

CaliFormer: Leveraging Unlabeled Measurements for Calibration with Self-supervised Learning

Haoyang Wang¹, Yuxuan Liu¹, Chenyu Zhao¹, Jiayou HE², Wenbo Ding¹, Xinlei Chen¹

¹Shenzhen International Graduate School, Tsinghua University ²Hong Kong University of Science and Technology

Abstract

Accurate calibration of low-cost sensors is essential for their utilization in environmental monitoring. SOTA methods¹ take recent past and close future measurements to achieve accurate calibration. However, these methods heavily rely on labeled data, which is hard to collect in the real world. In this work, We propose CaliFormer, a representation learning model that learn the time and spatial invariant knowledge from unlabeled measurements in self-supervised learning phase. A set of enhancements in pre-train and model architecture is proposed to facilitate the training of CaliFormer. Subsequently, the calibration model built on CaliFormer is trained with limited labeled data in supervised learning phase. We extensively evaluate the calibration system with a dataset measured by low-cost sensors. Results show that our method outperforms SOTA methods.

Background



Monitoring systems heavily depends on **measurements quality**

Research Question

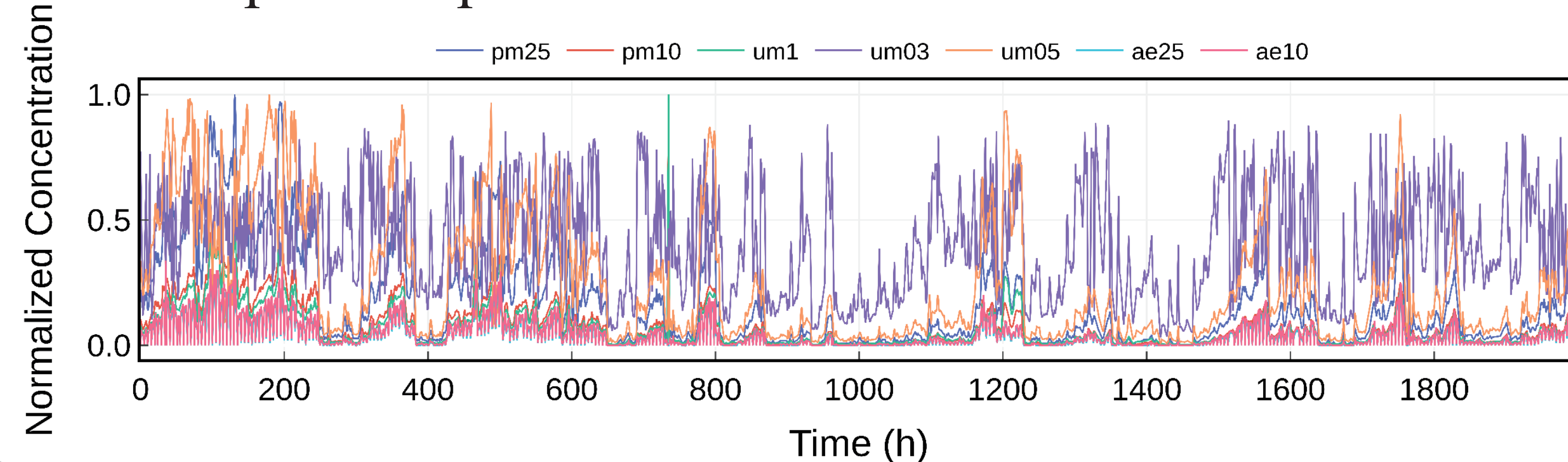
How to scale the **data-driven system** to large-scale environment sensing calibration with **limited labeled data** ?

$$\argmin_{\theta} \left(\sum_{t=1}^T \mathcal{L} \left(y_{\langle t-\delta_1, t+\delta_2 \rangle}, C_{\theta} \left(x_{\langle t-\delta_1, t+\delta_2 \rangle} \right) \right) \right)$$

Loss function Ground Truth Calibration model Measurement

Observation

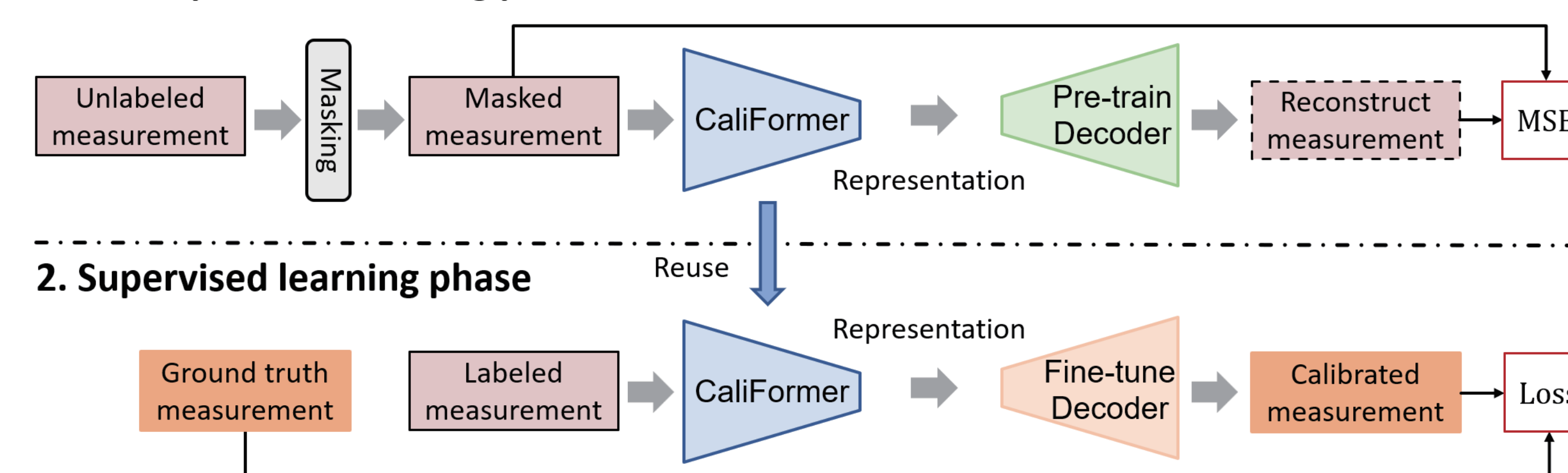
Temporal dependencies Correlation of measurements



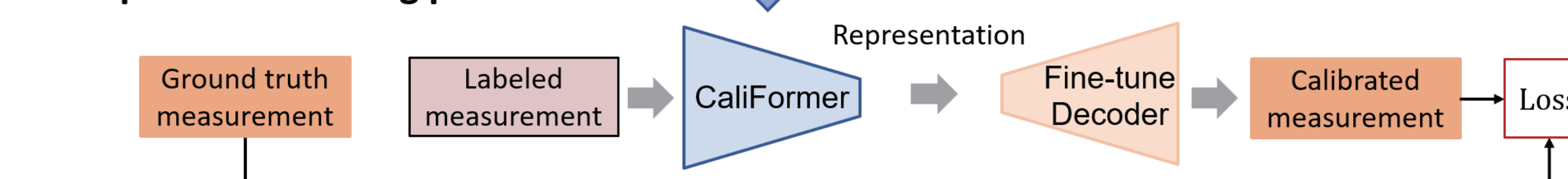
Model Design

1. Overview

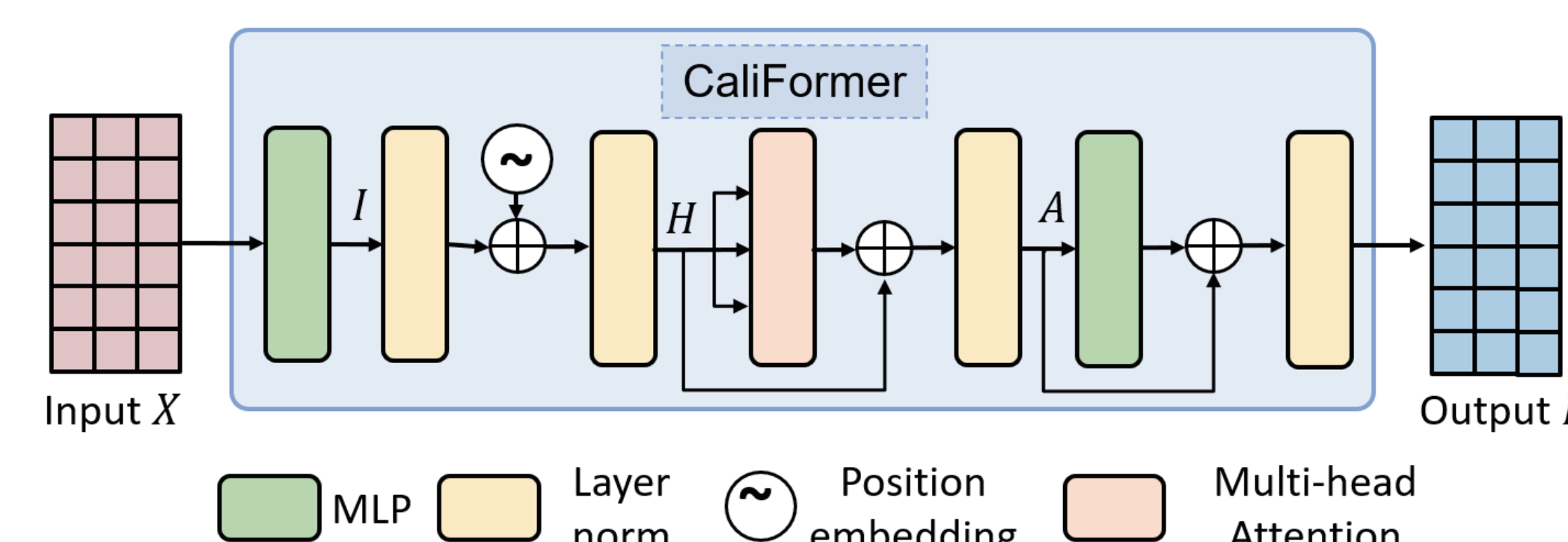
1. Self-supervised learning phase



2. Supervised learning phase

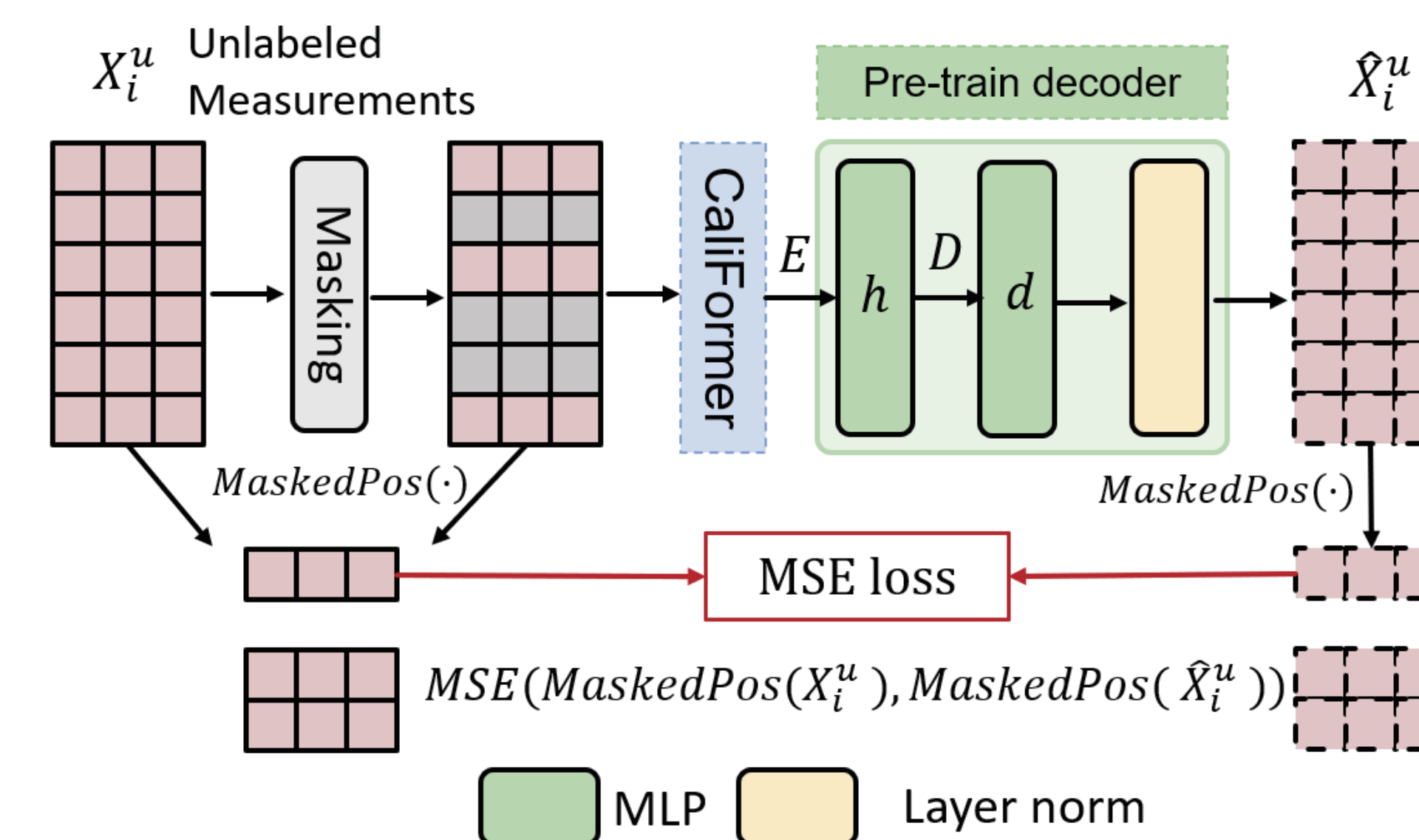


2. CaliFormer Design



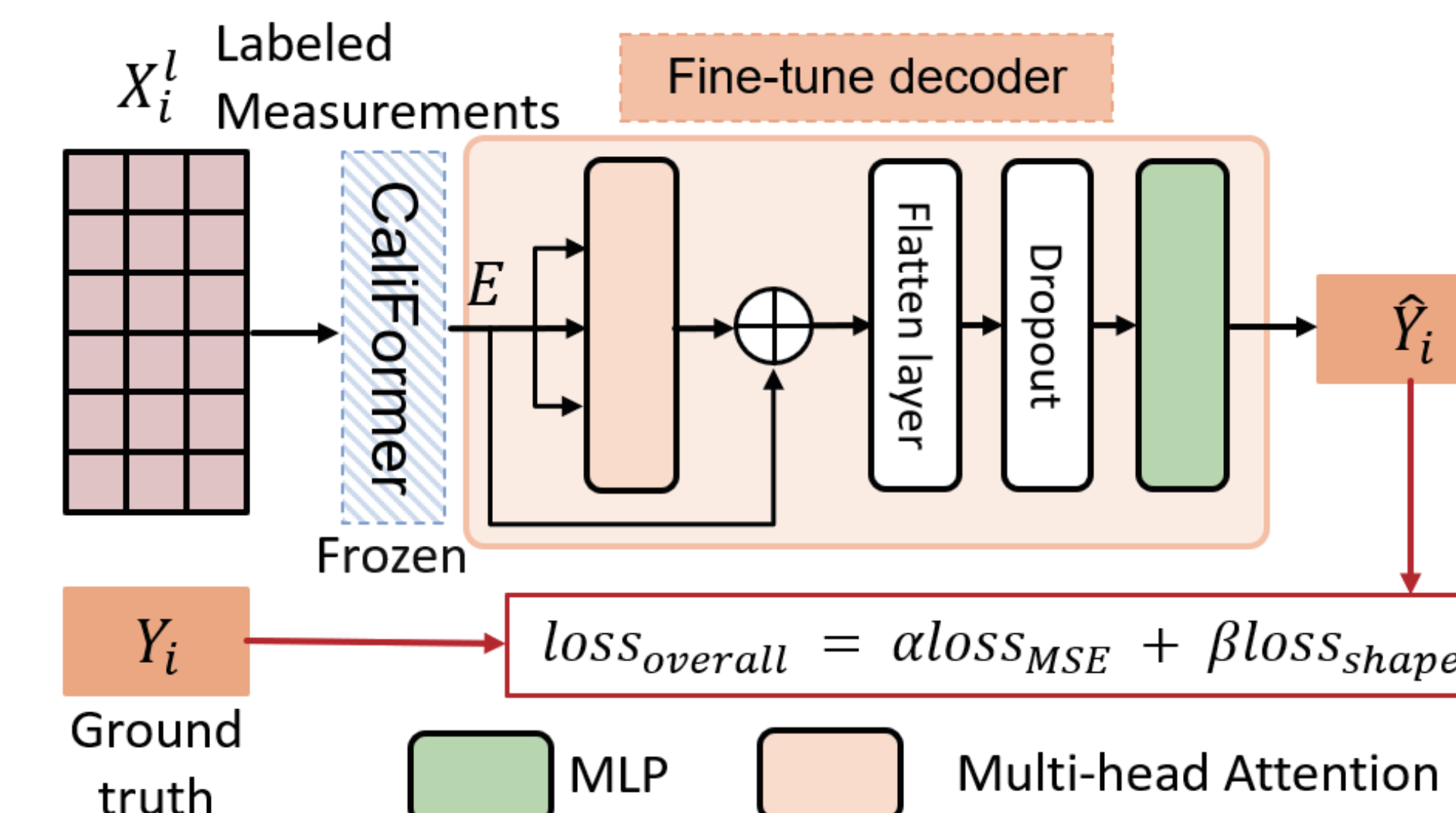
- Inspired by TransFormer² encoder architecture

3. Self-supervised learning phase



- Utilize unlabeled measurements to train the model

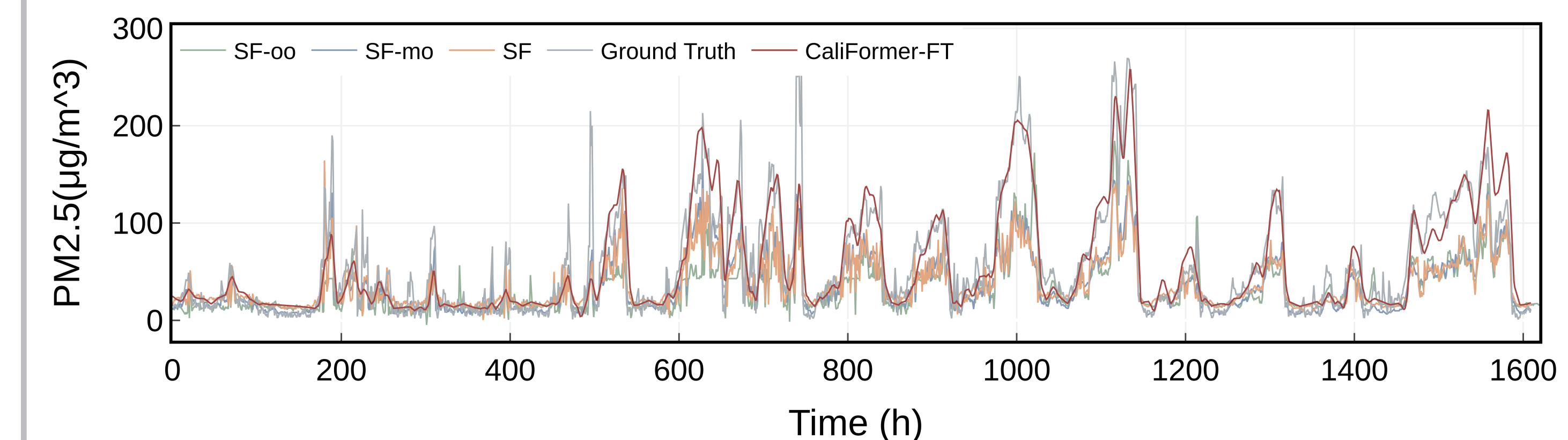
4. Supervised learning phase



- Utilize a small number of labeled measurements to train the calibration network

Experimental Results

1. Calibration result



Our method outperform the SOTA methods with 1% labeled data

2. Quantitative analysis

TABLE I
OVERALL PERFORMANCE WITH 1% LABELED DATA, WHICH IS SHOWN IN MAE($\mu\text{G}/\text{m}^3$).

Methods	Naive	SF-oo	SF-mo	SF	CaliFormer-FT
MAE	31.25	25.20	24.91	24.08	18.20

- Our method outperforms the SOTA method SF by 25%.

TABLE II
PERFORMANCE WITH DIFFERENT LABELING RATES, WHICH IS SHOWN IN MAE($\mu\text{G}/\text{m}^3$).

Labeling rate	0.5%	1%	2%	5%	10%	Average
Naive	31.25	31.25	31.25	31.25	31.25	31.25
SF-oo	29.89	25.20	24.68	21.93	21.78	24.70
SF-mo	29.72	24.91	22.33	21.77	20.86	23.92
SF	29.74	24.08	22.11	21.70	20.37	23.60
CaliFormer-FT	19.91	18.20	15.20	14.84	14.57	16.54

- The gain is significant with low labeling rate.

Conclusions

- To the best of our knowledge, CaliFormer is the first attempt to incorporate SSL into sensor calibration, which **necessitates significantly less labeled data**. This constitutes an advancement toward practical in-field sensor calibration.
- We develop the CaliFormer to process multi-modal sensor data. A set of enhancements in pre-training methodology and model architecture is proposed to facilitate the training of the calibration model.
- A prototype system is developed and compared with SOTA methods. Results demonstrate the effectiveness of the CaliFormer based calibration system.