VTC2025-Fall
CHENGDU
Connecting the Mobile World

BeamLLM: Vision-Empowered mmWave Beam Prediction with Large Language Models

Can Zheng, Korea University
Jiguang He, Great Bay University
Guofa Cai, Guangdong University of Technology
Zitong Yu, Great Bay University
Chung G. Kang, Korea University

Oct 22, 2025

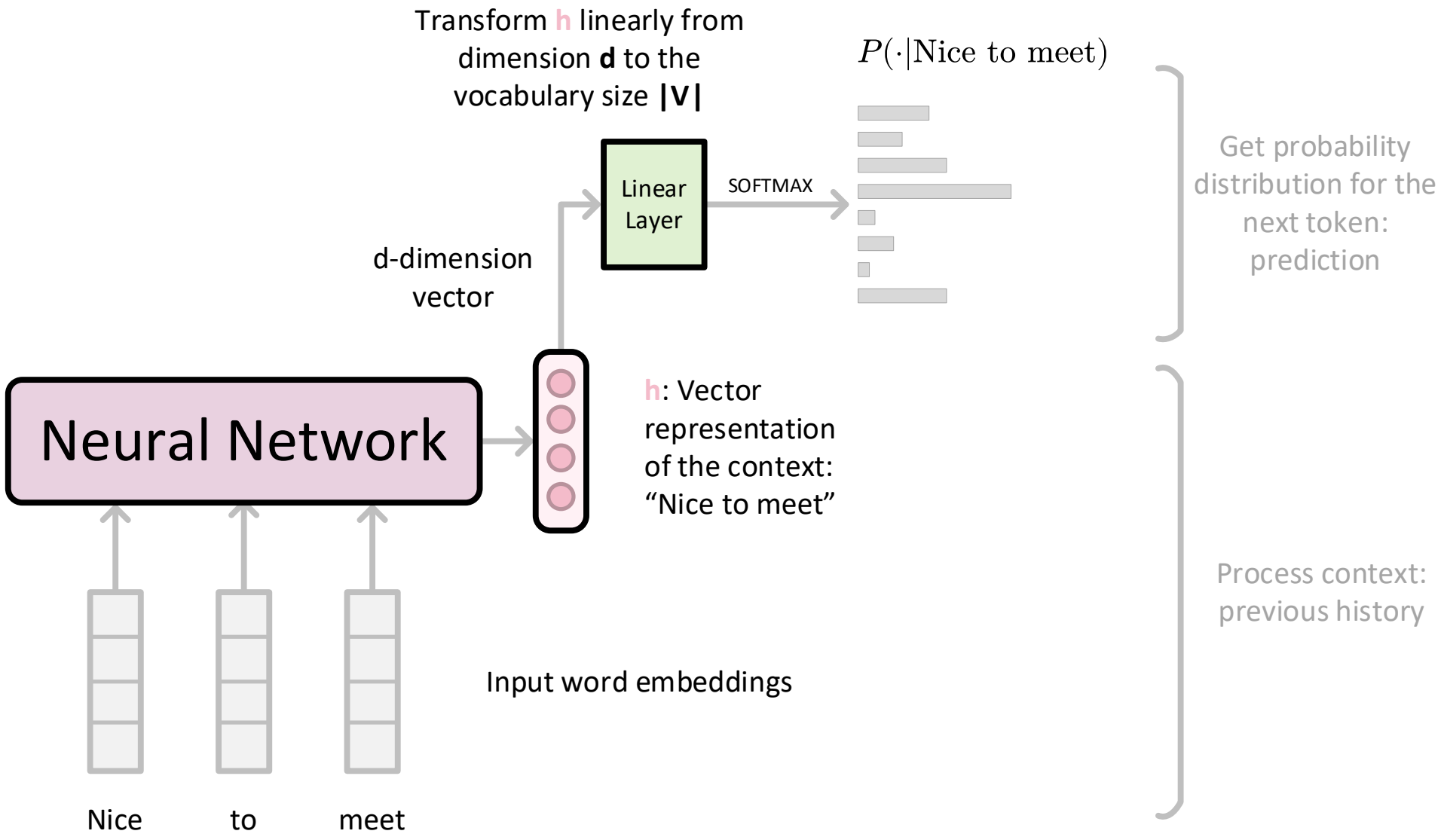
Outline

1. Introduction
2. System Model
3. Proposed BeamLLM Structure
3. Simulation Results
4. Conclusion

Introduction

- ✓ **Preliminary**
- ✓ **Related Works**
- ✓ **Motivation**

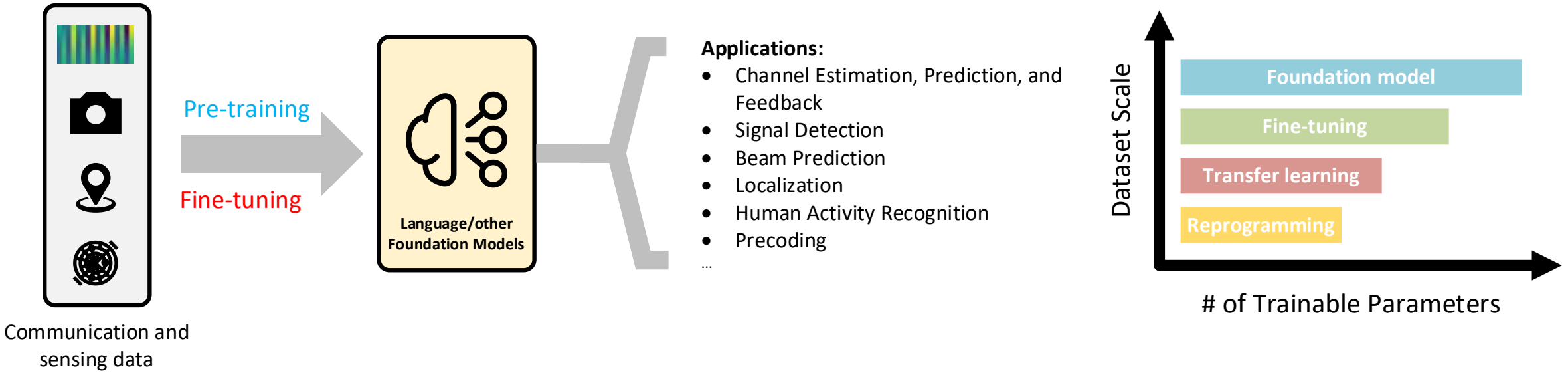
Preliminary: Introduction to Language Modeling



Inspired by: https://lena-voita.github.io/nlp_course/language_modeling.html

Background: LLM/LAM4PHY

Q1: How PHY tasks benefits from the recent advances in LLMs?



Q2: How can we let the LLMs to incorporate the modality of the various PHY layer tasks?

- **Foundation model:** Task-agnostic pre-training (on large-scale datasets)
- **Fine-tuning:** Train a specific model from the foundation model by minimizing the task-specific loss
- **Transfer learning:** In-domain knowledge transfer
- **Reprogramming:** Only requires training the inserted input transformation and output mapping layers

Jiajia Guo, Yiming Cui, Shi Jin, and Jun Zhang. "Large AI Models for Wireless Physical Layer." *arXiv preprint arXiv:2508.02314* (2025).

Motivation

Challenges

- **Millimeter-wave (mmWave) communication:** High data rates, but severe path loss; requires narrow beams for high gain; beam alignment.
- **Traditional beam management:** High overhead in high-mobility scenarios (e.g., V2X, UAVs).

Ideas

- **Sensing-assisted beam prediction:** Proactively predicts beam direction using sensory data (such as RGB images captured by base station cameras). Images provide information about the user's spatial location and surroundings, reducing reliance on beam training.
- **Pre-trained large model:** This system uses deep learning to learn the complex spatiotemporal relationship between visual input and future beam direction. This system leverages the generalization and predictive performance of the pre-trained large model to achieve stable beam prediction.

System Model

- ✓ Signal Model
- ✓ Problem Formulation

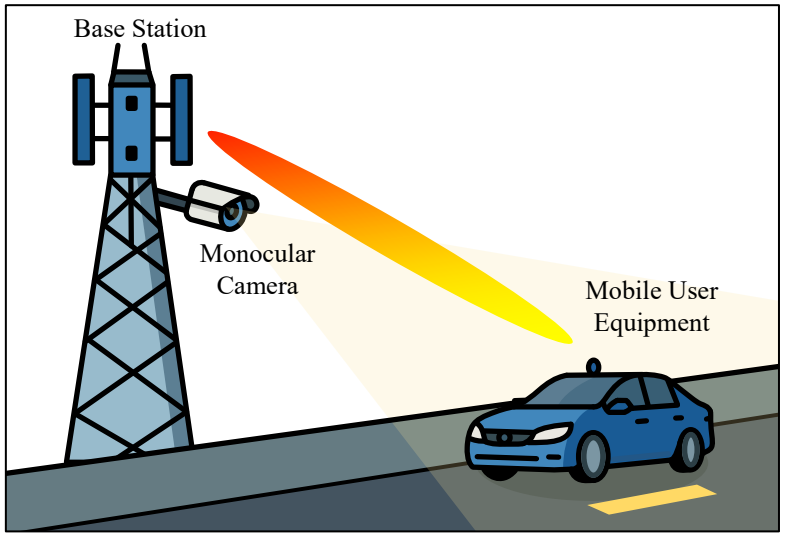
System Model

Signal Model

- Setup: BS with mmWave phased-array (N elements) and RGB camera; UE with single antenna.
- Beamforming Codebook: $\boldsymbol{F} = \{\boldsymbol{f}_1, \ldots, \boldsymbol{f}_M\}$ (M beams).
- Received Signal:

$$y[t] = \boldsymbol{h}^{\mathrm{H}}[t]\boldsymbol{f}_{m[t]}\, s[t] + n[t],$$

where $\boldsymbol{h}[t]$ is channel vector, $s[t]$ is symbol, $n[t]$ is noise.

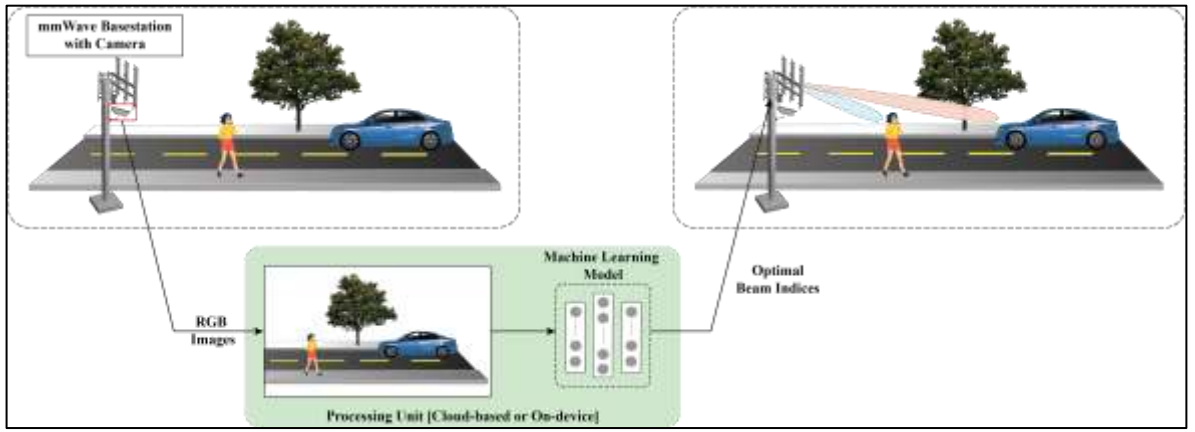


Problem Formulation

- Optimal Beam Selection:

$$\boldsymbol{f}_{m^*[t]} = \arg\max |\boldsymbol{h}^{\mathrm{H}}[t]\boldsymbol{f}_{m[t]}|^2.$$

- Goal: Predict optimal beams for $t \sim (t + T_{\mathrm{pred}} - 1)$ future steps using historical vision sensing data from $(t - T_{\mathrm{hist}}) \sim (t - 1)$.
- Why Vision? Cameras provide detailed, resource-free data; Leverage CV for enhanced functionality.

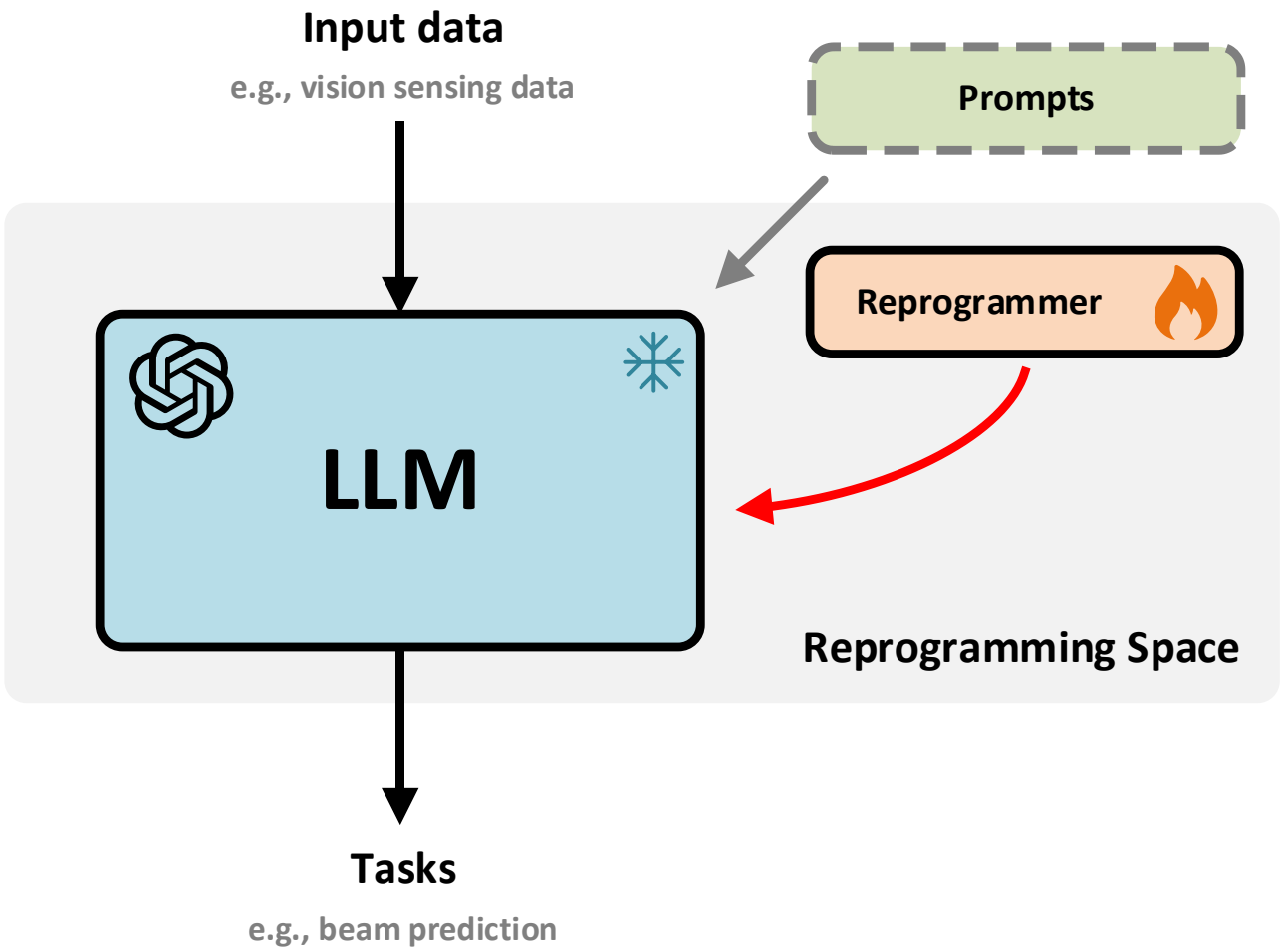


(https://www.deepsense6g.net/vision-aided-beam-prediction/)

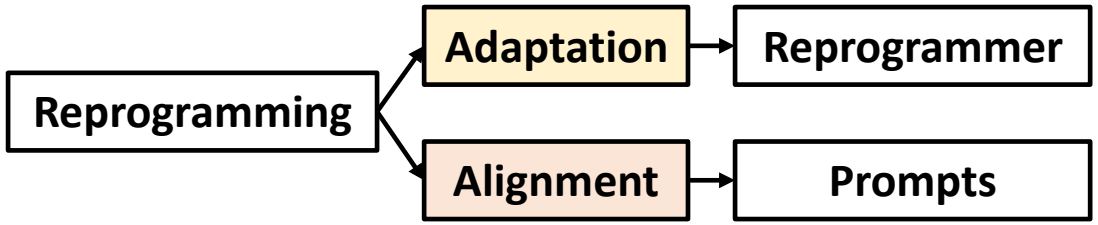
Proposed BeamLLM Structure

- ✓ **High-level Insights**
- ✓ **Proposed Structure**

High-level Insights (Backbone model: Time-LLM, ICLR 2024)

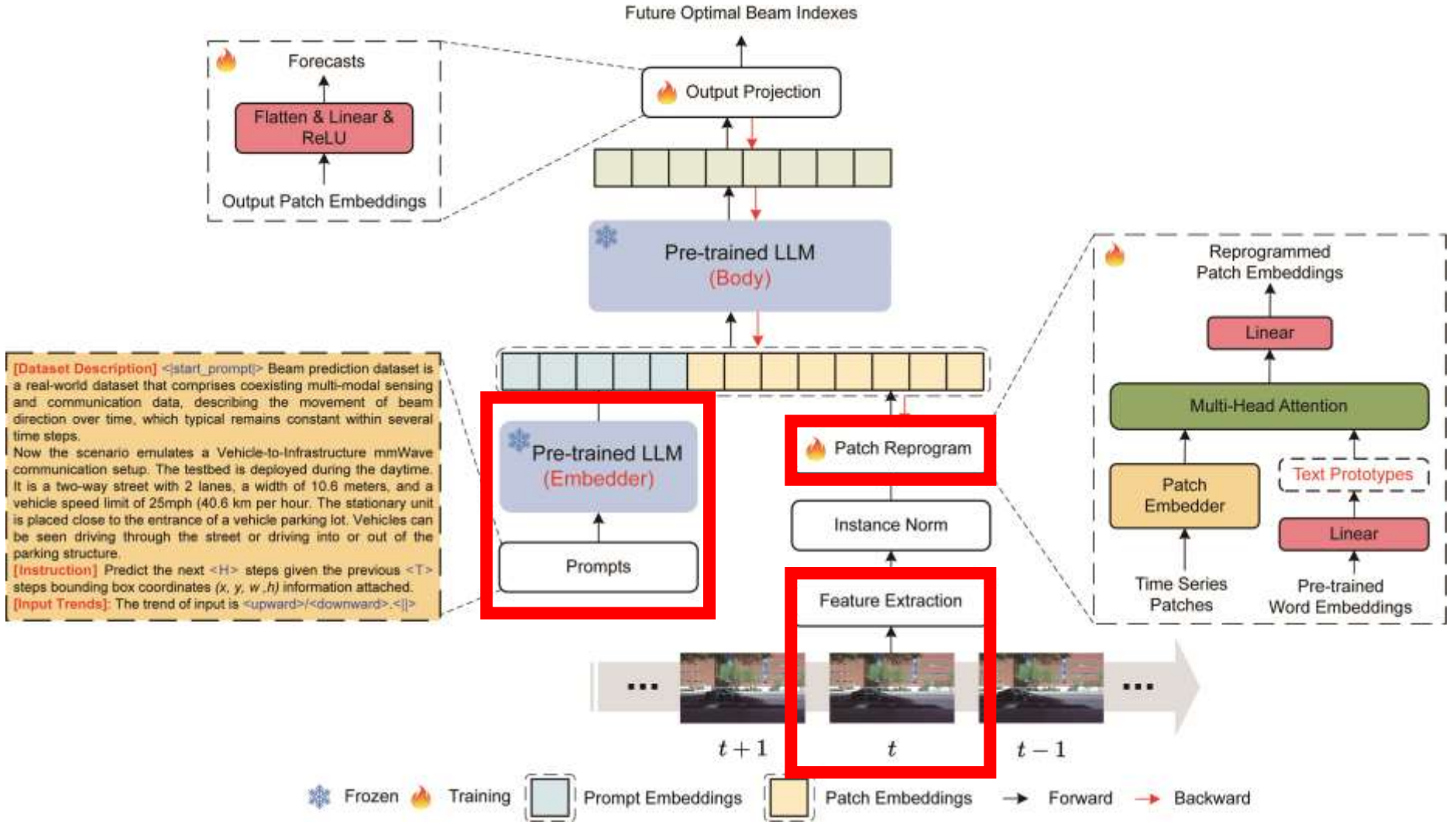


Frozen the parameters of pre-trained LLMs and train the **reprogrammer** to achieve domain transformation.



M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in Proc. International Conference on Learning Representations (ICLR), May. 2024, pp. 1–24.

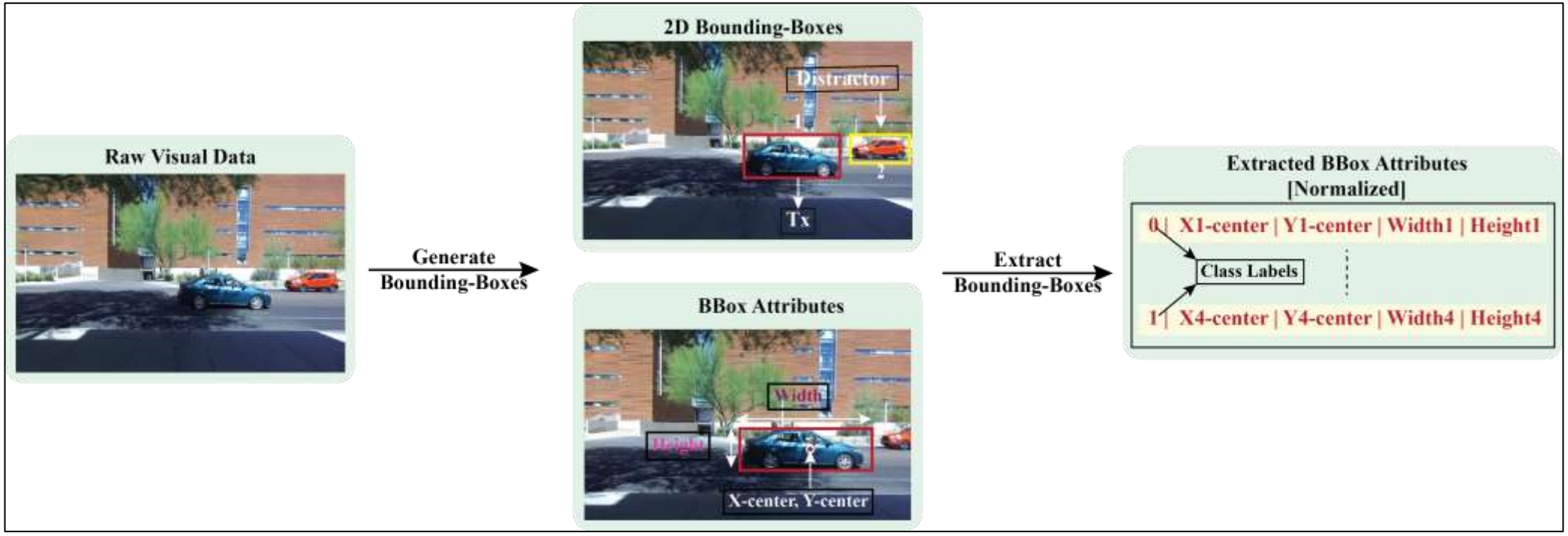
Proposed Structure: BeamLLM Overview



M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in Proc. International Conference on Learning Representations (ICLR), May. 2024.

BeamLLM – Feature Extraction (1/3)

- YOLO detects UE in each frame.
- From RGB image $\boldsymbol{X}_I$ → Bounding box $\boldsymbol{b} = [x_c, y_c, w, h]^{\mathrm{T}}$.
- Sequence: $\boldsymbol{B} = [\boldsymbol{b}[t - T_{\mathrm{hist}}], \ldots, \boldsymbol{b}[t-1]]$.
- Objective: Capture UE position/direction trend in the image.



(https://www.deepsense6g.net/scenario-8/)

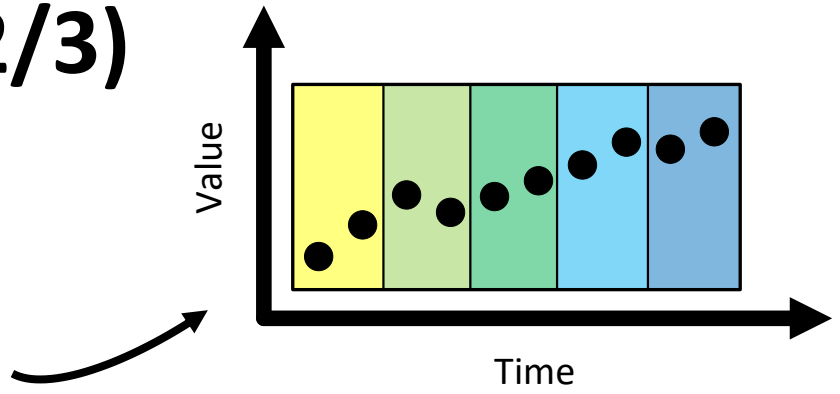
BeamLLM – Patch Reprogramming (2/3)

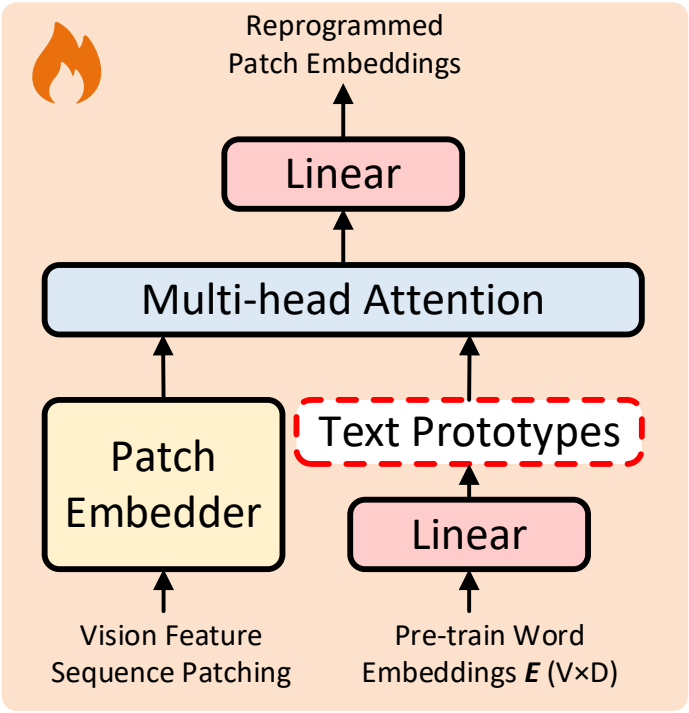
Challenge: LLM is designed for natural language, how to handle non-text visual features?

- **Patching[1]:** The input sequence is divided into multiple patches, which are shorter sequences from the original sequence, aiming to extract local semantic information from vision feature sequences
- **Input Embedding:** Segment $\boldsymbol{B}$ into patches and apply normalization
- **Patch Reprogramming:** Map visual features to natural language using cross-attention. Project word embeddings $\boldsymbol{E}$ to prototypes $\boldsymbol{E}'$. Compute $\boldsymbol{Q}$, $\boldsymbol{K}$, $\boldsymbol{V}$ for attention heads

$$\boldsymbol{Z}_k^{(i)} = \text{Attention}\left(\boldsymbol{Q}_k^{(i)}, \boldsymbol{K}_k^{(i)}, \boldsymbol{V}_k^{(i)}\right)$$

$$= \text{Softmax}\left(\frac{\boldsymbol{Q}_k^{(i)} \boldsymbol{K}_k^{(i)\mathrm{T}}}{\sqrt{d_k}}\right) \boldsymbol{V}_k^{(i)}.$$



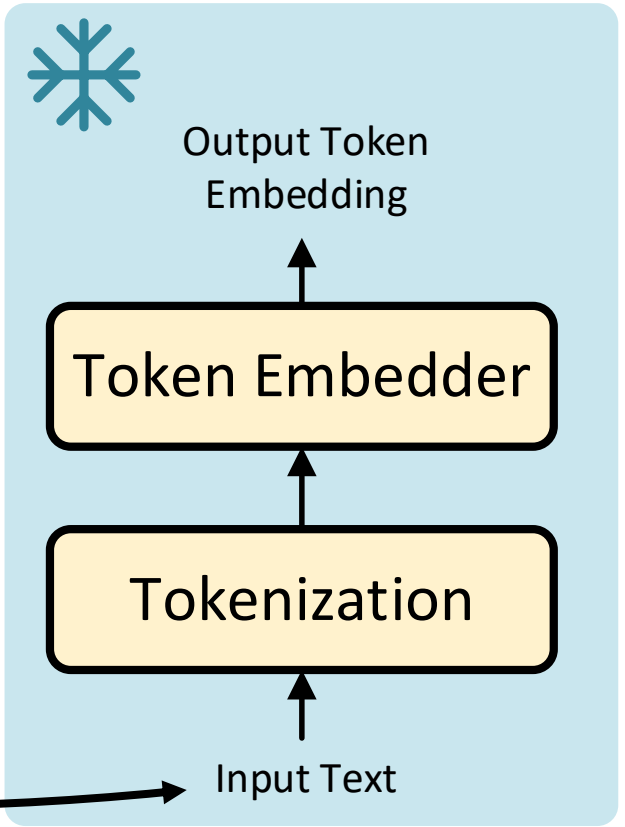


[1] Y. Q. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," in Proc. International Conference on Learning Representations (ICLR), May. 2023.
[2] M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, "Time-LLM: Time series forecasting by reprogramming large language models," in Proc. International Conference on Learning Representations (ICLR), May. 2024.

BeamLLM – Prompt-as-Prefix (PaP) (3/3)

- Natural language prompts as prefixes (dataset description, task instruction, input trends)
- In our work:

[Dataset Description] **<start prompt>** Beam prediction dataset is a real-world dataset that...
Now the scenario emulates a Vehicle-to-Infrastructure mmWave communication setup…
[Instruction] Predict the next **<T_{pred}>** steps given the previous **<T_{hist}>** steps bounding box coordinates (x, y, w ,h) information attached.
[Input Trends] The trend of input is **<upward> / <downward>. <| |>**



M. Jin, S. Wang, L. Ma, Z. Chu, J. Y. Zhang, X. Shi, P.-Y. Chen, Y. Liang, Y.-F. Li, S. Pan, and Q. Wen, “Time-LLM: Time series forecasting by reprogramming large language models,” in Proc. International Conference on Learning Representations (ICLR), May. 2024, pp. 1–24.

Simulation Results

- ✓ **Experimental Settings**
- ✓ **Dataset Visualization**
- ✓ **Simulation Results**

Simulation Results (1/2)

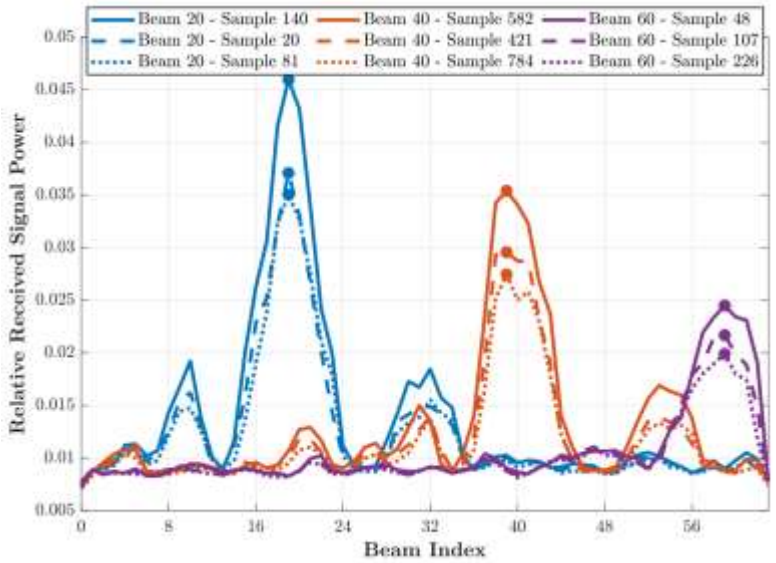
Experimental Settings

- **Dataset:** DeepSense 6G (Scenario 8).
- **Processing:** Split 70/10/20 (train/val/test); Sliding window size 13:
  - Standard: $T_{\text{hist}} = 8, T_{\text{pred}} = 5$;
  - Few-shot: $T_{\text{hist}} = 3, T_{\text{pred}} = 10$.
- **Baselines:** RNN, GRU, LSTM;
- **Parameters:** GPT-2 backbone.
- **Metrics:** Top-K Accuracy ($K = 1, 3$).

Findings:

- Beam index ↑, received power ↓
- Beam index ↑, direction →, distance ↑
- The same optimal beam: a very similar geographical location, but the received power may be different

Dataset Visualization



Beam Power

(a) Vision data visualization for the first 3 samples associated with optimal beam 20.

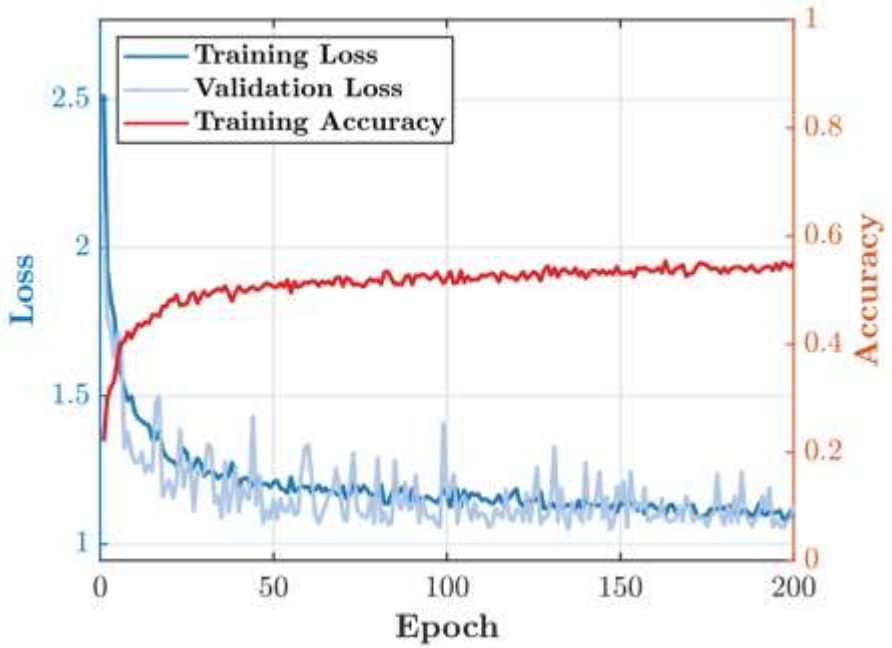
(b) Vision data visualization for the first 3 samples associated with optimal beam 40.

Vision

(c) Vision data visualization for the first 3 samples associated with optimal beam 60.

Simulation Results (2/2)

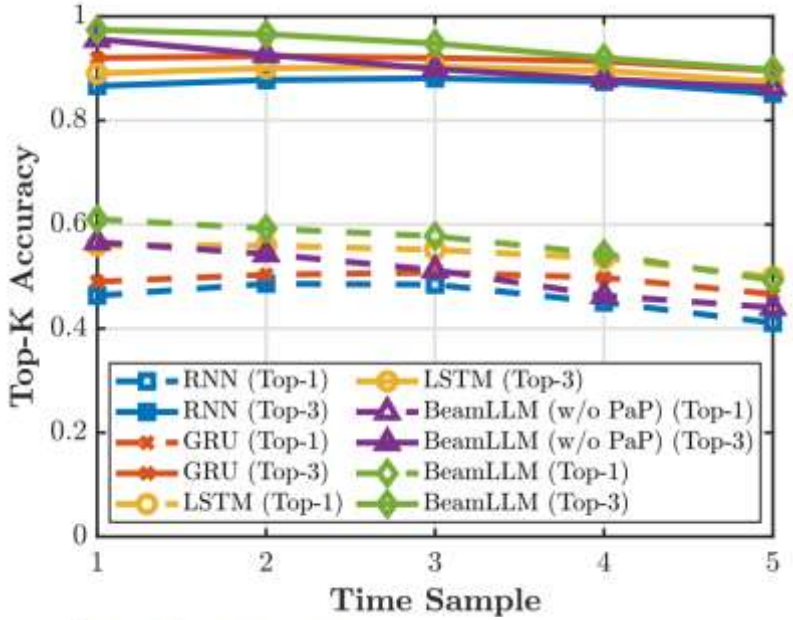
Training, validation loss, and accuracy



The number of parameters & inference time

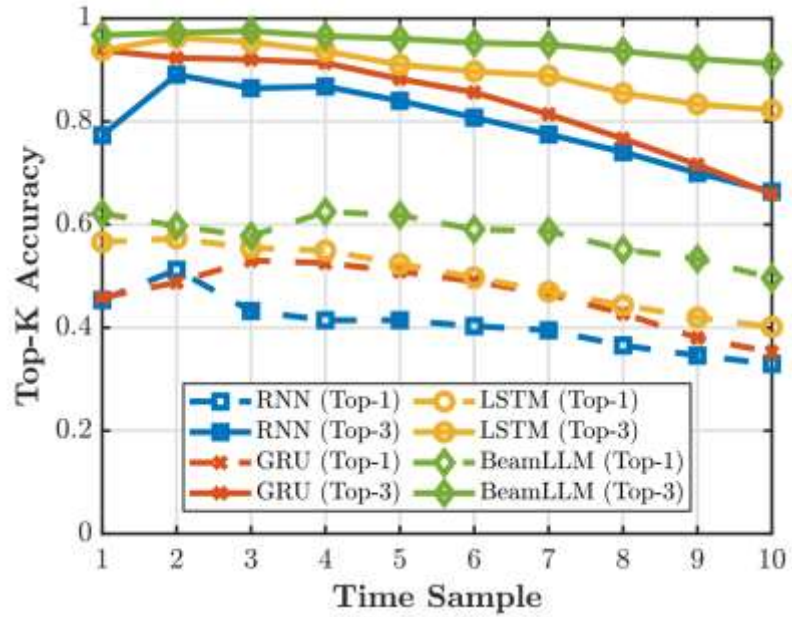
Models	# total params.	# trainable params.	Average inf. time (sec)
RNN	29,505	29,505	8.0×10^{-6}
GRU	79,425	79,425	3.7×10^{-5}
LSTM	104,385	104,385	1.9×10^{-5}
BeamLLM	178,303,798	53,863,990	7.5×10^{-4}

Top-K accuracy



Standard

(T_{hist}**:** $T_{\text{pred}} = 8:5$**)**



Few-shot

(T_{hist}**:** $T_{\text{pred}} = 3:10$**)**

Conclusions

- ✓ Summary
- ✓ Advantages
- ✓ Limitations
- ✓ Future Works

Conclusions

- **Summary:** BeamLLM leverages LLMs + CV for accurate, robust mmWave beam prediction; outperforms other DL methods
- **Advantages:** High performance; strong few-shot capability
- **Limitations:** Higher complexity/latency; vision-only; single-user case only
- **Future Work:** Lightweight designs; multimodal-sensing extension