7/10/22

A FINE-GRAINED AI-BASED MOBILE AIR QUALITY MONITORING AND FORECASTING SYSTEM

Nguyen Phi Le, Huynh Thi Thanh Binh, Do Phan Thuan, Le Minh Thuy (HUST)
Kien Nguyen (Chiba University), Nguyen Cam Ly (Toshiba Corp)

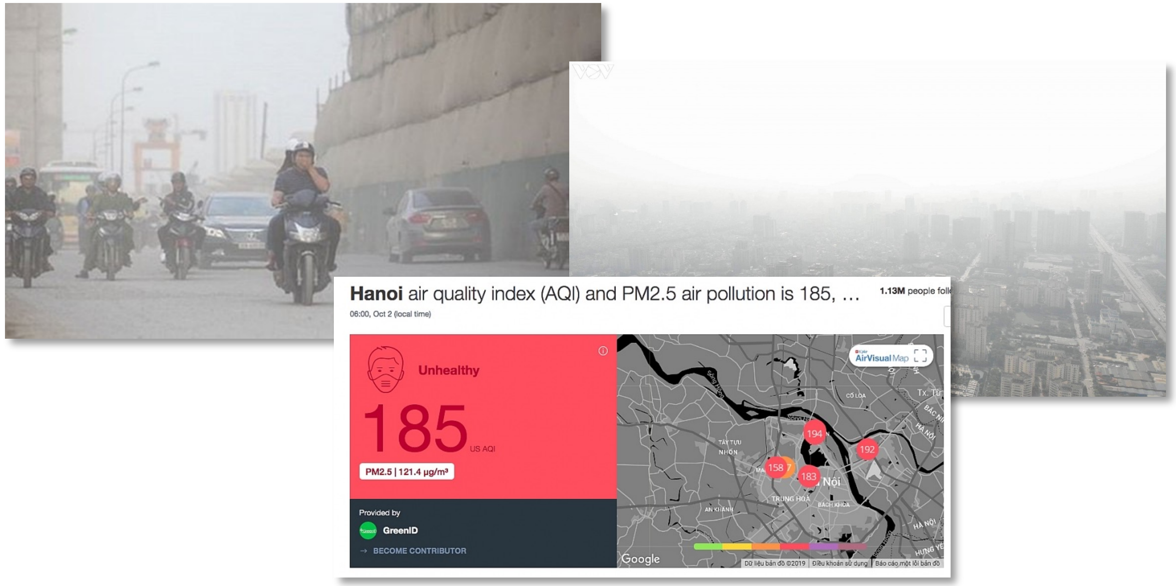
Agenda

- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

Agenda

- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

Motivation



- Hanoi's air quality has reached an emergency level
- We need to track the air quality indicators to help
  - civilians have a plan for protecting their health
  - experts investigate the causes and find solutions
  - policymakers come up with strategies in time

Traditional monitoring stations:

- Fixed locations
- High cost
  → limited number: only 50 stations in Hanoi
- Data is not public

Goals and Solutions

Fine-grained
Air quality map

Low cost

3. Utilize AI to fill in unmonitored regions

Even with mobile sensors, we cannot cover all regions → use AI to interpolate

4. Utilize AI to enhance the accuracy

Sensory data is not correct → use AI to calib

1. Use low cost sensors

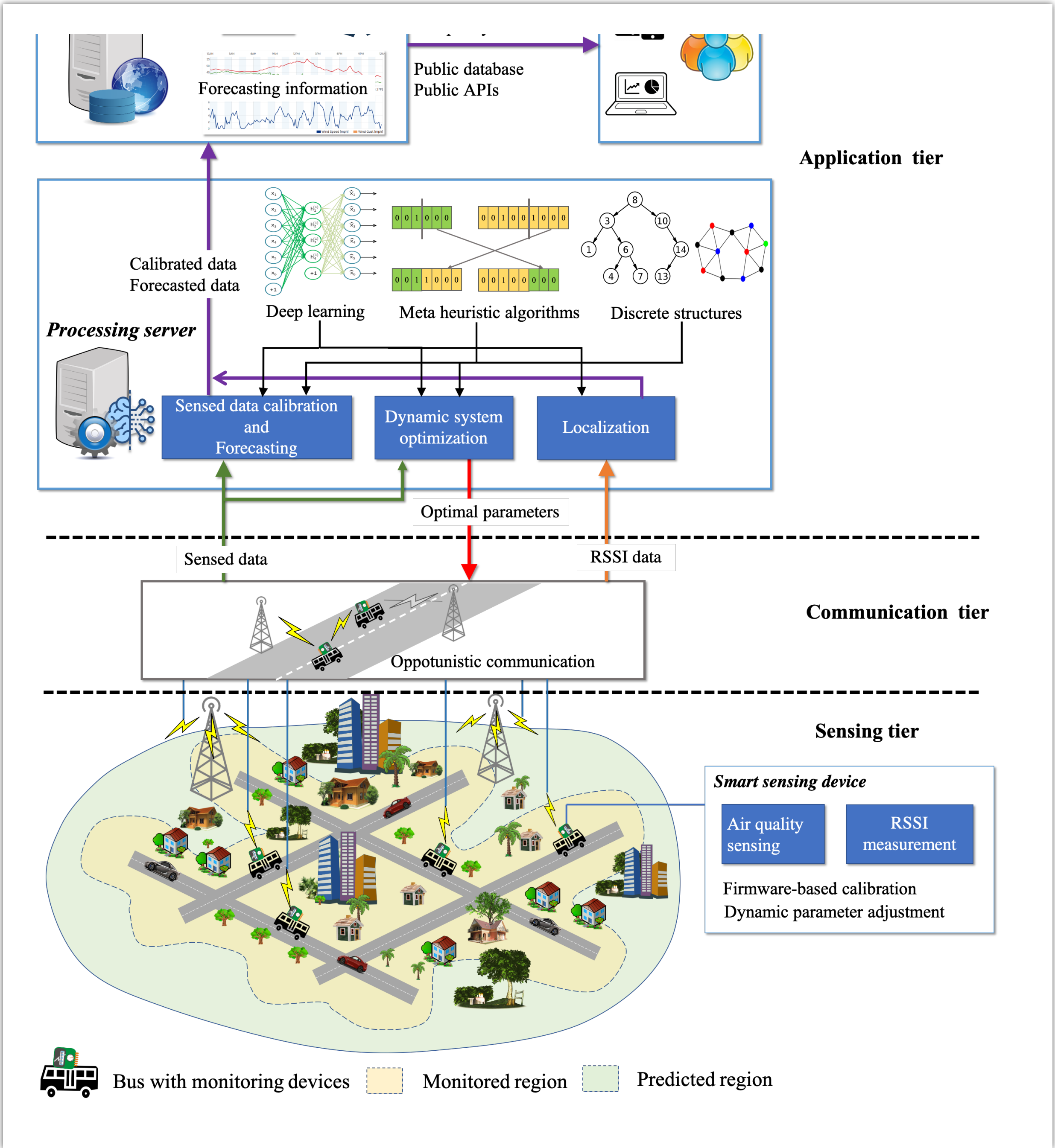
How many sensors do we need?
→ thousands → still expensive

2. Leverage the mobility of vehicles

Put the sensors on vehicle → we can monitor a large area with only one device

Agenda

- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work



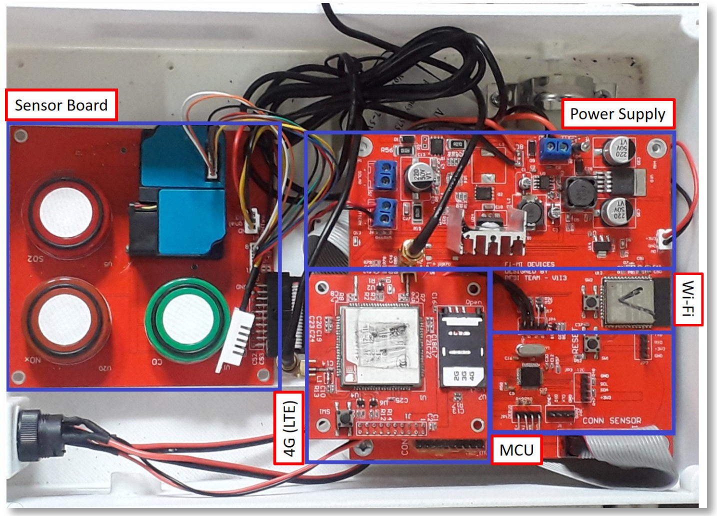
Fi-Mi Architecture

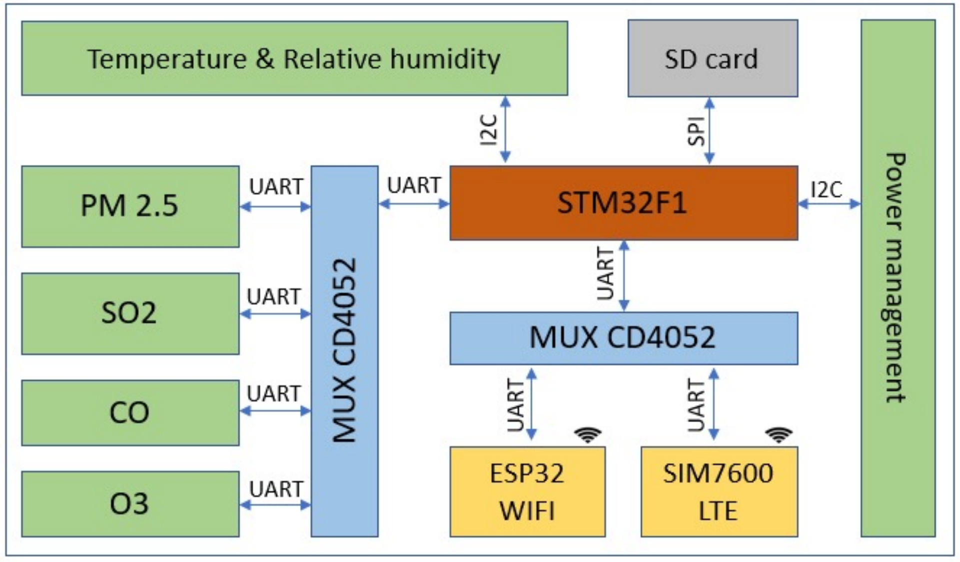
- **Sensing tier**
  - Collects **real-time** air quality data
  - Carried by air **monitoring devices** deployed on **vehicular devices** such as buses
- **Communication tier**
  - **Transfers data** between the monitoring devices and the servers
- **Application tier**
  - **Calibrates** and **Stores** the sensory data
  - **Predicts** air quality in **un-monitored regions**
  - **Forecasts** the **future** trend of the air quality
  - Provides information to users through **smartphone application** and a **web portal**

Agenda

- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

Hardware implementation





Block function of Fi-Mi device

- MCU STM32F103C6T8
- Collect sensor data
- Send data with Wi-Fi or LTE
- External memory and power management

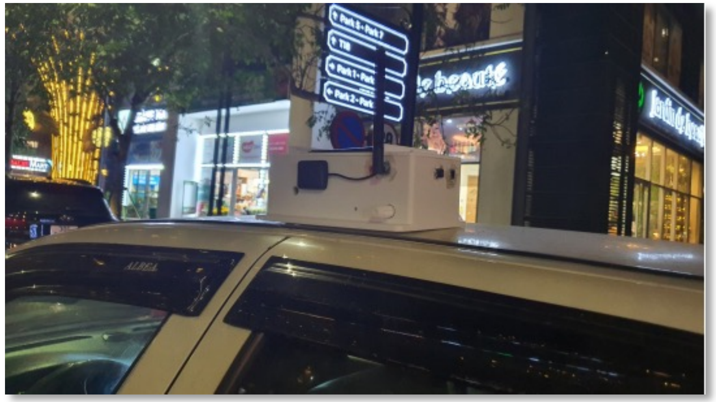
Experiments

Static experiment

	Average latency(s)	Delivery ratio
Sensory data reading	0.26	
Wi-Fi Transmission	0.32	97%
4G (LTE)Transmission	0.99	100%

Packet loss in trial

Communication channel	Sent packets	Received packets	Delivery ratio
Wi-Fi	1200	1198	99.8%
4G (LTE)	1200	1200	100%

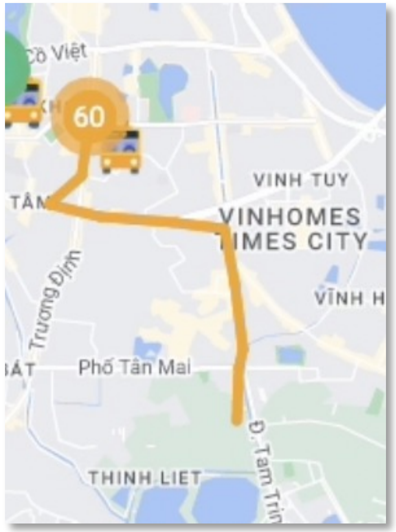


Fi-Mi monitoring device mounted on car

PM2.5 data reported by Fi-Mi device

Deployment on buses

- We already deployed two Fi-Mi devices on Hanoi buses from Apirl
- We will finish deploying 25 devices in July



Agenda

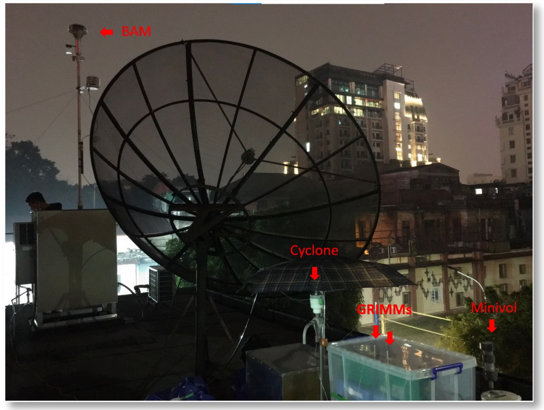
- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

AI core

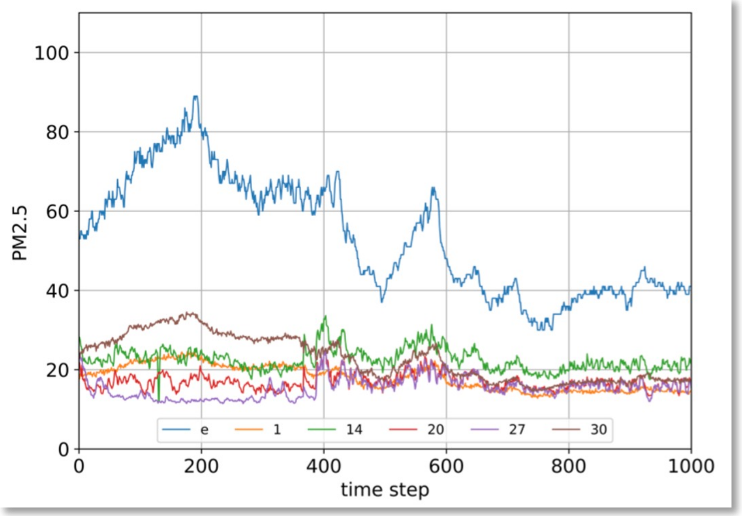
- Data calibration
- Air quality forecasting

Data calibration

- Reference device: GRIMM 107
- Observation
  - The gaps between reference device and Fi-Mi devices are large
  - Fi-Mi devices show different temporal patterns and covariate shifts
- Challenges
  - How to efficiently calibrate Fi-Mi devices?
  - Can we use a single model to calibrate all devices at the same time?



GRIMM 107



Reference device

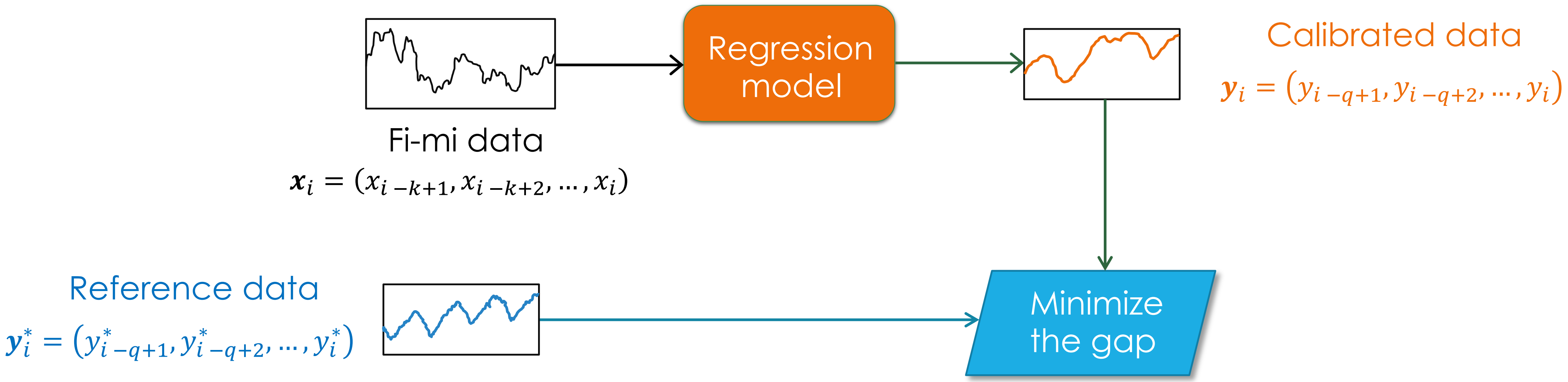
Fi-Mi devices

Single device calibration

- Problem formulation
  - Given
    - Data collected from a Fi-mi device
      - $\boldsymbol{x}_i = (x_{i-k+1}, x_{i-k+2}, \dots, x_i)$
    - Data collected from the reference device at the same location
      - $\boldsymbol{y}_i^* = \left(y_{i-q+1}^*, y_{i-q+2}^*, \dots, y_i^*\right)$
  - Objective
    - Calibrating the data of Fi-mi device to make it as close as possible to the reference data
      - $\boldsymbol{y}_i = \left(y_{i-q+1}, y_{i-q+2}, \dots, y_i\right) = f\left(\boldsymbol{x}_i = (x_{i-k+1}, x_{i-k+2}, \dots, x_i)\right)$
      - Minimize: $|\boldsymbol{y}_i^* - \boldsymbol{y}_i|$

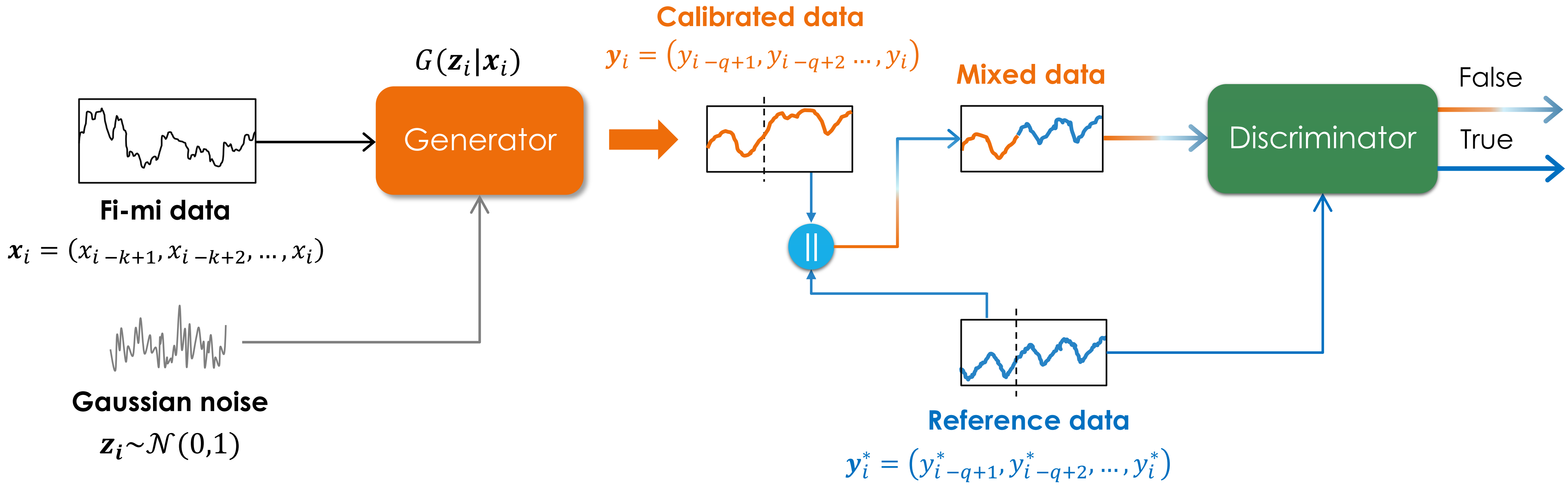
Single device calibration

- Common approach
  - Considering the calibration as a regression task
  - Using some common model architectures and training by MSE loss



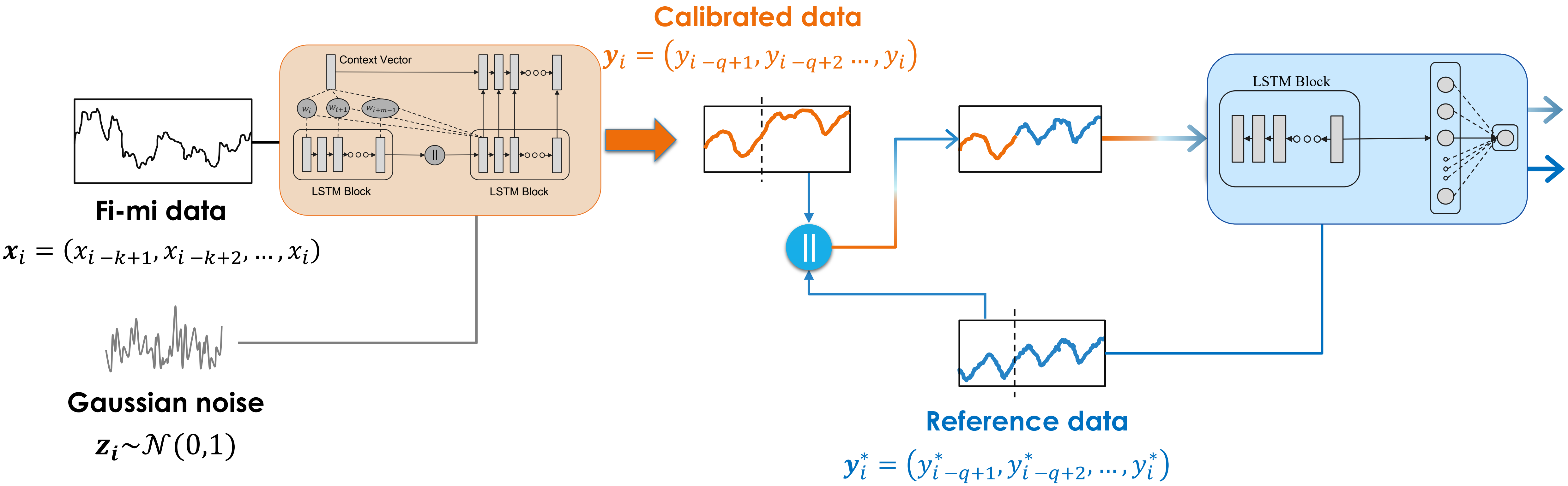
Single device calibration

- Generative Adversarial Learning approach



Single device calibration

- Generative Adversarial Learning approach



Single device calibration

- Generative Adversarial Learning approach: Preliminaries experiment results
  - Reference device: GRIMM 107 (operated and managed by INEST, HUST)

	RMSE	MAE	MAPE	CRPS
Our generator with MSE loss	6.591	2.758	9.085	2.758
XGBoost	29.899	20.986	115.222	20.986
GAN (our)	**5.551**	**2.449**	**9.208**	**2.212**

	RMSE	MAE	MAPE	CRPS
Our generator with MSE loss	1.198	0.862	3.466	0.862
XGBoost	2.454	2.247	15.563	2.454
GAN (our)	**1.110**	**0.849**	**2.937**	**0.633**

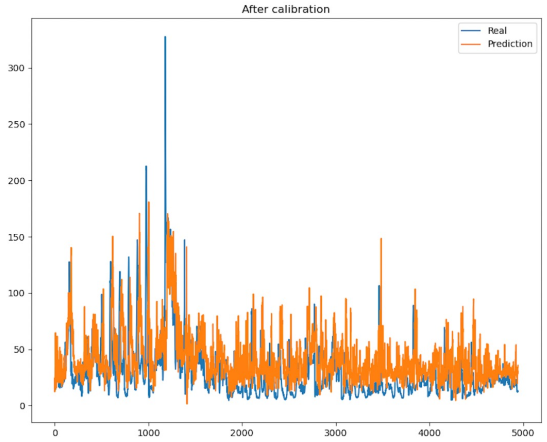
Single device calibration

- Enhancing the calibration accuracy
  - changes the **loss function**
  - uses decomposition method (SSA) to **elliminate the noise and outlier**
  - uses **attention mechanism** to select the most useful components

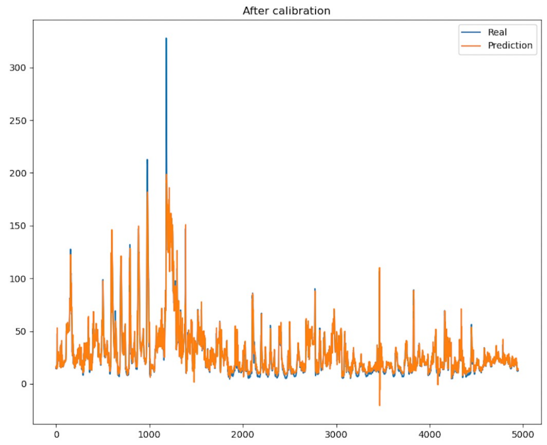
	RMSE	MAE	MAPE	CRPS
ForGAN	3.36	2.27	5.36	2.24
G only	3.76	2.75	6.42	2.75
G only + SSA Attention	3.29	2.1925	**5.22**	2.19
ForGAN + SSA Attention **(Our proposal)**	**3.20**	**2.190**	5.26	**2.14**
XGBoost	16.86	15.48	35.28	15.48
Linear Regression	17.23	16.006	36.70	16.006
Raw Data	20.46	19.3996	45.096	19.3996

groundtruth

calibrated data



XG Boost



Proposal

Multi-devices calibration

- Can we calibrate multiple devices at the same time?
  - Multi-task learning approach
- Challenges
  - Capturing common features of all devices
  - Identifying unique features of each device

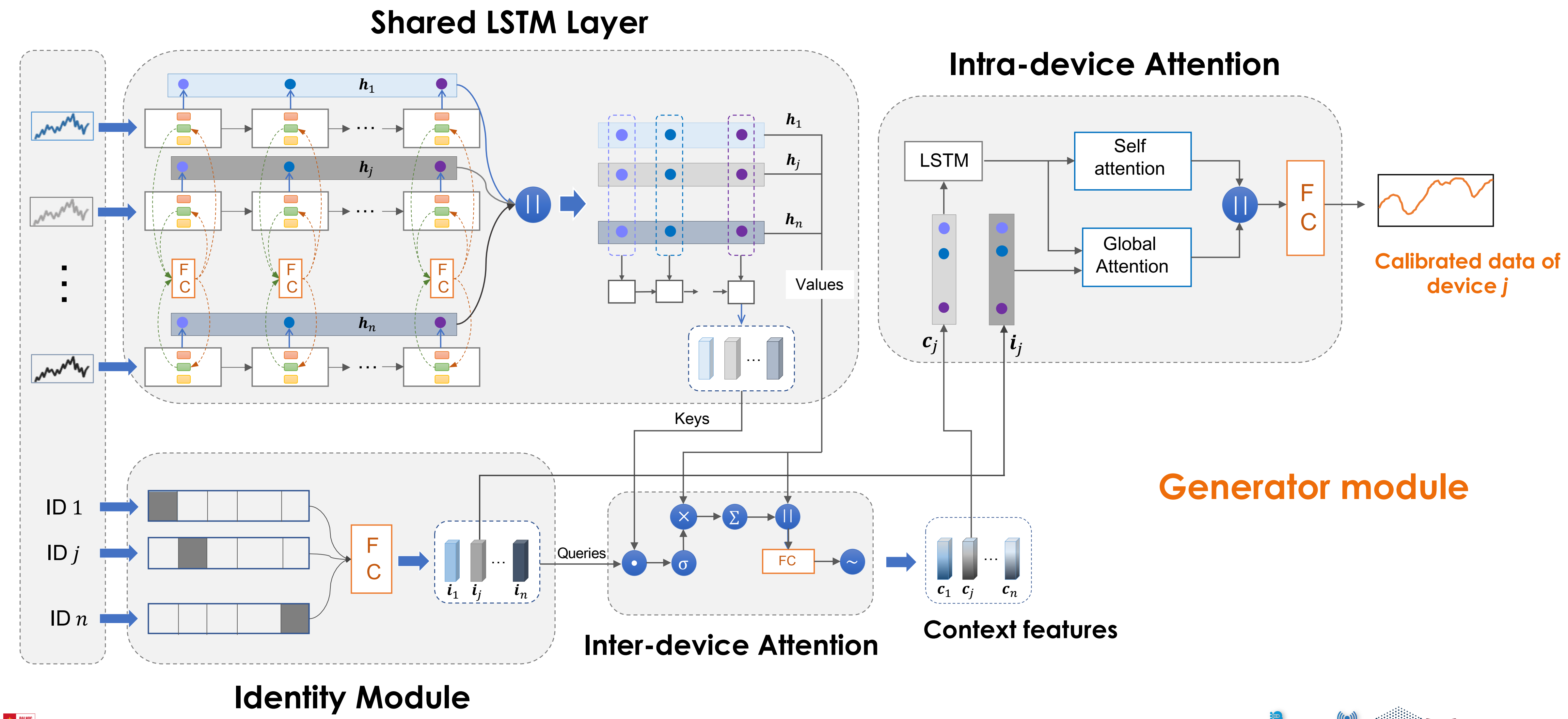
Multi-devices calibration

- Can we calibrate multiple devices at the same time?
  - Multi-task learning approach
- Challenges
  - **Capturing common features of all devices**
    - Shared LSTM layer
      - captures common features of the devices
    - Inter-device Attention block
      - learns relationship between the devices
  - Identifying unique features of each device

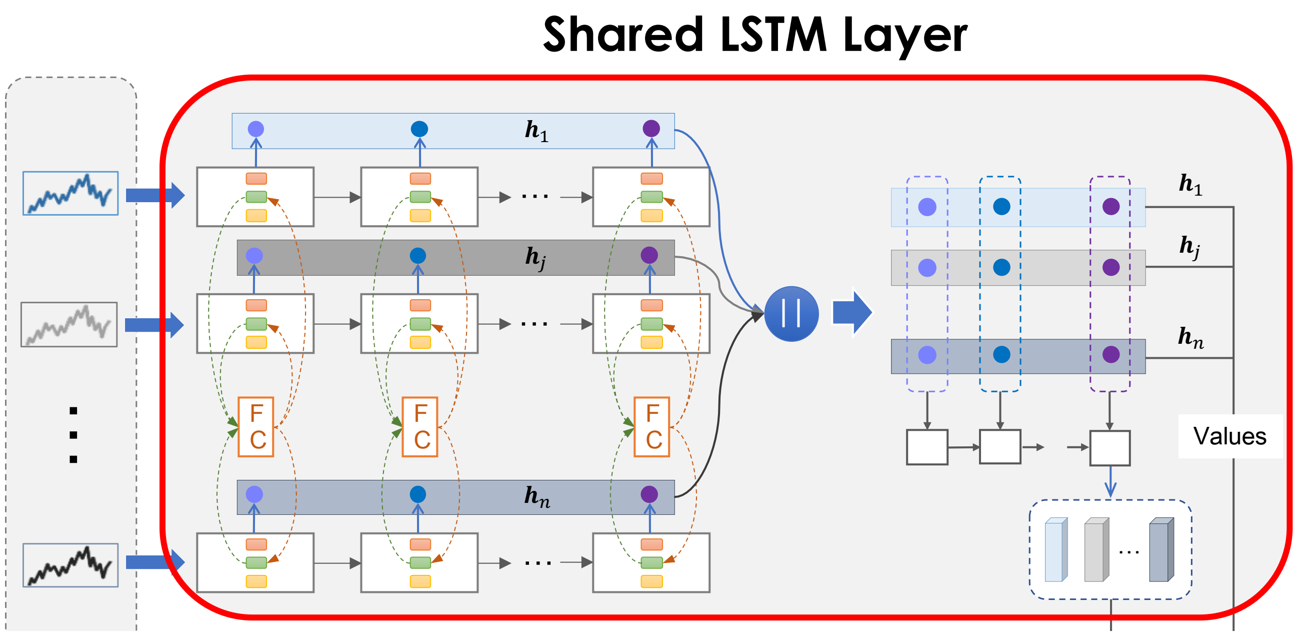
Multi-devices calibration

- Can we calibrate multiple devices at the same time?
  - Multi-task learning approach
- Challenges
  - Capturing common features of all devices
  - **Identifying unique features of each device**
    - Identity module
      - identifies the devices
    - Intra-device Attention block
      - learns characteristic inside every device

Multi-devices calibration



Multi-devices calibration



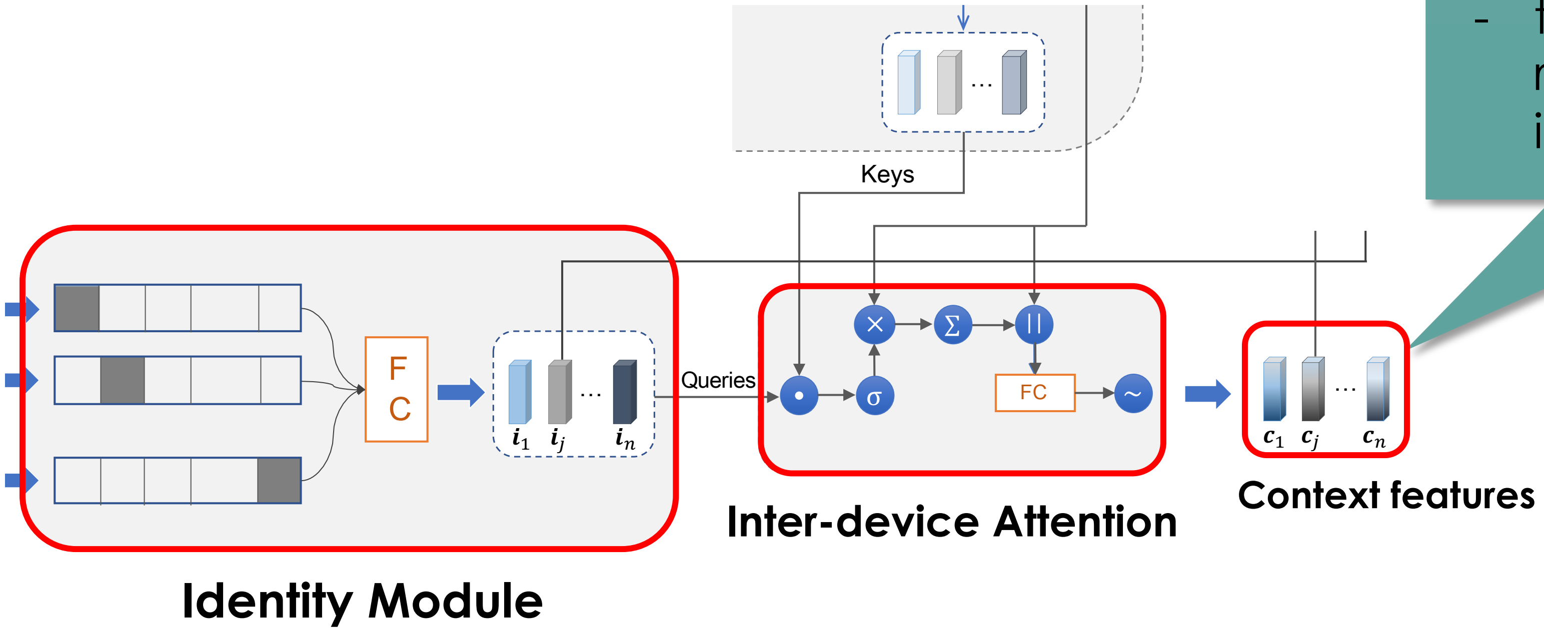
LSTM layer is **shared among all devices**

- **learns common features** of all devices
- leverages information from other devices to **enhance information** extracted for one device

Multi-devices calibration

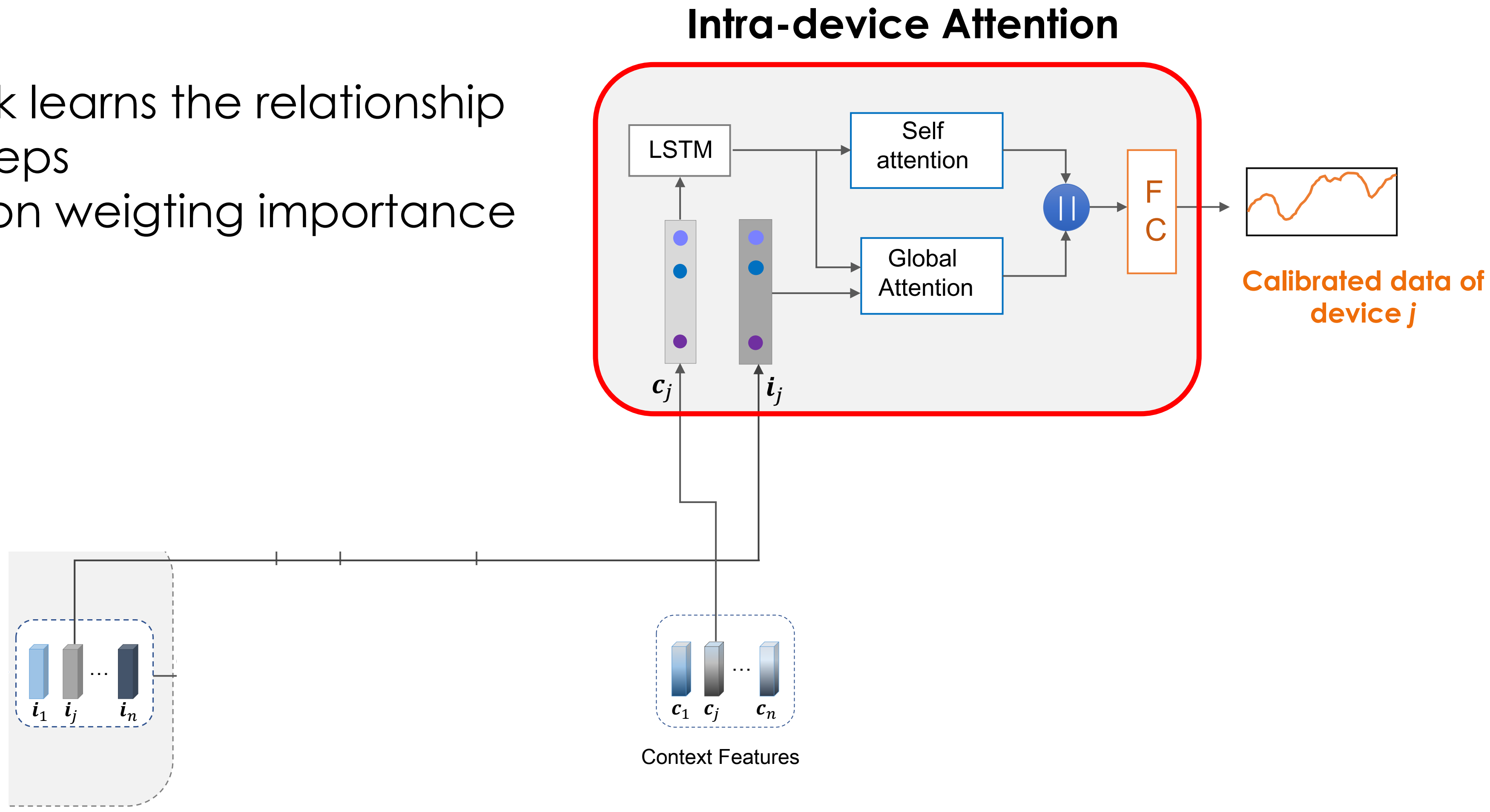
- Identity module helps to identify the device
- For each device, the inter-device attention block weights the impacts of other devices

Representing information of
- the device of interest
- the other devices that relate to the device of interest



Multi-devices calibration

- Self-attention block learns the relationship among the time steps
- The Global attention weigting importance of every time steps



Multi-devices calibration

Model	MSE	MAE	MAPE
CRNN [1]	53.71	5.2	15.79
AECRNN [1]	63.33	5.47	16.62
MTLGRU [2]	52.07	4.49	13.62
MSJF [3]	48.29	5.33	16.18
SPA [3]	38.43	5.09	15.45
FimiCalib-1model	30.51	4.47	**13.58**
FimiCalib-MSE (Our)	25.38	4.16	12.92
FimiCalib-cGAN (Our)	**20.17**	**3.47**	**10.59**

Reduces MAPE by 22% compared to the best baseline

[1] Cirstea, Razvan-Gabriel, et al. "*Correlated time series forecasting using multi-task deep neural networks.*" Proceedings of the 27th ACM international conference on information and knowledge management. 2018.
[2] Zhang, Kunpeng, et al. "*A multitask learning model for traffic flow and speed forecasting.*" IEEE Access 8 (2020): 80707-80715.
[3] Ma, Tao, and Ying Tan. "*Multiple stock time series jointly forecasting with multi-task learning.*" 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.

Multi-devices calibration

Model	Parameters	Inference time (s) (5 devices)
CRNN [1]	2072519	0.0032
AECRNN [1]	3920204	0.007
MTLGRU [2]	1010708	0.0019
MSJF [3]	516116	0.0007
SPA [3]	565268	0.0008
FimiCalib-Nmodel	3518611	0.0024
FimiCalib-cGAN (Our)	1820835	0.0022

Competitive parameter number and inference time

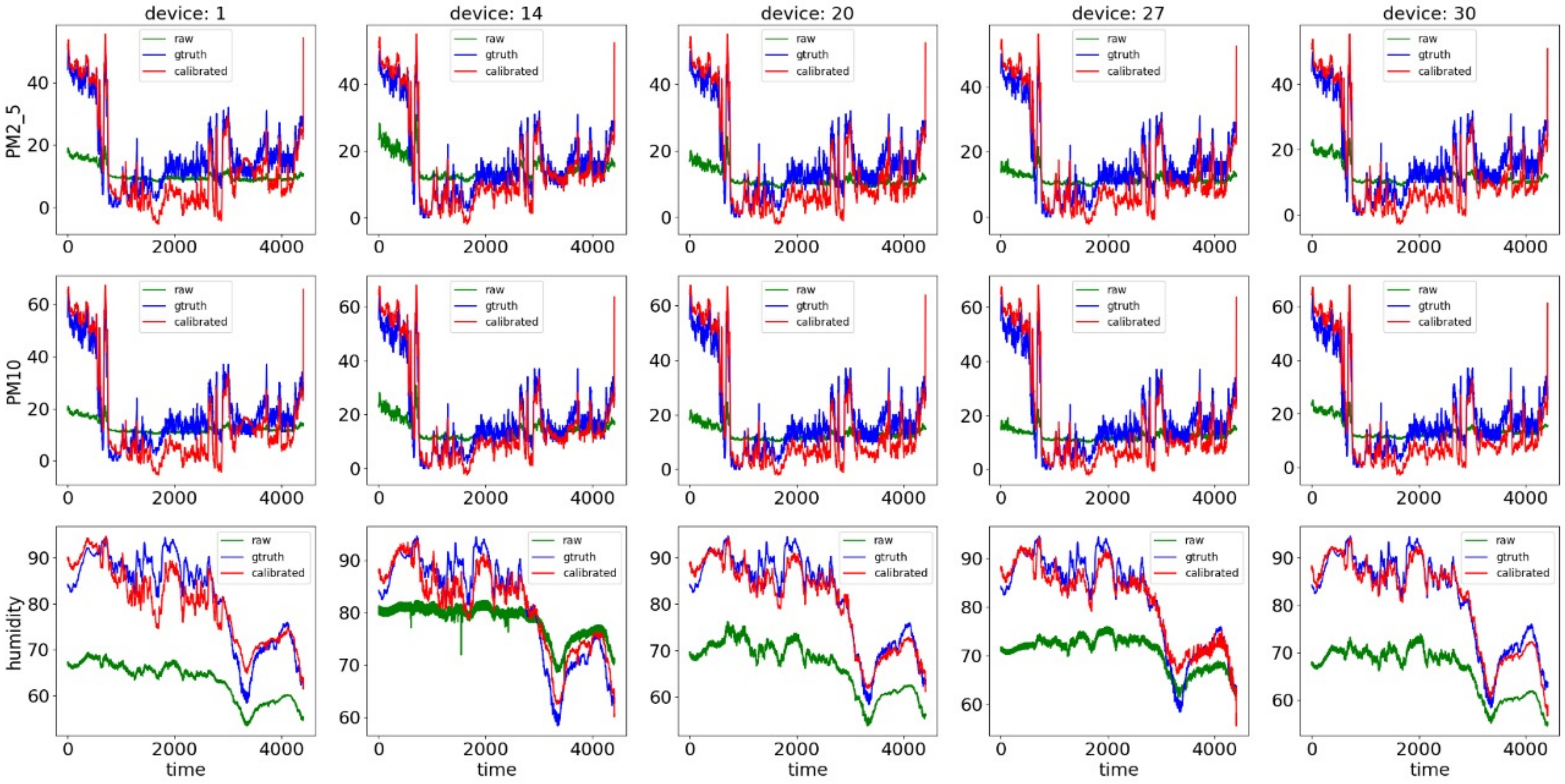
[1] Cirstea, Razvan-Gabriel, et al. "*Correlated time series forecasting using multi-task deep neural networks.*" Proceedings of the 27th ACM international conference on information and knowledge management. 2018.

[2] Zhang, Kunpeng, et al. "*A multitask learning model for traffic flow and speed forecasting.*" IEEE Access 8 (2020): 80707-80715.

[3] Ma, Tao, and Ying Tan. "*Multiple stock time series jointly forecasting with multi-task learning.*" 2020 International Joint Conference on Neural Networks (IJCNN). IEEE, 2020.

Multi-devices calibration

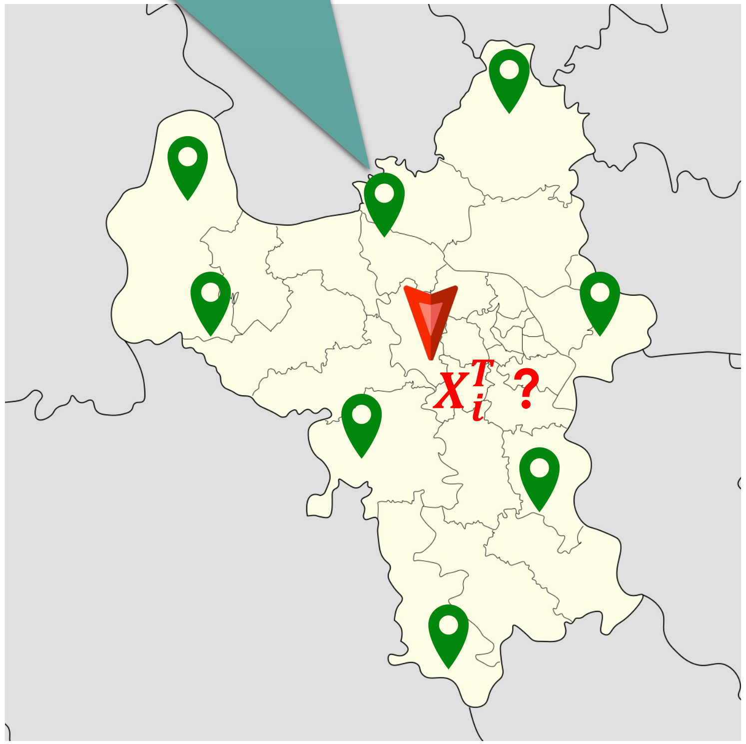
◦ Visualization of calibration results



Air quality interpolation

- Problem Formulation
  - **Given:** Data collected from n monitoring stations
    - Air quality indicators: $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$
    - Meteorology data: $\boldsymbol{M}_1, \ldots, \boldsymbol{M}_n$
  - **Objective:** air quality indicators at an arbitrary location S_x at the current time T
    - Currently, we focus on forecasting of PM2.5 as it is the most important air quality indicator
- Assumption: the meteorology data at S_x is available

Data collected from i-th station, at time step t
X_i^t: Air quality indicators
M_i^t: Meteorology data



Design principle

- Spatio-temporal dependency
  - Temporal dependency: current air quality value is often relevant to its historical data
  - Spatial dependency: air quality at a location often relates to that at neighboring locations
- Multi-modal information
  - Air quality data consists multivariate features (different air quality indicators, meteorology features)
  - Some of features might have more impact to PM2.5 than the others
- Interpolation capability
  - Lack of historical air quality data at the targeted location

Design principle

Design an architecture that capable of capturing both these information at the same time

- Spatio-temporal dependency
  - Temporal dependency: current air quality value is often relevant to its historical data
  - Spatial dependency: air quality at a location often relates to that at neighboring locations
- Multi-modal information
  - Air quality data consists multivariate features (different air quality indicators, meteorology features)
  - Some of features might have more impact to PM2.5 than the others
- Interpolation capability
  - Lack of historical air quality data at the targeted location

Design principle

- Spatio-temporal dependency
  - Temporal dependency: current air quality valu ... ta
  - Spatial dependency: air quality at a locati ... locations
- Multi-modal information
  - Air quality data consists multivariate features (different air quality indicators, meteorology features)
  - Some of features might have more impact to PM2.5 than the others
- Interpolation capability
  - Lack of historical air quality data at the targeted location

Effectively integrate the features into the model to boost the accuracy of PM2.5 prediction

Design principle

- Spatio-temporal dependency
  - Temporal dependency: current air quality value is often relevant to its historical data
  - Spatial dependency: air quality at a location often relates to that at neighboring locations
- Multi-modal information
  - Air quality data consists multivariate features meteorology features)
  - Some of features might have more i
- Interpolation capability
  - Lack of historical air quality data at the targeted location

modeling the correlation between the locations based on available stations and generalize to arbitrage places

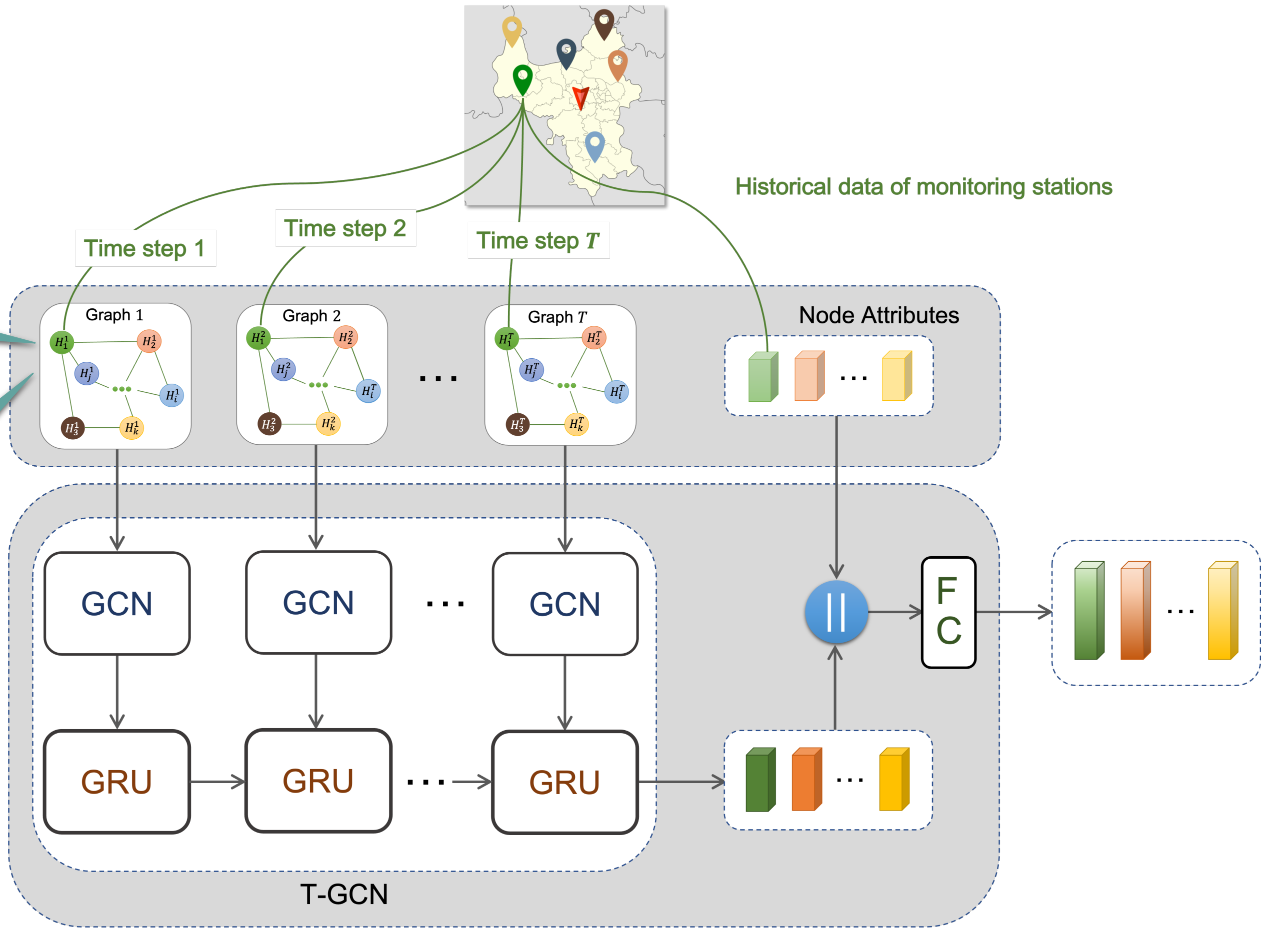
Our solution (1)

- Modelling the spatio-temporal dependency
  - Temporal graph convolution network (T-GCN) = Gated Recurrent Network (GRU) + Graph Convolutional Network (GCN)
    - GRU: strong capability in handling sequence data
    - GCN: effectively capture the relationship between nodes in spatial domain using node feature
  - Corrupt function utilizing both the global view corruption and feature level corruption.

Spatio-temporal graph representation learning

Represents information of all monitoring stations at a time step

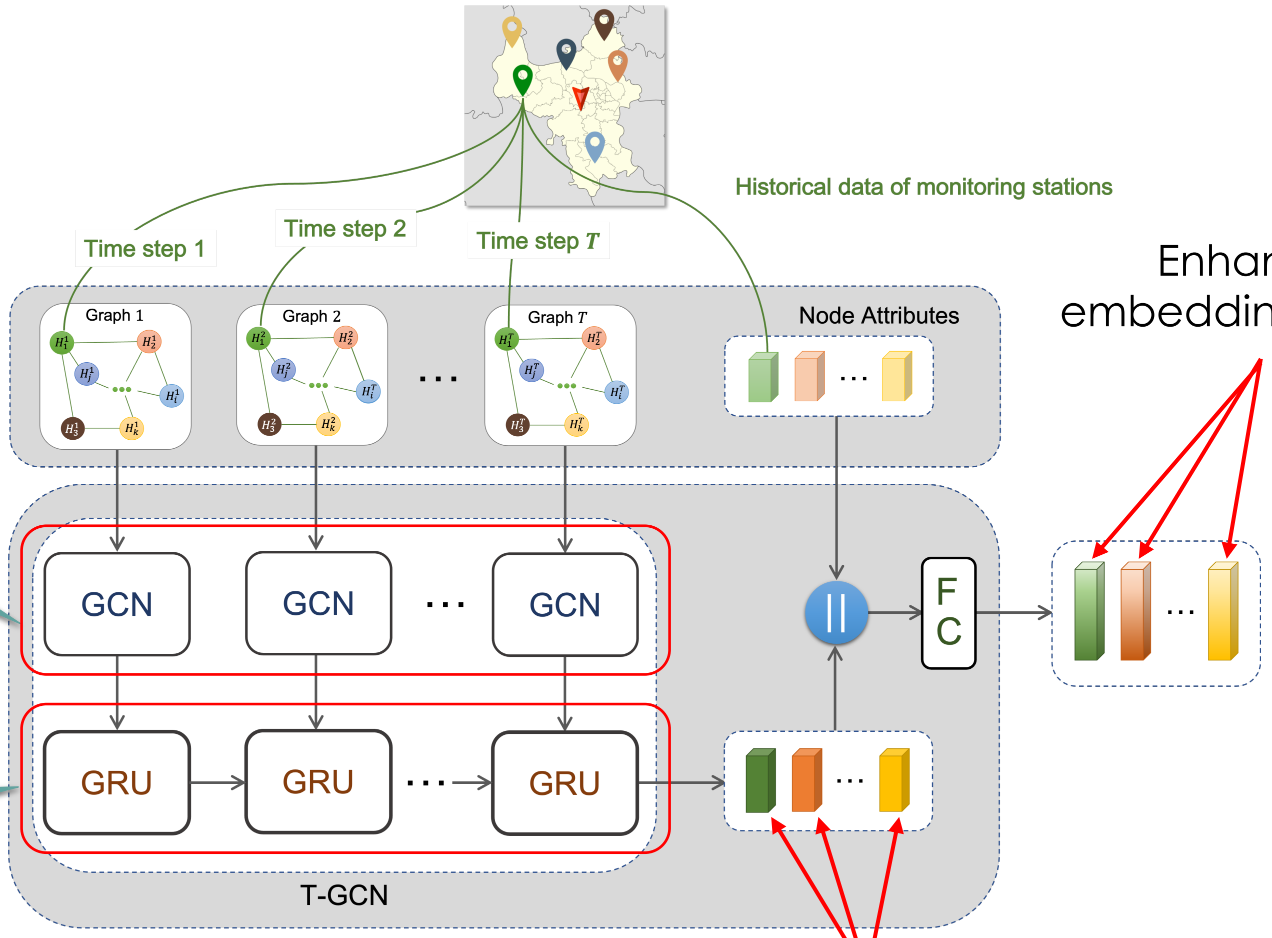
Nodes = {monitoring stations}
Atributes = {Aiquality indicators, meteorology}
Edges' weight = {inverse of distances}



Spatio-temporal graph representation learning

Models the spatial relationship between the monitoring stations

Models the temporal correlation of the data accross T timesteps



Enhanced embedding vectors

Each vector captures information of a monitoring station and its relationship with other stations

Our solution (2)

- Handling multi-modal information
  - Discuss with experts to find features which may affect PM2.5 the most
    - Wind speed, Win direction, Temperature, Pressure, Precipitation,
  - Propose
    - A data preprocessing process to extract useful information related to wind
      - analyzes the wind direction from the neighboring stations to the targeted location
      - calculates the wind speed at the targeted location based
    - A feature-aware attention mechanism
      - scores the importance of each feature
      - highlights the relevant features to PM2.5

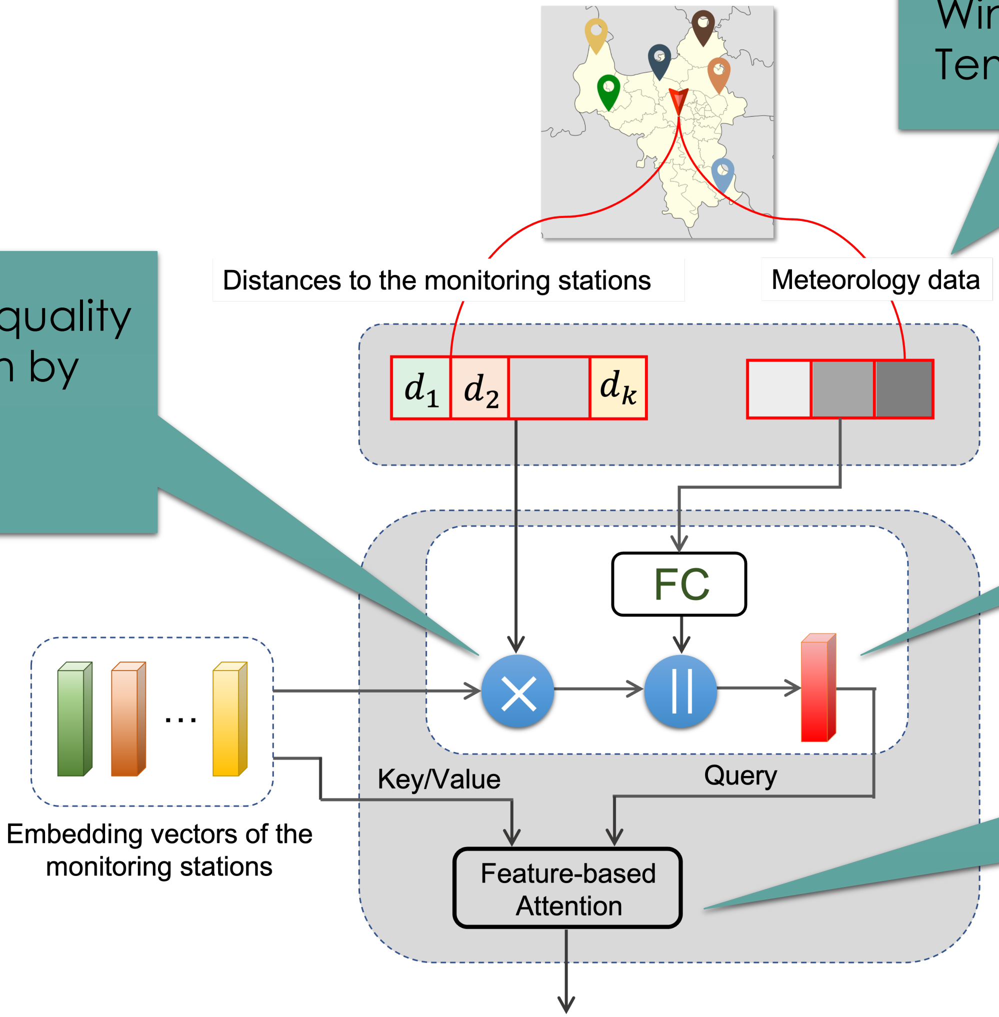
Handling multi-modal information

Wind speed, Win direction, Temperature, Pressure

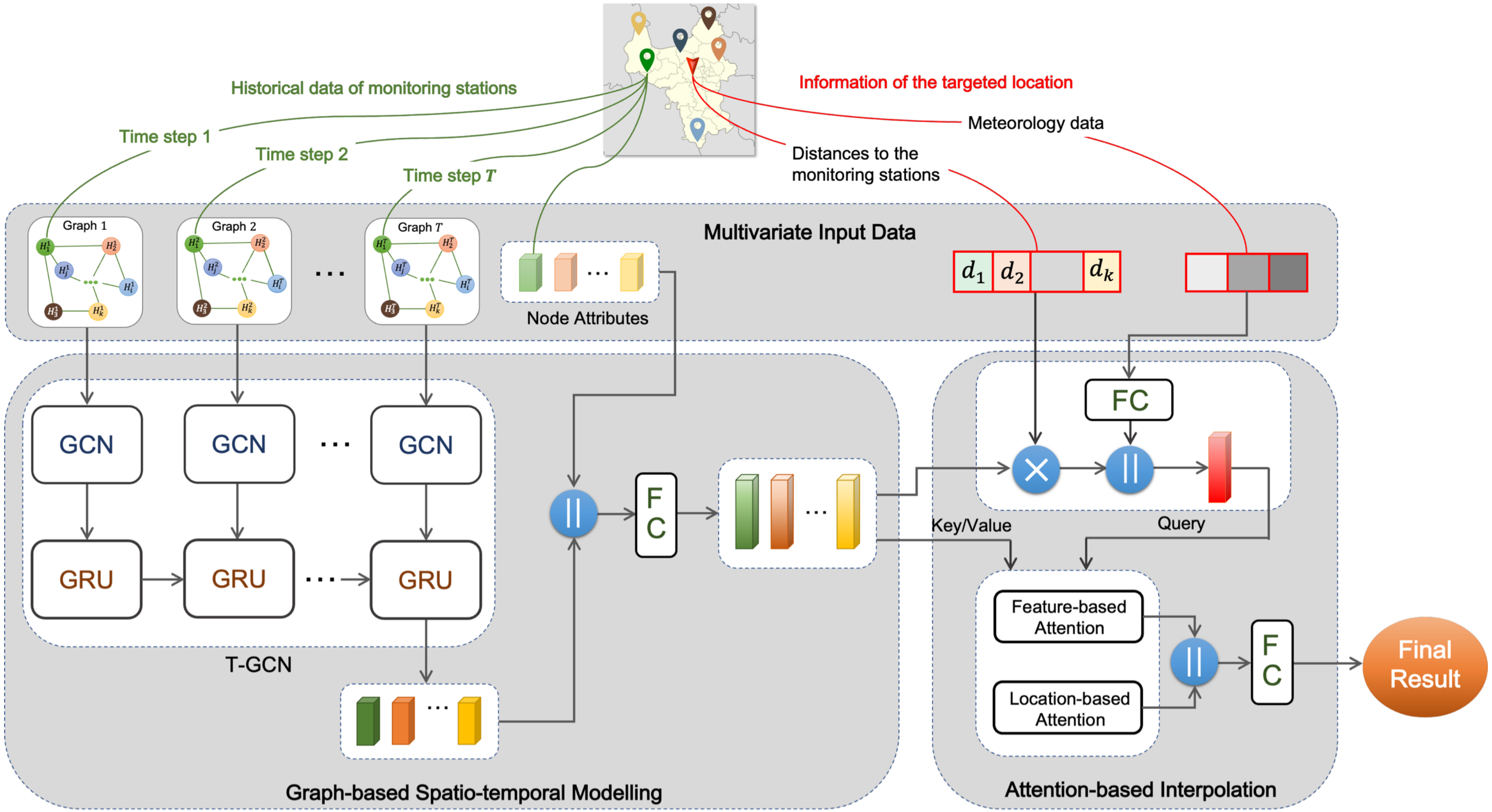
Roughly interpolate air quality at the targeted location by using information of the monitoring stations

Concatenate meteorology data and the interpolated data

Use attention mechanism to highlight most relevant features



Enhancing interpolation capability



Experiment results

- Datasets
  - Beijing Dataset
    - collects the air quality and meteorological information of 35 stations across Beijing
    - covers an area of 16,441 km^2
  - UK Dataset
    - collects the air quality and meteorological information of 141 stations in UK
    - covers an area of 242,295 km^2
- Research questions
  - Does our proposed model outperform the baseline methods?
  - How important is each design choice affect our model?
  - How does the strategy of selecting training station affect the model result?

Comparison with SOTA

Dataset	Model	MAE	RMSE	MAPE	R2
Beijing	BiLSTM-IDW	13.25	18.28	0.61	0.85
	KIDW-TCGRU	16.28	20.38	0.78	0.76
	GEDE	**10.80**	**16.7**	**0.55**	**0.871**
	IDW	11.62	17.74	0.72	0.86
UK	BiLSTM-IDW	2.59	3.6	0.39	0.152
	KIDW-TCGRU	2.85	4.18	0.52	0.31
	GEDE	**2.16**	**3.19**	**0.36**	**0.39**
	IDW	2.45	4.07	0.4	0.18

Our proposed method achieves the best performance

[BiLSTM-IDW] J. MA, Y. Ding, V. Gan, C. Lin, and Z. WAN, "Spatiotemporal prediction of pm2.5 concentrations at different time granularities using IDW-BLSTM," IEEE Access, vol. PP, pp. 1–1, 08 2019.
[KIDW-TCGRU] C. Guo, G. Liu, L. Lyu, and C.-H. Chen, "An unsupervised
pm2.5 estimation method with different spatio-temporal resolutions based on KIDW-TCGRU," IEEE Access, vol. 8, pp. 190 263– 850
190 276, 2020.

Ablation study

- GEDE-1: remove the local attention
- GEDE-2: remove the global attention
- GEDE-3: remove both two attention mechanisms
- GEDE-4: remove the GCN
- GEDE-5: remove the GRU units
- GEDE-6: remove the the graph module
- GEDE-7: remove the meteorology data of the targeted location
- GEDE-8: remove the node-feature corruption from the corrupting function

Dataset	Model	GEDE	GEDE-1	GEDE-2	GEDE-3	GEDE-4	GEDE-5	GEDE-6	GEDE-7	GEDE-8
Beijing	MAE	**10.21**	10.44	10.93	11.24	12.91	13.37	15.22	12.23	13.46
	RMSE	**15.22**	16.06	17.2	16.85	17.89	18.89	27.6	17.6	25.54
	MDAPE	**0.21**	0.21	0.25	0.54	0.31	0.31	0.37	0.28	0.33
	MAPE	**0.48**	0.48	0.52	0.25	0.71	0.7	1.02	0.6	0.83
	R2	**0.89**	0.88	0.86	0.86	0.85	0.83	0.82	0.85	0.85
UK	MAE	**2.16**	2.27	2.23	2.37	2.51	2.31	2.33	2.36	2.35
	RMSE	**3.19**	3.37	3.2	3.39	3.65	3.39	3.29	3.52	4.96
	MDAPE	**0.23**	0.24	0.24	0.39	0.28	0.26	0.26	0.24	0.24
	MAPE	**0.36**	0.39	0.39	0.28	0.48	0.37	0.42	0.39	0.39
	R2	**0.39**	0.34	0.31	0.32	0.18	0.29	0.272	0.22	0.21

Impacts of training station selection strategy

Dataset	Station selection method	MAE	RMSE	MDAPE	MAPE	R2
Beijing	Distance-based method	10.44	16.06	0.21	0.48	0.88
	Correlation-based method	**9.23**	**14.18**	**0.2**	**0.55**	**0.9**
	Random-based method	12.53	17.35	0.32	1.02	0.86
UK	Distance-based method	2.16	3.19	0.23	0.36	0.39
	Correlation-based method	**1.78**	**2.77**	**0.19**	**0.31**	**0.6**
	Random-based method	2.2	3.15	0.25	0.35	0.32

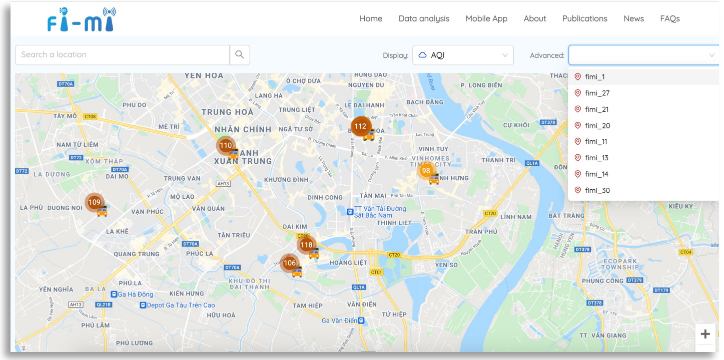
Selection the training stations by correlation achieves the best performance

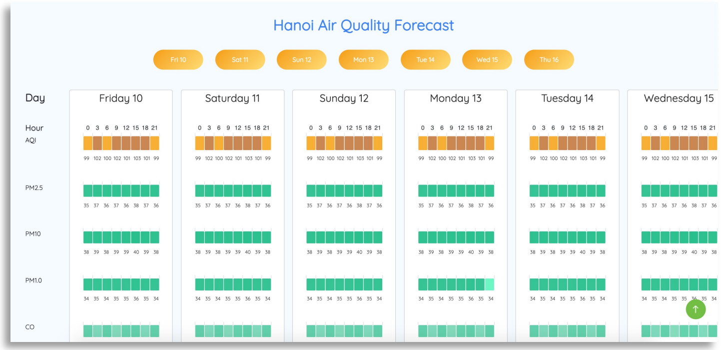
Agenda

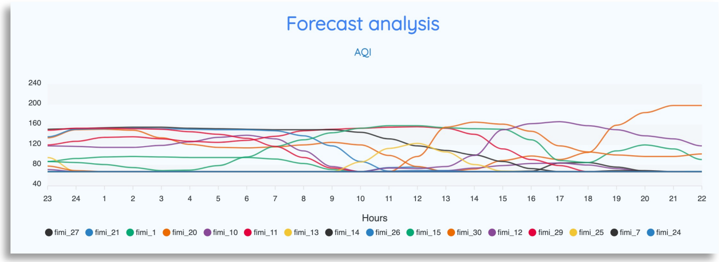
- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

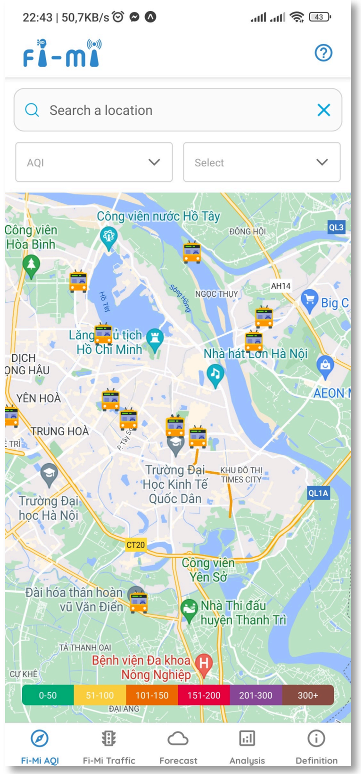
Software implementation

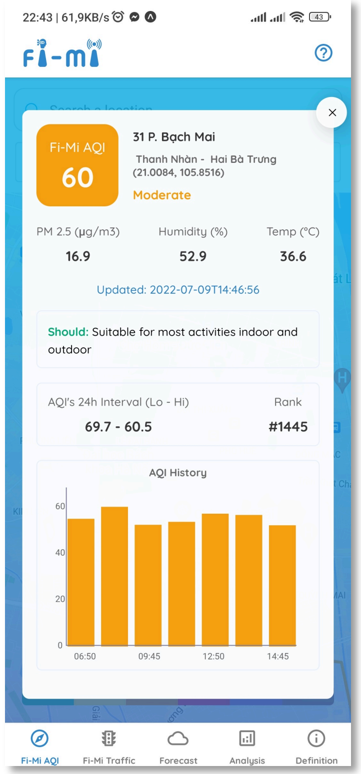
A Realtime air quality monitoring and forecasting webpage and smartphone app

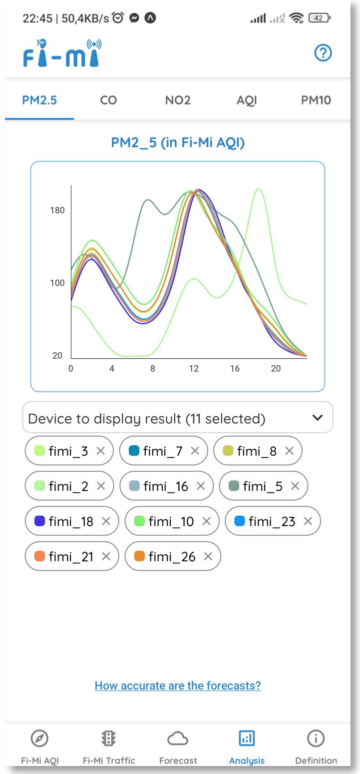


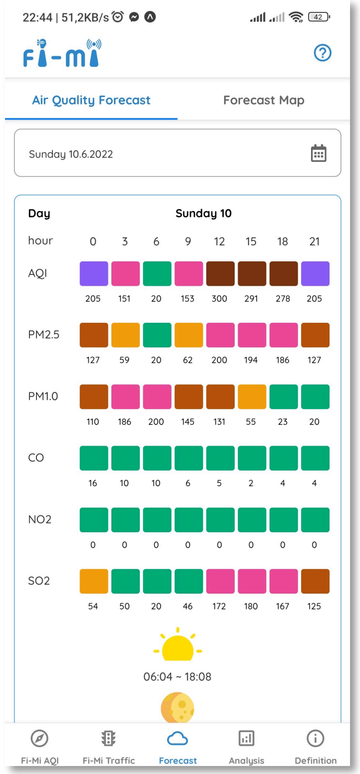












Software implementation

- Demo

Agenda

- Overview of the project
- Fi-Mi architecture
- Hardware implementation
- AI core
- Software implementation
- Conclusion and Future work

Conclusion and Future work

- What we have done
  - Proposed a mobile air quality monitoring system
  - Implemented 30 devices
  - Finished the testing phase, going to the real deployment
  - Implemented the software system
  - Proposed AI models for calibration and forecasting
- Future work
  - Real deployment
  - Test the proposed method on the real system


Thank You!