, Jinyang Li, Tianshi Wang, Shaohan Hu, Lu Su, Jiawei Han and Tarek Abdelzaher, "STFNets: Learning Sensing Signals from the Time-Frequency Perspective with Short-Time Fourier Neural Networks," In Proc. The Web Conference (WWW), San Francisco, CA, May 2019. 2. Dongxin Liu, Tianshi Wang, Shengzhong Liu, Ruijie Wang, Shuochao Yao, and Tarek Abdelzaher. "Contrastive self-supervised representation learning for sensing signals from the time-frequency perspective." In 2021 International Conference on Computer Communications and Networks (ICCCN), pp. 1-10. IEEE, 2021. 3. Yuan Gong, Cheng-I. Lai, Yu-An Chung, and James Glass. "Ssast: Self-supervised audio spectrogram transformer." In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 36, no. 10, pp. 10699-10709. 2022. 4. Setareh Rahimi Taghanaki, Michael Rainbow, and Ali Etemad. "Self-supervised human activity recognition with localized time-frequency contrastive representation learning." IEEE Transactions on Human-Machine Systems 53, no. 6 (2023): 1027-1037. 5. Denizhan Kara, Shengzhong Liu, Jinyang Li, Dongxin Liu, Tianshi Wang, Ruijie Wang, Yizhuo Chen, Yigong Hu, Tarek Abdelzaher, "FreqMAE: Frequency-Aware Masked Autoencoder for Multi-Modal IoT Sensing," In Proc. The	HW2 Debate (20 min) Student led talk (45 min + 10 min Q&A)

	2/26	Handling Spatial-Temporal IoT Data	Spatial Temporal Intro HW3 Out	Slide Set	2-page project proposal due
	3/3	Handling Spatial-Temporal IoT Data (Student Led)	G2	1. Yinghui Zhang, Hu An, Yaxuan Xing, Yang Liu, and Tiankui Zhang. "Learning temporal and spatial features jointly: A unified framework for space-time data prediction in industrial IoT networks." IEEE Sensors Journal 23, no. 16 (2023): 18752-18764. 2. Liu, Jing, et al. "Distributional and spatial-temporal robust representation learning for transportation activity recognition." Pattern Recognition (2023). 3. Porter Jenkins, Ahmad Farag, Suhang Wang, and Zhenhui Li. "Unsupervised representation learning of spatial data via multimodal embedding." In Proceedings of the 28th ACM international conference on information and knowledge management, pp. 1993-2002. 2019. 4. Yizhuo Chen, Tianchen Wang, You Lyu, Yanlan Hu, Jinyang Li, Tomoyoshi Kimura, Hongjue Zhao, Yigong Hu, Denizhan Kara, and Tarek Abdelzaher. "Spar: Self-supervised placement-aware representation learning for multi-node iot systems." arXiv e-prints (2025): arXiv-2505. 5. Tianchen Wang, Yizhuo Chen, Hongjue Zhao, You Lyu, Jinyang Li, Tomoyoshi Kimura, Yigong Hu et al. "On Network-Efficient Multimodal Multi-Vantage Foundation Models for Distributed Sensing." In 2025 IEEE 22nd International Conference on Mobile Ad-Hoc and Smart Systems (MASS), pp. 19-27. IEEE, 2025. 6. Pengrui Quan, Brian Wang, Kang Yang, Liying Han, and Mani Srivastava. "Benchmarking Spatiotemporal Reasoning in LLMs and Reasoning Models: Capabilities and Challenges." arXiv preprint arXiv:2505.11618 (2025).	HW3 Debate (20 min) br> Student led talk (45 min + 10 min Q&A)
Data Curation and "Faking"	3/5	Physical Data Curation and Augmentation	HW4 Out	Slide Set	
	3/10	Physical Data Curation and Augmentation (Student Led)	G6	1. Chenzhi Hu, Yatong Chen, Denizhan Kara, Shengzhong Liu, Tarek Abdelzaher, Fan Wu, Guihai Chen, "OpenMAE: Efficient Masked Autoencoder for Vibration Sensing with Open-Domain Data Enrichment," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (ACM IMWUT), also presented in UbiComp, Espoo, Finland, October 2025. 2. Jeongwoo Ju, Heechul Jung, Yoonju Oh, and Junmo Kim. "Extending contrastive learning to unsupervised coreset selection." IEEE Access 10 (2022): 7704-7715 3. Sungnyun KIm, Sangmin Bae, and Se-Young Yun. "Coreset sampling from open-set for fine-grained self-supervised learning." In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 7537-7547. 2023. 4. Haizhong Zheng, Elisa Tsai, Yifu Lu, Jiachen Sun, Brian R. Bartoldson, Bhavya Kailkhura, Atul Prakash, "ELFS: LABEL-FREE CORESET SELECTION WITHPROXY TRAINING DYNAMICS," ICLR 2025 5. Tianshi Wang, Jinyang Li, Ruijie Wang, Denizhan Kara, Shengzhong Liu, Davis Wertheimer, Antoni Martin, Raghu Ganti, Mudhakar Srivatsa, and Tarek Abdelzaher, "SudokuSens: Enhancing Deep Learning Robustness for IoT Sensing Applications using a Generative Approach," In Proc. ACM Sensys, Istanbul, Turkey, November 2023. 6. Tianshi Wang, Qikai Yang, Ruijie Wang, Dachun Sun, Jinyang Li, Yizhuo Chen, Yigong Hu, Chaoqi Yang, Tomoyoshi Kimura, Denizhan Kara, Tarek F. Abdelzaher, "Fine-grained Control of Generative Data Augmentation in IoT Sensing," In Proc. 38th Annual Conference on Neural Information Processing Systems (NeurIPS), Vancouver, Canada, December 2024.	HW4 Debate (20 min) Student led talk (45 min + 10 min Q&A)
	3/12	Project Elevator Talks			
Break	3/17 3/19	Spring Break			

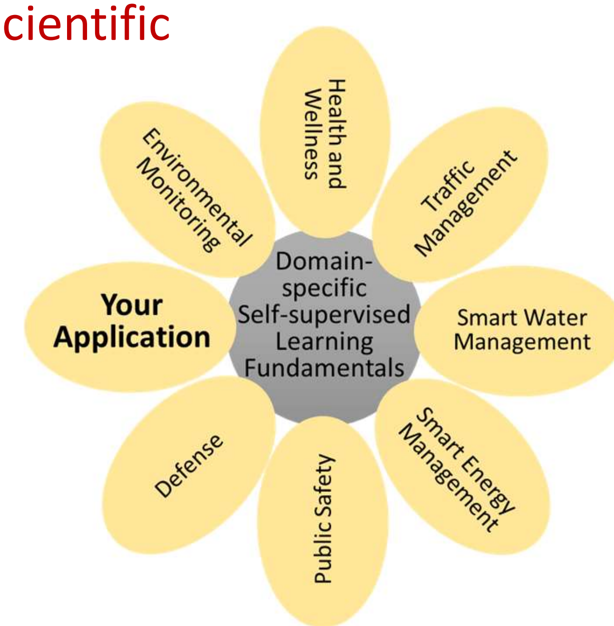
- Make learning more efficient for IoT data
- Increase the data size (curation and augmentation)

The Era of (Empirically Trained) ***Domain-Specific*** Foundation Models

CPS/IoT applications of the future will replace conventional scientific foundations with empirically trained foundation models



Conventional (scientific) foundation models

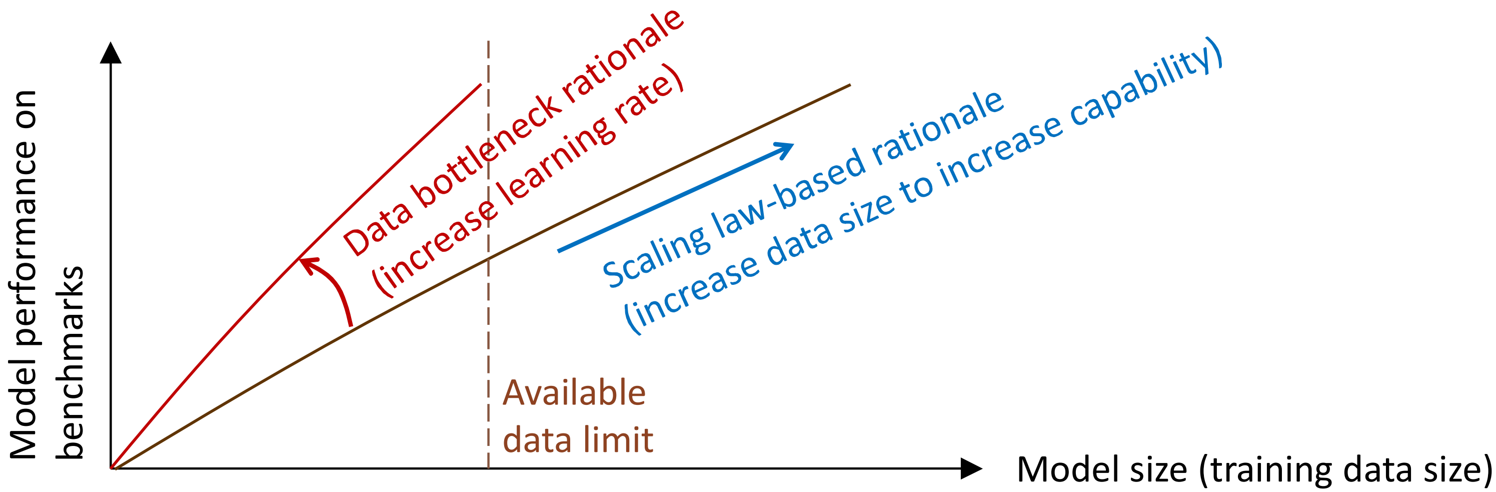


Tomorrow’s data driven foundation models and LLMs (for CPS/IoT applications)

The Data Bottleneck

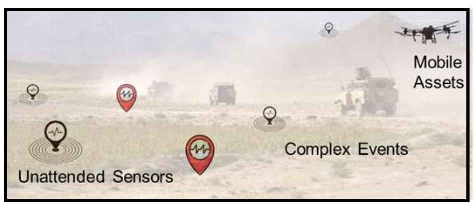
In domain-specific models, one can't rely on model scaling laws to improve model capabilities (because model size is limited by training data size)

Must re-think the training pipeline to *inject the "right amount" of application bias* such that the learning rate from limited data is accelerated.



Self-supervised Learning – The Principle of Operation

Unlabeled Data



Self-Supervised Pretraining

"Encoder" (Neural Network)

Encoding into a semantically well-organized latent space

Supported downstream Tasks

- Target/formation detection
- Classification/Identification
- Localization/Tracking
- Activity Recognition, Intent Prediction

Finetuning

Components of Self-Supervised Learning

Part I: The tokenizer

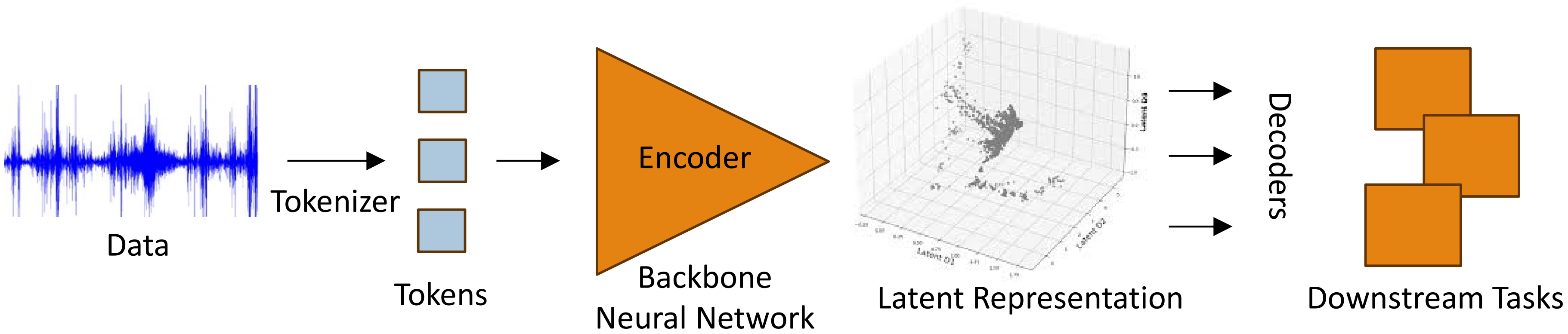
- Breaks the input stream into pieces to be individually encoded.

Part II: The backbone neural network model (or encoder)

- The neural network that converts the stream of input tokens into a latent representation

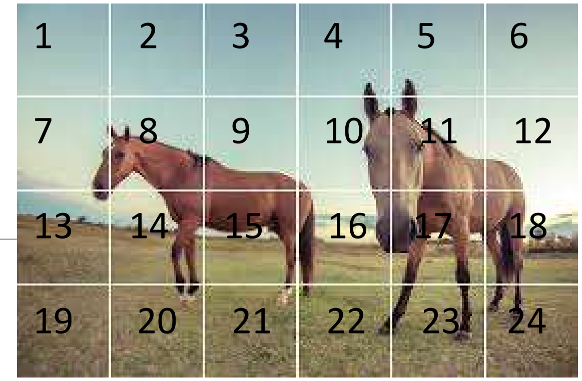
Part III: The pretext task(s)

- The task that trains the encoder



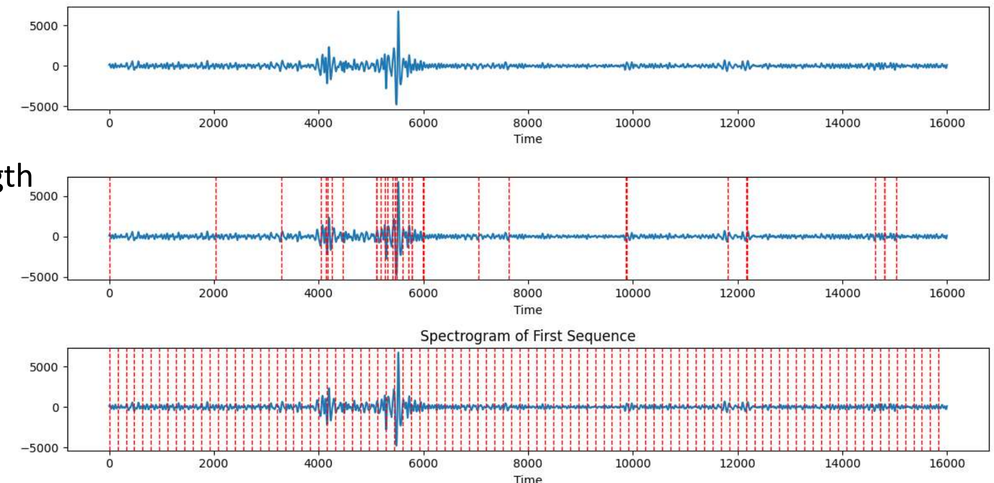
The Tokenizer

Multidimensional
time-series data



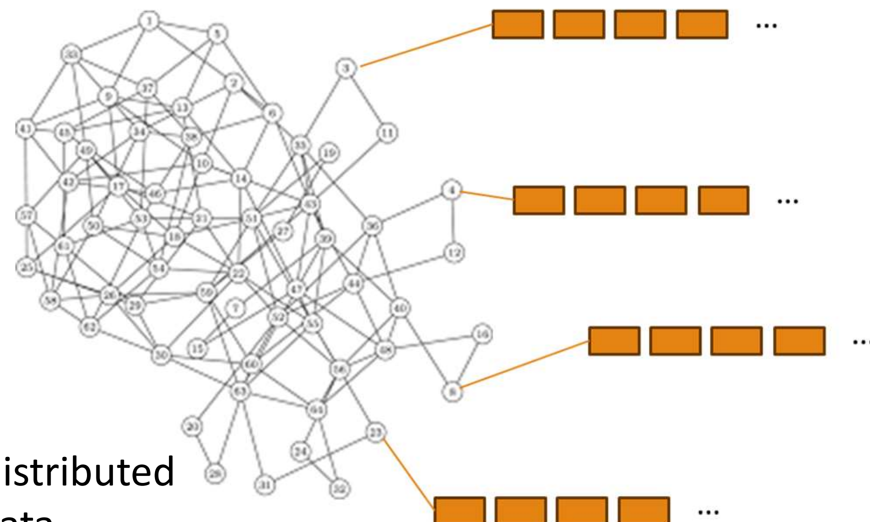
Variable length
Tokenization

Fixed length
Tokenization



Picture generated by Tommy Kimura

One-dimensional
time-series data



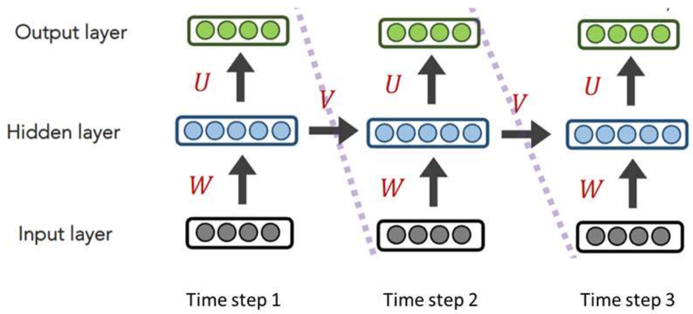
Spatially-distributed
or graph data

Foundation Models for CPS/IoT Data

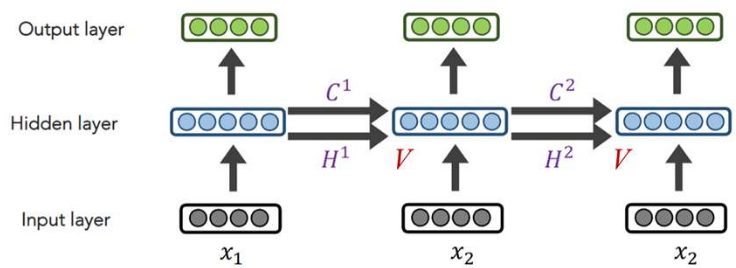
IoT Requirement #1:

Handling Time-Series Sensor Data

Neural Architecture for Time-Series Data

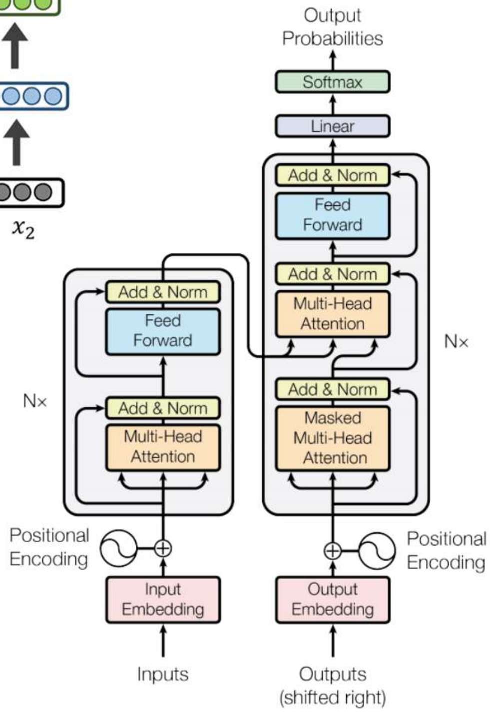


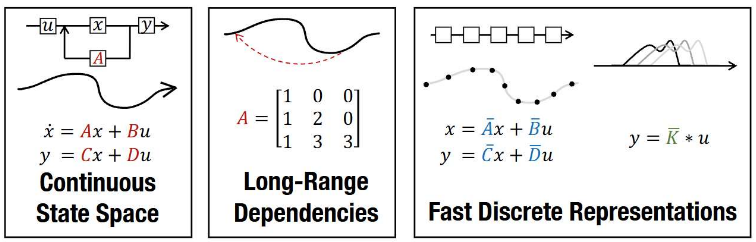
Recurrent Neural Networks



LSTM Models

Transformers





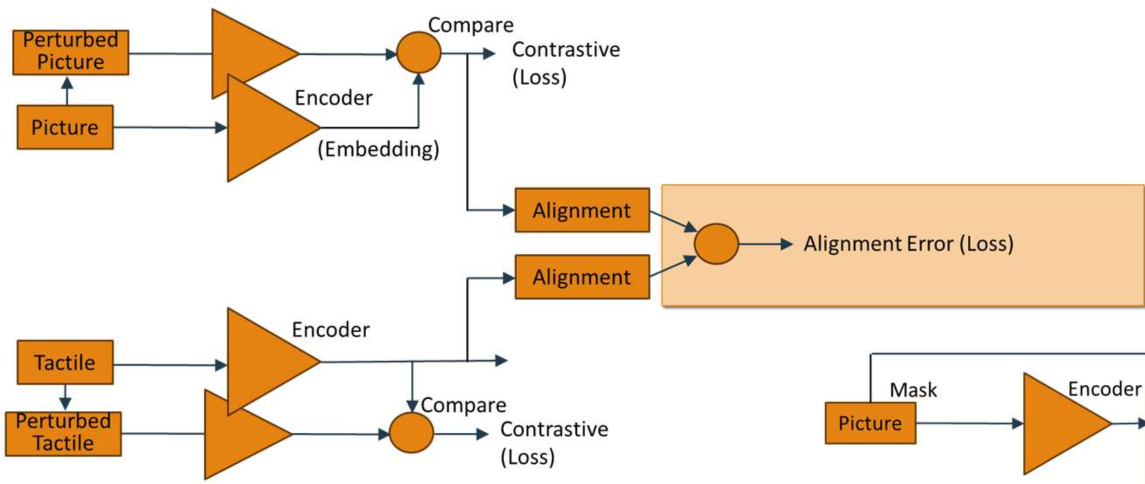
Structured State Space Models

Foundation Models for CPS/IoT Data

IoT Requirement #2:

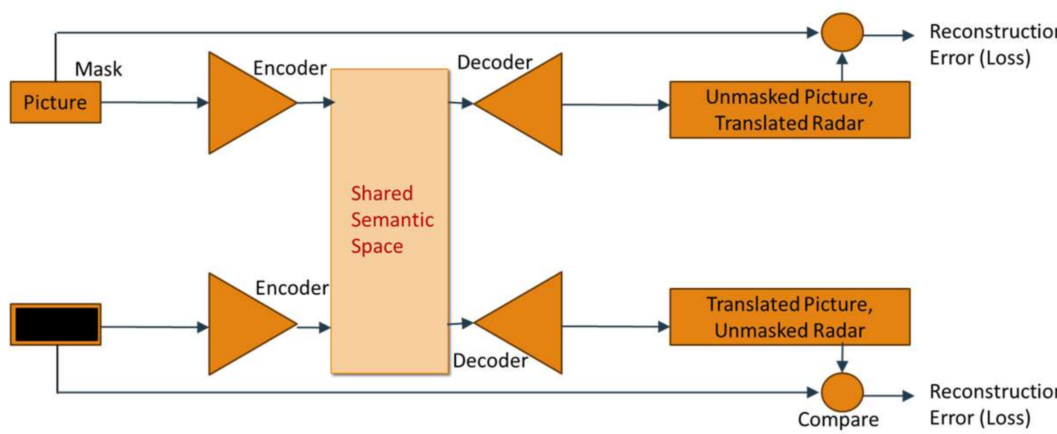
Handling Multimodal Sensor Data

Handling Multimodal Sensor Data



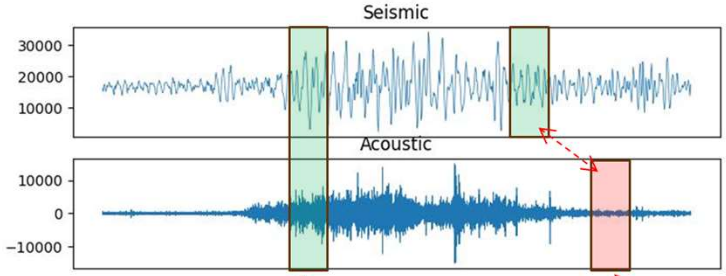
Modality alignment in contrastive learning

Modality alignment in masked auto-encoding



Handling Multimodal Sensor Data: Special IoT Considerations in SSL

Physical Event/Activity



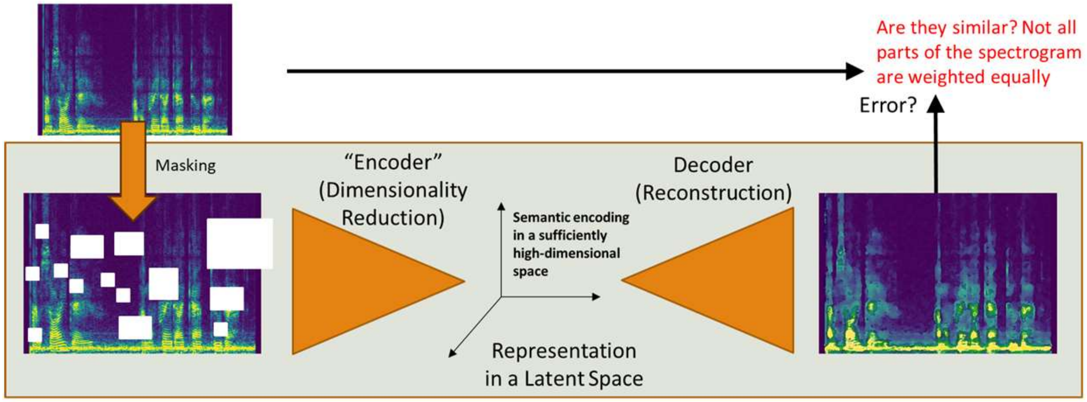
Multi-sensory Signature of Physical Event/Activity

Same time interval = similar

Different intervals = dissimilar

New notions of similarity

New masking policies

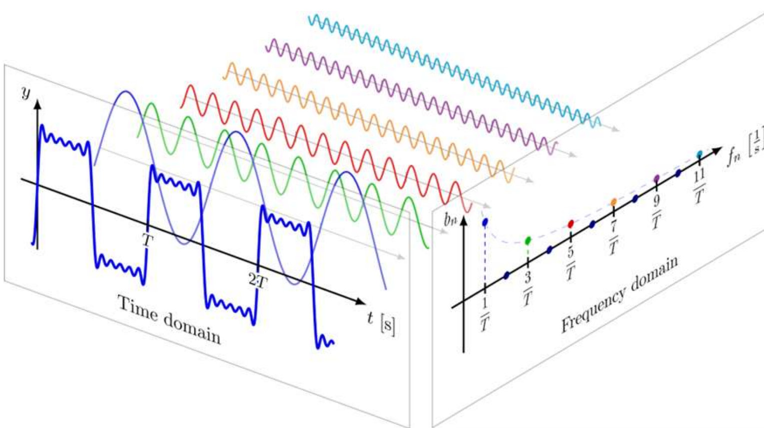


Foundation Models for CPS/IoT Data

Contribution #3:

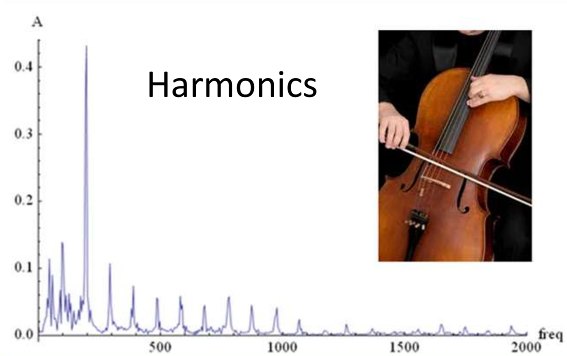
Exploiting Frequency Domain Insights

Accommodating Frequency Domain Inputs

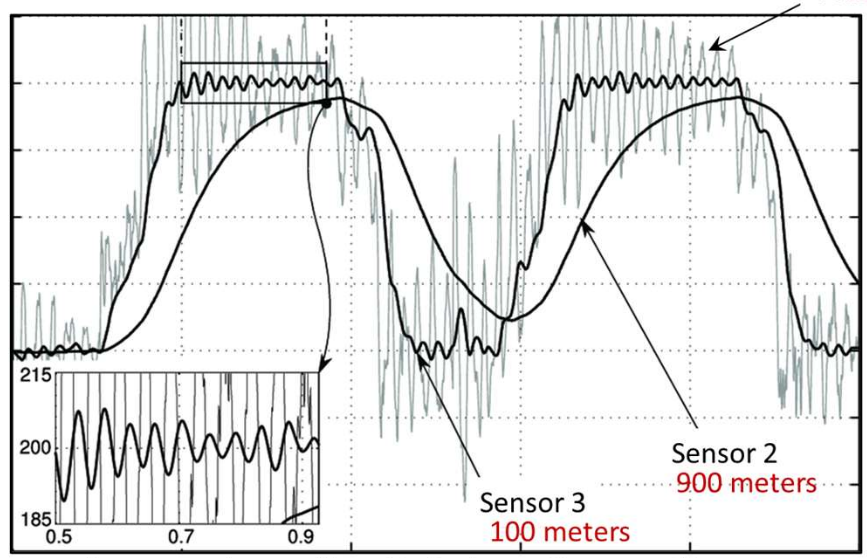


Fourier Transform

Harmonics



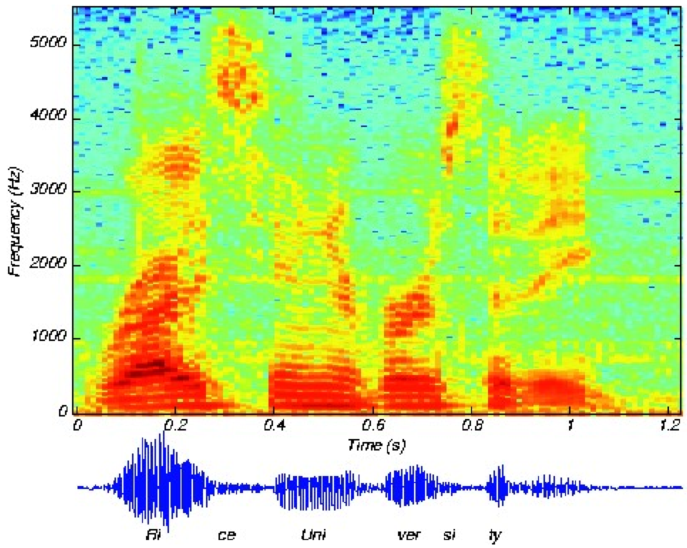
Signal propagation and low-pass filtering

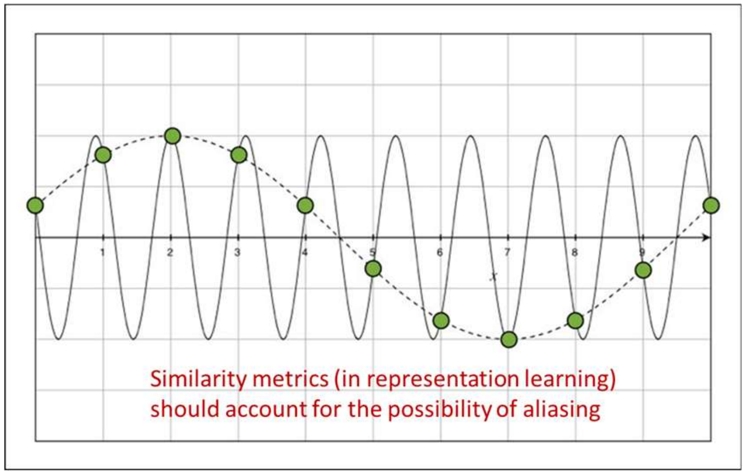


https://www.researchgate.net/figure/The-fundamental-limitation-of-linear-time-invariant-low-pass-filtering-the-length-of-a_fig5_260804992

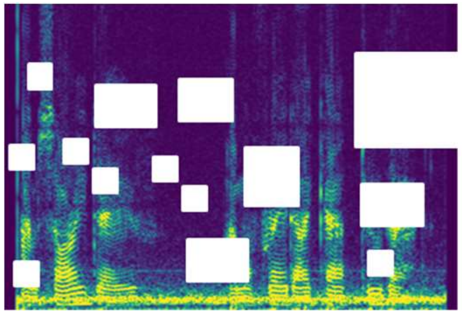
Accommodating Frequency Domain Inputs: Special IoT Considerations in SSL

Data augmentation considerations: No shift-invariance in spectrograms





Contrastive learning considerations: Similarity must account for aliasing



Masked auto-encoding considerations: Masking must account for semantic importance

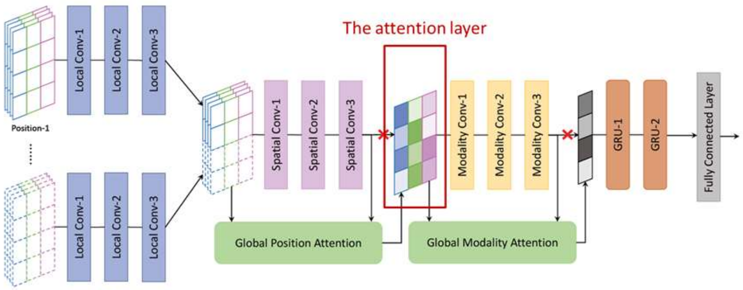
Foundation Models for CPS-IoT Data

Contribution #4:

Extensions to Distributed (Multi-Vantage) Sensing Data

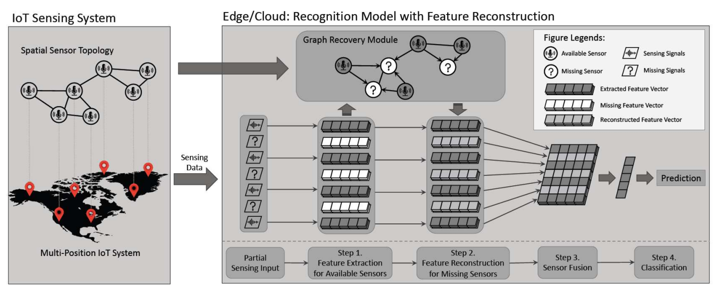
Extensions to Distributed (Spatial-Temporal) Sensing Data

Exploiting multiplicity/complementarity of vantage points



Overall encoder architecture

Exploiting structured relations among vantage points

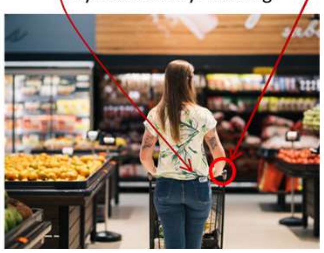


Exploiting geographic (spatial) locations of vantage points: duality of signal and location

Physical activity: walking

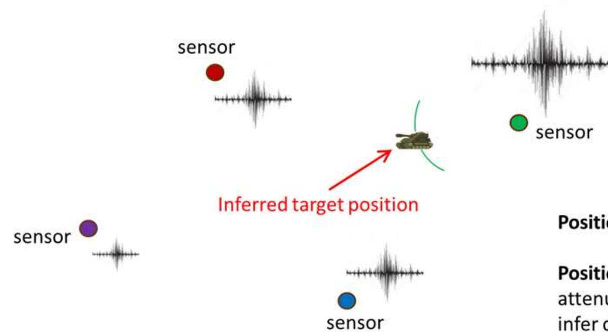
Wrist immobilized by cart.
Use phone not wrist-watch

Physical activity: walking



Wrist is free.
Wrist-watch OK

Physical activity: walking



Position encoding: GPS

Position/Signal Relation: Signal attenuates with distance → can infer distance from received signal

Foundation Models for CPS-IoT Data

Contribution #5:

Data Curation and Augmentation

Data Curation

Curation: How to select more useful data for training?

- How is this problem different for CPS/IoT data?

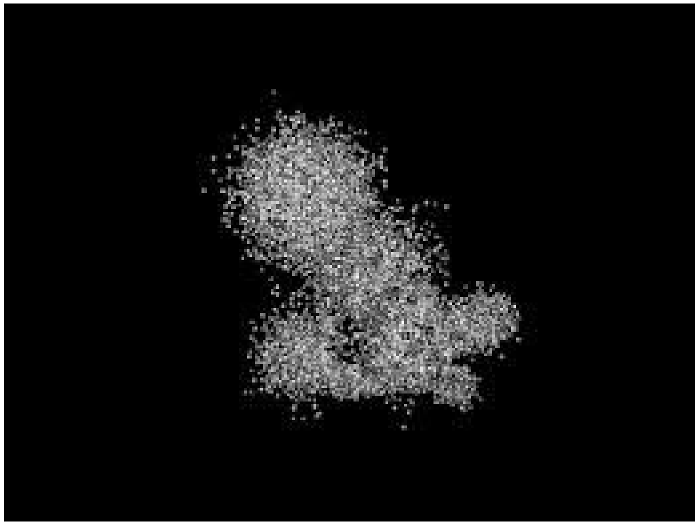
A Common Data “Selection” Problem: Coreset Selection for Training

What’s a core-set?

A Common Data "Selection" Problem: Coreset Selection for Training

What's a core-set?

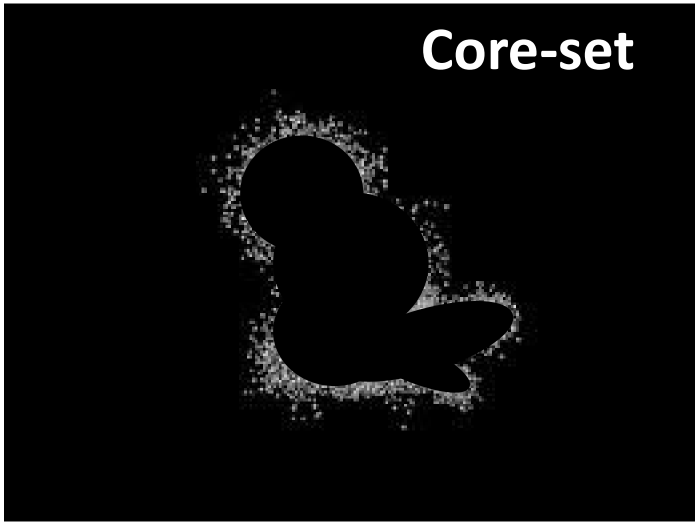
- A representative data subset, often for purposes of performing specific tasks
- Example: Find the "radius" of a point cloud



A Common Data "Selection" Problem: Coreset Selection for Training

What's a core-set?

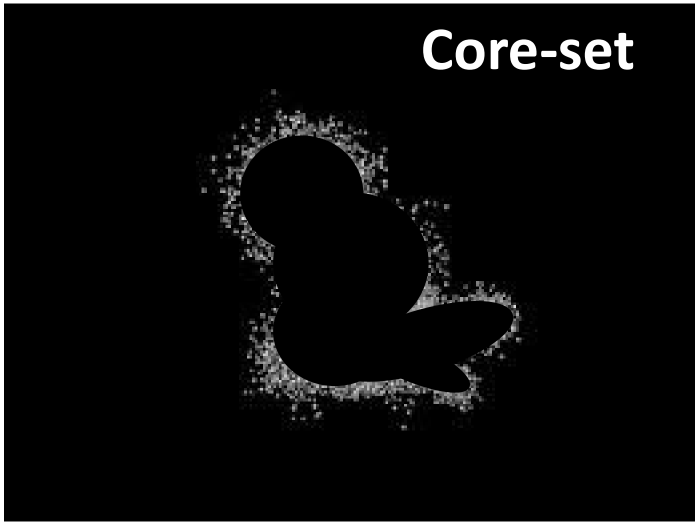
- A representative data subset, often for purposes of performing specific tasks
- Example: Find the "radius" of a point cloud



Property: The radius computed from the core-set is approximately the same as the radius computed from the entire set.

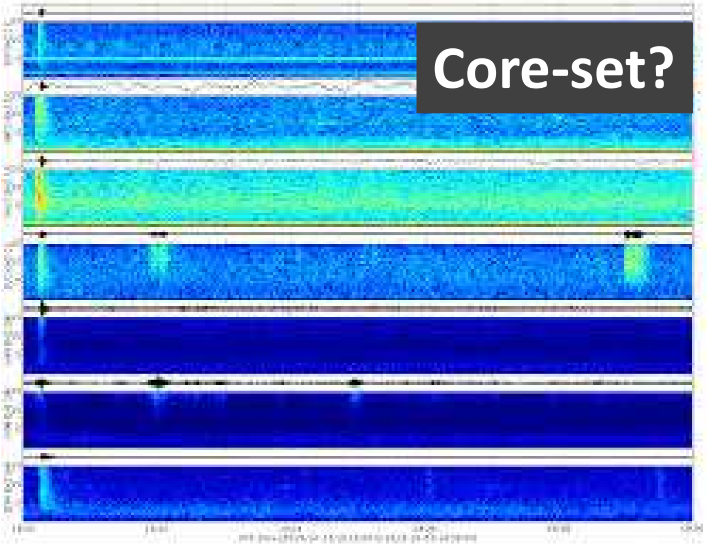
A Common Data “Selection” Problem: Coreset Selection for Training

How to select a core-set (from open domain data) with the property that training a neural network on the core-set produces the same quality analytics as training it on the original data?



A Common Data "Selection" Problem: Coreset Selection for Training

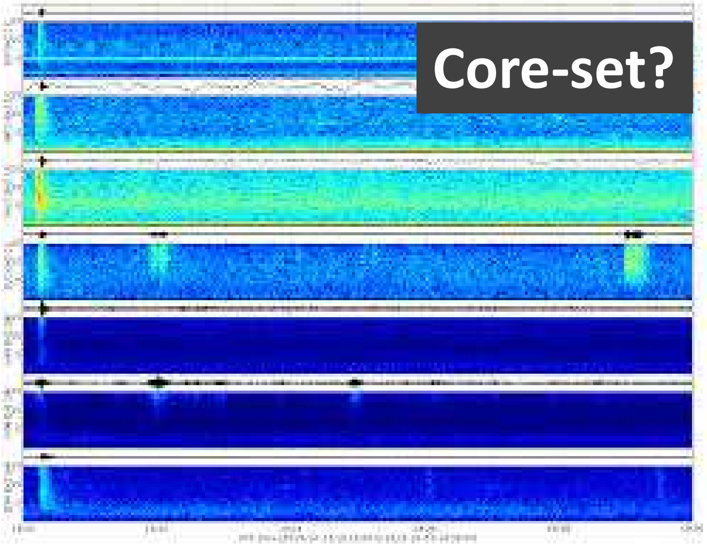
What's different about this problem in IoT settings?



A Common Data “Selection” Problem: Coreset Selection for Training

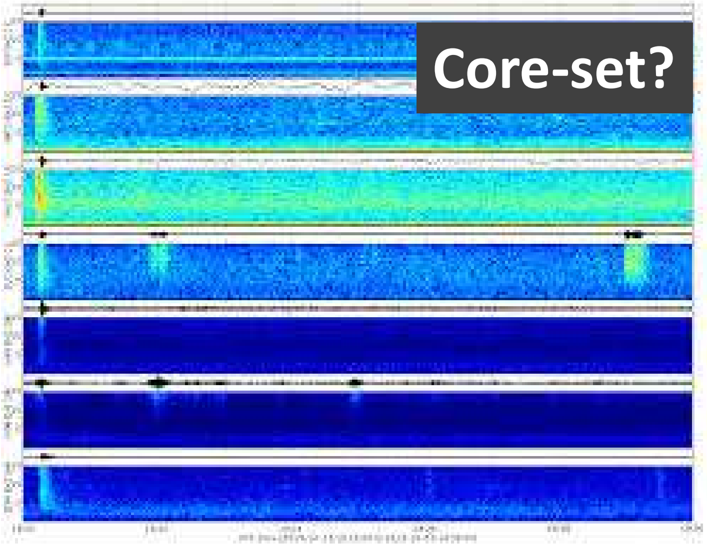
What’s different about this problem in IoT settings?

- Sensor data are mostly noise! Representative ≠ Informative!



The Data Curation Problem in IoT: *Informative* Coreset Selection for Training

May need to inject application bias to decide what data are informative and what data constitute noise when the data are neither interpretable nor labeled.



Data Augmentation for Training

Why do we need data augmentation?

Data Augmentation for Training

Why do we need data augmentation?

1) Fill in (categorical) gaps in observations

	Car	Bicycle
Freeway	✓	?
Dirt road	✓	✓

Example:

- We measured acoustic signatures of cars on a freeway
- We measured acoustic signatures of cars on dirt roads
- We measured acoustic signatures of bicycles on dirt roads
- Can we use generative AI to produces synthetic acoustic signatures of bicycles on a freeway?

Data Augmentation for Training

Why do we need data augmentation?

1) Fill in (categorical) gaps in observations

2) Increase the diversity of observations

Example: Synthetically vary parameters such as the intensity level of a human activity, speed of a car, etc.

Data Augmentation for Training

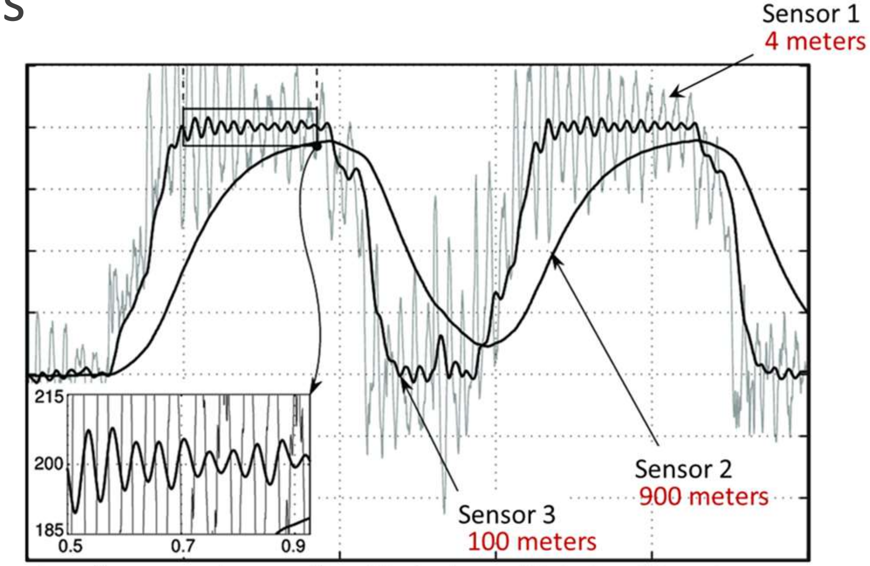
Why do we need data augmentation?

1) Fill in (categorical) gaps in observations

2) Increase the diversity of observations

3) Teach the AI (encoder) laws of nature

Example: Teach the AI propagation channel characteristics (e.g., low pass filtering) from synthetically augmented observations (e.g., real data + channel simulations)

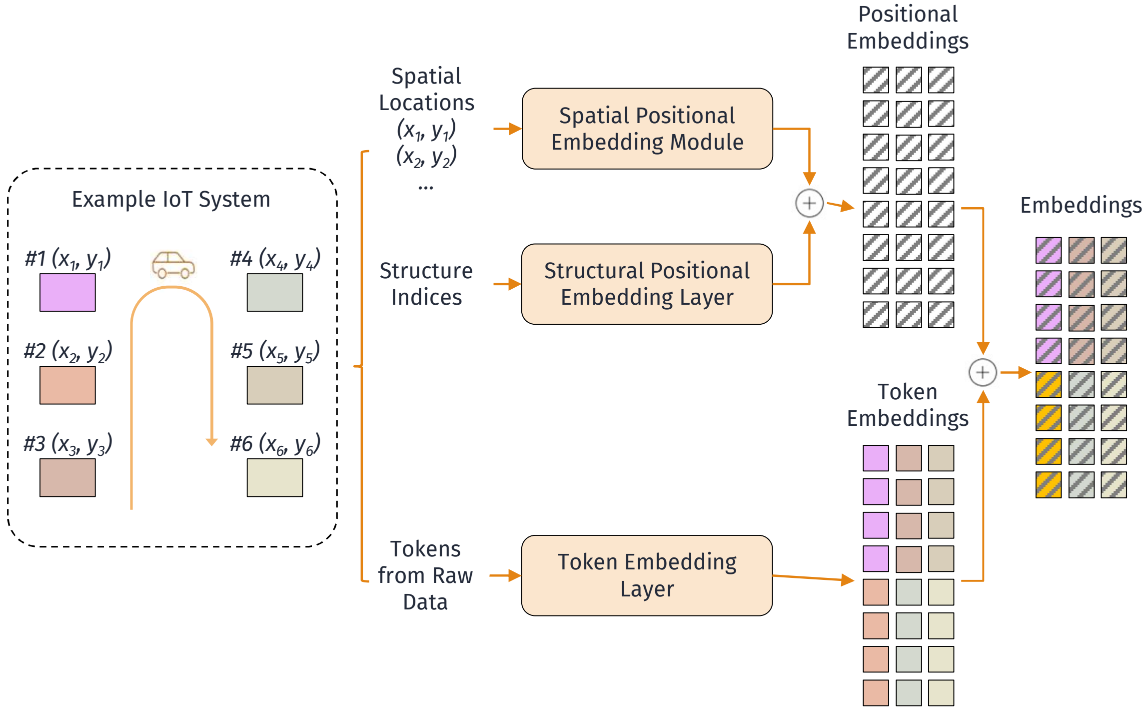


https://www.researchgate.net/figure/The-fundamental-limitation-of-linear-time-invariant-low-pass-filtering-the-length-of-a_fig5_260804992

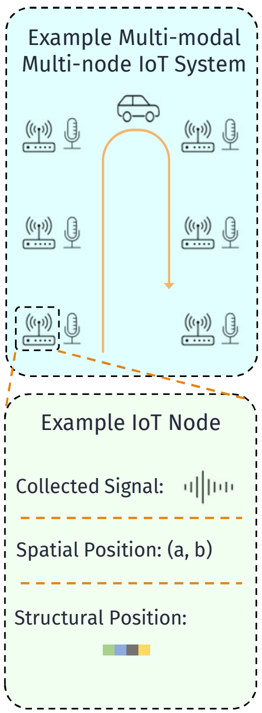
[6] Y Chen, T Wang, Y Lyu, Y Hu, J Li, T Kimura, H Zhao, Y Hu, D Kara, T. Abdelzaher, "SPAR: Self-supervised Placement-Aware Representation Learning for Multi-Node IoT Systems," arXiv preprint arXiv:2505.16936, **May 2025**.

Training Foundation Models from Multi-view Data

- **Key need:** must perform spatially-aware fusion of multi-vantage data streams
  - Must understand the implications of location and structural information on sensor data properties
- **Key Idea:** an extended Masked Auto-Encoder (MAE) with new reconstruction objectives that force learning the duality of signals and vantage points
  - Reconstruct masked signal given location, or reconstruct masked location given signal



Learning the Signal-Position Duality Improves Performance of Spatial Analysis



Moving vehicle localization experiments using self-supervised foundation models. Localization is based on 2-second seismic and acoustic data snippets collected from six nearby sensors.

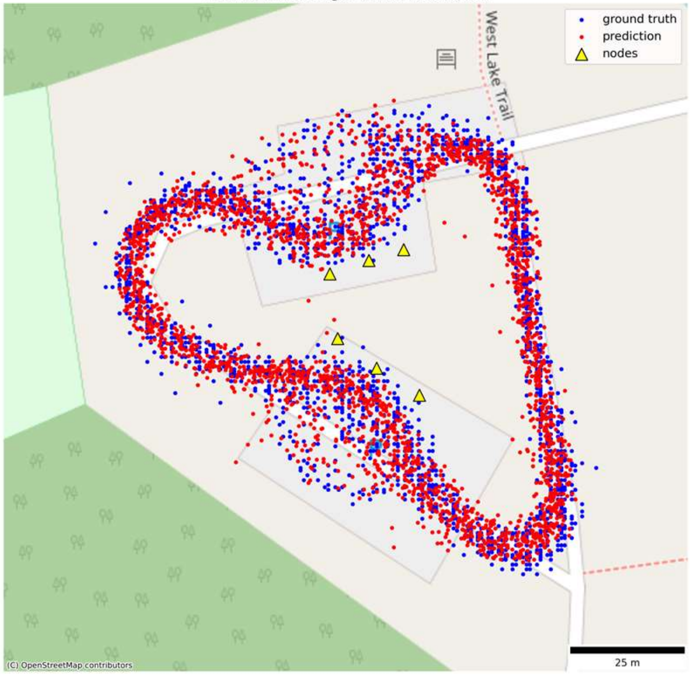
Method	M3N-VC Single-vehicle Localization					
	Label Ratio 1.0		Label Ratio 0.5		Label Ratio 0.2	
	MSE (m^2) (↓)	Dist. Err. (m) (↓)	MSE (m^2) (↓)	Dist. Err. (m) (↓)	MSE (m^2) (↓)	Dist. Err. (m) (↓)
CMC	51.11 ± 14.67	6.76 ± 0.75	71.81 ± 15.32	7.99 ± 0.64	111.37 ± 8.02	11.05 ± 0.57
Cosmo	38.40 ± 4.14	6.03 ± 0.21	53.12 ± 9.75	7.19 ± 0.40	97.08 ± 9.49	10.95 ± 0.57
SimCLR	34.40 ± 4.47	5.64 ± 0.25	45.14 ± 7.34	6.57 ± 0.08	74.53 ± 3.13	9.48 ± 0.17
AudioMAE	22.36 ± 0.49	5.40 ± 0.11	30.12 ± 2.97	6.33 ± 0.28	41.75 ± 3.30	7.47 ± 0.28
CAV-MAE	18.85 ± 0.41	5.06 ± 0.04	22.90 ± 0.82	5.58 ± 0.12	24.84 ± 0.33	5.78 ± 0.10
FOCAL	32.43 ± 4.68	5.37 ± 0.22	40.84 ± 2.82	6.20 ± 0.19	69.62 ± 5.62	8.50 ± 0.35
FreqMAE	29.61 ± 2.85	5.36 ± 0.16	42.06 ± 14.44	6.25 ± 0.70	91.40 ± 35.32	9.15 ± 1.27
PhyMask	28.02 ± 5.91	5.29 ± 0.33	33.74 ± 2.18	5.85 ± 0.12	64.36 ± 4.70	8.44 ± 0.36
SPAR	**12.98 ± 0.11**	**4.20 ± 0.07**	**15.07 ± 1.03**	**4.51 ± 0.09**	**21.36 ± 0.62**	**5.40 ± 0.04**

Learning the Signal-Position Duality Improves Performance of Spatial Analysis

Moving vehicle localization experiments using self-supervised foundation models. Localization is based on 2-second seismic and acoustic data snippets collected from six nearby sensors.

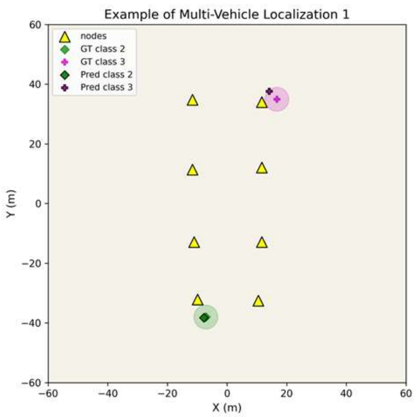
Method	M3N-VC Multi-vehicle Joint Classification and Localization			
	mAP@4m (%) (↑)	mAP@6m (%) (↑)	mAP@8m (%) (↑)	mAP@10m (%) (↑)
CMC	0.06 ± 0.05	0.48 ± 0.36	1.61 ± 1.10	3.62 ± 2.19
Cosmo	0.16 ± 0.05	1.66 ± 0.23	4.77 ± 0.72	9.52 ± 1.20
SimCLR	0.31 ± 0.14	2.22 ± 0.58	6.53 ± 1.24	13.07 ± 2.08
AudioMAE	1.39 ± 0.48	6.96 ± 1.42	17.11 ± 3.24	28.98 ± 4.01
CAV-MAE	22.12 ± 2.94	52.08 ± 4.16	73.41 ± 3.24	85.36 ± 1.78
FOCAL	0.08 ± 0.05	0.82 ± 0.40	2.94 ± 1.04	6.82 ± 1.99
FreqMAE	0.24 ± 0.01	1.67 ± 0.32	5.34 ± 0.99	11.31 ± 1.49
PhyMask	0.08 ± 0.03	0.88 ± 0.24	3.04 ± 0.74	6.64 ± 1.46
SPAR	**41.57** $\pm$ **2.69**	**71.82** $\pm$ **3.69**	**86.28** $\pm$ **1.77**	**92.99** $\pm$ **0.79**

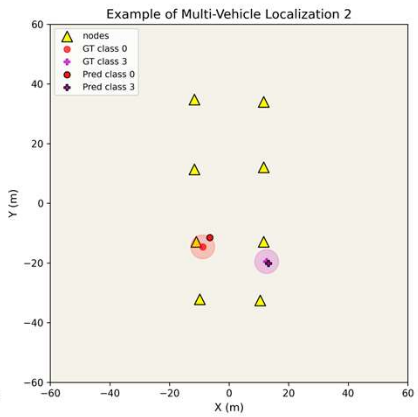
Visualization of Single-Vehicle Localization

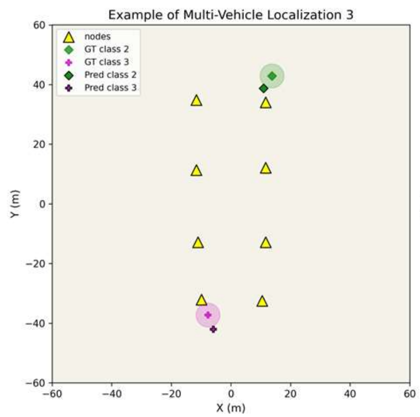


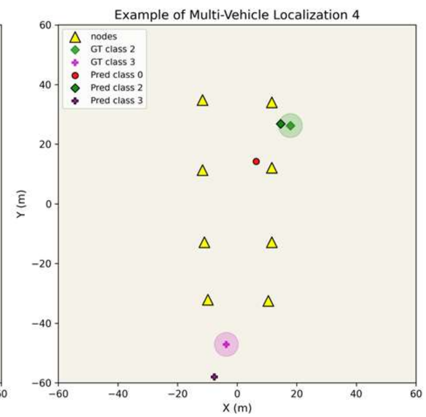
Example: Multi-Vehicle Joint Classification and Localization

Example shows that the approach is able to jointly localize and classify multiple vehicle types that are simultaneously present in the field









[7] Tianchen Wang, Yizhuo Chen, Hongjue Zhao, You Lyu, Jinyang Li, Tomoyoshi Kimura, Yigong Hu, Denizhan Kara, Maggie Wigness, Jeffrey Twigg, Tarek Abdelzaher, "On Network-Efficient Multimodal Multi-Vantage Foundation Models for Distributed Sensing," In Proc. 22nd IEEE International Conference on Mobile Ad-Hoc and Smart Systems (IEEE MASS), Chicago, IL, October 2025.

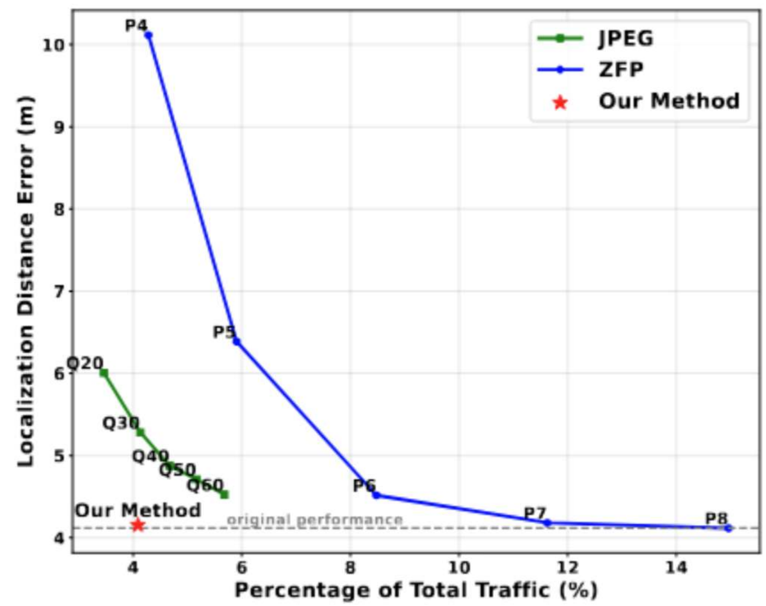
Adaptive Lossy Compression of Sensor Data for Multi-vantage Foundation Model Inference

Key observation:

- The encoder projects input sensor data into a latent representation (an embedding) that extracts useful higher-level semantics.
- Lossy data compression changes the projection of data embedding in the latent space. More loss implies bigger changes.

Approach: Minimize the maximum change in data embedding location across different sensor streams

Implication: Sensor streams not relevant to the higher-level semantics can be compressed more (because they contribute less to the latent semantic representation).



[8] Jinyang Li, Yizhuo Chen, Ruijie Wang, Tomoyoshi Kimura, Tianshi Wang, You Lyu, Hongjue Zhao, Binqi Sun, Shangchen Wu, Yigong Hu, Denizhan Kara, Beitong Tian, Klara Nahrstedt, Suhas Diggavi, Jae H Kim, Greg Kimberly, Guijun Wang, Maggie Wigness, Tarek Abdelzaher, "RestoreML: Practical Unsupervised Tuning of Deployed Intelligent IoT Systems," In Proc. DCoSS-IoT, Lucca, Italy, **June 2025**.
Best Paper Award.

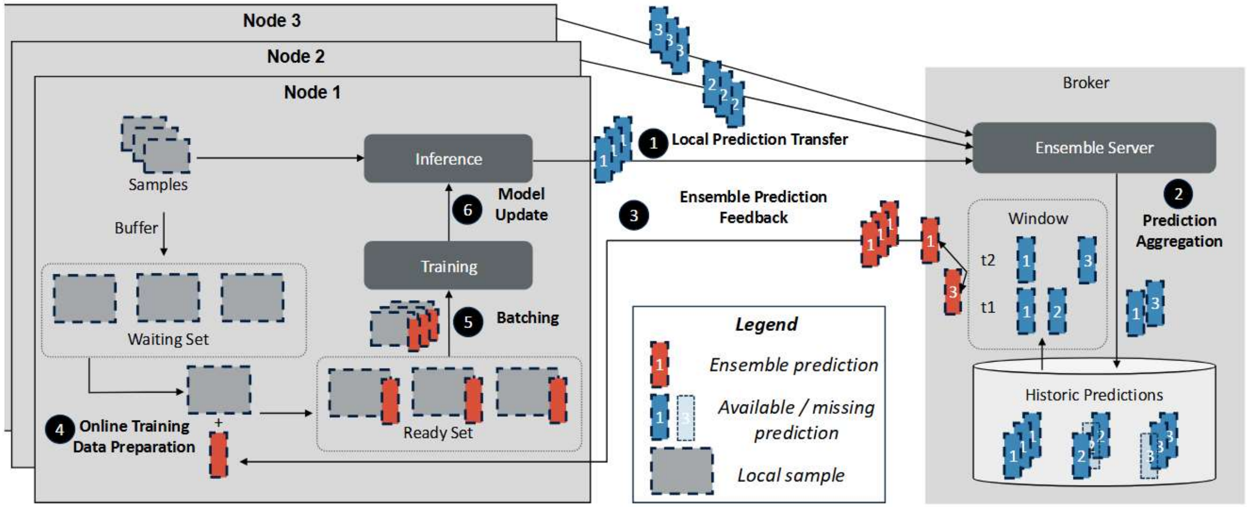
Fine-tuning Nodes in the Field

Can a group of heterogeneous deployed sensors jointly improve each other's AI models in the field without using labeled data?

Yes → jointly vote on (i.e., guess) labels of new samples, then use those guesses as soft labels for fine-tuning…

BUT…

Must be careful how to batch new samples to avoid catastrophic forgetting and avoid bias to specific classes



[8] Jinyang Li, Yizhuo Chen, Ruijie Wang, Tomoyoshi Kimura, Tianshi Wang, You Lyu, Hongjue Zhao, Binqi Sun, Shangchen Wu, Yigong Hu, Denizhan Kara, Beitong Tian, Klara Nahrstedt, Suhas Diggavi, Jae H Kim, Greg Kimberly, Guijun Wang, Maggie Wigness, Tarek Abdelzaher, "RestoreML: Practical Unsupervised Tuning of Deployed Intelligent IoT Systems," In Proc. DCoSS-IoT, Lucca, Italy, **June 2025**.
Best Paper Award.

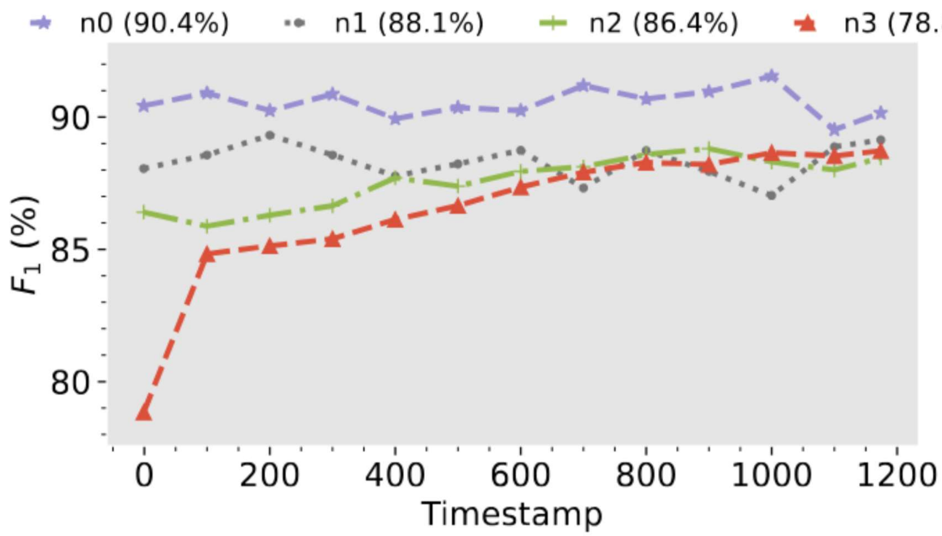
Fine-tuning Nodes in the Field

Can a group of heterogeneous deployed sensors jointly improve each other's AI models in the field without using labeled data?

Yes → jointly vote on (i.e., guess) labels of new samples, then use those guesses as soft labels for fine-tuning…

BUT…

Must be careful how to batch new samples to avoid catastrophic forgetting and avoid bias to specific classes



Note: Accuracy of new (red) node improves without degrading accuracy of other nodes

Foundation Models for Battlefield Data

Contribution #4:

Efficiency Extensions

[9] Ashitabh Misra, Nurani Saoda, Tarek Abdelzaher, "Latency-Constrained Input-Aware Quantization of Time Series Inference Workflows at the Edge," In Proc. IEEE Conference on Computer Communications (Infocom), London, UK, **May 2025**.

Input-Aware Model Quantization

Can we train an adaptive quantization scheme to adaptively quantize the AI model at run-time in an input-aware fashion to reduce the fidelity of less important inputs while retaining the fidelity of more important ones?

Architecture	Method	Bit-widths known during training ($\alpha \pm \sigma$)				Bit-widths not known during training ($\alpha \pm \sigma$)			
		$\sim$ 6 bits	$\sim$ 8 bits	$\sim$ 14 bits	$\sim$ 26 bits	$\sim$ 5 bits	$\sim$ 7 bits	$\sim$ 13 bits	$\sim$ 23 bits
FCN	Bit-Mixer	54.2 ± 17.6	62.1 ± 11.73	70.8 ± 7.2	74.2 ± 4.3	43.9 ± 11.5	58.2 ± 10.2	64.2 ± 7.3	65.3 ± 8.4
	CoQuant	64.4 ± 11.6	71.2 ± 4.6	70.3 ± 6.3	77.1 ± 3.3	62.6 ± 5.7	66.3 ± 3.4	71.3 ± 2.9	72.2 ± 2.8
	RobustQuant	64.6 ± 14.7	68.2 ± 6.4	69.3 ± 6.7	78.1 ± 6.2	59.4 ± 4.2	61.4 ± 3.1	68.3 ± 4.1	71.4 ± 3.1
	DQNet	71.2 ± 7.6	72.5 ± 4.5	73.4 ± 5.2	77.2 ± 5.5	68.6 ± 6.1	69.4 ± 4.4	71.3 ± 3.2	73.1 ± 2.1
	ReactQuant	$\mathbf{73.4 \pm 4.5}$	$\mathbf{74.4 \pm 3.1}$	$\mathbf{76.8 \pm 4.6}$	$\mathbf{78.8 \pm 3.3}$	$\mathbf{69.1 \pm 4.1}$	$\mathbf{74.3 \pm 3.1}$	$\mathbf{72.7 \pm 4.6}$	$\mathbf{75.6 \pm 3.2}$
ResNet18	Bit-Mixer	65.9 ± 10.4	73.8 ± 7.2	74.3 ± 11.1	81.1 ± 6.7	59.2 ± 12.6	64.7 ± 11.6	69.0 ± 9.3	73.1 ± 8.1
	CoQuant	78.4 ± 12.5	79.8 ± 15.9	84.4 ± 9.8	87.5 ± 6.1	68.5 ± 8.8	76.8 ± 3.5	79.6 ± 3.1	83.2 ± 2.8
	RobustQuant	79.3 ± 9.1	83.3 ± 9.5	85.8 ± 9.1	86.5 ± 8.3	69.2 ± 9.3	72.1 ± 7.2	81.8 ± 4.7	82.7 ± 4.3
	DQNet	81.5 ± 11.3	84.6 ± 7.2	84.2 ± 6.5	87.6 ± 5.2	71.8 ± 7.3	74.1 ± 6.1	78.8 ± 5.2	84.6 ± 3.2
	ReactQuant	$\mathbf{85.2 \pm 6.4}$	$\mathbf{87.7 \pm 6.5}$	$\mathbf{87.8 \pm 5.7}$	$\mathbf{88.1 \pm 4.5}$	$\mathbf{79.2 \pm 5.3}$	$\mathbf{83.8 \pm 2.7}$	$\mathbf{83.1 \pm 3.3}$	$\mathbf{85.5 \pm 2.1}$
EBN	Bit-Mixer	54.1 ± 9.2	68.0 ± 9.1	69.8 ± 7.2	69.6 ± 6.3	57.1 ± 11.2	61.3 ± 9.2	63.3 ± 8.5	66.0 ± 6.7
	CoQuant	58.3 ± 8.2	69.2 ± 8.6	71.2 ± 5.6	74.5 ± 3.7	58.8 ± 7.2	67.0 ± 5.1	68.9 ± 2.8	73.9 ± 1.2
	RobustQuant	63.7 ± 7.1	72.3 ± 7.1	73.8 ± 4.2	75.1 ± 4.1	62.1 ± 4.8	64.1 ± 4.9	66.4 ± 5.9	73.1 ± 2.1
	DQ-Net	65.3 ± 7.2	71.3 ± 5.3	74.1 ± 5.51	74.6 ± 3.9	64.1 ± 3.2	70.3 ± 4.2	70.1 ± 3.5	72.7 ± 3.7
	ReactQuant	$\mathbf{71.7 \pm 4.5}$	$\mathbf{76.4 \pm 5.2}$	$\mathbf{75.2 \pm 3.7}$	$\mathbf{76.1 \pm 5.1}$	$\mathbf{66.7 \pm 2.5}$	$\mathbf{71.5 \pm 3.2}$	$\mathbf{72.2 \pm 4.1}$	$\mathbf{75.1 \pm 3.6}$

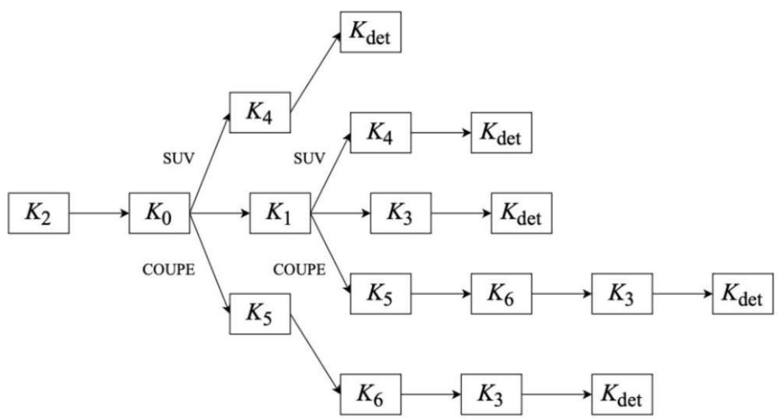
[10] Tarek Abdelzaher, Sanjoy Baruah, Alan Burns, Yigong Hu, "Timely Classification of Hierarchical Classes," In Proc. 46th IEEE Real-Time Systems Symposium (IEEE RTSS) , Boston, MA, **December 2025**.

Latency-Optimal Hierarchical (Mixture of Experts) Classification

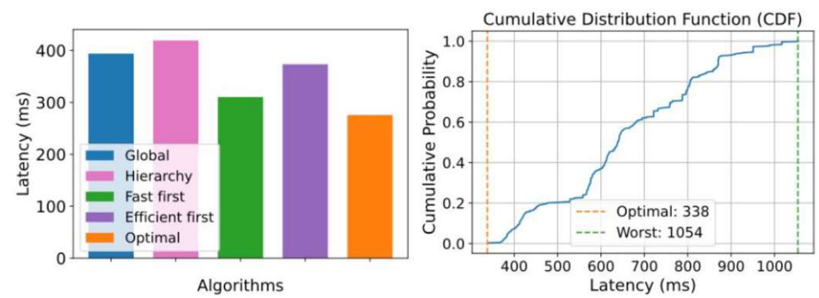
Given a choice of experts (classifiers), in what order to try them to minimize time to successful classification?

Classifiers	Intermediate		Global		SUV	COUPE		
	K_0	K_1	K_2	K_3	K_4	K_5	K_6	K_{det}
Modality, params	Both, 129698	Both, 356610	Both, 130469	Both, 1217109	Acoustic, 80355	Acoustic, 80355	Both, 129955	-
Success rate	76.1	87.6	66.8	99.9	1	91.1	94.9	1
Execution time	80.8	317.0	104.7	869.9	80.9	80.9	104.5	10000

Classifiers for the vehicle detection case study (Section V-A).



Optimal cascade for the vehicle detection case Study



(a) Latency comparison of the algorithms. (b) Latency distribution of all cascades.

Vehicle detection case study results

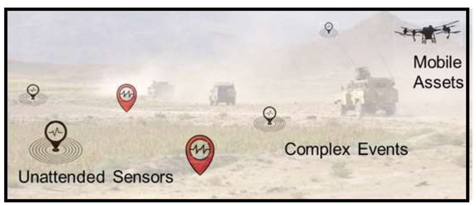
Foundation Models for Battlefield Data

Other Topics

Background: Foundation Models – The Principle of Operation

Unlabeled Data

Self-Supervised Pretraining



"Encoder" (Neural Network)

Encoding into a semantically well-organized latent space

Supported downstream Tasks

Target/formation detection

Classification/Identification

Localization/Tracking

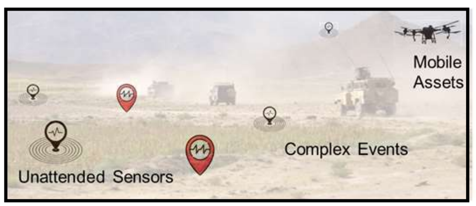
Activity Recognition, Intent Prediction

Finetuning

Background: Foundation Models – The Principle of Operation ++

Unlabeled Data

Self-Supervised Pretraining



"Encoder" (Neural Network)

Encoding into a semantically well-organized latent space

Supported downstream Tasks

- Target/formation detection
- Classification/Identification
- Localization/Tracking
- Activity Recognition, Intent Prediction

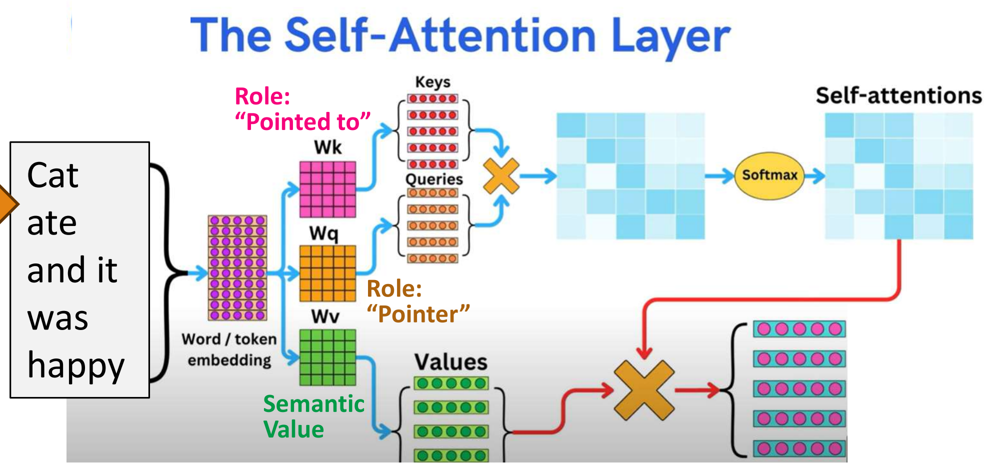
Finetuning

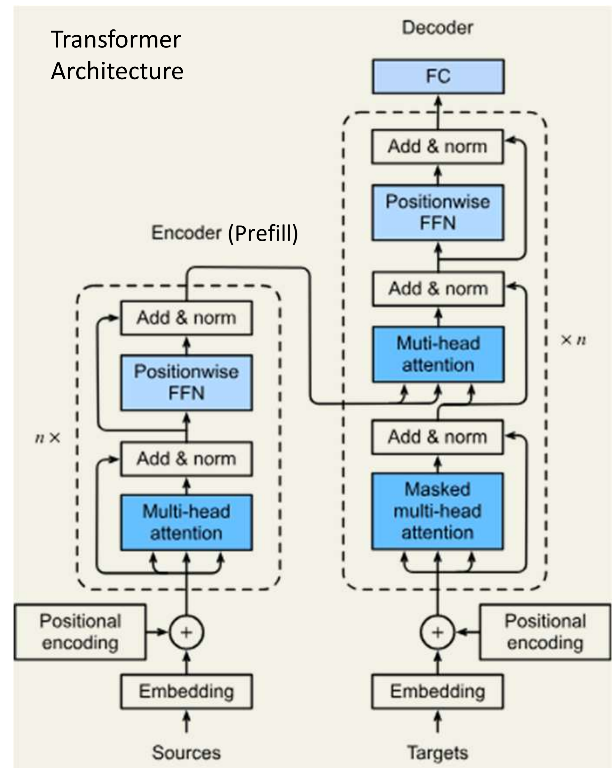
++*How to interpret elements of data sequences in proper context? (The interpretation of each element might depend on what happened earlier)*

Sequences, Attention, and Transformers

A key to encoding the inputs properly is to *interpret them in context*.

Attention matrix: what other tokens should I consider in interpreting the current token?

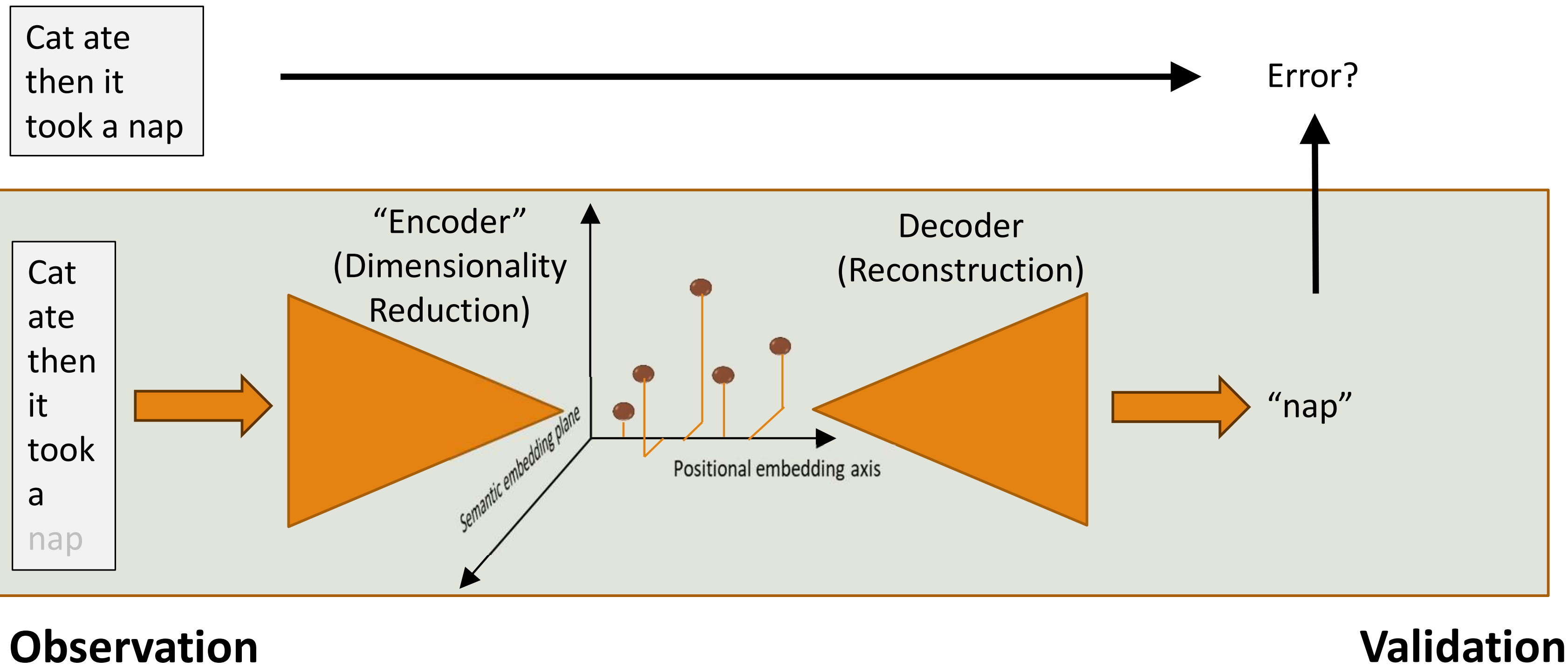




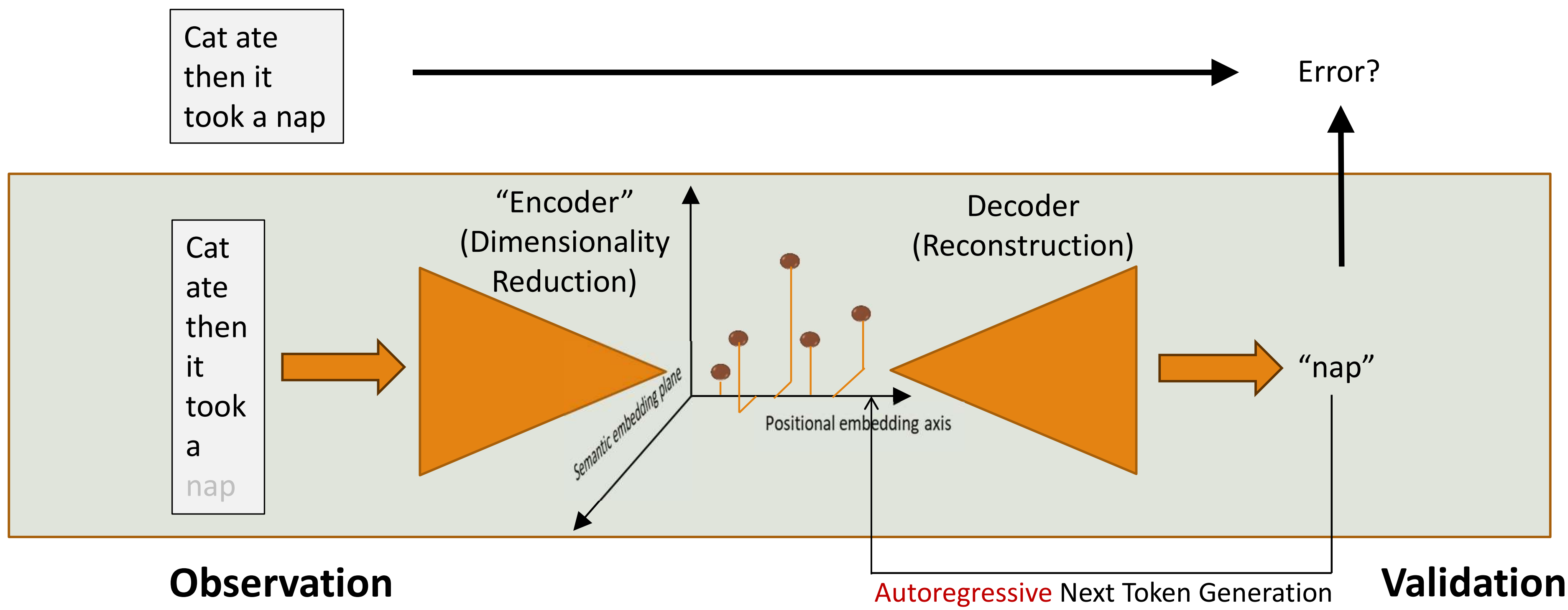
The KV cache holds the embeddings of recently/frequently used tokens

Common Self-Supervised Pretraining Approaches

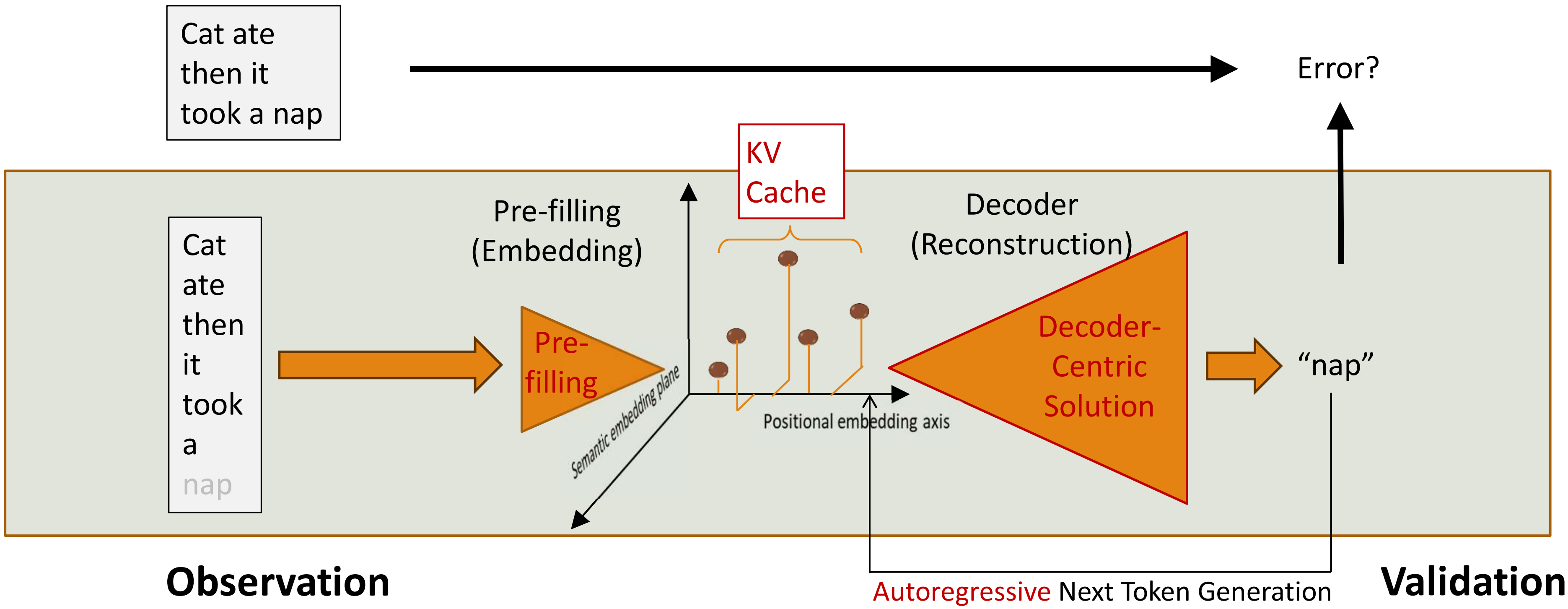
Next Token Prediction



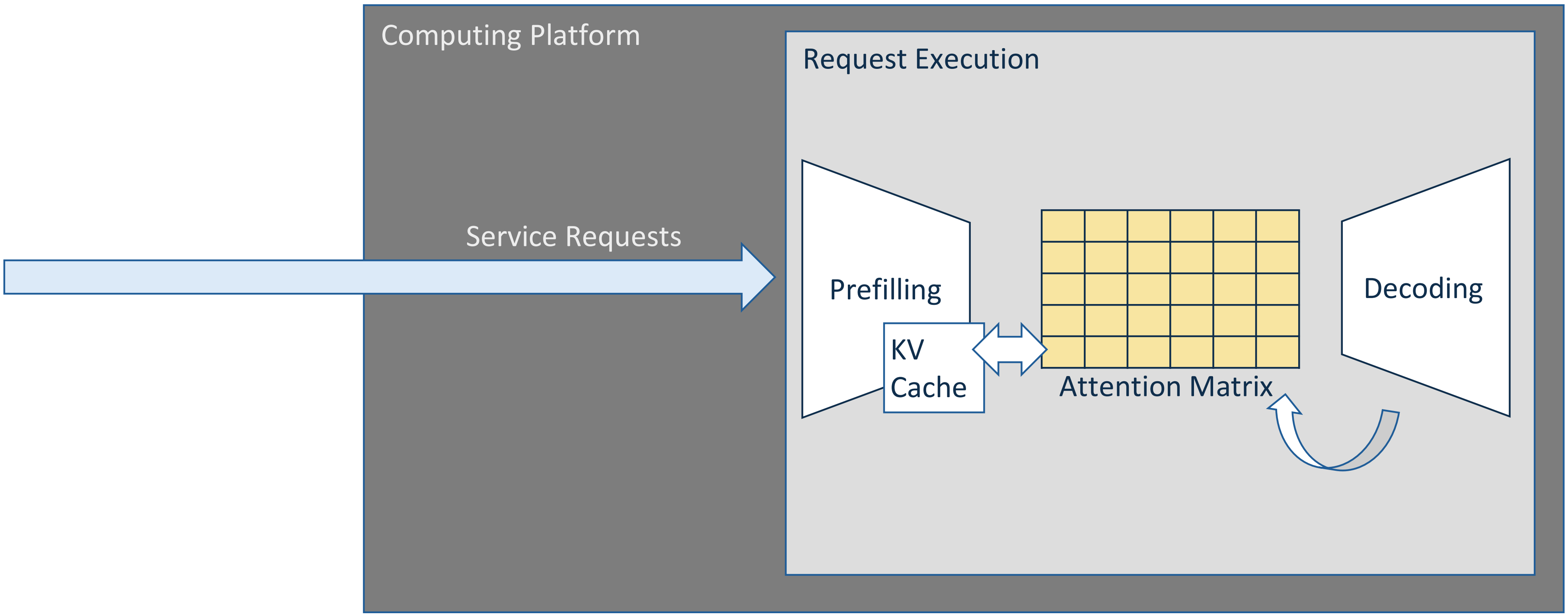
Common Self-Supervised Pretraining Approaches
Next Token Prediction



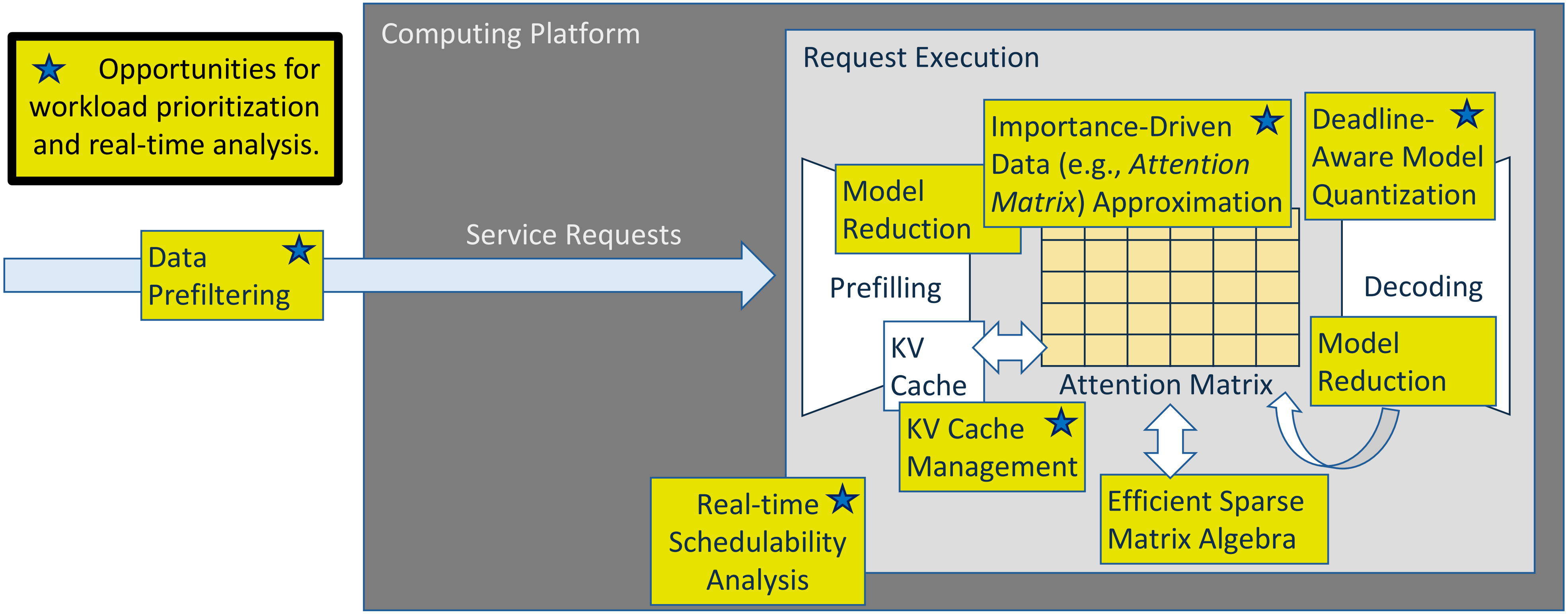
Common Self-Supervised Pretraining Approaches
Next Token Prediction



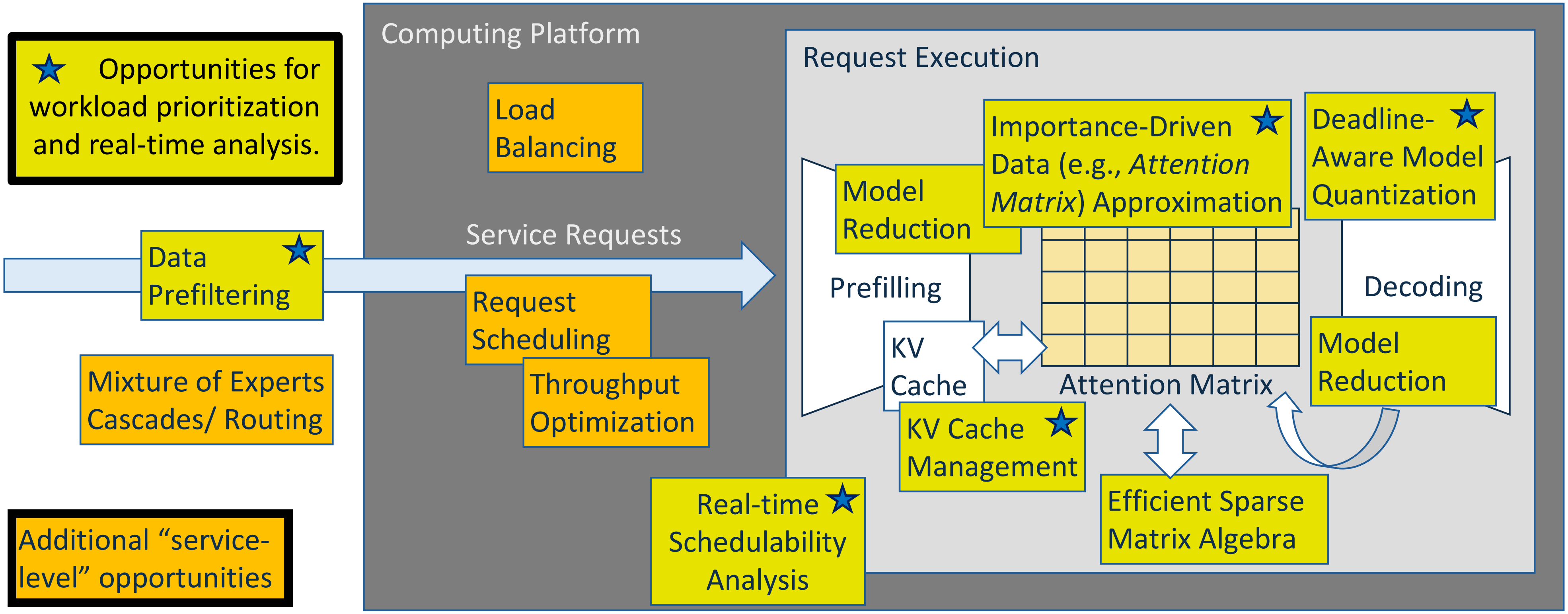
An Abstraction of Modern Foundation Model Execution



An Abstraction of Modern Foundation Model Execution



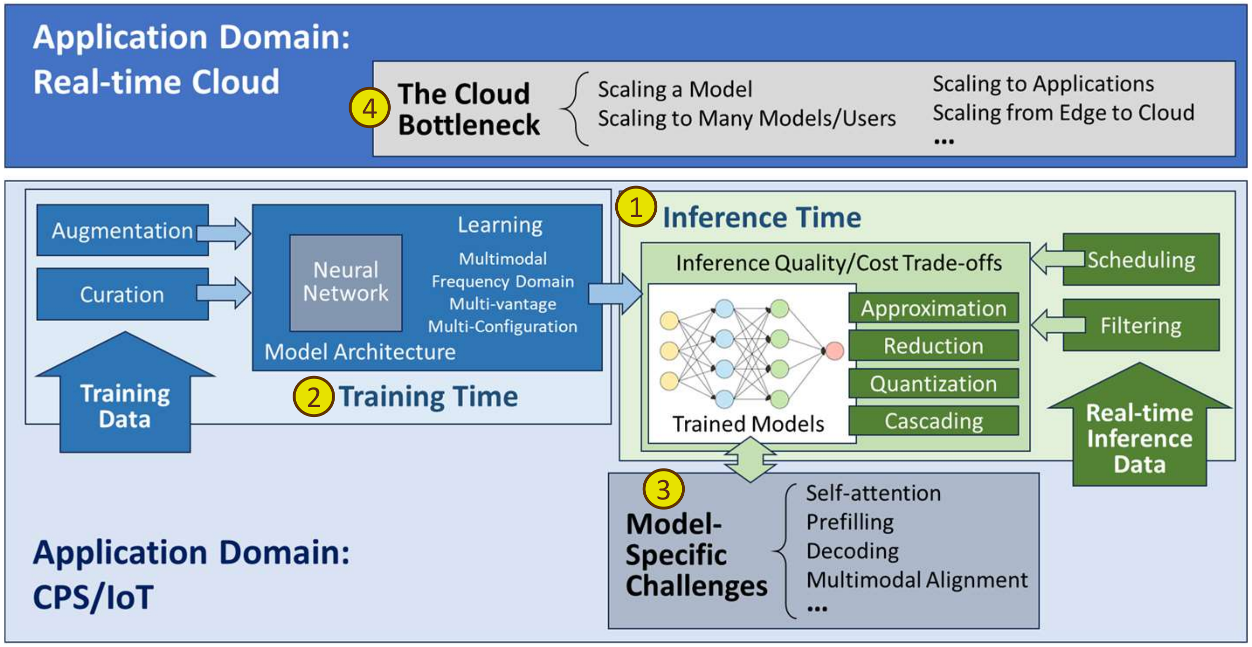
An Abstraction of Modern Foundation Model Execution (as a Service)



[11] T. Abdelzaher, Y. Hu, D. Kara, T. Kimura, A. Misra, V. Ramani, O. Tardieu, T. Wang, M. Wigness, A. Youssef, "The Bottlenecks of AI: Challenges for Embedded and Real-Time Research in a Data-Centric Age," Accepted to Special Issue of the Journal of Real-Time Systems, Vol. 61, Issue 2, **June 2025**

For More on the Foundation Model Bottlenecks Topic

T. Abdelzaher, Y. Hu, D. Kara, T. Kimura, A. Misra, V. Ramani, O. Tardieu, T. Wang, M. Wigness, A. Youssef, "The Bottlenecks of AI: Challenges for Embedded and Real-Time Research in a Data-Centric Age," Accepted to Special Issue of the Journal of Real-Time Systems, Vol. 61, Issue 2, June 2025



https://link.springer.com/article/10.1007/s11241-025-09452-w

A Foundation Model for Multivantage Sensing

Summary

- **Key Insight: We enhanced the state of the art on foundation model training and inference for CPS-IoT applications in four respects:**
    - Improved learning and inference from multimodal sensor data
    - Exploitation of frequency domain characteristics of physical signals
    - Extentions of foundation model training to multi-vantage sources
    - Improved efficiency of AI models

- **Next Steps:**
    - A middleware to support edge AI
    - Extend to embodied AI applications
    - Improve robustness of edge AI data-to-decision workflows to domain shift, re-tasking, reconfiguration, etc.

Publications Covered in this Talk (The Last 12 Months):

1. Chenzhi Hu, Yatong Chen, Denizhan Kara, Shengzhong Liu, Tarek Abdelzaher, Fan Wu, Guihai Chen, "OpenMAE: Efficient Masked Autoencoder for Vibration Sensing with Open-Domain Data Enrichment," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (ACM IMWUT), also presented in UbiComp, Espoo, Finland, **October 2025**.
2. Tianchen Wang, Yizhuo Chen, Hongjue Zhao, You Lyu, Jinyang Li, Tomoyoshi Kimura, Yigong Hu, Denizhan Kara, Maggie Wigness, Jeffrey Twigg, Tarek Abdelzaher, "On Network-Efficient Multimodal Multi-Vantage Foundation Models for Distributed Sensing," In Proc. 22nd IEEE International Conference on Mobile Ad-Hoc and Smart Systems (IEEE MASS), Chicago, IL, **October 2025**.
3. Denizhan Kara, Tomoyoshi Kimura, Dachun Sun, Jinyang Li, Yizhuo Chen, Yigong Hu, Hongjue Zhao, Joydeep Bhattacharyya and Tarek Abdelzaher, "DiffPhys: Differential Physics Augmentations for Enhanced Representations," In Proc. *34th International Conference on Computer Communications and Networks (ICCCN)*, Tokyo, Japan, **August 2025**.
4. T. Abdelzaher, Y. Hu, D. Kara, T. Kimura, A. Misra, V. Ramani, O. Tardieu, T. Wang, M. Wigness, A. Youssef, "The Bottlenecks of AI: Challenges for Embedded and Real-Time Research in a Data-Centric Age," Special Issue of the *Journal of Real-Time Systems*, Vol. 61, Issue 2, **July 2025**
5. Jinyang Li, Yizhuo Chen, Ruijie Wang, Tomoyoshi Kimura, Tianshi Wang, You Lyu, Hongjue Zhao, Binqi Sun, Shangchen Wu, Yigong Hu, Denizhan Kara, Beitong Tian, Klara Nahrstedt, Suhas Diggavi, Jae H Kim, Greg Kimberly, Guijun Wang, Maggie Wigness, Tarek Abdelzaher, "RestoreML: Practical Unsupervised Tuning of Deployed Intelligent IoT Systems," In Proc. DCoSS-IoT, Lucca, Italy, **June 2025**. **(Best Paper Award)**
6. Y Chen, T Wang, Y Lyu, Y Hu, J Li, T Kimura, H Zhao, Y Hu, D Kara, T. Abdelzaher, "SPAR: Self-supervised Placement-Aware Representation Learning for Multi-Node IoT Systems," arXiv preprint arXiv:2505.16936, **May 2025**.
7. Ashitabh Misra, Nurani Saoda, Tarek Abdelzaher, "Latency-Constrained Input-Aware Quantization of Time Series Inference Workflows at the Edge," In Proc. IEEE Conference on Computer Communications (Infocom), London, UK, **May 2025**.
8. Tomoyoshi Kimura, Xinlin Li, Osama Hanna, Yatong Chen, Yizhuo Chen, Denizhan Kara, Tianshi Wang, Jinyang Li, Xiaomin OUYANG, Shengzhong Liu, Mani Srivastava, Suhas Diggavi, Tarek F. Abdelzaher, "InfoMAE: Pairing-Efficient Cross-Modal Alignment with Informational Masked Autoencoders for IoT Signals," In Proc. *ACM TheWebConference (WWW)*, Sydney, Australia, **April 2025**.
9. Tomoyoshi Kimura, Yizhuo Chen, Denizhan Kara, Jinyang Li, Tianshi Wang, Ruijie Wang, Joydeep Bhattacharyya, Jae Kim, Prashant Shenoy, Mani Srivastava, Maggie Wigness, Tarek Abdelzaher, "The Case for Micro Foundation Models to Support Robust Edge Intelligence," In Proc. *10th IEEE International Conference on Collaboration and Internet Computing (IEEE CIC)*, Washington, DC, **November 2024**.
10. Denizhan Kara, Tomoyoshi Kimura, Yatong Chen, Jinyang Li, Ruijie Wang, Yizhuo Chen, Tianshi Wang, Shengzhong Liu, Lance Kaplan, Joydeep Bhattacharyya, Tarek Abdelzaher, "PhyMask: An Adaptive Masking Paradigm for Efficient Self-Supervised Learning in IoT," In Proc. *22nd ACM Conference on Embedded Networked Sensor Systems (SenSys)*, Hangzhou, China, **November 2024**.
11. Tomoyoshi Kimura, Jinyang Li, Tianshi Wang, Yizhuo Chen, Ruijie Wang, Denizhan Kara, Maggie Wigness, Joydeep Bhattacharyya, Mudhakar Srivatsa, Shengzhong Liu, Mani Srivastava, Suhas Diggavi, Tarek Abdelzaher, "VibroFM: Towards Micro Foundation Models for Robust Multimodal IoT Sensing," In Proc. *21st IEEE International Conference on Mobile Ad-Hoc and Smart Systems (MASS)*, Seoul, South Korea, **September 2024**.