CaliFormer: Leveraging Unlabeled Measurements to Calibrate Sensors with Self-supervised Learning

Haoyang Wang*
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
haoyang-22@mails.tsinghua.edu.cn

Yuxuan Liu*
Tsinghua-Berkeley Shenzhen Institute, Tsinghua University
Shenzhen, China
yuxuan-l21@mails.tsinghua.edu.cn

Chenyu Zhao
Shenzhen International Graduate School, Tsinghua University
Shenzhen, China
zhaocy22@mails.tsinghua.edu.cn

Jiayou He
School of Engineering, Hong Kong University of Science and Technology
Hong Kong, China
jhecf@connect.ust.hk

Wenbo Ding
Shenzhen International Graduate School, Tsinghua University
Pengcheng Laboratory
RISC-V International Open Source Laboratory
Shenzhen, China
ding.wenbo@sz.tsinghua.edu.cn

Xinlei Chen†
Shenzhen International Graduate School, Tsinghua University
Pengcheng Laboratory
RISC-V International Open Source Laboratory
Shenzhen, China
chen.xinlei@sz.tsinghua.edu.cn

ABSTRACT

Accurate calibration of low-cost sensors is critical for improving their potential in environmental monitoring. State-of-the-art (SOTA) methods based on supervised learning commonly calibrate sensor measurements with ground truth from the immediate past or future. However, these techniques rely heavily on labeled data which is challenging to obtain in real-world scenarios. Thus, this paper introduces CaliFormer, a novel representation learning model using self-supervised learning to extract time- and spatial-invariant knowledge from unlabeled measurements. Moreover, we propose pre-training enhancements and model architecture modifications to help train CaliFormer. We then fine-tune the calibration model with the learned representations, which is supervised by limited labeled data. Finally, we comprehensively evaluate our calibration model with a dataset collected by low-cost sensors. Results show that our model outperforms other SOTA calibration methods significantly.

CCS CONCEPTS

• **Theory of computation → Models of learning**; • **Computing methodologies → Spatial and physical reasoning**; • **Computer systems organization → Sensors and actuators**.

KEYWORDS

Air Pollution; Sensor calibration; Self-supervised learning

*Both authors contributed equally to this research.
†Xinlei Chen is the corresponding author.

UbiComp/ISWC '23 Adjunct, October 8–12, Cancun, Quintana Roo, Mexico

ACM ISBN 979-8-4007-0200-6/23/10…$15.00
https://doi.org/10.1145/3594739.3612917

ACM Reference Format:
Haoyang Wang, Yuxuan Liu, Chenyu Zhao, Jiayou He, Wenbo Ding, and Xinlei Chen. 2023. CaliFormer: Leveraging Unlabeled Measurements to Calibrate Sensors with Self-supervised Learning. In *Adjunct Proceedings of the 2023 International Joint Conference on Pervasive and Ubiquitous Computing and the 2023 International Symposium on Wearable Computers (UbiComp/ISWC '23 Adjunct), October 8–12, 2023, Cancun, Quintana Roo, Mexico.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3594739.3612917

1 INTRODUCTION

In the era of Artificial Intelligence of Things (AIoT), environmental sensors have gained significant momentum thanks to their potential in large-scale sensing applications, such as urban air quality assessment[1, 2], smart transportation monitor[3] and industrial gas leakage detection[4]. These sensors are popular for their low cost and low power consumption[5, 6]. However, the low quality of the sensor measurements is still a general concern[7, 8]. The quality is easily susceptible to various environmental factors, such as thermal noise and temperature drift[9, 10].

Consequently, many in-field calibration techniques have emerged recently to address the challenges of low-cost sensors[11, 12], aiming to improve the sensing accuracy via periodic sensor calibration. The application scenarios include air pollution monitoring[13], mobile crowdsensing[14], multi-task scheduling[15–17], and multi-agent coordination[18, 19]. These techniques can be divided into 3 calibration paradigms. In specific, some techniques adopt a *one-to-one* calibration paradigm. Each sensor measurement is calibrated by the corresponding feature vector at a specific time step[20]. While other methods utilize a *many-to-one* calibration paradigm where feature vectors are extracted from several recent past time steps[21]. This paradigm exhibits more pronounced effectiveness in capturing fluctuating environmental patterns. In contrast, more recent calibration strategies exploit feature vectors from both the recent past and the immediate future, which is called *many-to-many* calibration paradigm[22]. It refines sensor readings by adding samples gradually, enabling real-time calibration and SOTA performance.

Most of these techniques are designed based on supervised learning, where a large number of labeled data is highly required to train

the calibration model effectively[23]. Nevertheless, it is challenging to collect adequate labeled data in the real world[24]. First, the ground truth is only available at a few sparsely scattered monitoring stations, making the labeled data acquisition high-cost and time-consuming[25]. Second, to obtain a generalized model, the wide range of sensor models and usage situations necessitate collecting labeled data under diverse combinations of multiple sensor models and usage scenarios.

Therefore, the research question of this paper is: **How to scale the data-driven methods to large-scale environment sensing calibration with limited labeled data?** To address the problem, we design a calibration model based on representation learning to extract time- and spatial-invariant knowledge shared among tasks, such as the temporal dependencies among measurements[26]. The learned representations can be employed to train a task-specific inference model with limited labeled data[27, 28]. We train the representation learning model through self-supervised learning leveraging a large amount of unlabeled data[29, 30]. Our key insight is to derive generalizable representations from abundant unlabeled data, instead of relying on scarce labeled data[26, 31]. Notably, it is straightforward to collect unlabeled data from widely distributed sensors[32, 33].

Our representation learning model design mainly focuses on how to extract representations from unlabeled sensor measurements. Inspired by similarities between text and measurements, and the effectiveness of Transformer in sequence-to-sequence prediction tasks, we propose a Transformer encoder-based model called **CaliFormer** to process air pollution sensor measurements [34, 35]. An auxiliary task of reconstructing a sensor measurement sequence is then established to pre-train CaliFormer and a decoder together[36]. In this process, CaliFormer learns the time- and spatial-invariant knowledge in sensor measurements effectively. Then, the fine-tune decoder is trained with parameter-frozen CaliFormer on limited labeled data. To demonstrate our calibration model, we conduct experiments on a dataset with PM2.5 values measured by low-cost sensors. Results show that our model outperforms SOTA methods.

In summary, this paper provides the subsequent contributions:

- To the best of our knowledge, CaliFormer is the first attempt to incorporate self-supervised learning into sensor calibration to overcome the challenge of limited labeled data.
- Drawing inspiration from Transformer, we develop CaliFormer to process sensor measurements. A set of enhancements in pre-training methodology and model architecture are proposed to help train the calibration model effectively.
- Our method is compared with SOTA methods through experiments. The results demonstrate the performance of the calibration model based on CaliFormer.

The rest of this paper is organized as follows: Section 2 formulates the issue of sensor calibration. Section 3 presents the calibration framework and design specifics. Section 4 reports the experiments and results. Finally, Section 5 concludes the paper.

2 PROBLEM DEFINITION

This paper examines an affordable sensor that operates in conjunction with a reference station. This low-cost sensor is unable to generate precise measurements because of temperature drift and other extraneous factors. In contrast, the reference station is trustworthy, having the capacity to produce accurate and reliable measurements. During the time period T, we have compiled measurements from both devices. The low-cost sensor measurements are portrayed as a time sequence $X \in \mathbb{R}^{|T| \times d}$, where $|T|$ denotes the duration of the time series, and d refers to the magnitude of the measurements generated by the low-cost sensor during each time step. On the other hand, the measurements from the reference station comprise a time series $Y \in \mathbb{R}^{|T| \times 1}$, where $|T|$ represents the time series length. In this paper, we regard the measurements from the reference station to be the *ground truth*. The central objective of sensor calibration is to acquire a *calibration model* $C_\theta : \mathbb{R}^{p \times d} \rightarrow \mathbb{R}^{q \times 1}$, capable of decreasing the deviation between the calibrated value $\hat{Y} = C_\theta(X)$ and the ground-truth Y obtained from the reference station. Here, p denotes the input length, q represents the output length, and θ denotes trainable parameters.

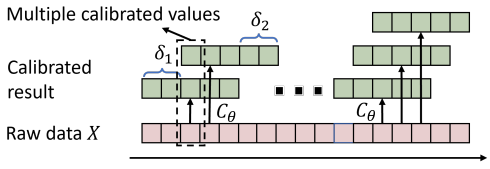

Figure 1: In many-to-many calibration, the model generates multiple calibrated values for the same input.

Let $x_t \in X$ denote the measurement obtained from the low-cost sensor at time step t, whereas $y_t \in Y$ symbolizes the measurement derived from the reference station during time step t, precisely referred to as the ground truth. The integration of uncalibrated low-cost sensor measurements obtained from time windows in the past $\langle t - \delta_1, t \rangle$ and the future $\langle t, t + \delta_2 \rangle$ as inputs to the calibration model C_θ, according to previous studies [22], assumes a crucial role in the attainment of accurate calibration of the present measurement. An enhanced level of accuracy is attainable when obtaining information about future measurements $\langle t, t + \delta_2 \rangle$ with respect to time t. Consequently, this observation motivates the application of a many-to-many calibration model, which has multiple inputs and outputs. As shown in Fig.1, this model yields multiple calibrated values for the same input, ultimately resulting in improved calibration as additional measurements become available. The objective function of the many-to-many model is

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

where symbol $\mathcal{L}$ signifies the discrepancy between the calibrated outcomes and the ground truth.

3 DESIGN

3.1 Overview

As Fig. 2 illustrates, the framework consists of two phases: the self-supervised learning phase and the supervised learning phase. Meanwhile, this framework includes three components, *CaliFormer*, *pre-train decoder*, and *fine-tune decoder*.

Self-supervised learning phase. During this stage, the CaliFormer and pre-trained decoder are trained using unlabeled data.

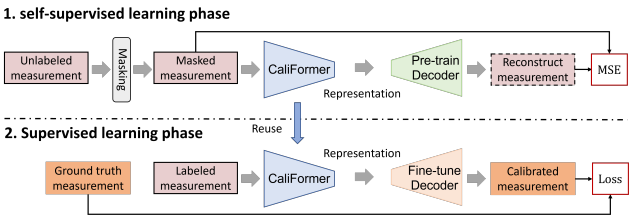

Figure 2: Framework overview.

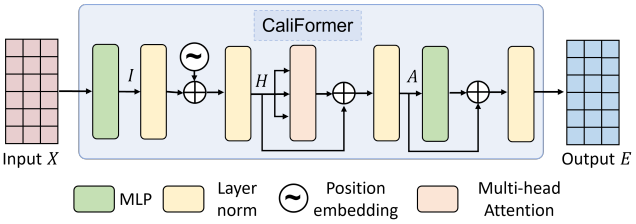

Figure 3: CaliFormer architecture.

Specifically, we mask a portion of the unlabeled sensor measurements and input the masked result into the CaliFormer, which generates high-level representations. The CaliFormer and pre-trained decoder work together to reconstruct the unlabeled sensor measurements by learning the time- and spatial-invariant knowledge among them. This phase fully utilizes a substantial amount of unlabeled data to extract high-level representation accordingly.

Supervised learning phase. During this phase, we implement a parameter freeze on the CaliFormer and establish a connection with the fine-tune decoder. The limited labeled representations, which are processed by the CaliFormer are utilized for training the fine-tune decoder. Upon completion of the network training, the CaliFormer and fine-tune decoder work in tandem to calibrate the sensor measurement.

Subsequently, in the following section, we present a comprehensive exposition of the architectural design of the CaliFormer. Next, we explicate two training phases and their corresponding decoders.

3.2 CaliFormer Design

As illustrated in Fig.3, the CaliFormer receives sequential sensor measurements within a time window of δ and generates high-level representations. Specifically, the input of CaliFormer is denoted as $X \in \mathbb{R}^{|\delta| \times d}$, where $|\delta|$ represents the total number of sensor measurements in a time window and d corresponds to the dimension of each input sensor measurement. The high-level representation generated by CaliFormer is denoted by E, a matrix of size $|\delta| \times h$, and the h is the hidden dimension large than d.

Initially, we employ a multilayer perceptron (MLP) to facilitate the mapping of the original sensor measurement sequence X to a concealed representation $I \in \mathbb{R}^{|\delta| \times h}$, wherein the latter incorporates more implied features than the former. Afterwards, layer normalization is employed to standardize the concealed representation I. Then, positional embedding is included to the concealed representation to preserve positional information. Subsequently, the concealed representations are referred to as H post the second normalization layer. Ultimately, an attention-based block receives H as its input and yields the definitive high-level representations E. The computation proceeds of this attention-based block are as follows.

$$\begin{aligned} A &= \text{LayerNorm}(\text{MultiAttn}(H) + H), \\ E &= \text{LayerNorm}(\text{MLP}(A) + A). \end{aligned} \tag{2}$$

The self-attention layer denoted as $\text{MultiAttn}(\cdot)$, is comprised of A attention heads. In each attention head, the query, key, and value dimensions are all equivalent to h. It effectively captures the contextual relations of the input sensor measurements with positional embedding. The $\text{MLP}(\cdot)$ comprises an MLP containing two fully-connected layers and a Gaussian Error Linear Unit (GELU) activation layer. Both the input and output dimensions of the $\text{MLP}(\cdot)$ are h. Ultimately, the CaliFormer generates the high-level representations $E \in \mathbb{R}^{|\delta| \times h}$ for the input sensor measurement sequence X. In CaliFormer, the number of attention head A is set to 4, the total number of sensor measurement $|\delta|$ is set to 12, and the hidden dimension h are set to 72.

3.3 Learning Representation

To capture time and spatial invariant knowledge among sensor measurements, we introduce a self-supervised learning-based approach. Our method leverages the unlabeled sensor measurements, which reconstruct masked patches using a pre-train model composed of CaliFormer and the pre-train decoder. Specifically, we randomly mask patches of the sensor measurements, and jointly train CaliFormer and the pre-train decoder to reconstruct the masked patches while learning from the visible sensor measurements. This reconstruction task trains the CaliFormer to capture time and spatial invariant knowledge among the sensor measurements. In this context, we define $\boldsymbol{X^u}$ as the set of unlabeled sensor measurement sequences, where $X_i^u \in \boldsymbol{X^u}$ refers to the i-th sequence of unlabeled sensor measurements.

3.3.1 Masking. As illustrated in Fig.4, we mask a high proportion of data with the Span Masking mechanism in unlabeled sensor measurement sequences to create a challenging condition in this phase [37]. In this case, we can train the model effectively and the model is forced to learn the time- and spatial-invariant knowledge. All the masked data are replaced with 0, and the mask ratio is set to 0.15. The masked position set is denoted by I, which is utilized in the loss function of pre-train model training.

3.3.2 Pre-train model. A visual representation of the structure of the pre-train model is depicted in Figure 4., which consists of CaliFormer and the pre-train decoder. The pre-train decoder aims to reconstruct the masked sensor measurements with high-level representation E output from CaliFormer. The reconstructed result of unlabeled sensor measurement sequence X_i^u is denoted as $\hat{X}_i^u$. The computation proceeds of the pre-train decoder are as follows.

$$\begin{aligned} D &= \text{MLP}_1(\text{GELU}(E)), \\ \hat{X}_i^u &= \text{LayerNorm}(\text{MLP}_2(D)), \end{aligned} \tag{3}$$

where $\text{MLP}_1(\cdot)$ and $\text{MLP}_2(\cdot)$ denote single fully-connected layers with h and d units, respectively. The $\text{GELU}(\cdot)$ is the activation function. Finally, we obtain the reconstructed result $\hat{X}_i^u$ of unlabeled sensor measurement sequence X_i^u.

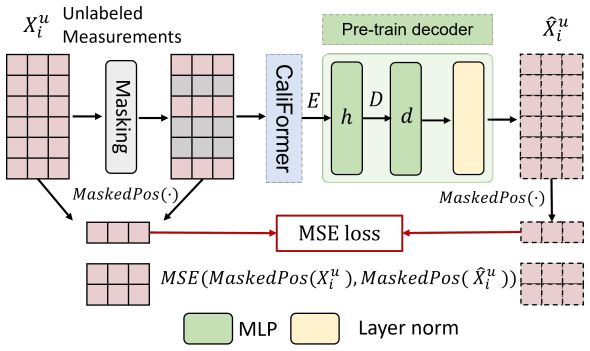

Figure 4: Self-supervised training process and the architecture of pre-train decoder.

3.3.3 Reconstruction. The reconstruction of the masked sensor measurement sequences is a regression task. During the training of the pre-train model, we utilize the difference between masked measurements and reconstructed measurements as a loss function. Specifically, we utilize the Mean Square Error (MSE) function to calculate the pre-training loss as follows.

$$loss = \frac{1}{|\boldsymbol{X^u}|} \sum_{i=1}^{|\boldsymbol{X^u}|} \text{MSE}(\text{MaskedPos}(X_i^u), \text{MaskedPos}(\hat{X}_i^u)), \quad (4)$$

where $|\boldsymbol{X^u}|$ is the number of unlabeled sensor measurement sequences, and the MaskedPos(·) selects the measurements that correspond to the set I of masked positions. Thus the loss function is computed using only measurements in masked positions. In this phase, the pre-train model is updated with the Adam optimizer.

3.4 Task-specific Model: Supervised Learning Phase

In order to utilize the time- and spatial-invariant knowledge among the sensor measurements learning from unlabeled measurements for sensor calibration, we design a fine-tune model consisting of the learned CaliFormer and fine-tune decoder with supervised learning. Figure 5 illustrates the workflow of supervised learning. The input sensor measurement sequences are not masked in this phase, and the parameters of learned CaliFormer are frozen. This procedure trains the fine-tune decoder with labeled sensor measurement sequences which learns to map the representations generated by CaliFormer to ground truth measurements. Let $\boldsymbol{X^l}$ denote the labeled sensor measurement sequences and $\hat{Y}$ denote the calibrated measurement sequences. For i-th labeled sensor measurement sequence $X_i^l \in \boldsymbol{X^l}$, there exists a corresponding ground truth sequence $Y_i \in \boldsymbol{Y}$, where $\boldsymbol{Y}$ is the set of ground-truth measurement sequences generated by the reference station, and $Y_i \in \mathbb{R}^{|\delta|\times 1}$. The calibrated result of X_i^l is denoted as $\hat{Y}_i \in \mathbb{R}^{|\delta|\times 1}$.

3.4.1 Fine-tune model. Fig. 5 depicts the interconnections between the frozen CaliFormer and the fine-tune decoder that comprises four basic components: the self-attention layer, the flatten layer, the dropout layer, and an individual fully-connected MLP layer. The self-attention layer is defined by a hidden dimension of 36

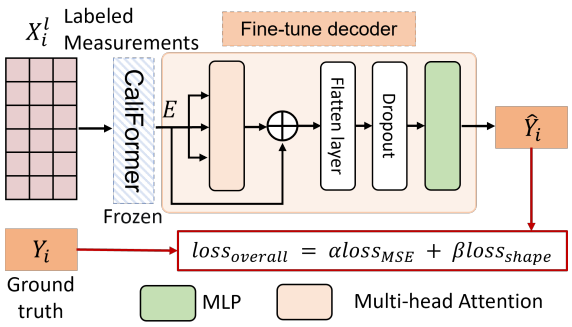

Figure 5: Supervised training process and fine-tune decoder.

and an input size of h, with 4 attention heads contributing to its architecture. The dropout layer aids in preventing over-fitting, and its associated drop rate is set at 0.5. A single fully-connected MLP layer, consisting of 12 hidden units, is subsequently constructed, which generates the corresponding calibrated result $\hat{Y}$. The length of the output result, denoted as $|\delta|$, is identical to that of the input, and is set at 12 in harmony with Section 3.2.

3.4.2 Calibration. The calibration of the sensor measurement sequences is a regression task with multi outputs. During the training of the fine-tune model, we utilize the MSE between calibrated measurements and origin measurements as a part of the loss function,

$$loss_{MSE} = \frac{1}{|Y|} \sum_{i=1}^{|Y|} MSE(Y_i, \hat{Y}_i), \quad (5)$$

where $|Y|$ is the number of calibrated sensor measurement sequences. As discussed in [22], sensor measurement sequences with varying characteristics can exhibit similar MSE. Consequently, it becomes imperative to consider the loss between their respective time-series shapes. Thus, we calculate the degree of shape distortion between $\hat{Y}$ and Y with the differentiable loss function based on dynamic-time warping (DTW) [38]. We incorporate this into the final loss function as the "shape loss" term, designated by $loss_{shape}$. The overall loss of the fine-tune model is as follows.

$$loss_{overall} = \alpha loss_{MSE} + \beta loss_{shape}, \quad (6)$$

where coefficients α and β represent adjustable trade-off parameters, employed to effectively balance the respective importance of two constituent components of the loss function.

4 EVALUATION

4.1 Setup

4.1.1 Dataset. The Beijing Data Set is constituted by PM2.5 (particulate matter with a diameter less than 2.5μm) measurements recorded at seven different locations throughout Beijing [22]. Hourly measurements are acquired over the time interval spanning from March 2018 to May 2019. Low-cost sensors, positioned alongside the relevant government reference stations, are utilized to gather data. Seven different feature measurements are mapped and used for training the models, with a total of 60,450 samples being employed during the experimental phase.

Table 1: Overall performance with 1% labeled data, which is shown in MAE($\mu g/m^3$).

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

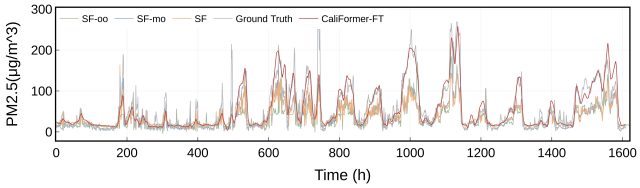

Figure 6: Calibration results from different methods with 1% labeled set. CaliFormer-FT achieves the best results.

4.1.2 Preprocessing. The sensor measurements are partitioned into three subsets: the training set (60%), validation set (20%), and test set (20%). Following this, the previously established training set is divided into a labeled set (constituting 1% of the training set) and an unlabeled set (constituting 99% of the training set), with the labeling rate reflecting the proportion of labeled data in the training set. The self-supervised learning phase relied upon the usage of the unlabeled set, which is instrumental in the training of the pre-train decoder and CaliFormer. The validation set is integrated into the model selection process. Then the labeled set is subsequently employed to train the fine-tune model through supervised learning. The validation set is once again involved in the process of model selection. We evaluate our trained model on the test set.

4.1.3 Methods in comparison. We evaluate our system by comparing it with other methods.

Naive. No calibration method is used. We directly use the raw measurements from low-cost sensors.

SensorFormer (SF) [22]. SF is the SOTA many-to-many calibration method. The method provides immediate calibration and gradual refining with additional measurements, which has a similar architecture with fine-tune decoder.

SensorFormer-mo (SF-mo) [22]. SF-mo is the many-to-one version of SF, which means this method does not use $loss_{shape}$ in the loss function.

SensorFormer-oo (SF-oo) [22]. SF-oo is the one-to-one version of the state-of-the-art calibration method.

CaliFormer-FT. We propose it for CaliFormer with a fine-tune decoder, respectively trained in the self-supervised and supervised learning phases.

4.1.4 Implementation. We developed the methods using Python and PyTorch and trained them on a server with 4 NVIDIA RTX A6000 GPUs, 128GB memory, and an Intel(R) Xeon(R) Gold 6242R 3.4GHz CPU. Both two phases use a learning rate of 0.001 and a batch size of 64. In the supervised training phase, all methods are trained with the same hyper-parameters and labeled set. CaliFormer is trained for 320 epochs on the unlabeled set in the self-supervised learning phase, while in the supervised learning phase, the fine-tune decoder and other methods are trained with the labeled set over 200 epochs. The α is 0.15 and β is 1 accroding to [22]. Finally, the calibrated result of each time step is obtained by averaging the values of all time windows that include this step.

Table 2: Performance with different labeling rates, which is shown in MAE($\mu g/m^3$).

Labeling rate	0.5%	1%	2%	5%	10%	Average
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**

4.1.5 Metric. The Mean Absolute Error (MAE) calculates the absolute differences between the calibrated measurements and the ground truth, serving as a quantitative indicator of their similarity. Hence, the MAE is employed as the evaluation metric for assessing the accuracy of all methods.

4.2 Result

4.2.1 Overall performances. Table 1 illustrates the performance on PM2.5 calibration between CaliFormer-FT and different baselines. The calibration performance is shown in MAE($\mu g/m^3$). The labeling rate of the experiments is 1%, which means only 1% of the labeled dataset is utilized for training the networks. The CaliFormer-FT outperforms the SOTA method SF by 25%, demonstrating the effectiveness of the CaliFormer-FT. This is because the CaliFormer extracts effective representation from unlabeled data. Compared to other baselines, CaliFormer-FT performs better, especially during the peak periods as shown in Fig 6.

4.2.2 Impact of labeling rate. We compare the performance between CaliFormer-FT and other baselines across a range of labeling rates, spanning from 0.5% to 10%. Table 2 illustrates the results on PM2.5 calibration of different methods. All methods achieve better results with an increasing labeling rate. The performance of CaliFormer-FT is consistently better than other baselines in all cases. The experiment results suggest that CaliFormer can learn high-level representations from unlabeled data and fine-tune decoder can achieve better performance with trained CaliFormer. When the labeling rate is low, the increase is substantial.

5 CONCLUSION

This study introduces CaliFormer, a new model for many-to-many sensor calibration, that leverages self-supervised learning to extract high-level representations from unlabeled data. The proposed model aims to aid the sensor calibration mission by taking advantage of these learned representations. The effectiveness of CaliFormer in the downstream sensor calibration mission is thoroughly evaluated through extensive experiments, and compared against SOTA methods, showing a significant reduction in the required effort for ground truth collection.

6 ACKNOWLEDGMENTS

This paper was supported by the National Key R&D program of China (2022YFC3300703), Guangdong Innovative and Entrepreneurial Research Team Program (2021ZT09L197), Shenzhen 2022 Stabilization Support Program (WDZC20220811103500001), and Tsinghua Shenzhen International Graduate School Cross-disciplinary Research and Innovation Fund Research Plan (JC20220011).

REFERENCES

[1] Yuxuan Liu, Xinyu Liu, Fanhang Man, Chenye Wu, and Xinlei Chen. Fine-grained air pollution data enables smart living and efficient management. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 768–769, 2022.

[2] Xiangxiang Xu, Xinlei Chen, Xinyu Liu, Hae Young Noh, Pei Zhang, and Lin Zhang. Gotcha ii: Deployment of a vehicle-based environmental sensing system. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 376–377, 2016.

[3] Jaminur Islam, Jose Paolo Talusan, Shameek Bhattacharjee, Francis Tiausas, Sayyed Mohsen Vazirizade, Abhishek Dubey, Keiichi Yasumoto, and Sajal K Das. Anomaly based incident detection in large scale smart transportation systems. In *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*, pages 215–224. IEEE, 2022.

[4] Alibek Kopbayev, Faisal Khan, Ming Yang, and Syeda Zohra Halim. Gas leakage detection using spatial and temporal neural network model. *Process Safety and Environmental Protection*, 160:968–975, 2022.

[5] KM Talluru, V Kulandaivelu, N Hutchins, and I Marusic. A calibration technique to correct sensor drift issues in hot-wire anemometry. *Measurement Science and Technology*, 25(10):105304, 2014.

[6] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Shijia Pan, Jiayou He, Hae Young Noh, Lin Zhang, and Pei Zhang. Pga: Physics guided and adaptive approach for mobile fine-grained air pollution estimation. In *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers*, pages 1321–1330, 2018.

[7] Xinlei Chen, Xiangxiang Xu, Xinyu Liu, Hae Young Noh, Lin Zhang, and Pei Zhang. Hap: Fine-grained dynamic air pollution map reconstruction by hybrid adaptive particle filter. In *Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM*, pages 336–337, 2016.

[8] Xinlei Chen, Susu Xu, Xinyu Liu, Xiangxiang Xu, Hae Young Noh, Lin Zhang, and Pei Zhang. Adaptive hybrid model-enabled sensing system (hmss) for mobile fine-grained air pollution estimation. *IEEE Transactions on Mobile Computing*, 21(6):1927–1944, 2020.

[9] Chengzhao Yu, Ji Luo, Rongye Shi, Xinyu Liu, Fan Dang, and Xinlei Chen. St-icm: spatial-temporal inference calibration model for low cost fine-grained mobile sensing. In *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, pages 910–912, 2022.

[10] Yifei Sun, Yuxuan Liu, Ziteng Wang, Xiaolei Qu, Dezhi Zheng, and Xinlei Chen. C-ridge: Indoor co2 data collection system for large venues based on prior knowledge. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 1077–1082, 2022.

[11] Xinyu Liu, Xinlei Chen, Xiangxiang Xu, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. Delay effect in mobile sensing system for urban air pollution monitoring. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–2, 2017.

[12] Ke Hu, Vijay Sivaraman, Blanca Gallego Luxan, and Ashfaqur Rahman. Design and evaluation of a metropolitan air pollution sensing system. *IEEE Sensors Journal*, 16(5):1448–1459, 2015.

[13] Xinyu Liu, Xiangxiang Xu, Xinlei Chen, Enhan Mai, Hae Young Noh, Pei Zhang, and Lin Zhang. Individualized calibration of industrial-grade gas sensors in air quality sensing system. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*, pages 1–2, 2017.

[14] Huadong Ma, Dong Zhao, and Peiyan Yuan. Opportunities in mobile crowd sensing. *IEEE Communications Magazine*, 52(8):29–35, 2014.

[15] Xinlei Chen, Susu Xu, Jun Han, Haohao Fu, Xidong Pi, Carlee Joe-Wong, Yong Li, Lin Zhang, Hae Young Noh, and Pei Zhang. Pas: Prediction-based actuation system for city-scale ridesharing vehicular mobile crowdsensing. *IEEE Internet of Things Journal*, 7(5):3719–3734, 2020.

[16] Xuecheng Chen, Haoyang Wang, Zuxin Li, Wenbo Ding, Fan Dang, Chengye Wu, and Xinlei Chen. Deliversense: Efficient delivery drone scheduling for crowdsensing with deep reinforcement learning. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pages 403–408, 2022.

[17] Zuxin Li, Fanhang Man, Xuecheng Chen, Baining Zhao, Chenye Wu, and Xinlei Chen. Tract: Towards large-scale crowdsensing with high-efficiency swarm path planning. In *Adjunct Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing and the 2022 ACM International Symposium on Wearable Computers*, pages 409–414, 2022.

[18] Haoyang Wang, Xuecheng Chen, Yuhan Cheng, Chenye Wu, Fan Dang, and Xinlei Chen. H-swarmloc: Efficient scheduling for localization of heterogeneous mav swarm with deep reinforcement learning. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, pages 1148–1154, 2022.

[19] Jiawei Guo, Haoyang Wang, Wei Liu, Guosheng Huang, Jinsong Gui, and Shaobo Zhang. A lightweight verifiable trust based data collection approach for sensor–cloud systems. *Journal of Systems Architecture*, 119:102219, 2021.

[20] Matthieu Puigt, Farouk Yahaya, Gilles Delmaire, Gilles Roussel, et al. In situ calibration of cross-sensitive sensors in mobile sensor arrays using fast informed non-negative matrix factorization. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3515–3519. IEEE, 2021.

[21] Haomin Yu, Qingyong Li, Yangli-ao Geng, Yingjun Zhang, and Zhi Wei. Airnet: A calibration model for low-cost air monitoring sensors using dual sequence encoder networks. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 1129–1136, 2020.

[22] Yun Cheng, Olga Saukh, and Lothar Thiele. Sensorformer: Efficient many-to-many sensor calibration with learnable input subsampling. *IEEE Internet of Things Journal*, 9(20):20577–20589, 2022.

[23] Sen Qiu, Hongkai Zhao, Nan Jiang, Zhelong Wang, Long Liu, Yi An, Hongyu Zhao, Xin Miao, Ruichen Liu, and Giancarlo Fortino. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Information Fusion*, 80:241–265, 2022.

[24] Yingzhou Lu, Huazheng Wang, and Wenqi Wei. Machine learning for synthetic data generation: a review. *arXiv preprint arXiv:2302.04062*, 2023.

[25] Lei Yan, Mehrdad Sheikholeslami, Wenlong Gong, Wei Tian, and Zuyi Li. Challenges for real-world applications of nonintrusive load monitoring and opportunities for machine learning approaches. *The Electricity Journal*, 35(5):107136, 2022.

[26] Ning Liu, Zhiyuan Wu, Guodong Li, Xinyu Liu, Yue Wang, and Lin Zhang. Maic: Metalearning-based adaptive in-field calibration for iot air quality monitoring system. *IEEE Internet of Things Journal*, 9(17):15928–15941, 2022.

[27] Huatao Xu, Pengfei Zhou, Rui Tan, Mo Li, and Guobin Shen. Limu-bert: Unleashing the potential of unlabeled data for imu sensing applications. *GetMobile: Mobile Computing and Communications*, 26(3):39–42, 2022.

[28] Pengfei Liu, Weizhe Yuan, Jinlan Fu, Zhengbao Jiang, Hiroaki Hayashi, and Graham Neubig. Pre-train, prompt, and predict: A systematic survey of prompting methods in natural language processing. *ACM Computing Surveys*, 55(9):1–35, 2023.

[29] Jingya Zhou, Ling Liu, Wenqi Wei, and Jianxi Fan. Network representation learning: from preprocessing, feature extraction to node embedding. *ACM Computing Surveys (CSUR)*, 55(2):1–35, 2022.

[30] Sichun Luo, Yuanzhang Xiao, Xinyi Zhang, Yang Liu, Wenbo Ding, and Linqi Song. Perfedrec++: Enhancing personalized federated recommendation with self-supervised pre-training. *arXiv preprint arXiv:2305.06622*, 2023.

[31] Danyang Li, Jingao Xu, Zheng Yang, Qian Zhang, Qiang Ma, Li Zhang, and Pengpeng Chen. Motion inspires notion: self-supervised visual-lidar fusion for environment depth estimation. In *Proceedings of the 20th Annual International Conference on Mobile Systems, Applications and Services*, pages 114–127, 2022.

[32] Jieming Bian, Zhu Fu, and Jie Xu. Fedseal: Semi-supervised federated learning with self-ensemble learning and negative learning. *arXiv preprint arXiv:2110.07829*, 2021.

[33] Xiaojin Jerry Zhu. Semi-supervised learning literature survey. 2005.

[34] Weiwei Jiang. Internet traffic prediction with deep neural networks. *Internet Technology Letters*, 5(2):e314, 2022.

[35] Xinrui Zhang, Xingyuan Liang, Hai Wang, Shuai Wang, and Tian He. Patr: Periodicity-aware trajectory recovery for express system via seq2seq model. In *GLOBECOM 2022-2022 IEEE Global Communications Conference*, pages 486–491. IEEE, 2022.

[36] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[37] Mandar Joshi, Danqi Chen, Yinhan Liu, Daniel S. Weld, Luke Zettlemoyer, and Omer Levy. Spanbert: Improving pre-training by representing and predicting spans, 2020.

[38] Vincent Le Guen and Nicolas Thome. Shape and time distortion loss for training deep time series forecasting models. *Advances in neural information processing systems*, 32, 2019.