Anti-Aging Calibration Methodology with User Log-Oriented Anomaly Detection for Wi-Fi Fingerprinting Localization

Kohei Yamamoto
Ubiquitous Computing and Networking Lab

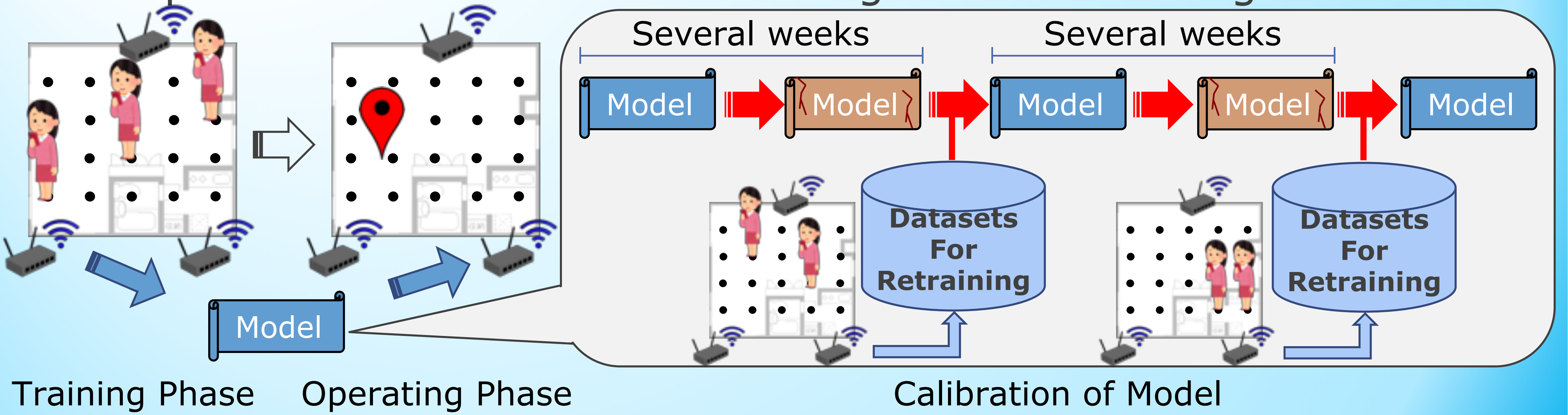
Motivation & State of the Art

Wi-Fi Fingerprinting localization is major

But accuracy of model deteriorate with age

- Caused by environmental changes

Requires **calibration** of model using transfer learning



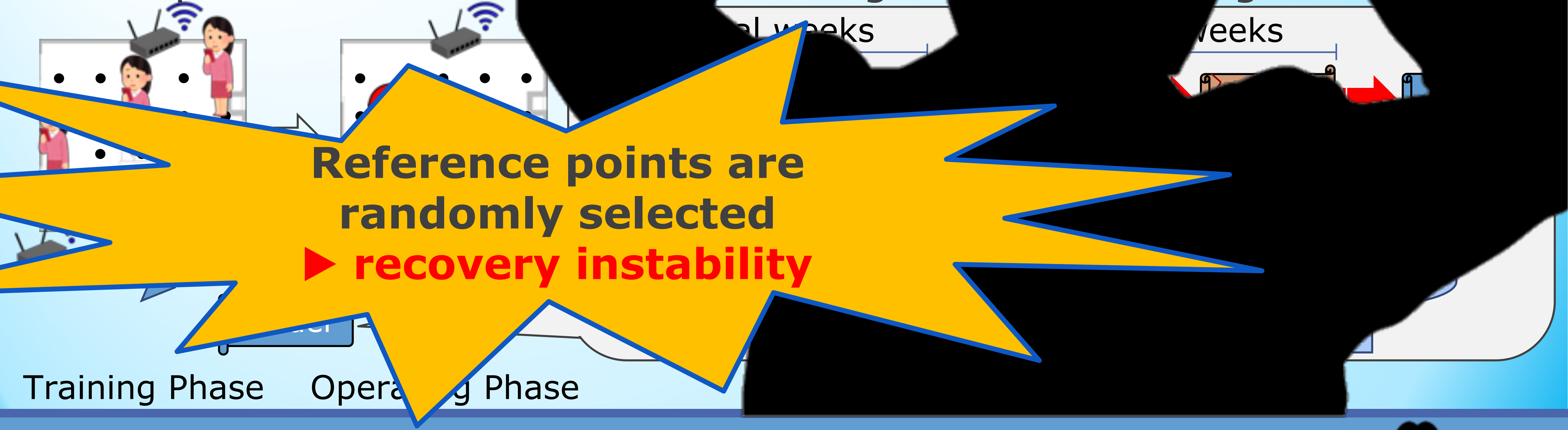
Motivation & State of the Art

Wi-Fi Fingerprinting localizati

But accuracy of model d

- Caused by environment

Requires **calibratio** el using tra ng



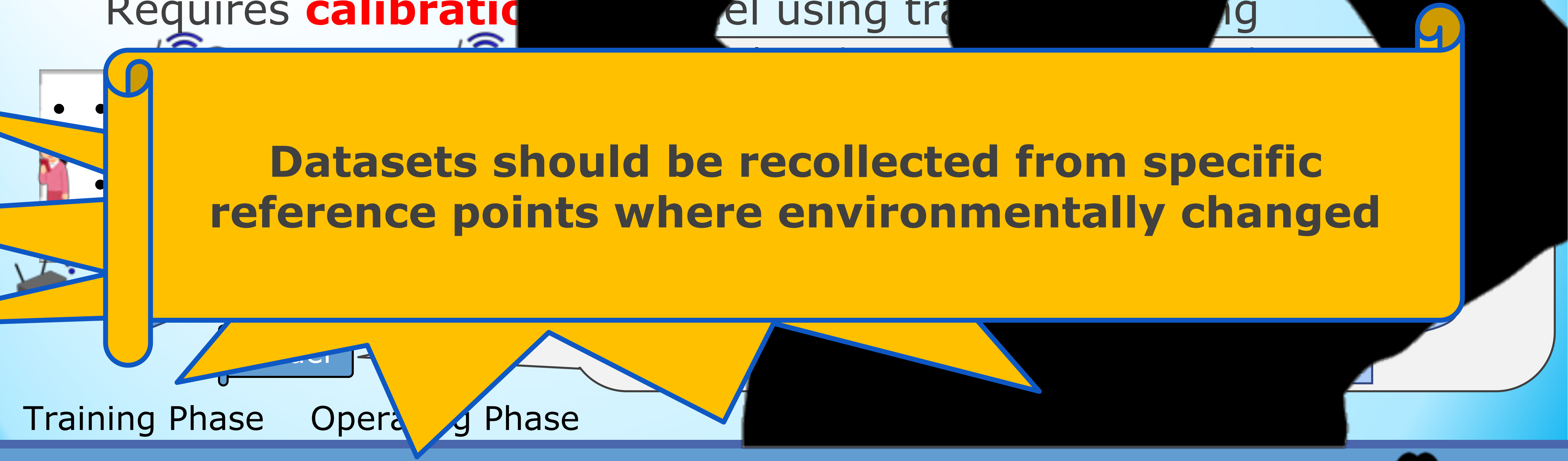
Motivation & State of the Art

Wi-Fi Fingerprinting localizati

But accuracy of model d

- Caused by environment

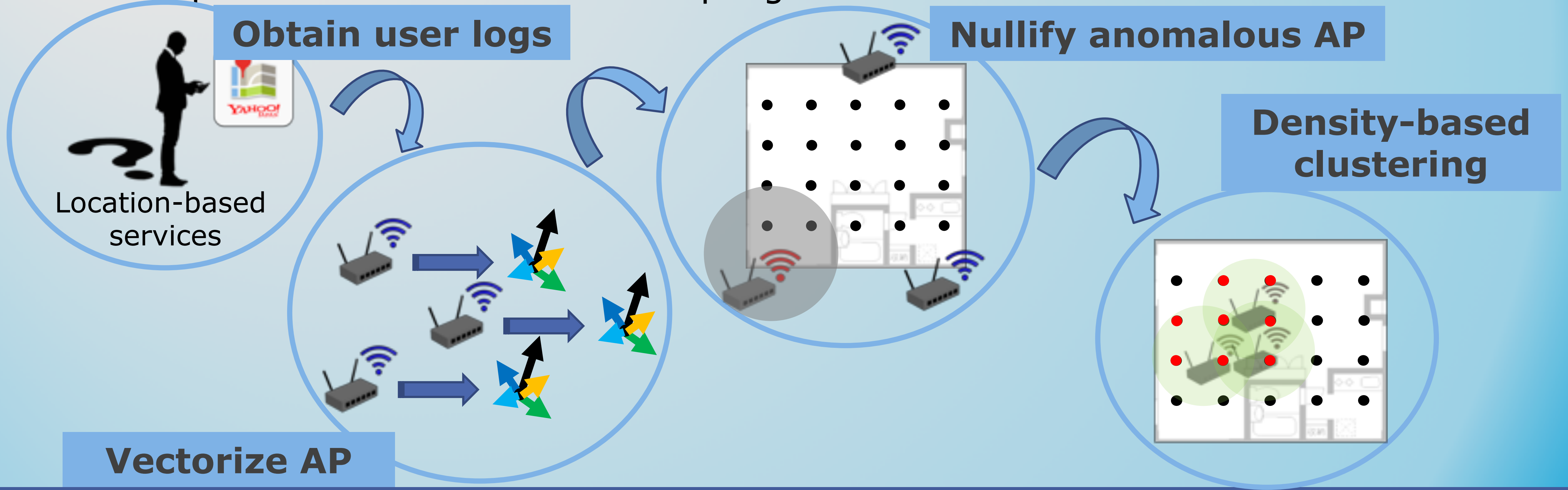
Requires **calibratio** el using tra g



No-sweat Detective (汗をかかない探偵)

Detects environmentally-changed reference points as anomaly

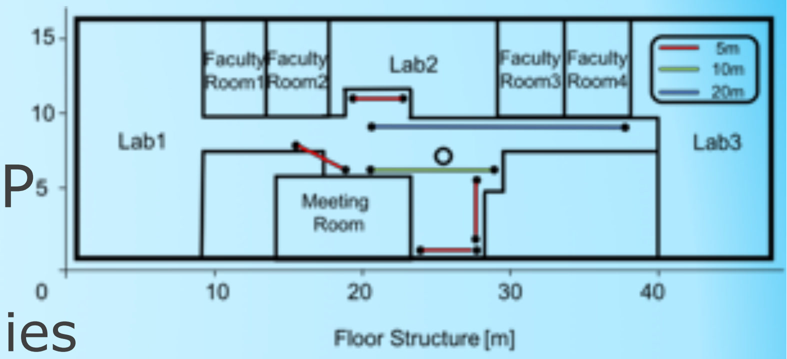
- Performs with **no effort** using a **user log**
- **Stabilizes** the calibration with **less deterioration**
  - Compared to that of random-sampling calibration with same amount of datasets



Laboratory Test-Bed

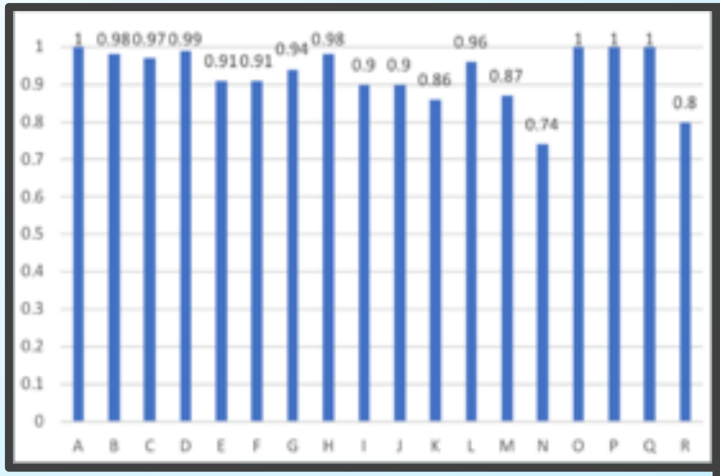
SETTING

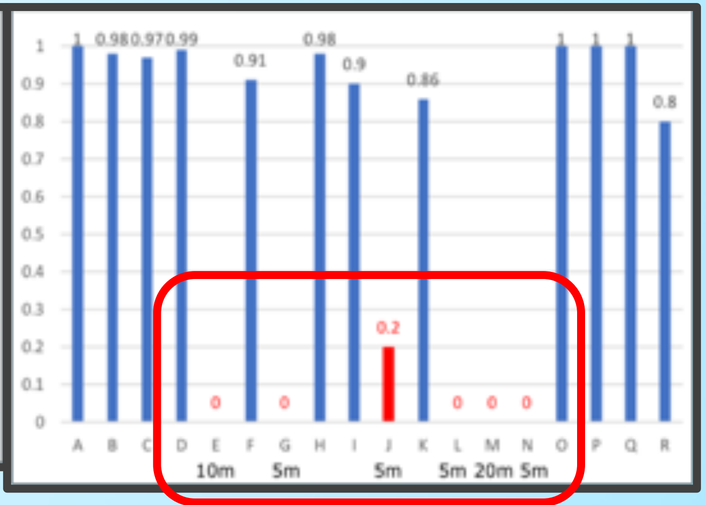
Simulate environmental changes by displacement of APs
- Four **5m**, one **10m**, and one **20m** displacements
- Validate whether No-Sweat Detective can detect anomalies



RESULTS

No-sweat Detective could detect APs that were affected by environmental distortion





Normal

Anomalous

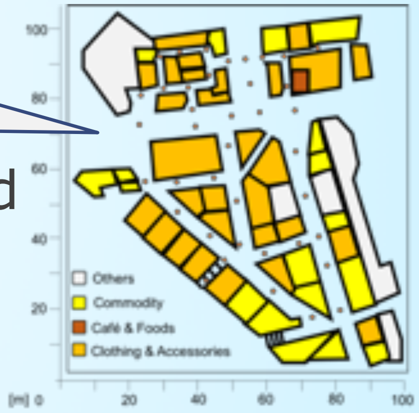
APs without / APs with environmental changes

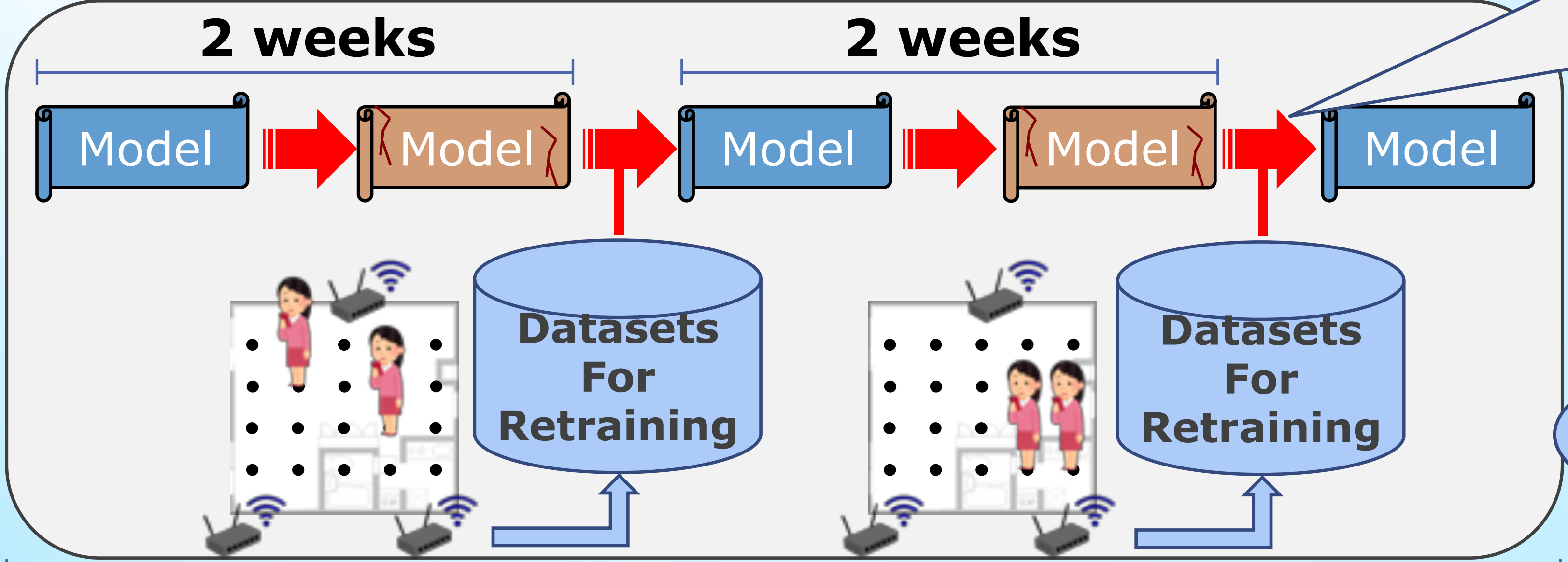
Real World Test-Bed

SETTING

- Evaluate accuracy of model at the end
- Slide amount of reference points
  - 10%, 20%, 30%, 60%, 100%

Underground district in Osaka





5-month-long observation

Baseline (never calibrated)
- **NonTrain**

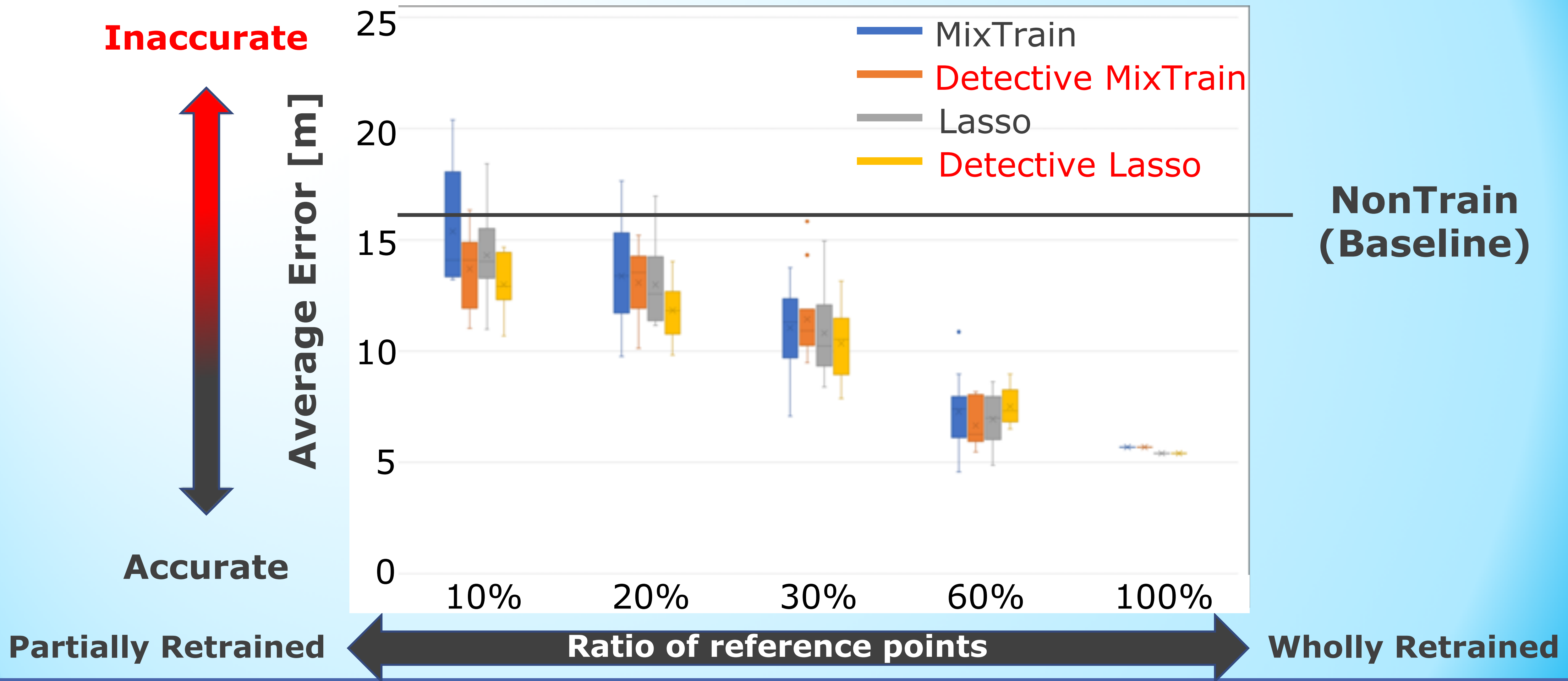
Transfer Learning[1]
- **MixTrain**
- **Lasso**
- **Detective MixTrain**
- **Detective Lasso**

Random-sampling

MixTrain and Lasso combined with our methods

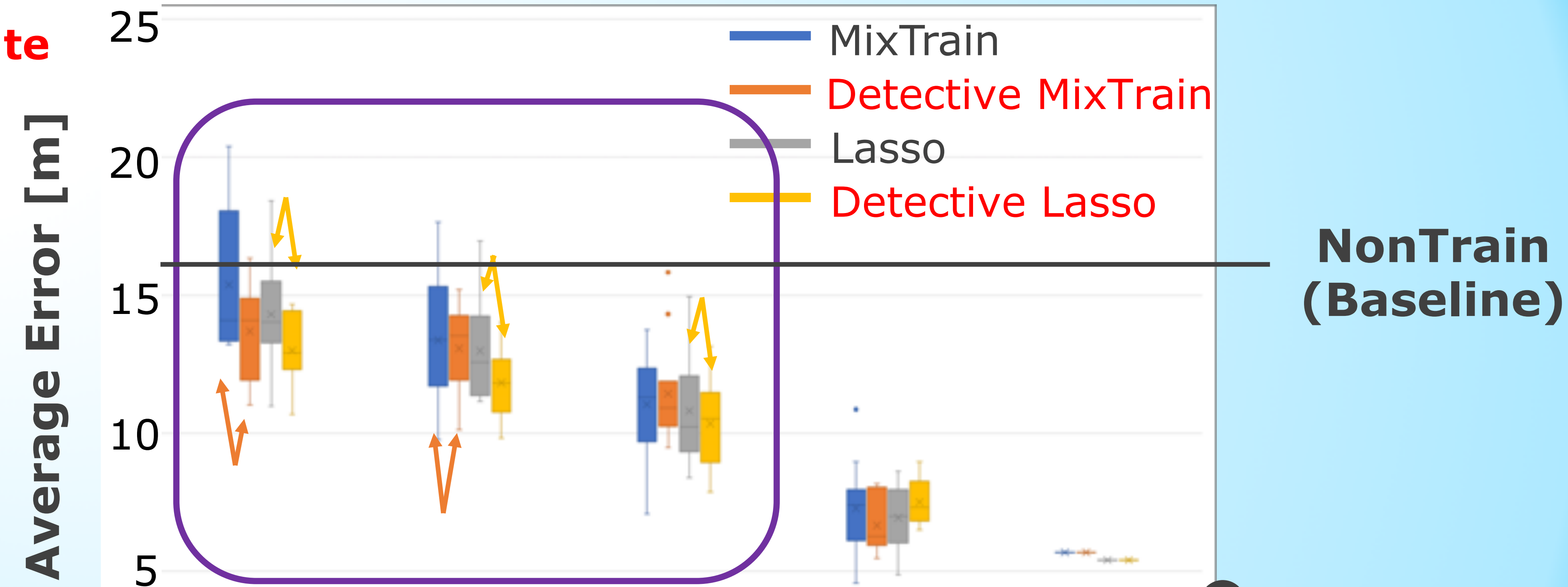
[1]: Pengcheng Wu and Thomas G. Dietterich. "Improving SVM Accuracy by Training on Auxiliary Data Sources."

Results After 10 Trials



Results After 10 Trials

Inaccurate



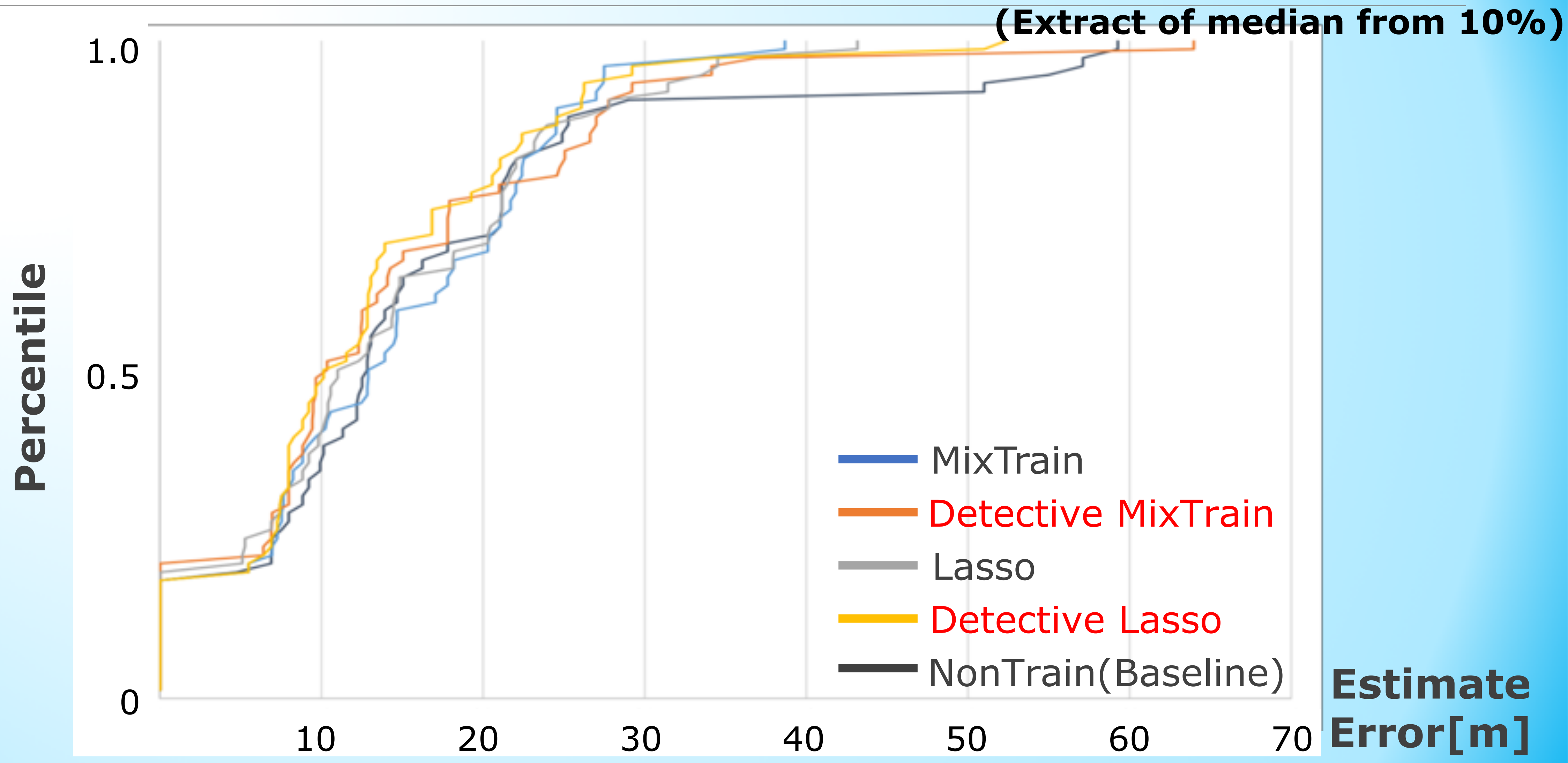
Accurate

Calibration were stabilized with less deterioration

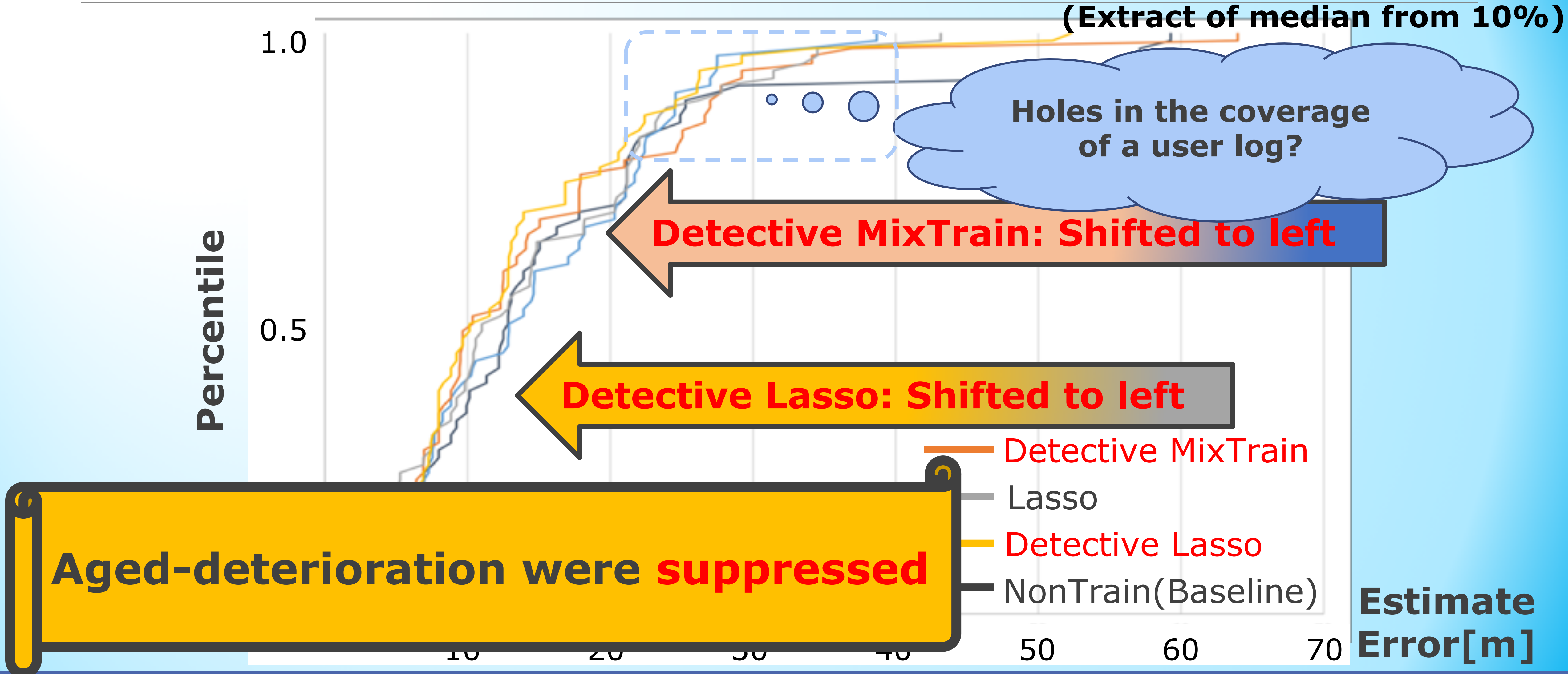
Partially Retrain

Wholly Retrained

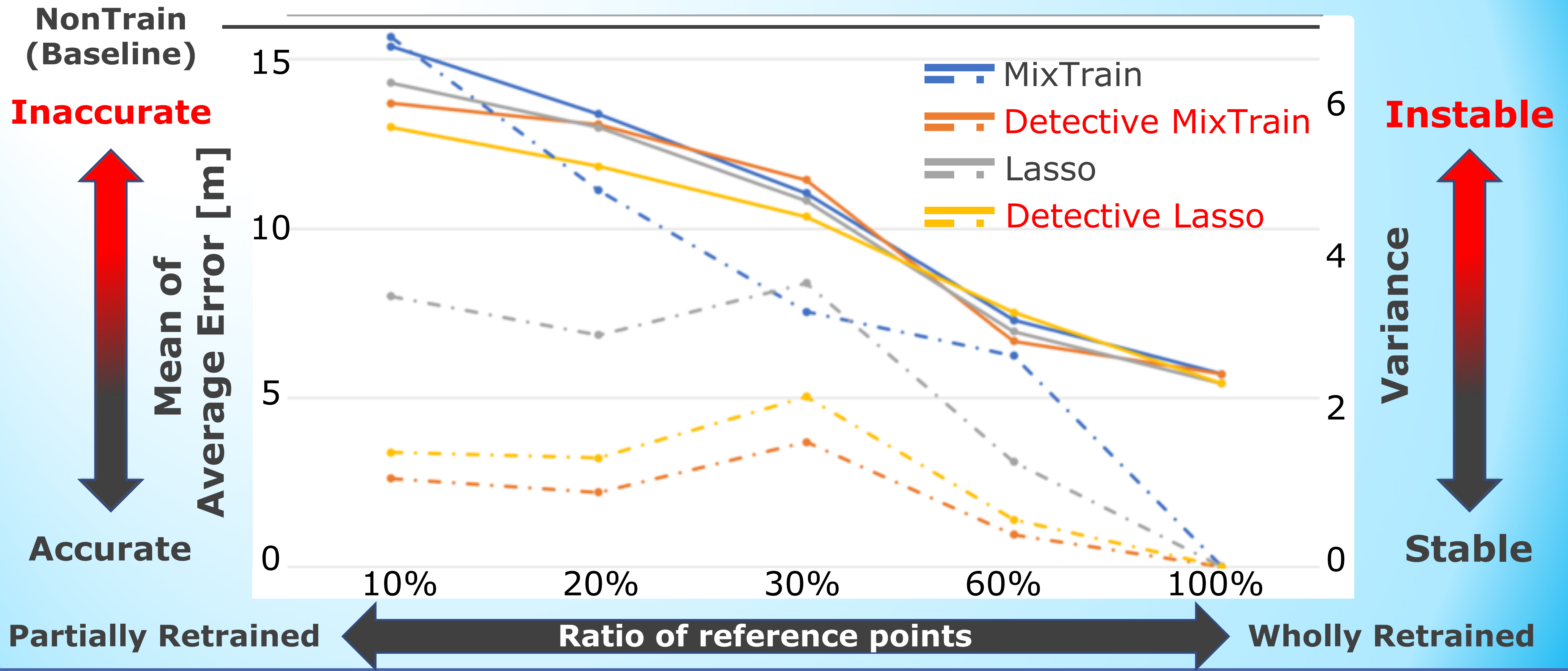
Cumulative Distribution of Estimate Error



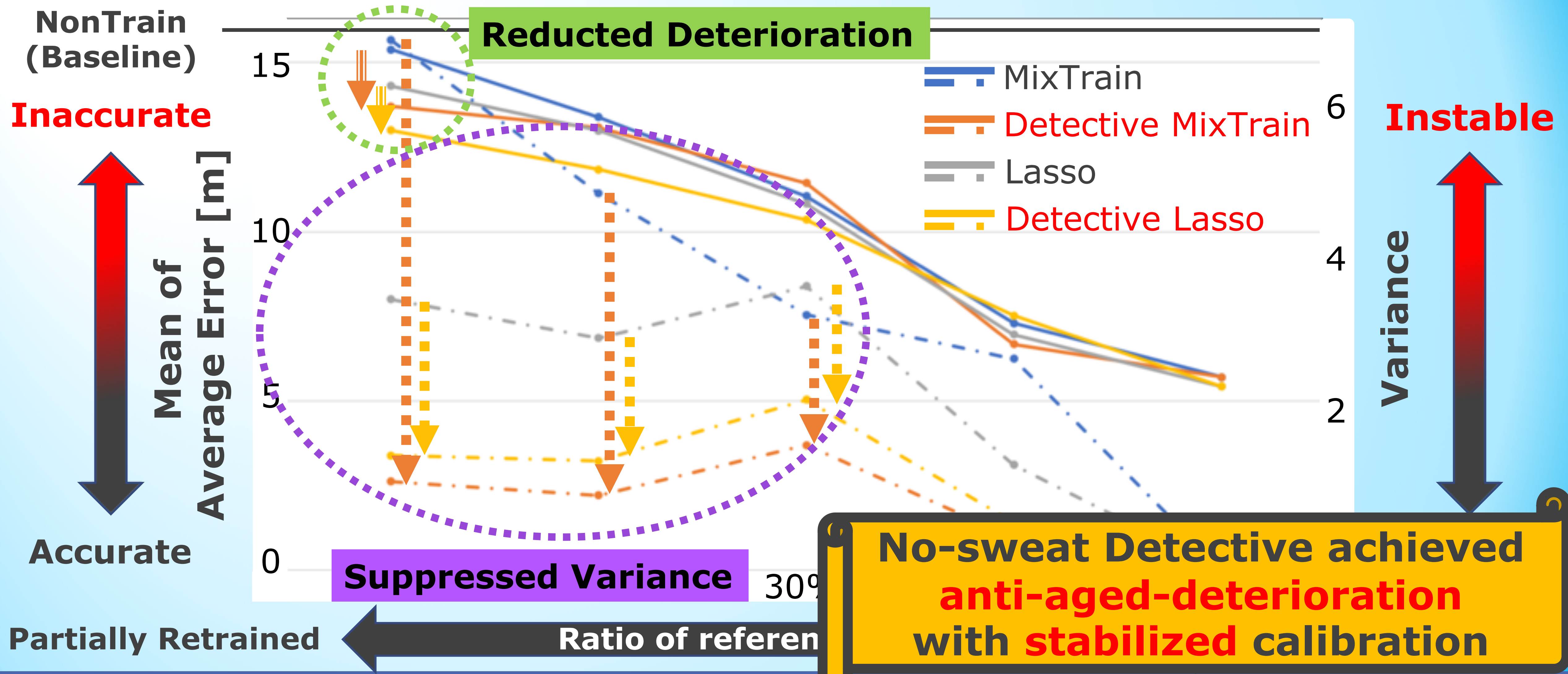
Cumulative Distribution of Estimate Error



Means & Variance



Means & Variance



No-sweat Detective achieved
anti-aged-deterioration
with **stabilized** calibration

Summary and Outlook

Observation-based model for localization deteriorates with age

Transfer learning is employed lately

- Datasets for retraining model are randomly sampled

No-sweat Detective

- Identifies reference points where environmentally changed with no effort

Achieved anti-aged-deterioration with stabilized calibration

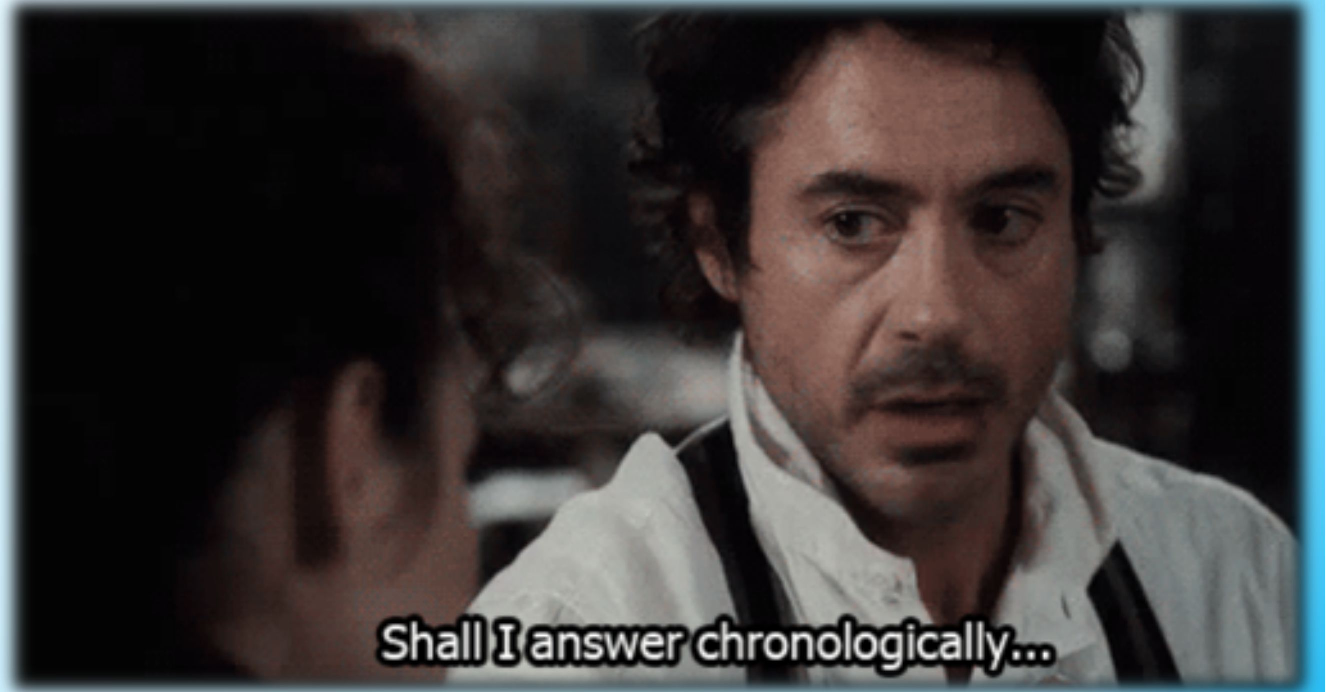
- In comparison with existing random-sampling transfer learnings

Outlook

- Substitute a user log for datasets to retrain model in order to materialize completely autonomous calibration

Thank you!

Questions?



moi@ubi.cs.ritsumei.ac.jp
http://koheiyamamoto.net/

©Sherlock Holmes

Supplements

Motivation

Indoor position is now of considerable importance for IoT

- To capture user contexts and her interests

Fingerprinting localization based on Wi-Fi RSSI is major

- Workflow consists of training and operational phases
  I. Primary signal model created from primary datasets of labeled fingerprint
  II. Estimate her location by comparing her current fingerprint with model

But accuracy of model deteriorates with age

- Caused by environmental changes
  - Movement of objects, construction of obstacles, automatic power adjustment, etc.

Requires **calibration** of model at certain periodic cycle

- Laborious effort and time-consuming overhead

State of the Art 1/2

Reconstruction of model with less or no effort

- TuRF, QRFC, and UnLoc can obtain labeled fingerprint installing sensors

Detection of signal changes in wireless networks

- Song et al. showed detecting node redeployment focusing neighborship
- Ohara et al. showed detecting environmental changes observing CSI (Wi-Fi channel state information)
- Meng et al. proposed probabilistic algorithm to detect distortion

State of the Art 1/2

Reconstruction of model with less or no effort

- TuRF, QRFC, and UnLoc can obtain labeled fingerprint installing sensors

Forcing user to put on sensors is far from practical use

Detection of signal changes in wireless networks

- Song et al. showed detecting node redeployment focusing neighborship

Forcing environment to have installment is not pragmatic

- Ohara et al. showed detecting environmental changes observing CSI (Wi-Fi channel state information)

Requires pairs of connection and CSI is scarcely used

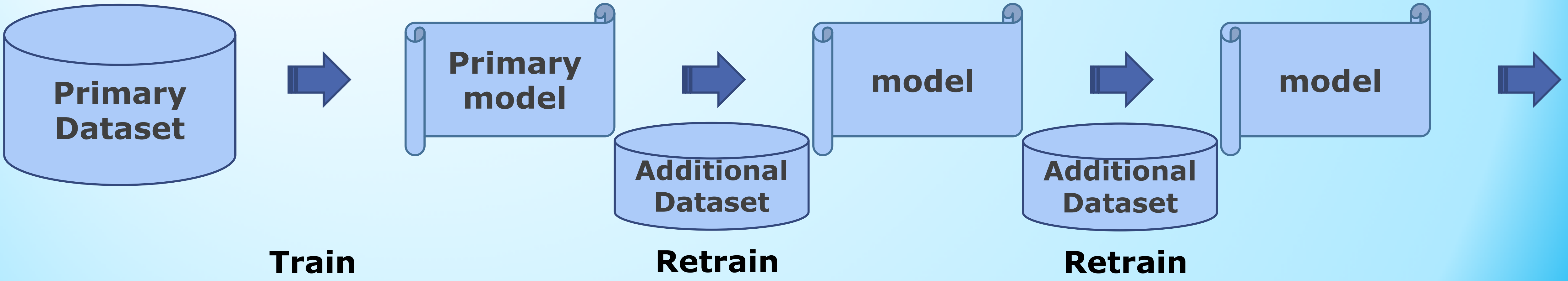
- Meng et al. proposed probabilistic algorithm to detect distortion

Only proven in test-bed and does not target calibration

State of the Art

Major current is to employ **transfer learning**[1)]

- Reflects current Wi-Fi environment in model
- Retrains model with adding a small amount of labeled fingerprint
  - Yang et al. proved higher accuracy with much less calibration effort
  - Yin et al. materialized regression and model-tree based algorithm
  - Tian et al. coped with performance degradation of model using SVM regression analysis
- Remarkably **suppressing amount of dataset** to calibrate model



[1]: Pengcheng Wu and Thomas G. Dietterich. "Improving SVM Accuracy by Training on Auxiliary Data Sources."

State of the Art

Major current is to employ **transfer learning**[1)]

- Reflects current Wi-Fi environment in model
- Retrains model with adding a small amount of labeled fingerprint
  - Yang et al. proved higher accuracy with much less calibration effort
  - Yin et al. materialized regression and model-tree based algorithm
  - Tian et al. coped with performance degradation of model using SVM regression analysis
- Remarkably suppressing amount of dataset to calibrate model

Prim
Data

These methods pick up additional dataset (labeled fingerprint) randomly or comprehensively

- Spilled into destabilization of accuracy recovery at every calibration
- Recovery rate heavily relies on selection of labeled fingerprint

Train Retrain Retrain

[1]: Pengcheng Wu and Thomas G. Dietterich. "Improving SVM Accuracy by Training on Auxiliary Data Sources."

State of the Art

Major current is to employ **transfer learning**[1)]

- Reflects current Wi-Fi environment in model
- Retrains model with adding a small amount of labeled fingerprint
  - Yang et al. proved higher accuracy with much less calibration effort
  - Yin et al. materialized regression and model-tree based algori
  - Tian et al. coped with performance d tion of m M regression analysis
- Remarkabl ing am m

Prim

rint

Tra train Retrain

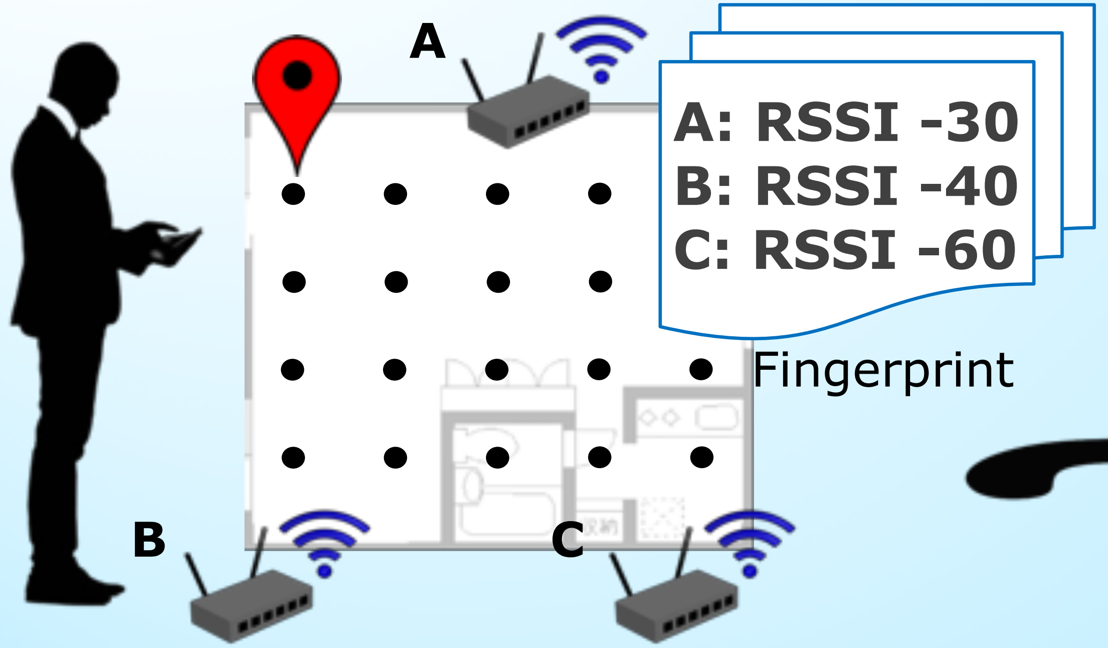
Labeled fingerprint should be convergently collected from specific (anomalous) reference points

[1]: P g Wu and Thomas G. Dietterich. “Improving SVM Accuracy by Training on Auxiliary Data Sources.”

Two Types of Fingerprint

LABELED

- Used for training primary model
- Small amount used for retraining model
- Manual collection by administrator



UNLABELED

- Collected from user of location services
- Automatically but not labeled

Two Types of Fingerprint

Identify reference points where environmentally changed

UNLABELED

- Collected from user of location services
- Automatically but not labeled

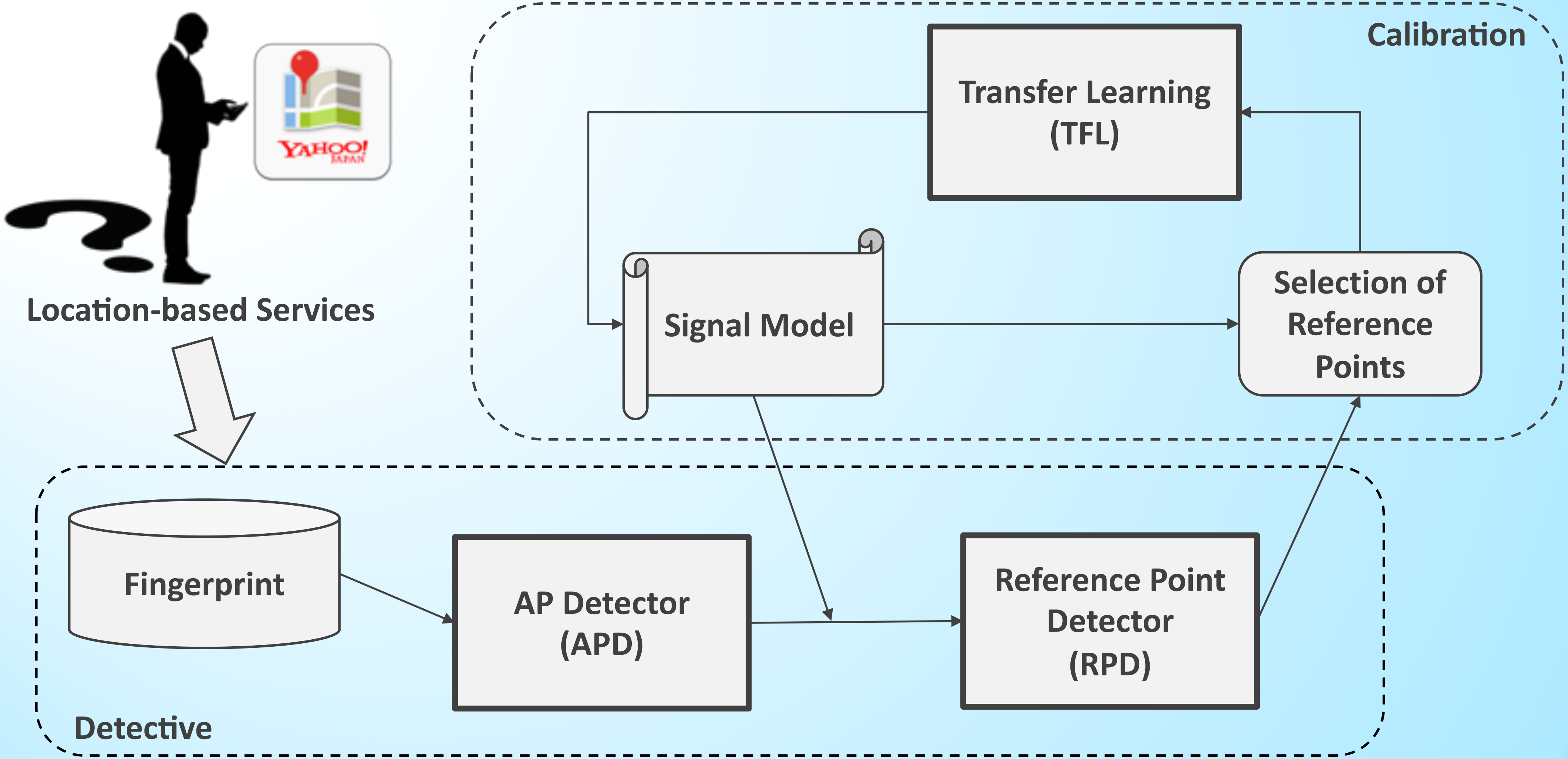
Plenty

Methods

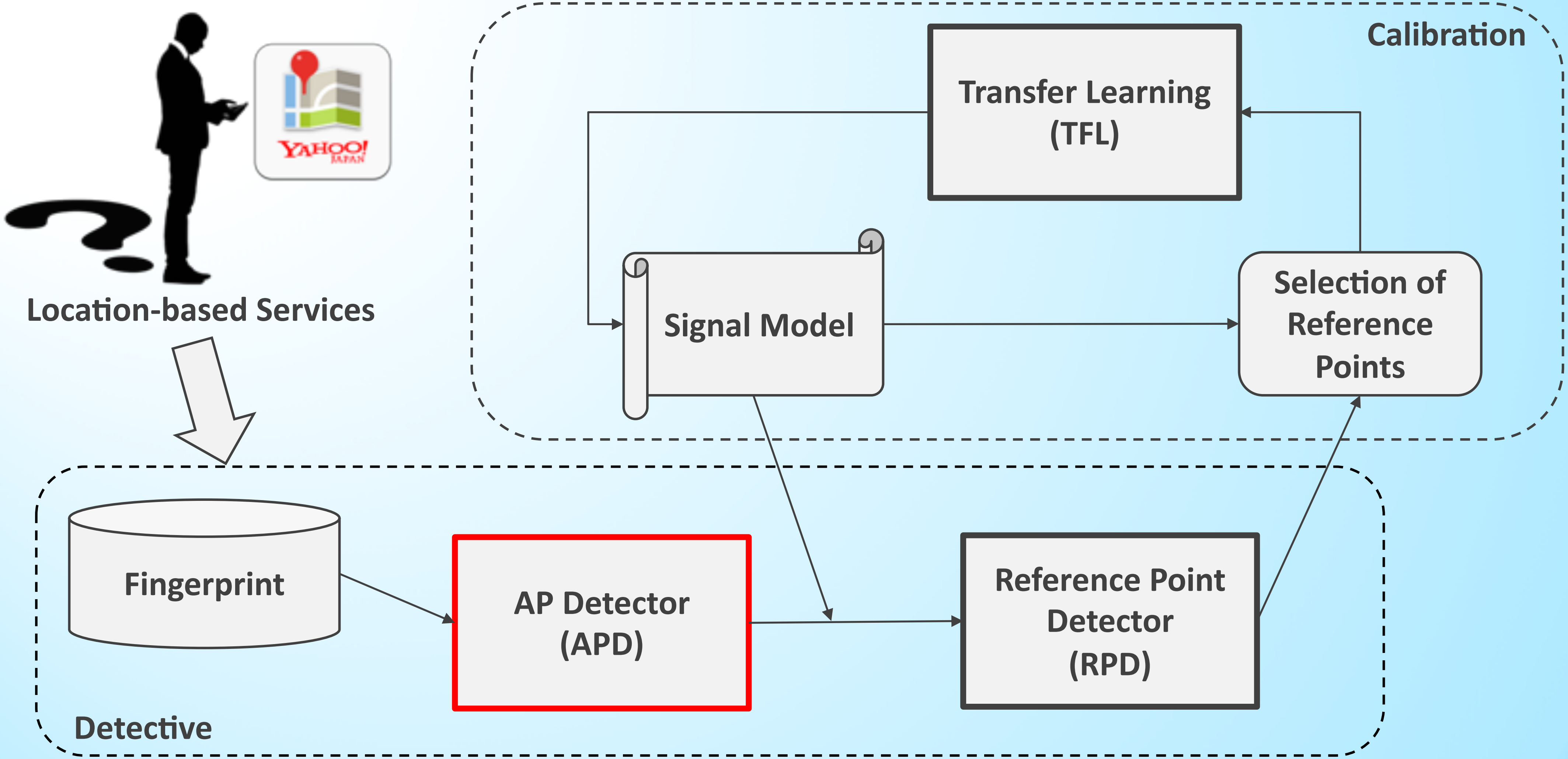
Identify reference points where environmentally changed as anomaly

- Build a system named **No-sweat Detective** (汗をかかない探偵)
- Perform with **no effort** using unlabeled fingerprint
- Work strongly even in a **real world** with **long**-period observation
- Achieve **higher accuracy recovery** in comparison with existing transfer learning methods using the **same amount** of labeled fingerprint

No-sweat Detective



No-sweat Detective



AP Detector (APD)

Detects anomalous AP

- Utilizes co-occurrence and analyzes relative position with other AP

Workflow

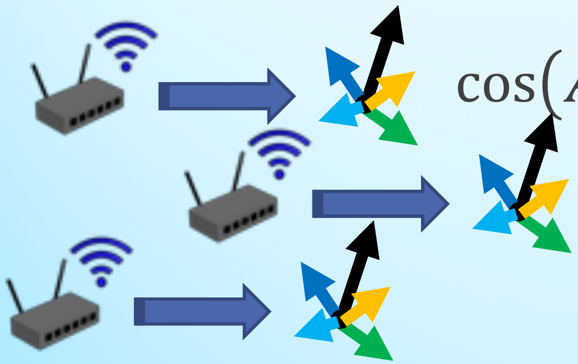
I. Picks up unlabeled fingerprint having maximum RSSI over ***vecFilt***

$$(\mathrm{R1}_{\mathrm{r1}},\mathrm{R2}_{\mathrm{r2}},\mathrm{R3}_{\mathrm{r3}},\cdots,\mathrm{Rx}_{\mathrm{rx}})\ \{\boldsymbol{vecFilt} < \max(\mathrm{r1},\mathrm{rx})\}\ [\mathrm{R{:}\ AP,\ r{:}\ RSSI}] \cdots (1)$$

II. Vectorizes unlabeled fingerprint in sparse space with ***vecWidth***

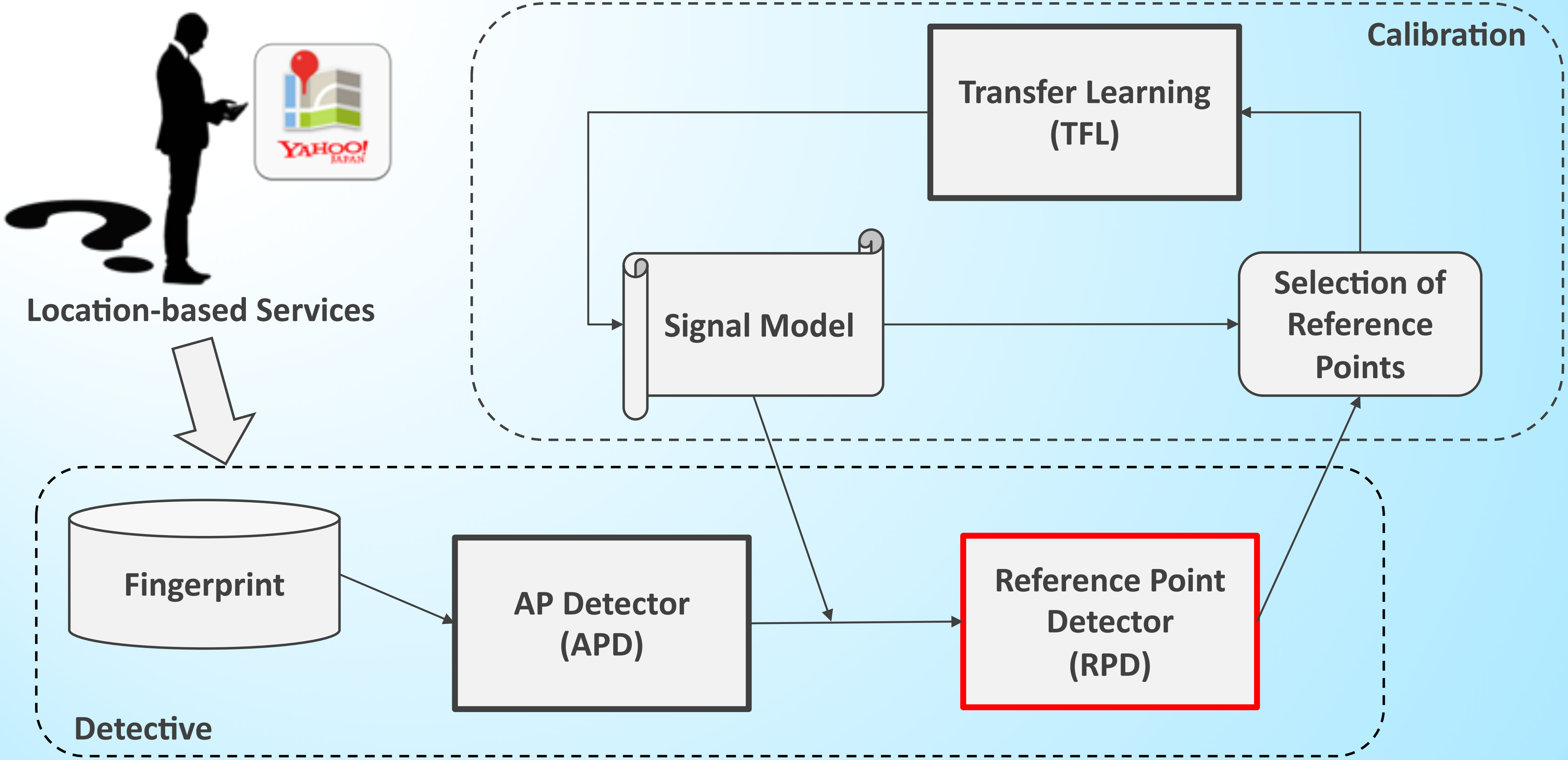
$$\vec{A} = (\mathrm{R1}_{\mathrm{r1}},\mathrm{R2}_{\mathrm{r2}},\mathrm{R3}_{\mathrm{r3}},\cdots,\mathrm{Rx}_{\mathrm{rx}})\ \{\boldsymbol{vecFilt} < \max(\mathrm{r1},\mathrm{rx}),\ \mathbf{vecWidth} < \mathrm{rx}\} \cdots (2)$$

III. Calculates similarity of vector model



$$\cos(\vec{A},\vec{A'}) = \frac{\vec{A}\cdot\vec{A'}}{|\vec{A}||\vec{A'}|} = \frac{\vec{A}}{|\vec{A}|}\cdot\frac{\vec{A'}}{|\vec{A'}|} = \frac{\sum_{i=1}^{|V|} A_i A'_i}{\sqrt{\sum_{i=1}^{|V|} A_i^2}\cdot\sqrt{\sum_{i=1}^{|V|} A'^2_i}} \cdots (3)$$

No-sweat Detective



Reference Point Detector (RPD)

Detects anomalous Reference Point

- Environmentally changed around detected Wi-Fi source

Workflow

I. Singularly detected:
   i. Roughly estimates location where each unlabeled fingerprint were observed by ignoring distorted radio wave
   ii. Gets weighted average center
   iii. Recollect labeled fingerprint at reference points where within radius εm from center

II. Plurally detected:
   - Employ DBSCAN to Wi-Fi source location retrieved from past model with radius εm
     - Regarding collective changes hardly caused by displacement of Wi-Fi source

Reference Point Detector (RPD)

Detects anomalous Reference Point

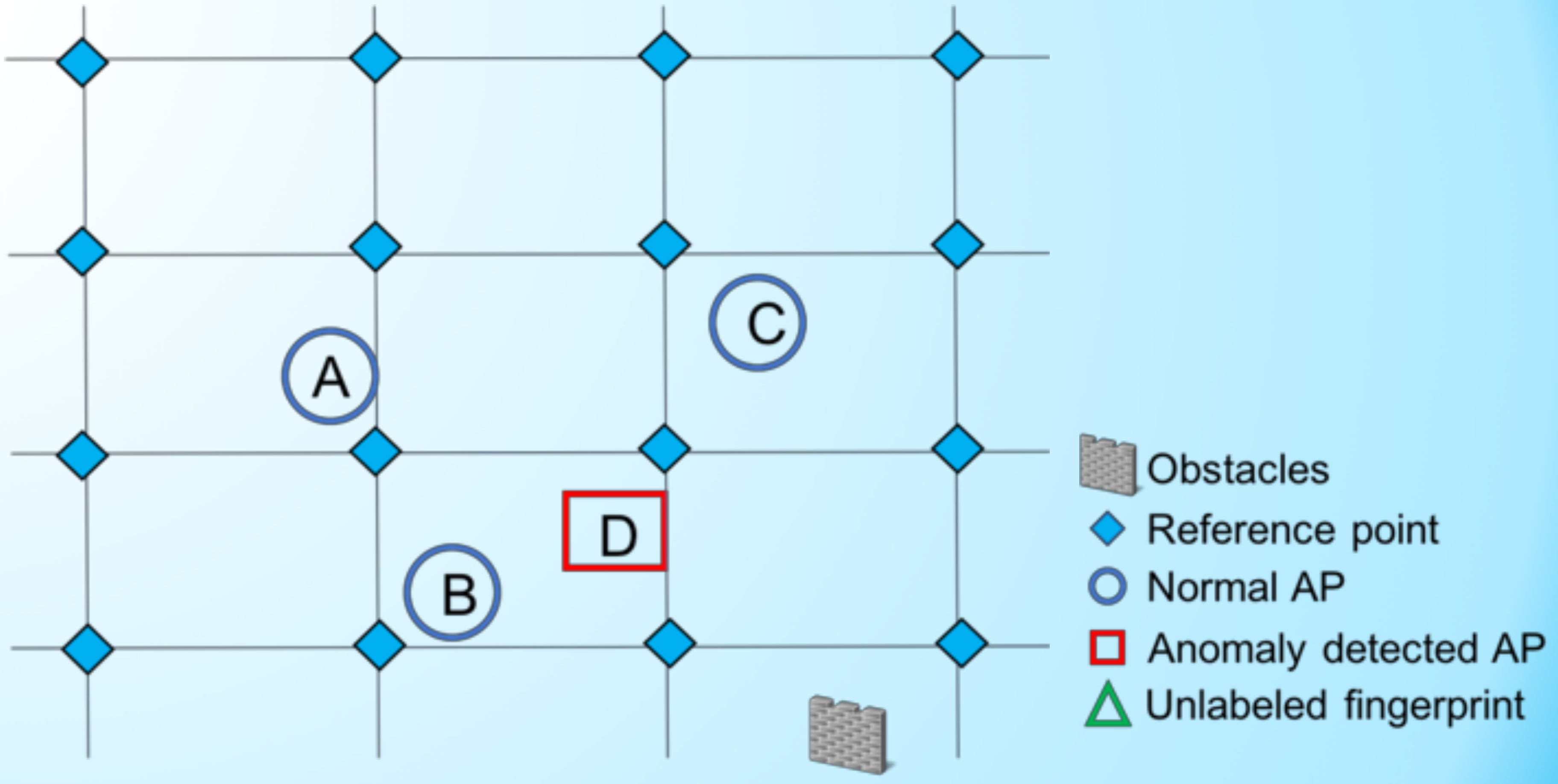
- Environmentally changed around detected Wi-Fi source

Workflow

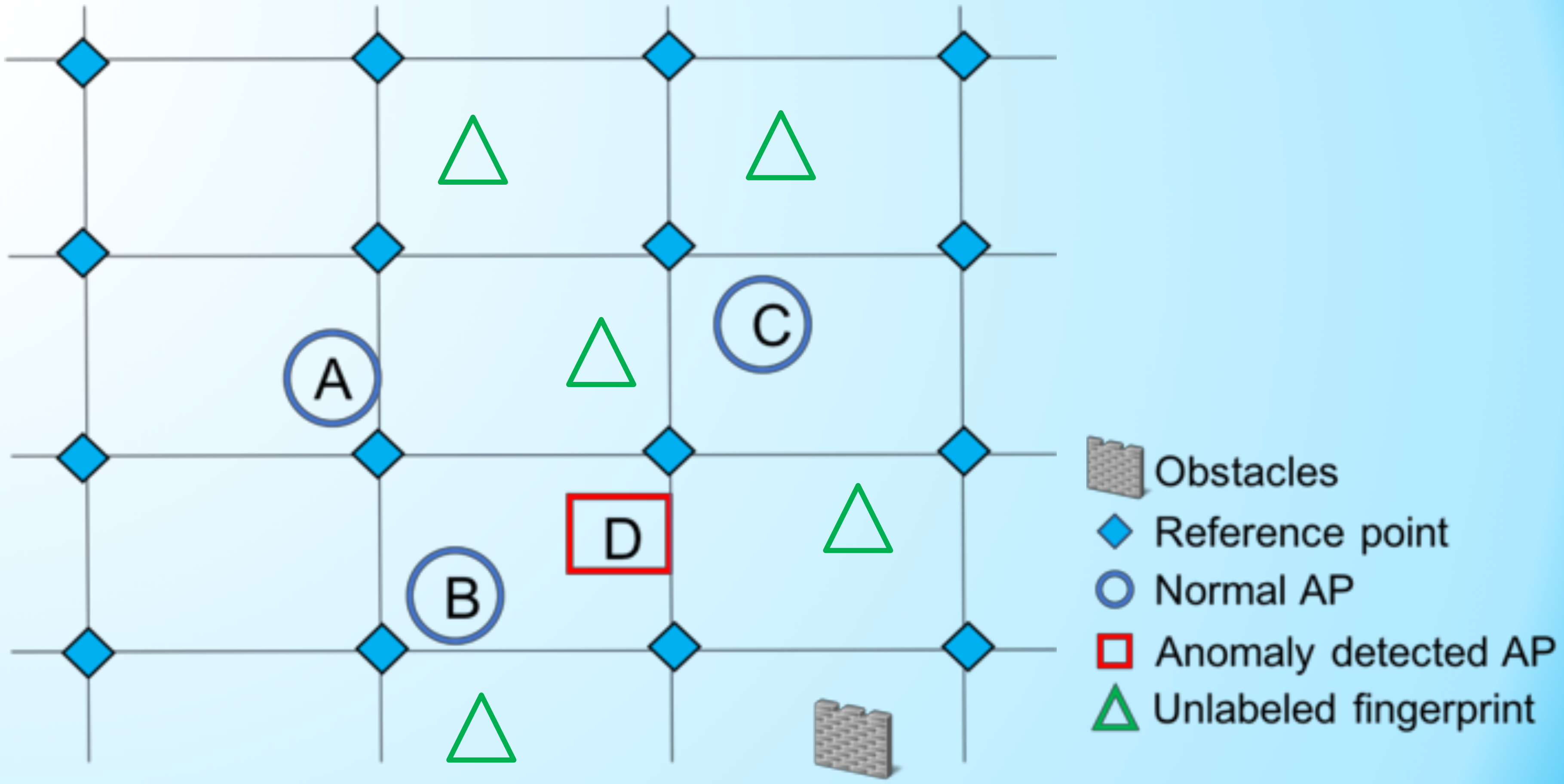
I. Singularly detected:
   i. Roughly estimates location where each unlabeled fingerprint were observed by ignoring distorted radio wave
   ii. Gets weighted average center
   iii. Recollect labeled fingerprint at reference points where within radius εm from center

II. Plurally detected:
   - Employ DBSCAN to Wi-Fi source location retrieved from past model with radius εm
     - Regarding collective changes hardly caused by displacement of Wi-Fi source

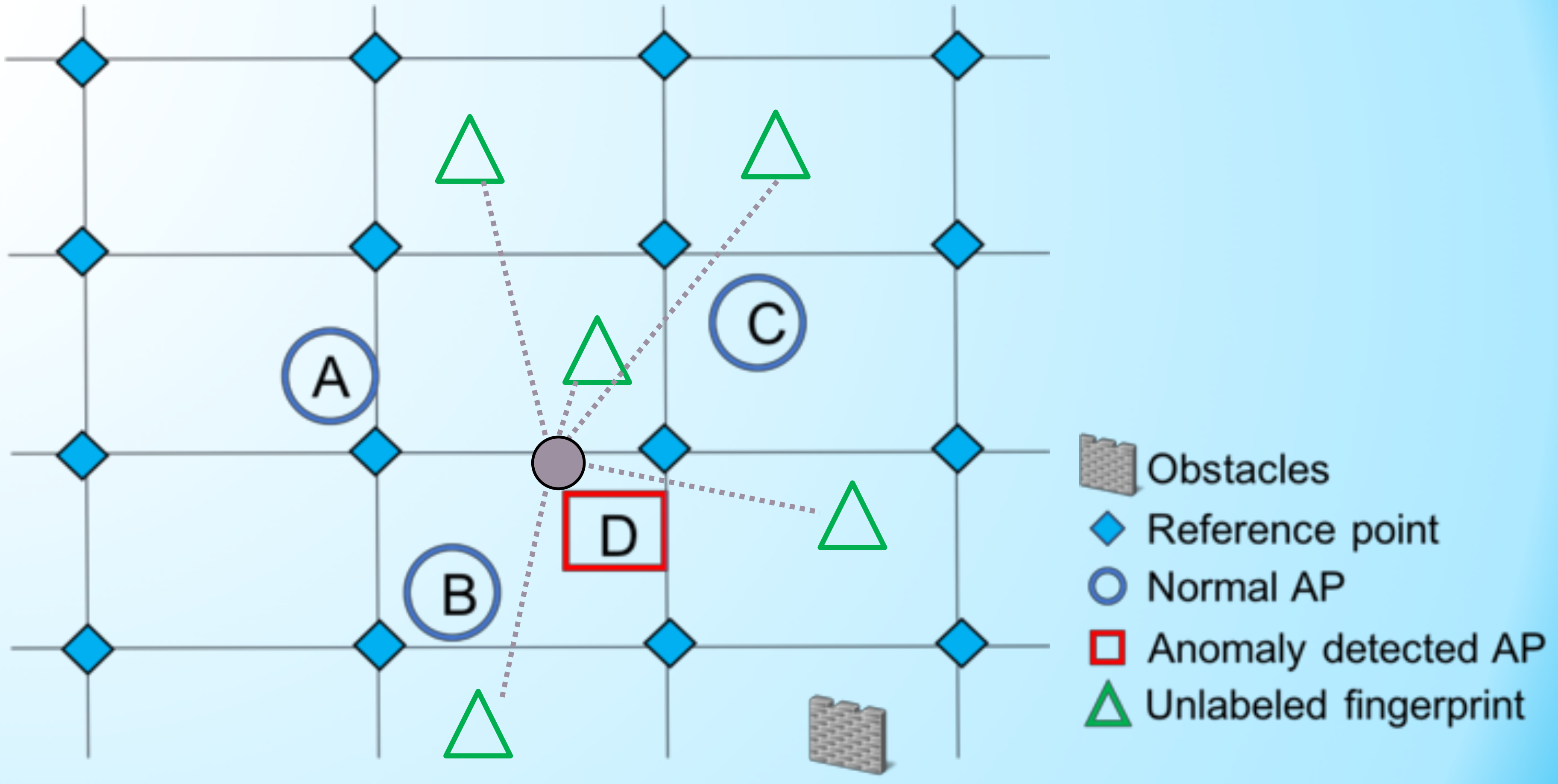
When Singularly Detected



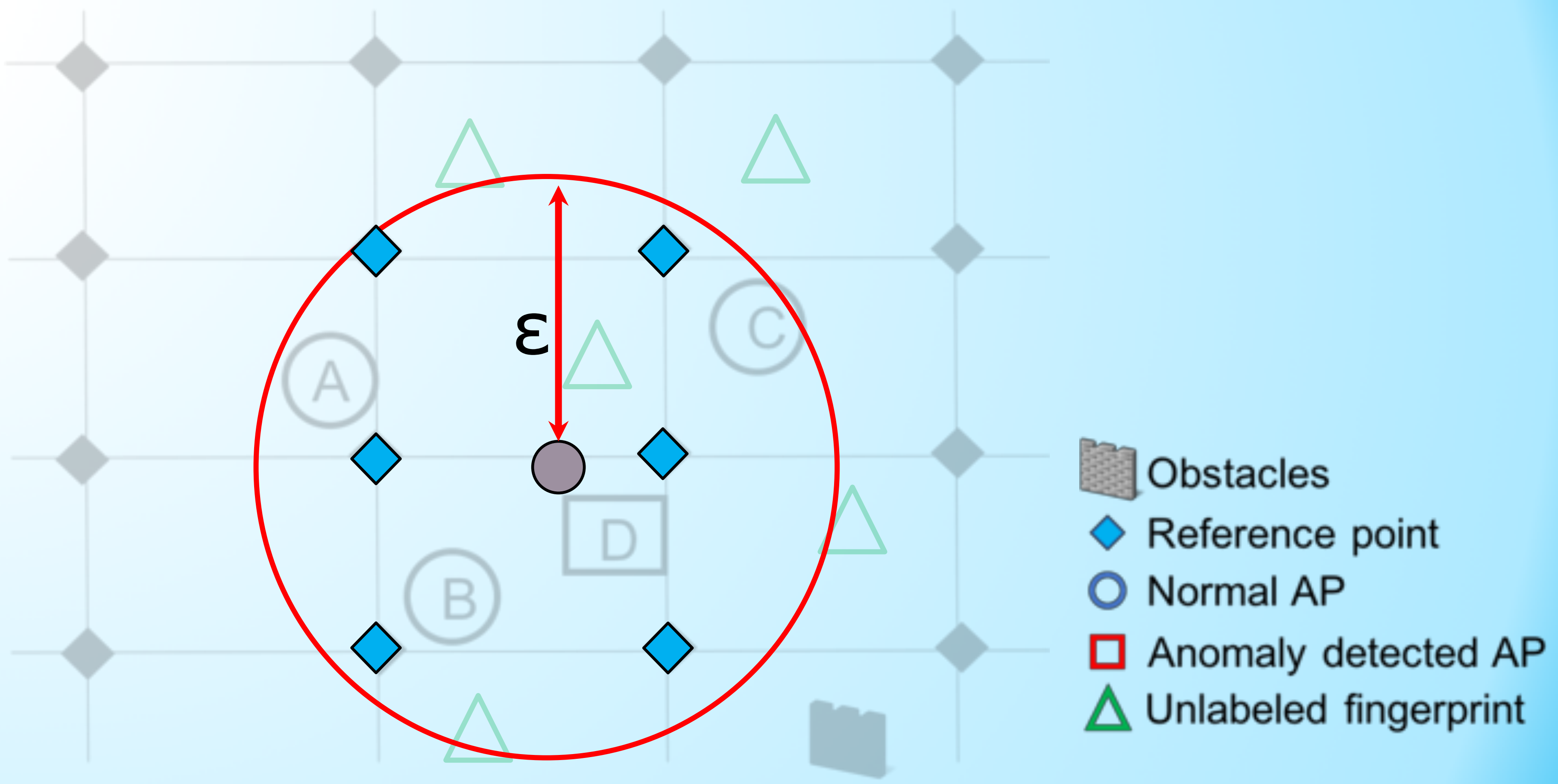
When Singularly Detected



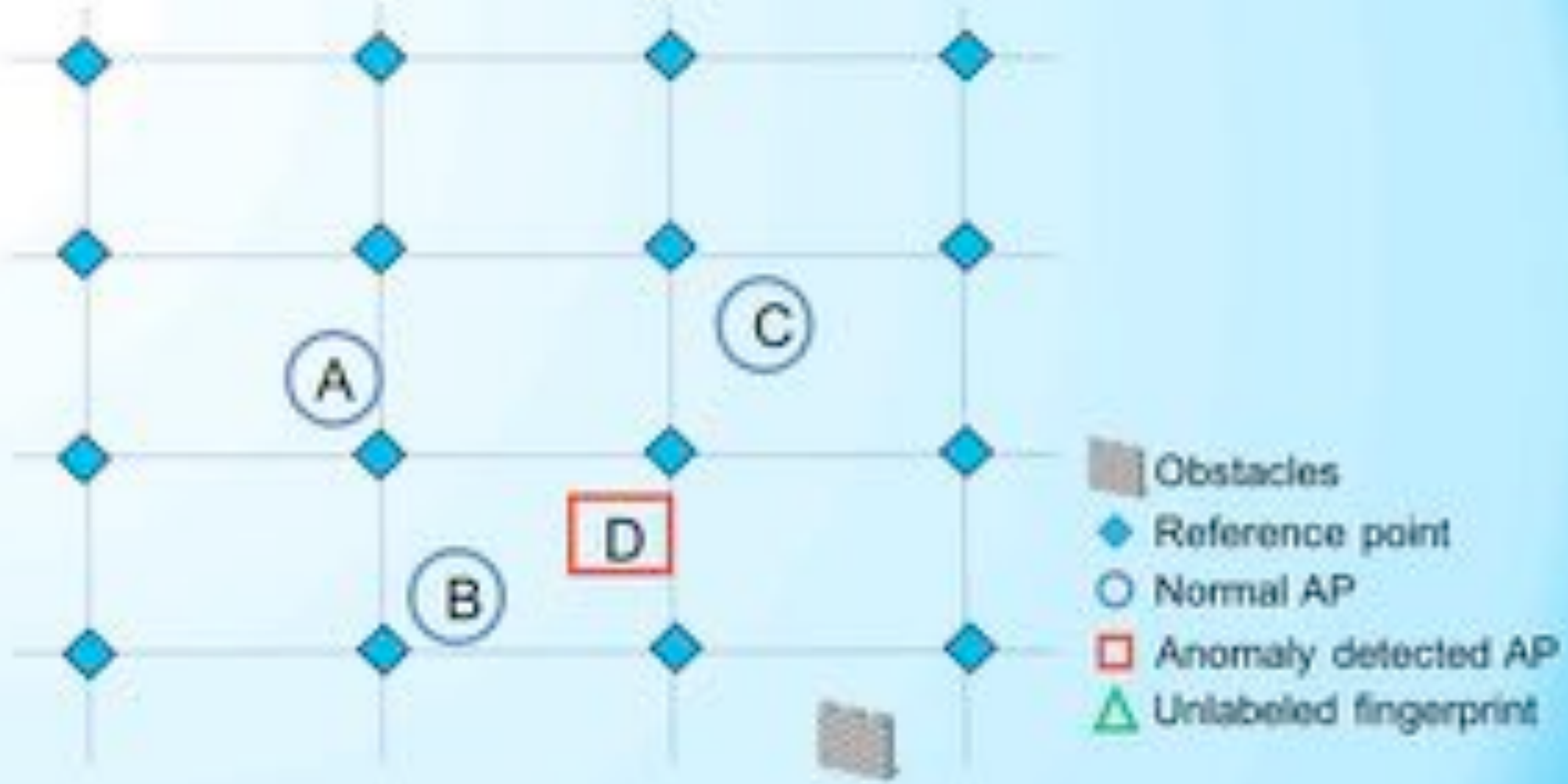
When Singularly Detected



When Singularly Detected



When Singularly Detected



Reference Point Detector (RPD)

Detects anomalous Reference Point

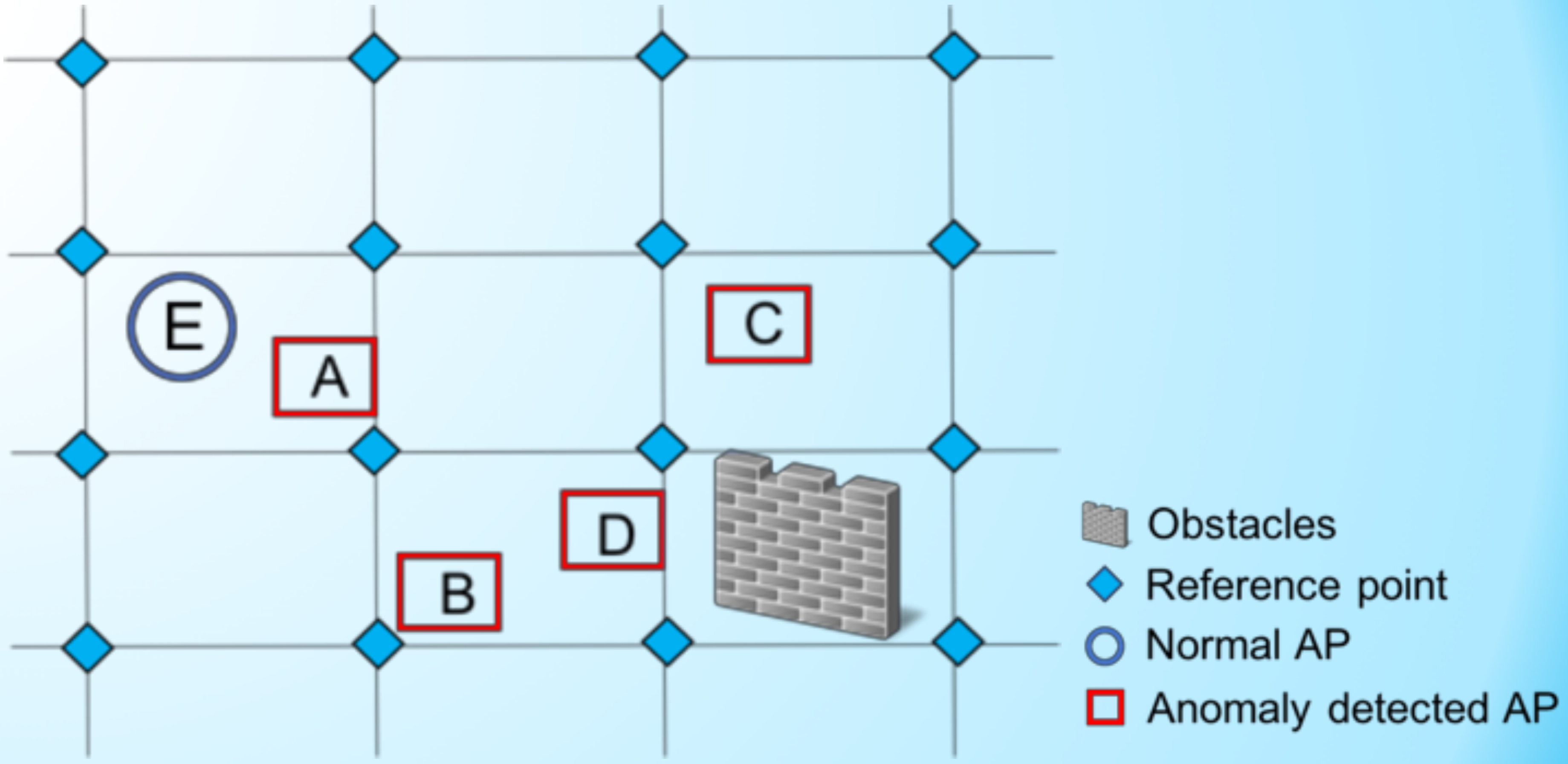
- Environmentally changed around detected Wi-Fi source

Workflow

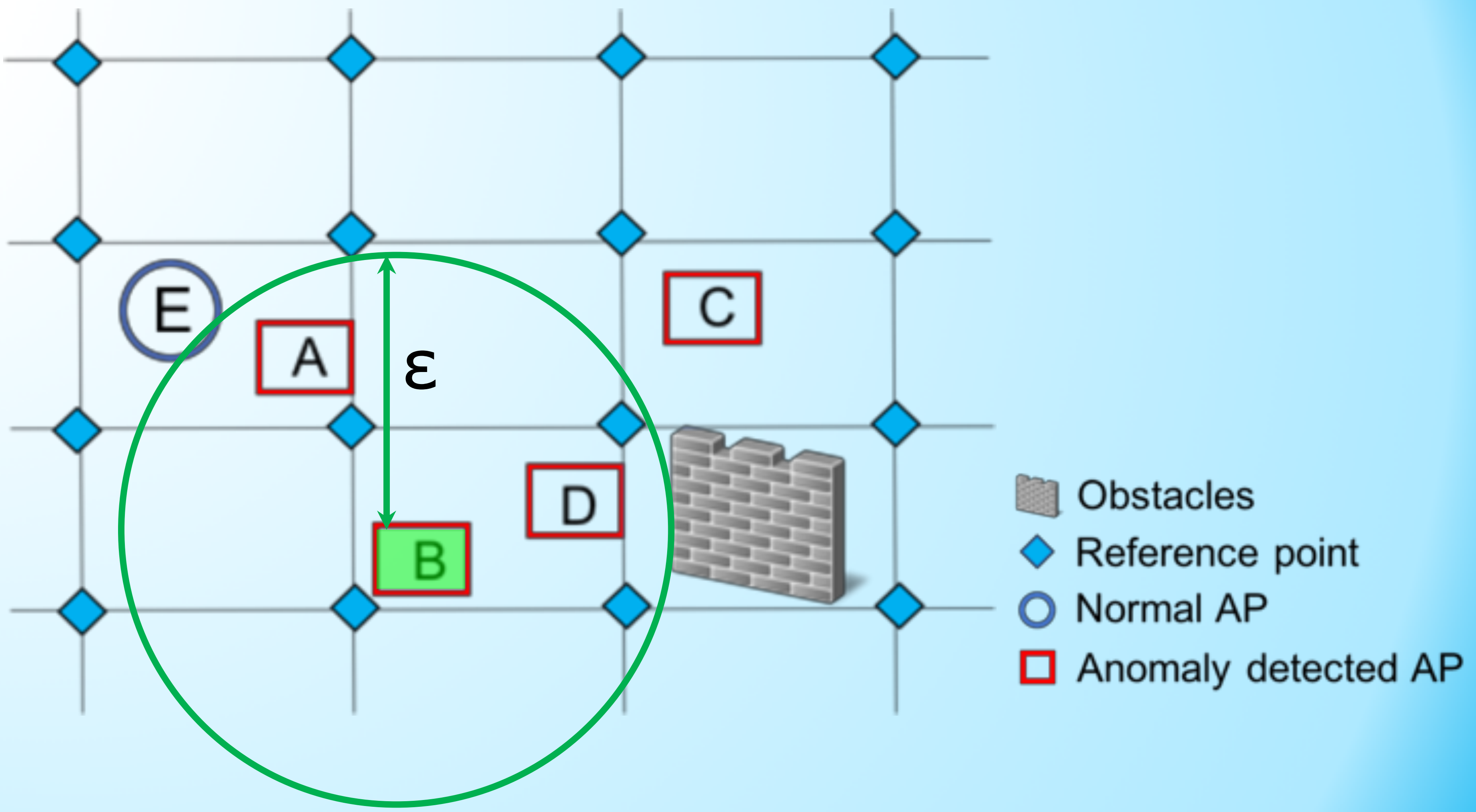
I. Singularly detected:
   i. Roughly estimates location where each unlabeled fingerprint were observed by ignoring distorted radio wave
   ii. Gets weighted average center
   iii. Recollect labeled fingerprint at reference points where within radius εm from center

II. Plurally detected:
   - Employ DBSCAN to Wi-Fi source location retrieved from past model with radius εm
     - Regarding collective changes hardly caused by displacement of Wi-Fi source

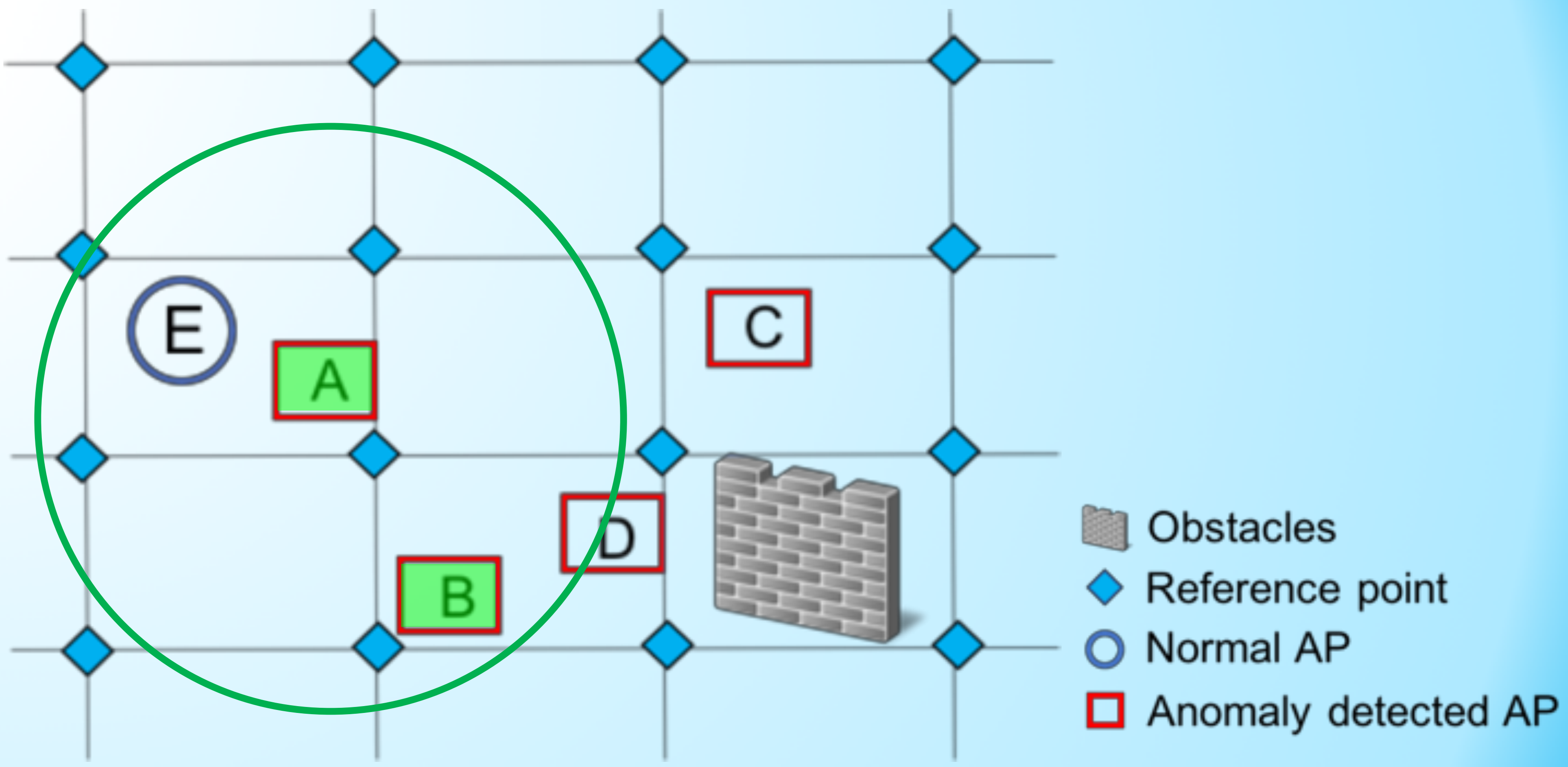
When Plurally Detected



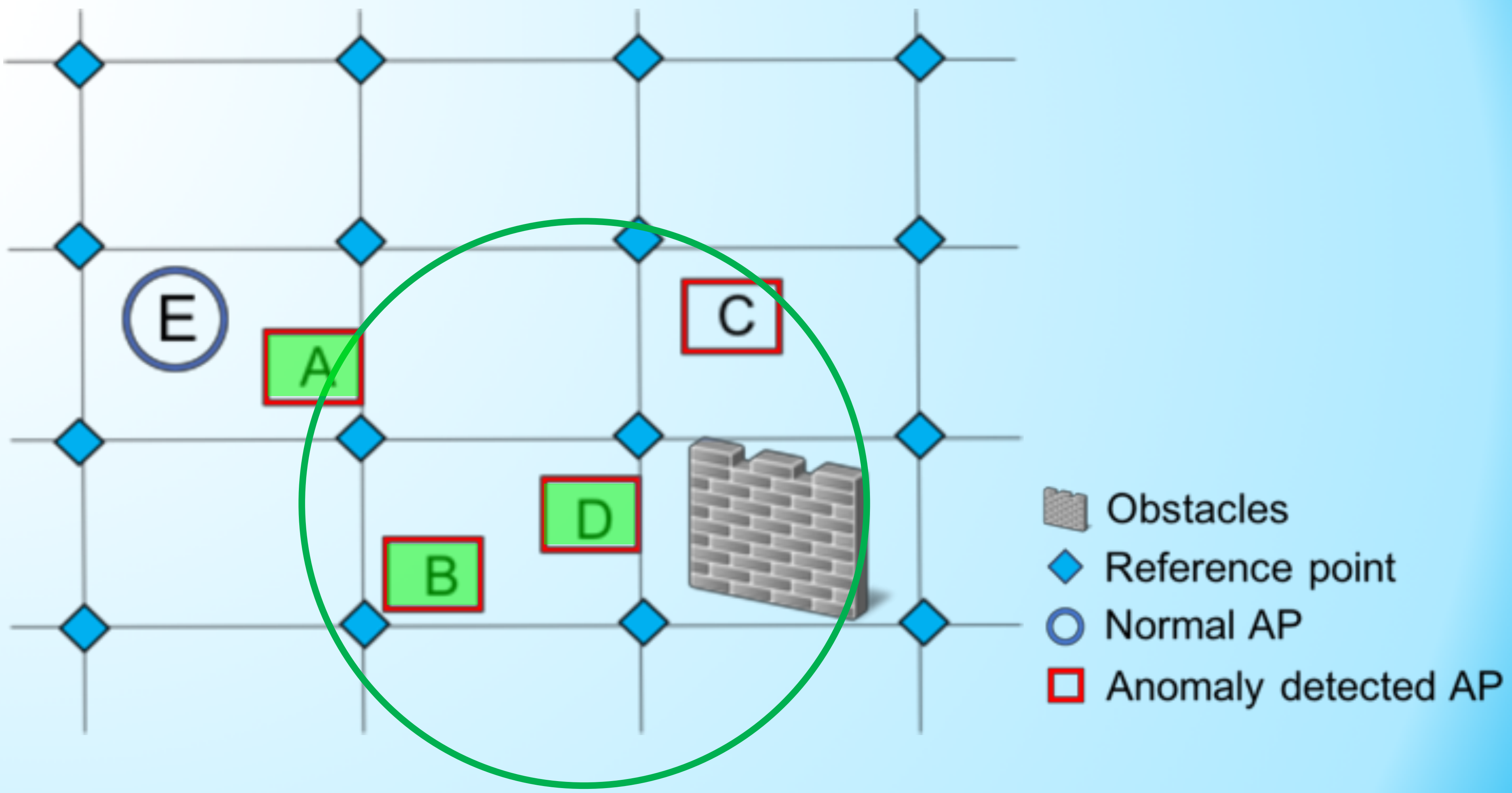
When Plurally Detected



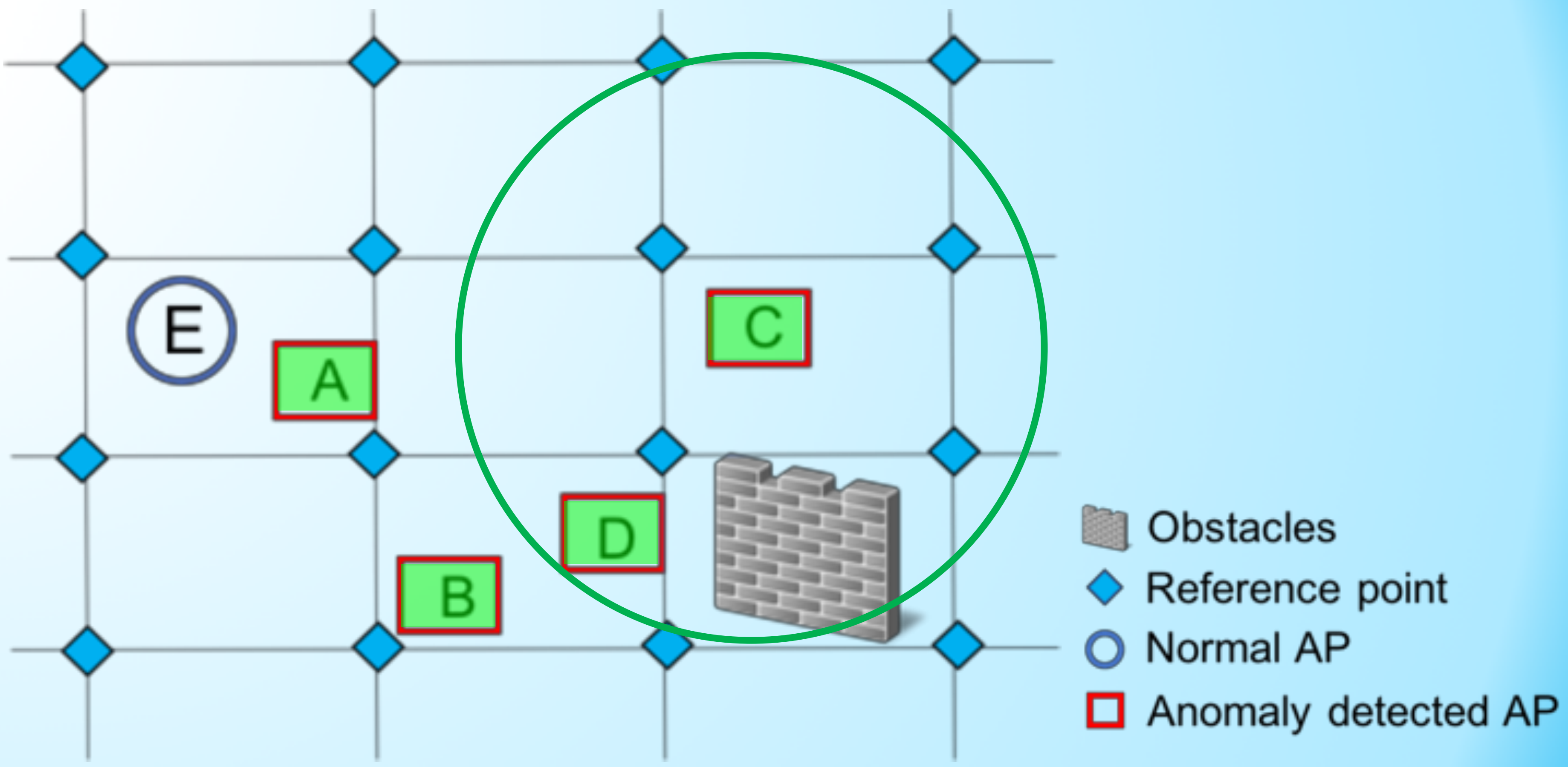
When Plurally Detected



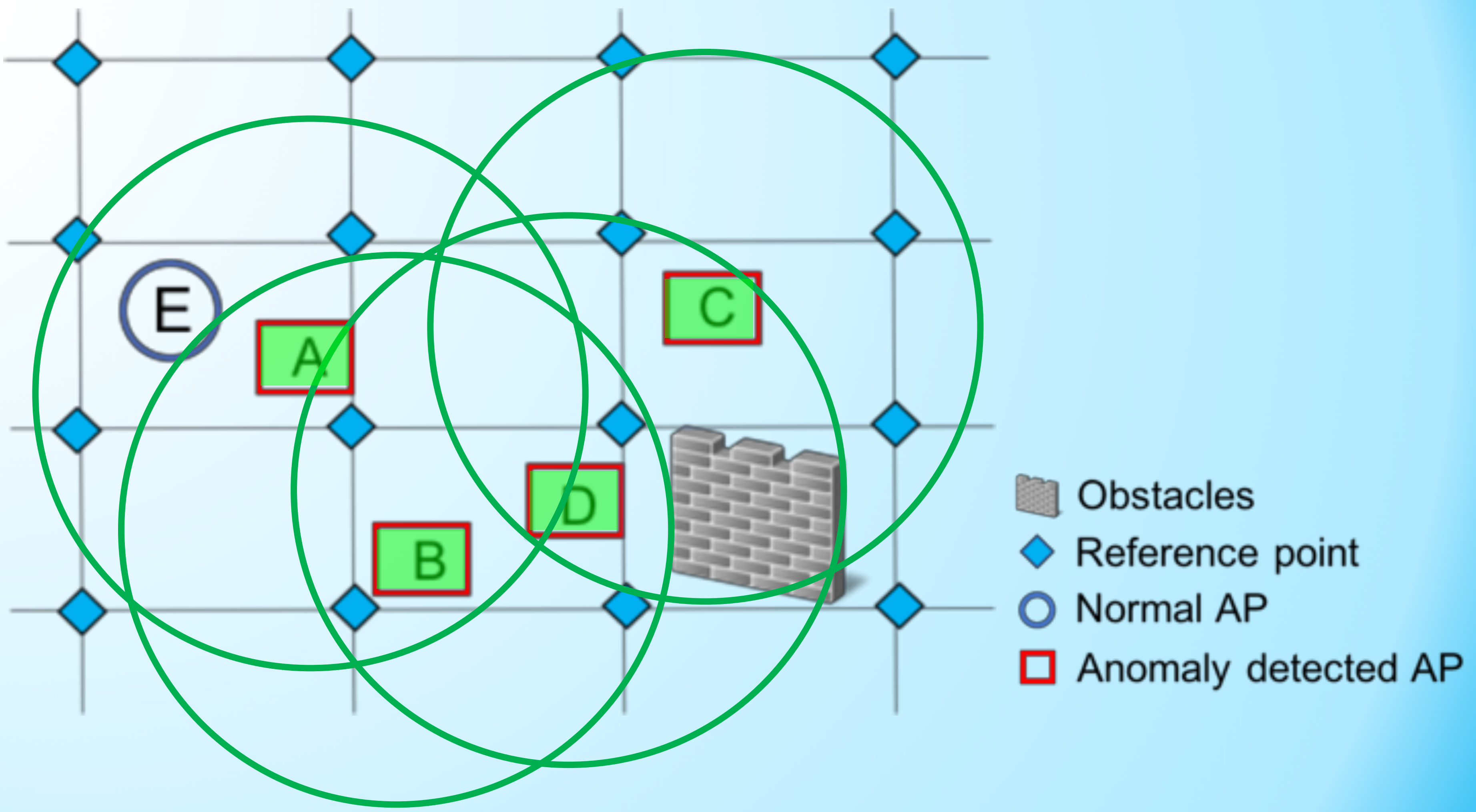
When Plurally Detected



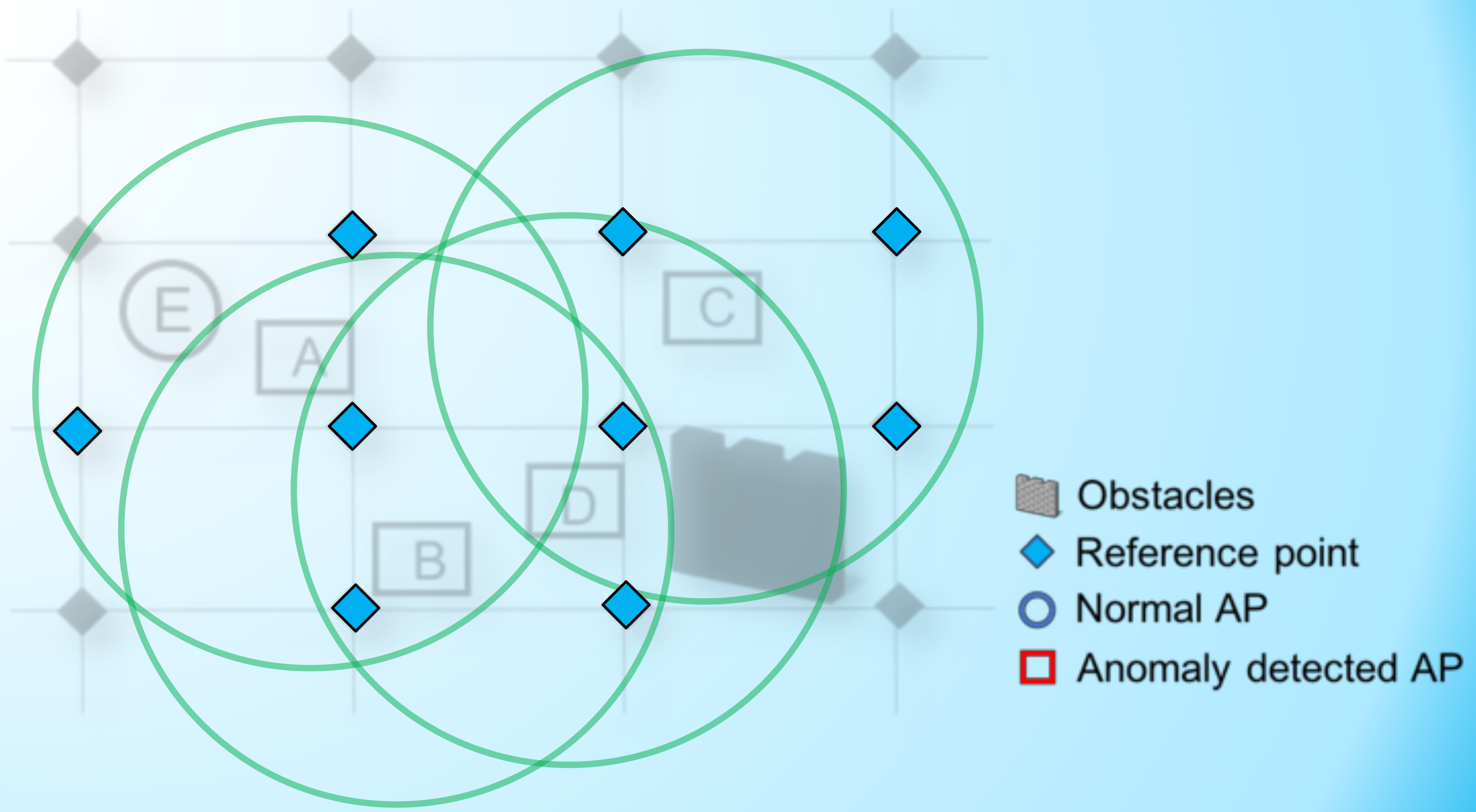
When Plurally Detected



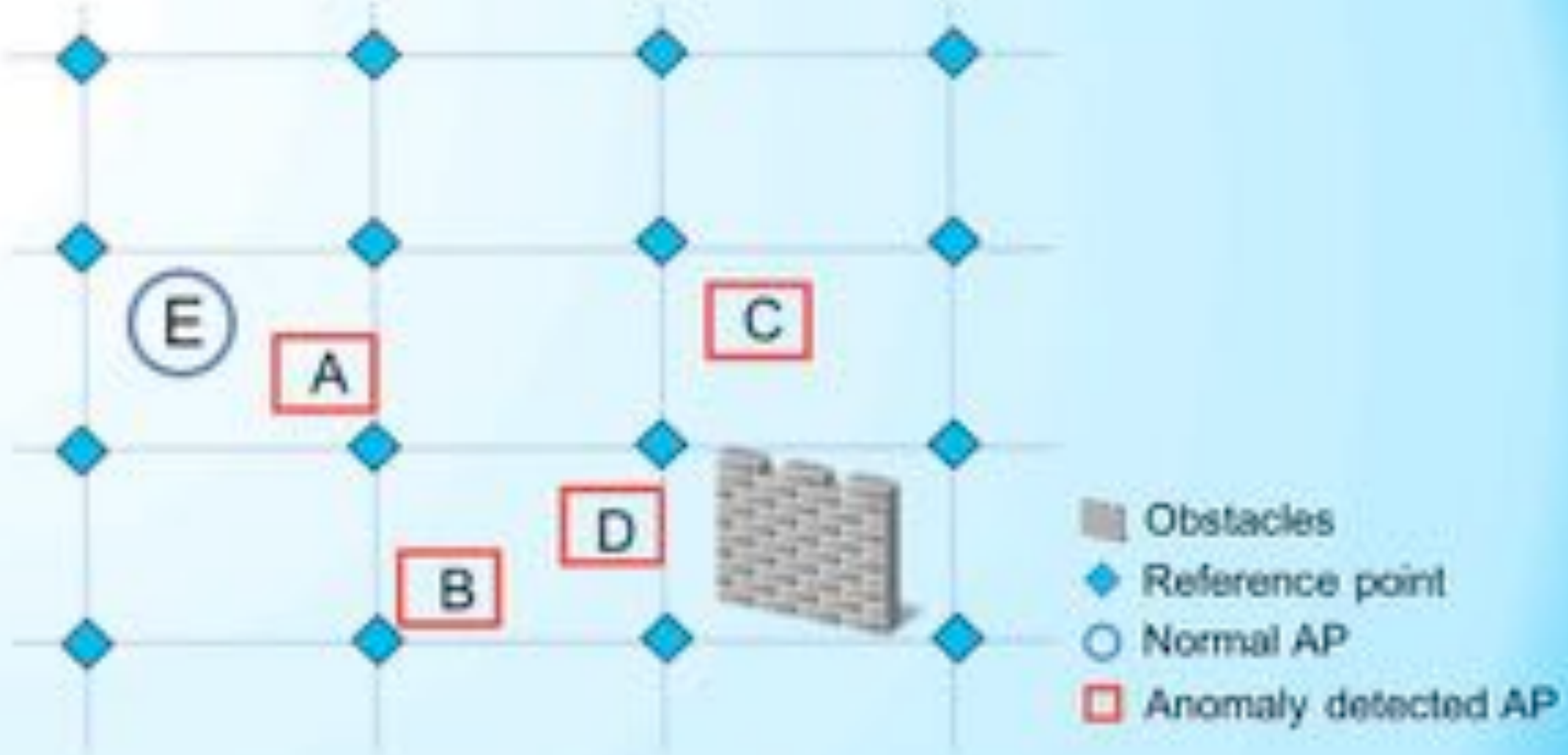
When Plurally Detected



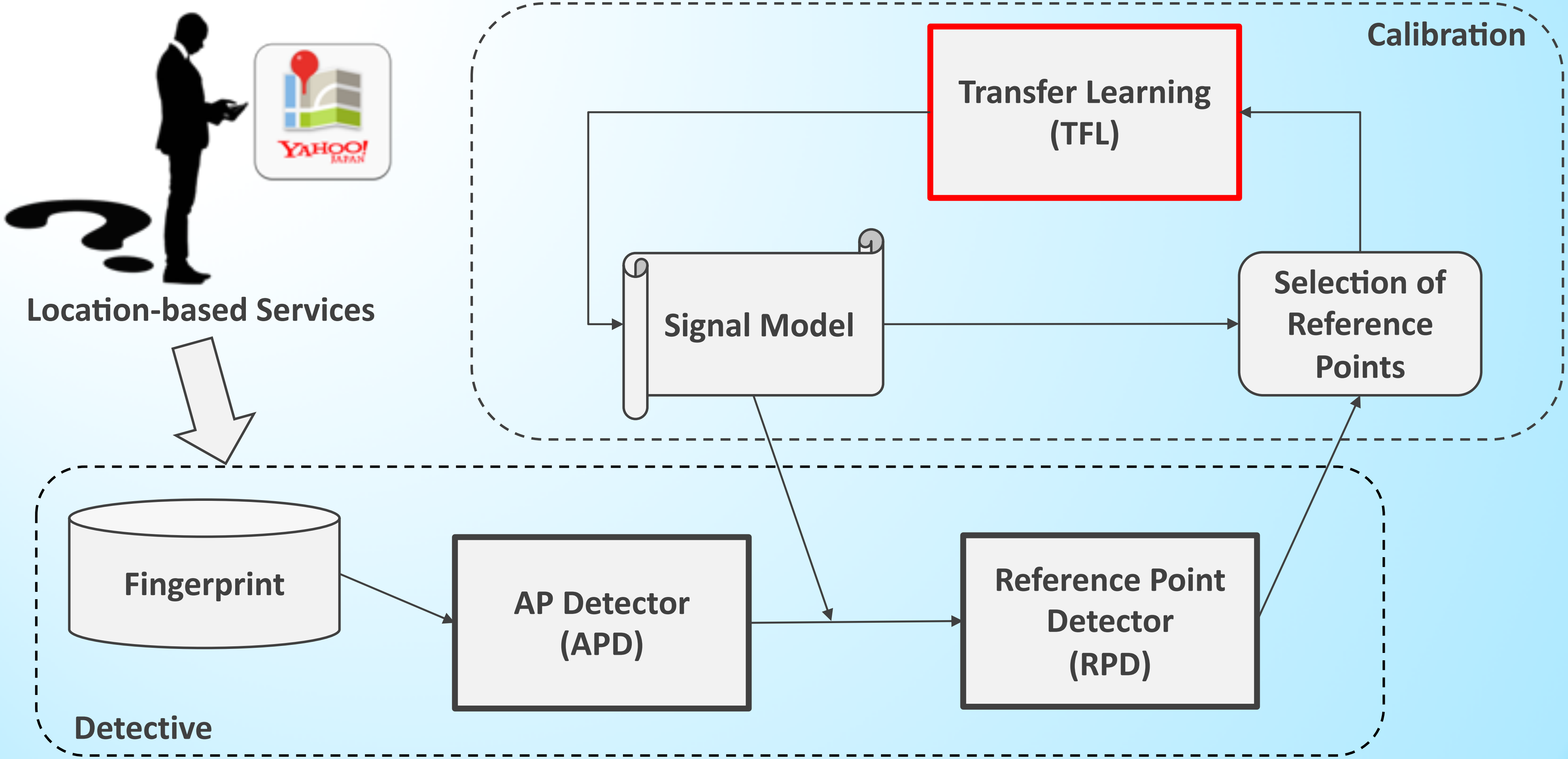
When Plurally Detected



When Plurally Detected



No-sweat Detective



Transfer Learning (TFL)

TFL can be any colors and replaceable

- Testify No-sweat Detective's performance by being applied to MixTrain and Lasso transfer methods[1)]

MixTrain method

- Closer to basis of transfer learning
- Learns parameters $\boldsymbol{\theta}$ itself utilizing all the dataset from primary to additional dataset at each calibration
- Updates model adding regularization term of L1 norm ($\sum_{i=1}^{|\theta|} |\theta i|$)
- Simply for keeping weights given to features from being hyper-activated by usual L1 norm

Lasso method

- Learns parameters from variation of parameters
- Does not store anything other than previous parameters and additional dataset
- Learns from L1 norm and regularization term $\left(\sum_{i=1}^{|\theta|} \left| \theta_i^{(k-1)} - \theta_i^{(k)} \right| \right)$
- Regularization minimizing variation between $\boldsymbol{\theta}$ at period $\boldsymbol{k-1}$ and $\boldsymbol{\theta}$ at period $\boldsymbol{k}$

[1]: Pengcheng Wu and Thomas G. Dietterich. "Improving SVM Accuracy by Training on Auxiliary Data Sources."

Evaluation

Tested in two different situations

- Laboratory dataset
  - Testify anomaly detection modules
  - Simulate environmental changes
- Underground dataset
  - Testify performance in real world scenario
  - With five-month-long labeled fingerprint and unlabeled fingerprint from Umechika-navi[1)]

[1]: http://www.umechikanavi.jp/

Evaluation

Tested in two different situations

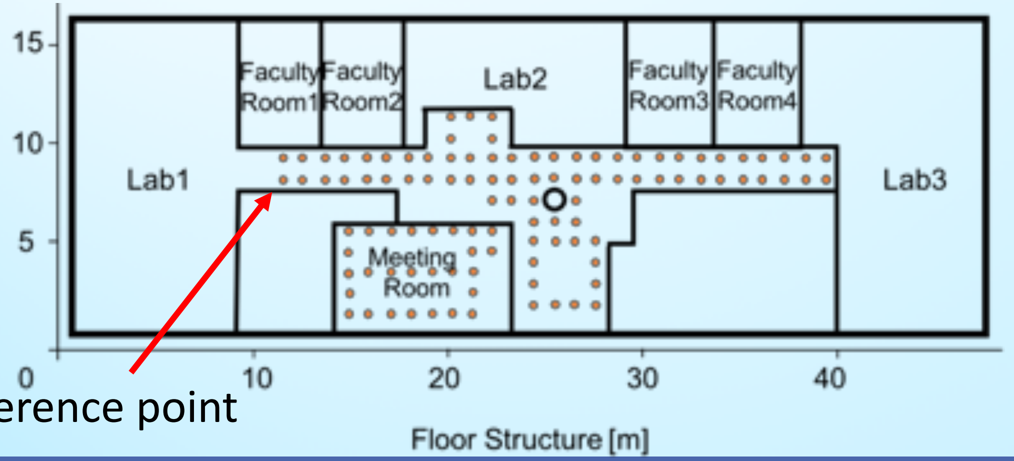
- Laboratory dataset
  - Testify anomaly detection modules
  - Simulate environmental changes
- Underground dataset
  - Testify performance in real world scenario
  - With five-month-long labeled fingerprint and unlabeled fingerprint from Umechika-navi[1)]

[1]: http://www.umechikanavi.jp/

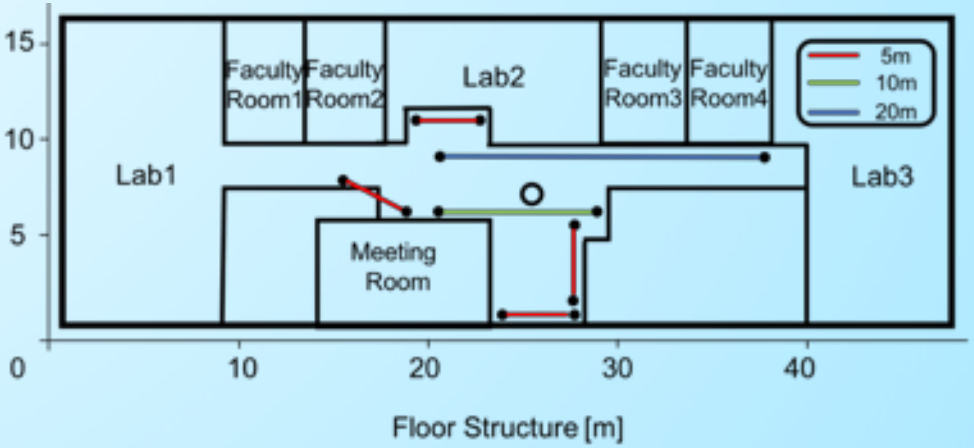
Laboratory Dataset

Details:

- Set 105 reference points within $1m^2$ of floor covering $348m^2$ (12m×29m)
  - Meeting room and hallway used as test-bed
- 2,100 labeled fingerprint in total
  - One scan per point and repeatedly ten times a day with Nexus5 over two days
- Simulate environmental changes by displacement of AP
  - Four 5m, one 10m, and one 20m displacements
  - Observe how environmental changes affect similarity of vector models

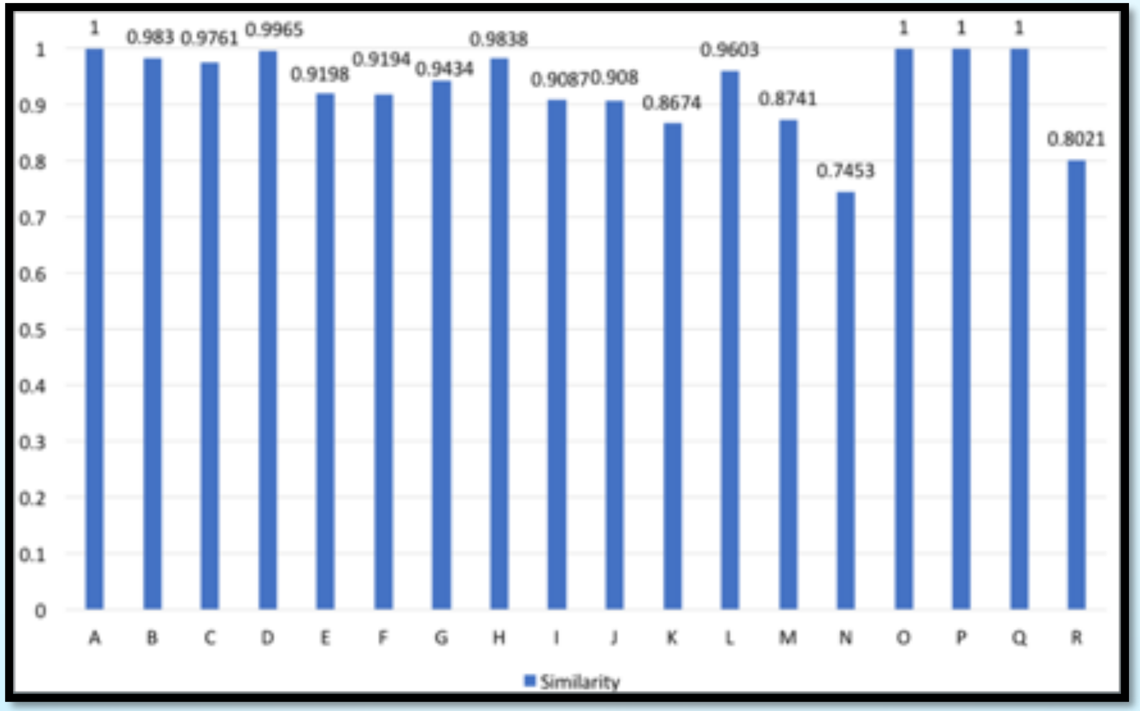


Reference point

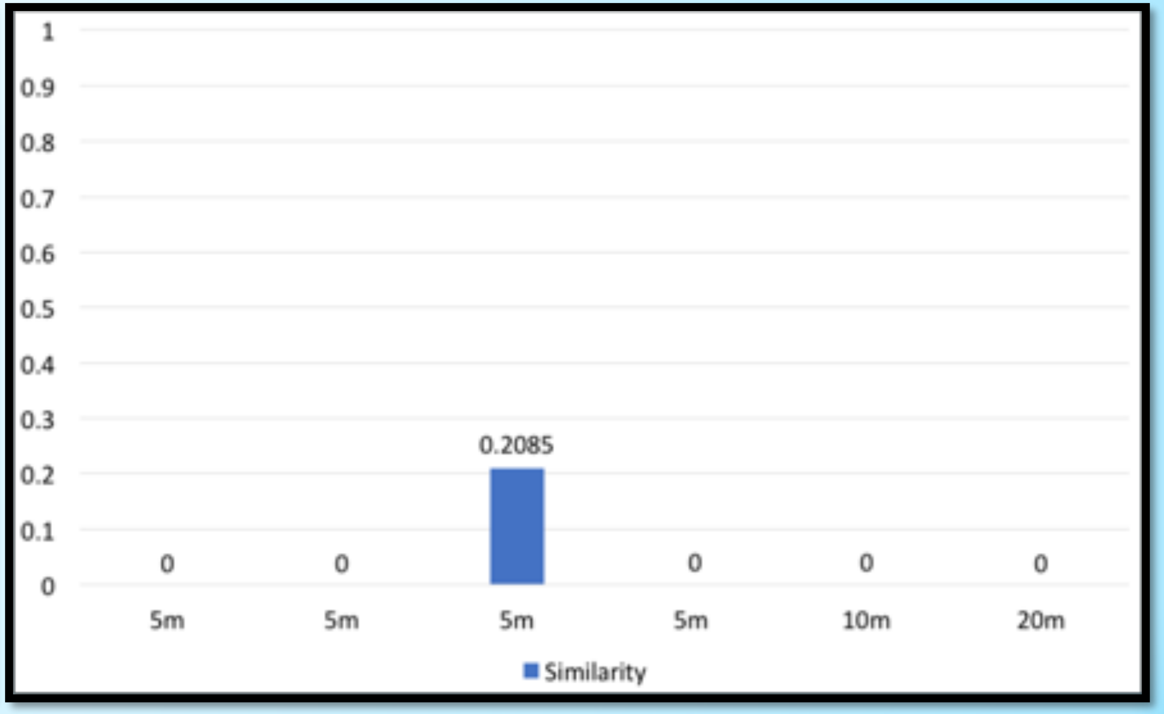


Results – Laboratory Dataset

18 vector models were created in both days



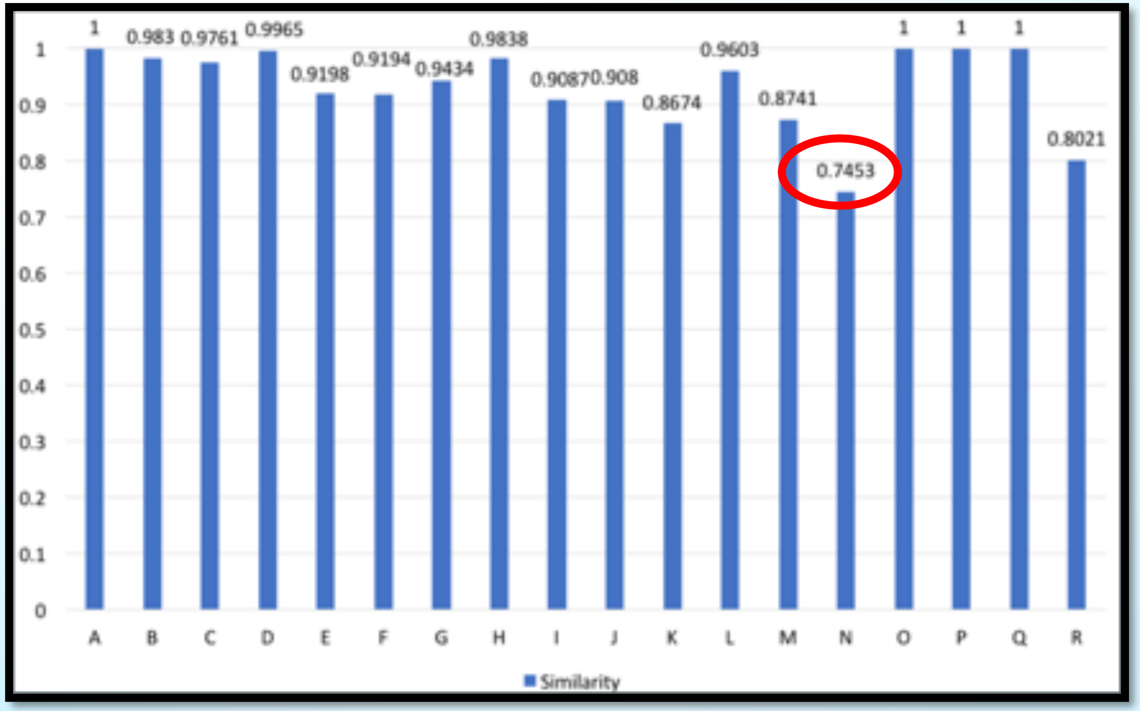
Non-distorted 18 AP's similarity



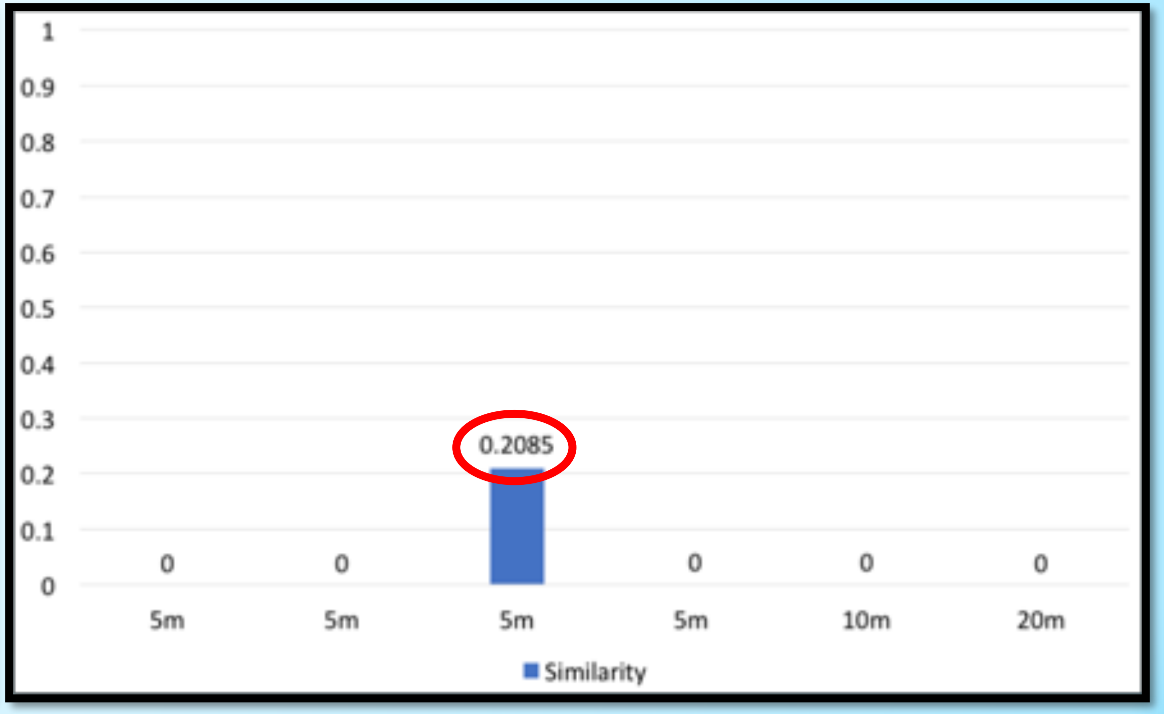
Distorted 5 AP's similarity

Results – Laboratory Dataset

18 vector models were created in both days



Non-distorted 18 AP's similarity



Distorted 5 AP's similarity

Validated significant difference in similarity
No-sweat Detective could detect environmental distortion

Evaluation

Tested in two different situations

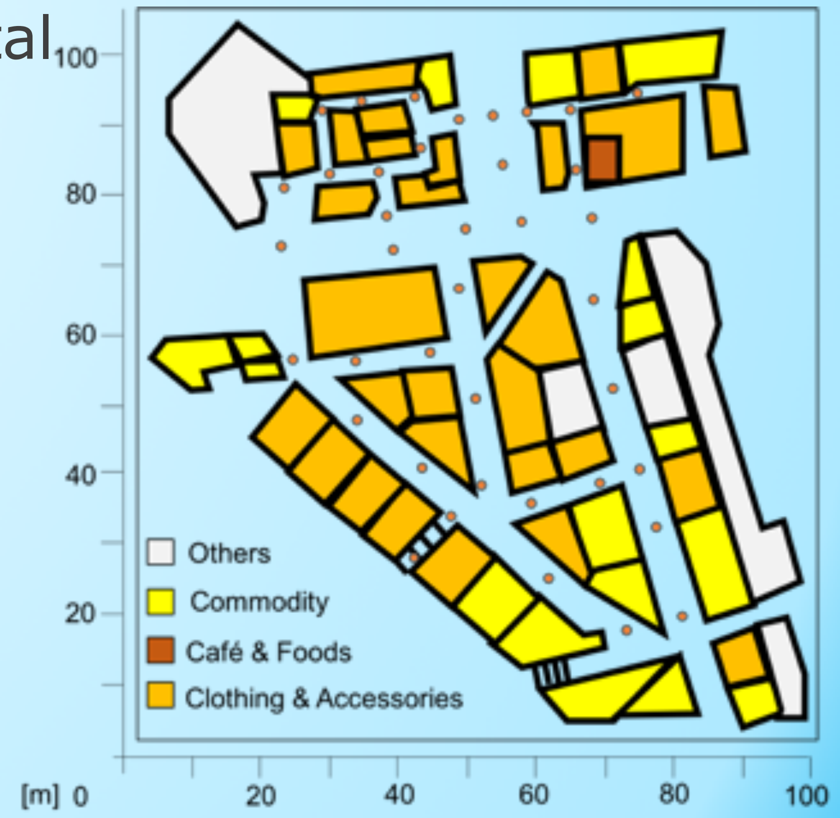
- Laboratory dataset
  - Testify anomaly detection modules
  - Simulate environmental changes
- Underground dataset
  - Testify performance in real world scenario
  - With five-month-long labeled fingerprint and unlabeled fingerprint from Umechika-navi[1)]

[1]: http://www.umechikanavi.jp/

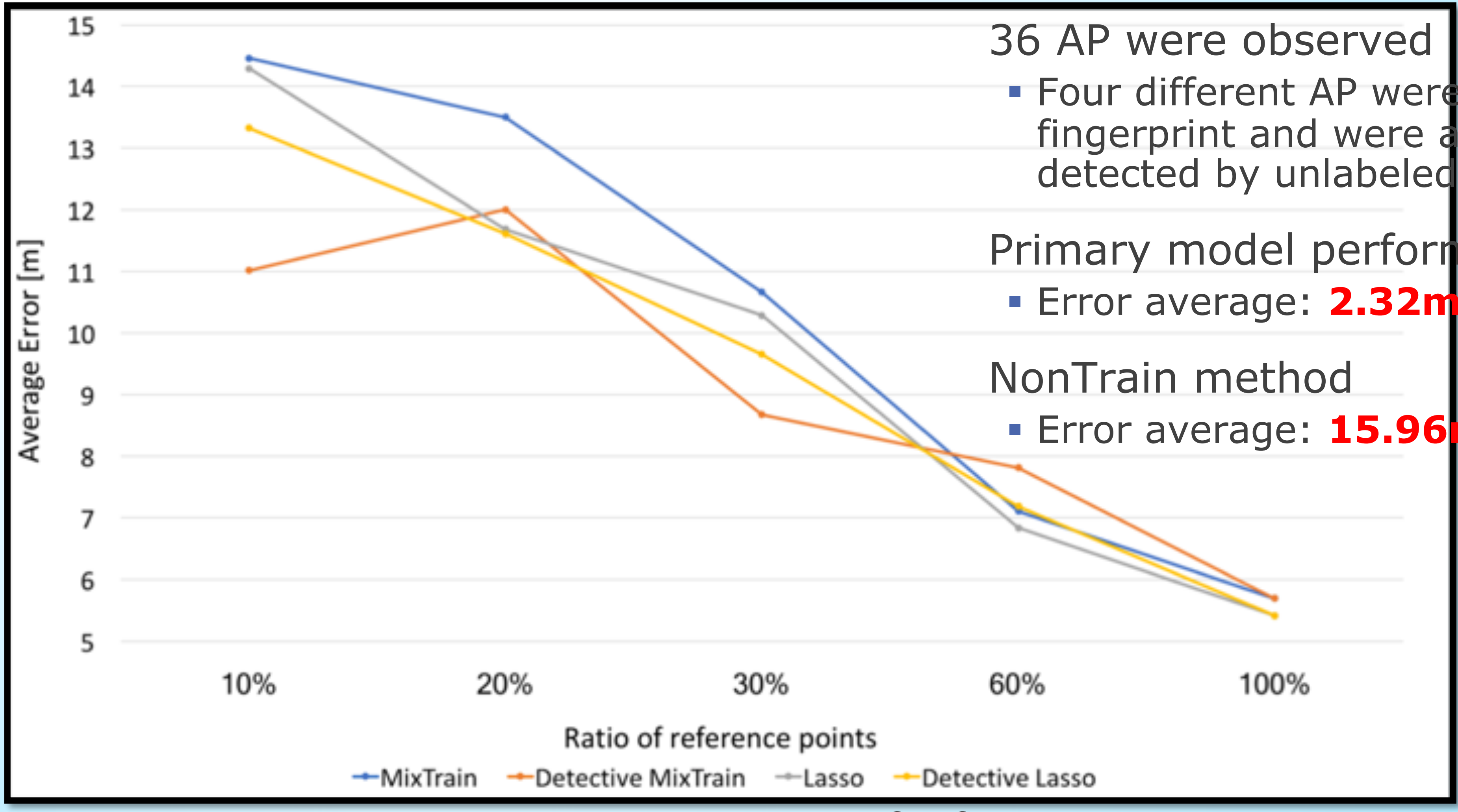
Underground Dataset

Details:

- Set 39 reference points at 8m intervals covering $348m^2$ (71m×65m)
  - Underground district where infamous for murderous congestion of people
- 2,693 labeled, 764 unlabeled fingerprint in total
  - Six scans per point with Nexus5 every two weeks
- Retrain model every two weeks
  - Sliding amount of reference points
    - 10%, 20%, 30%, 60%, 100%
  - Apply No-sweat Detective to conventional methods
    - MixTrain method
    - Lasso method
  - Validate average error of final model
    - NonTrain method (never trained) as base-line



Results – Underground Dataset 2/3



36 AP were observed

- Four different AP were detected by labeled fingerprint and were all covered by 10 AP detected by unlabeled fingerprint

Primary model performance

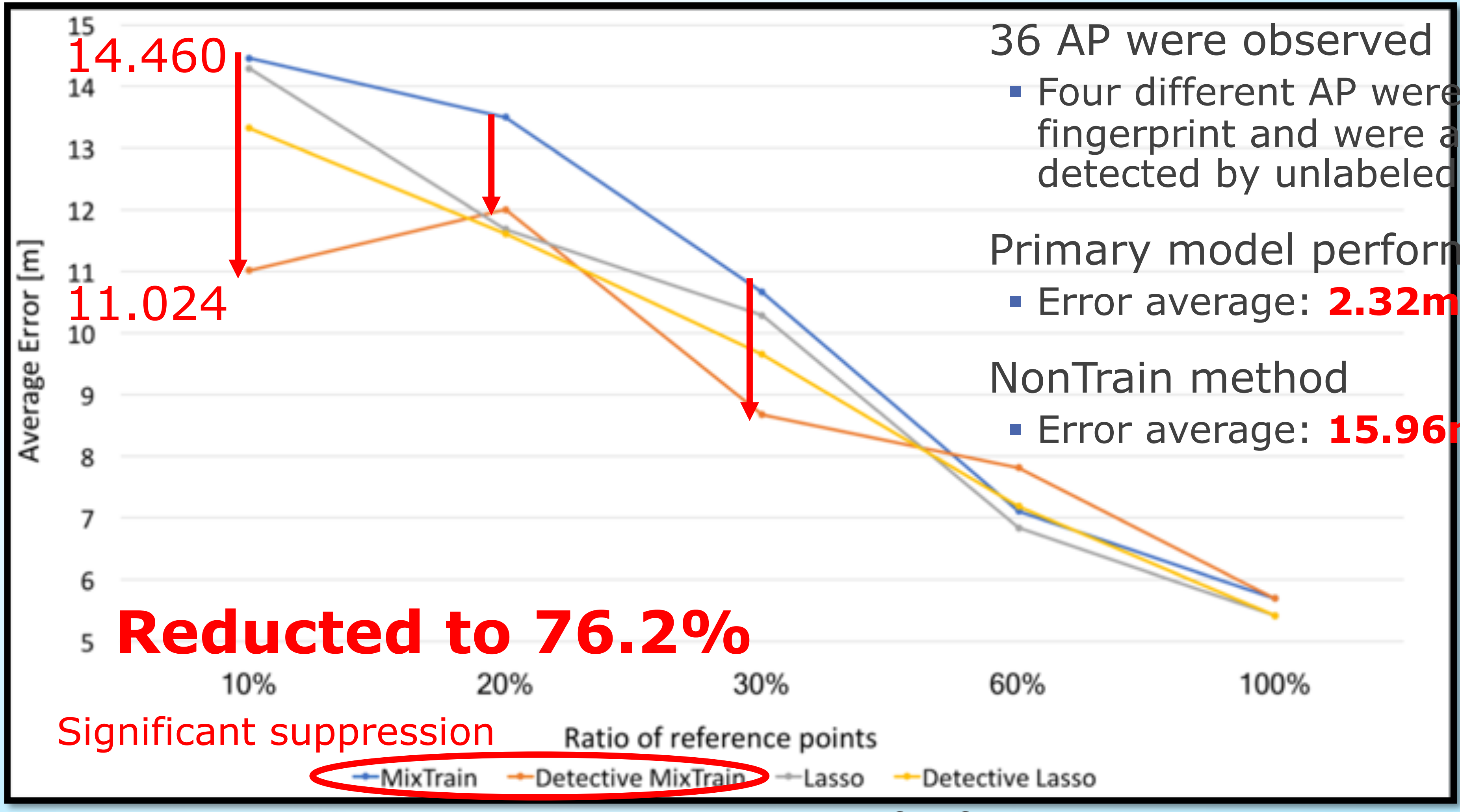
- Error average: **2.32m**

NonTrain method

- Error average: **15.96m**

Average error varying in amount of reference points

Results – Underground Dataset 2/3



Average error varying in amount of reference points

36 AP were observed

- Four different AP were detected by labeled fingerprint and were all covered by 10 AP detected by unlabeled fingerprint

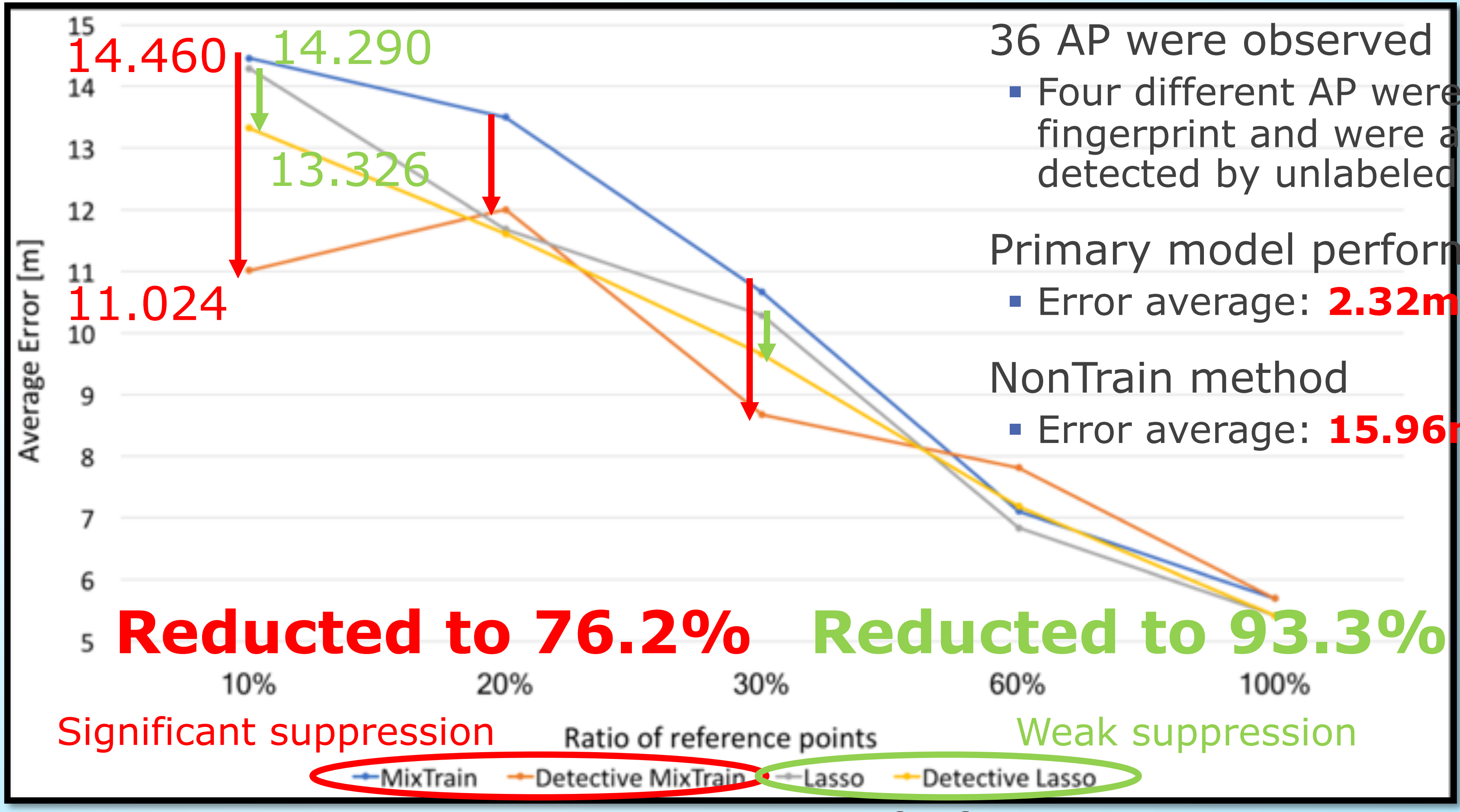
Primary model performance

- Error average: **2.32m**

NonTrain method

- Error average: **15.96m**

Results – Underground Dataset 2/3



36 AP were observed

- Four different AP were detected by labeled fingerprint and were all covered by 10 AP detected by unlabeled fingerprint

Primary model performance

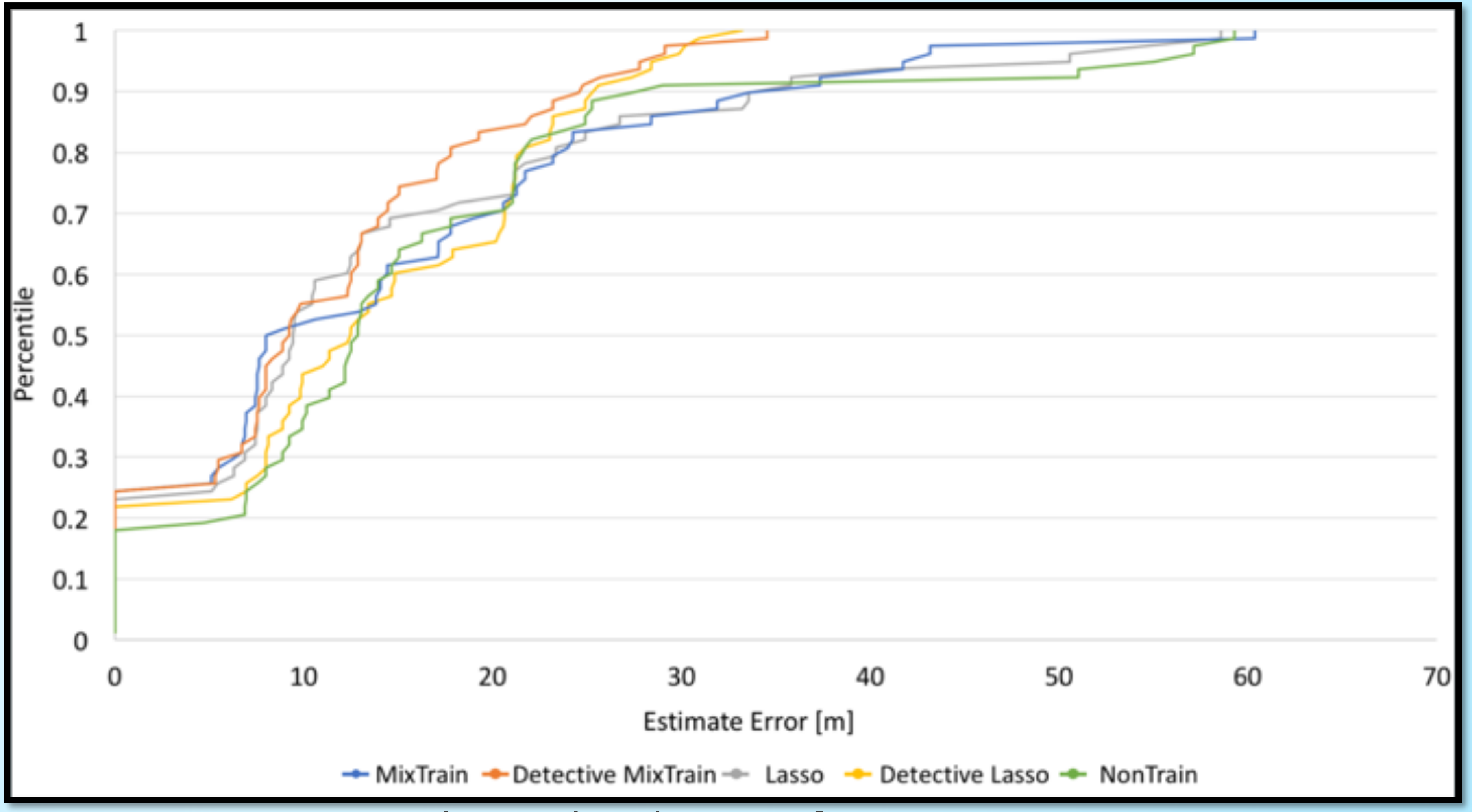
- Error average: **2.32m**

NonTrain method

- Error average: **15.96m**

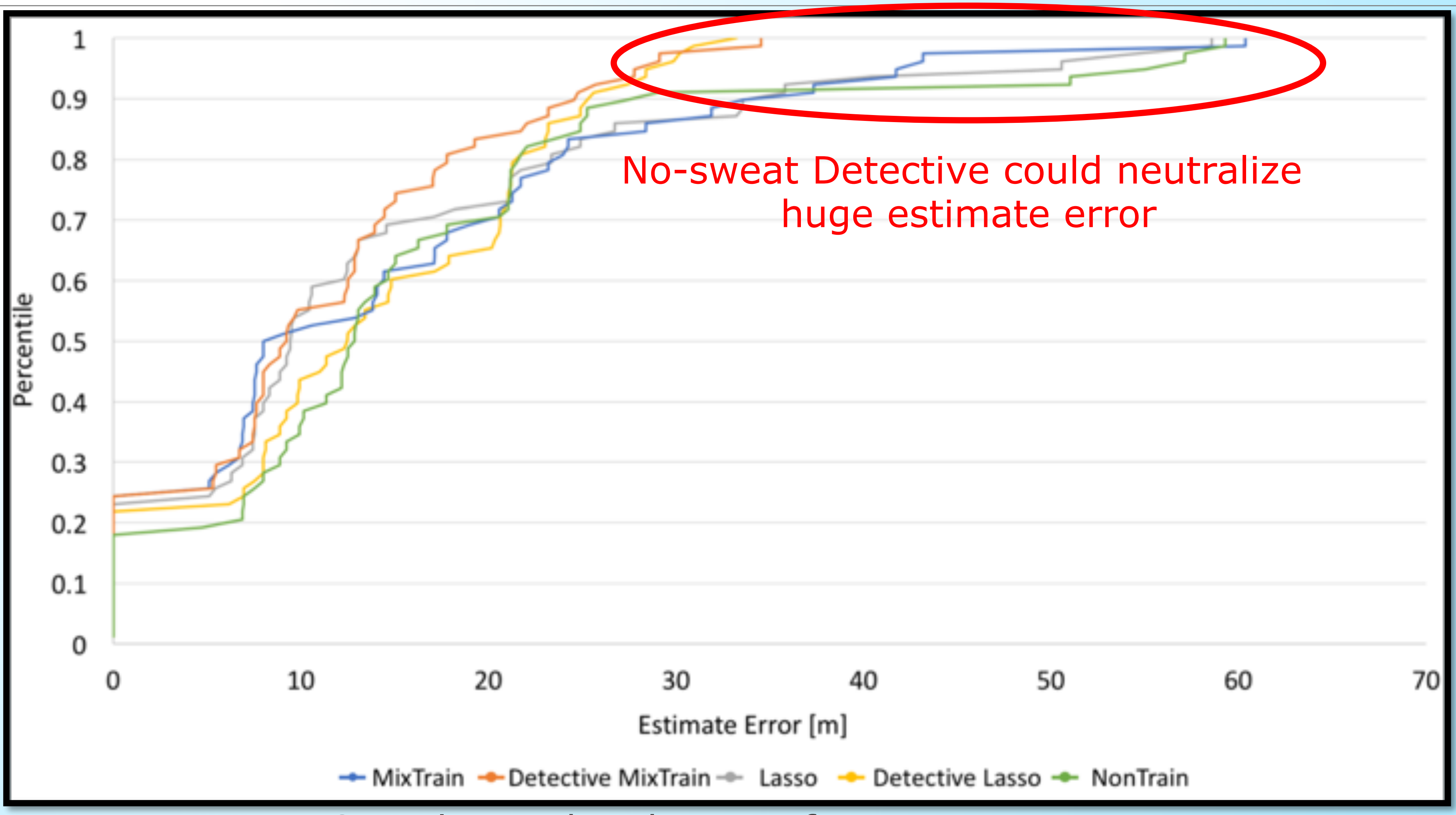
Average error varying in amount of reference points

Results – Underground Dataset 3/3



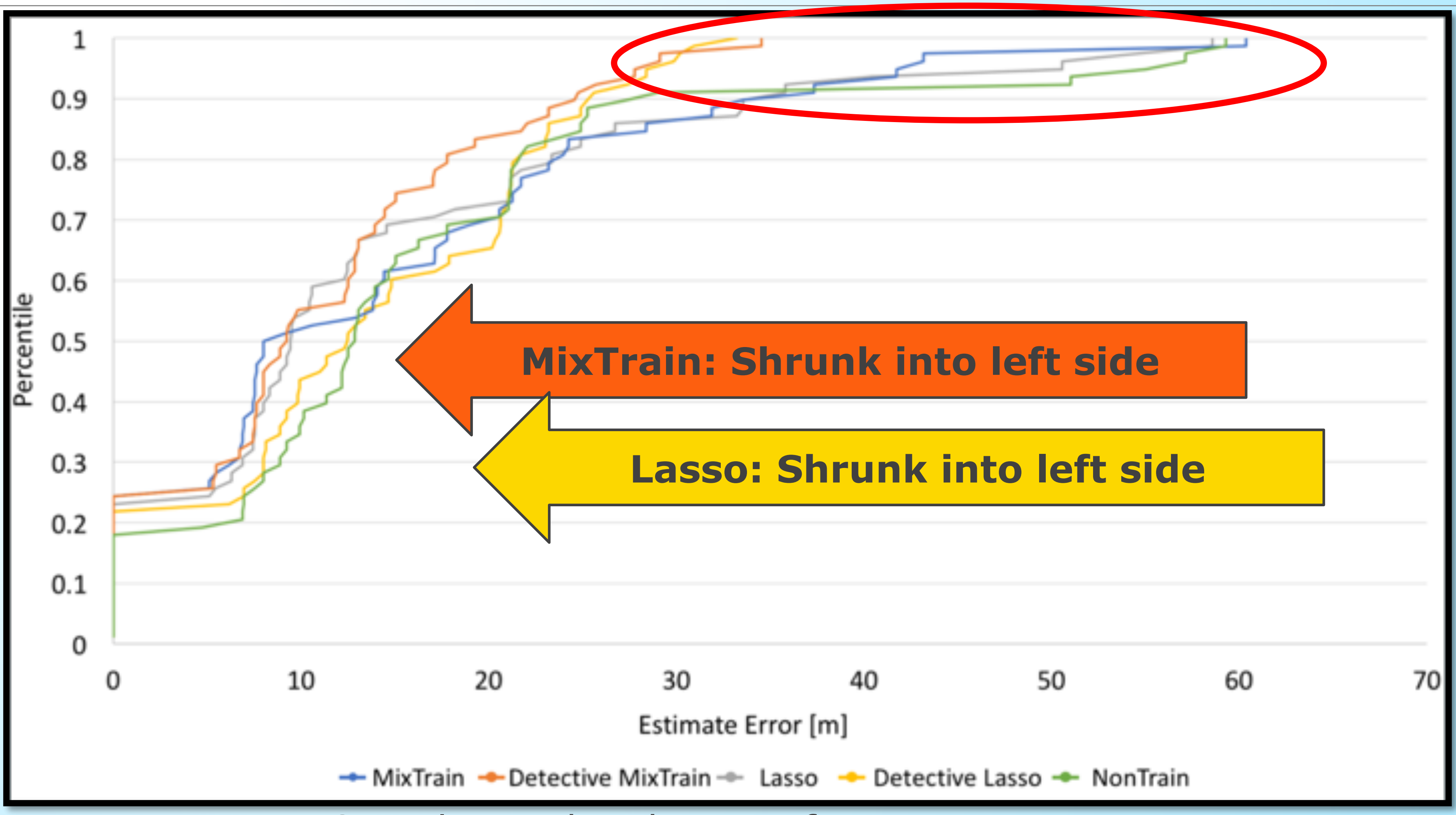
Cumulative distribution of estimate error

Results – Underground Dataset 3/3



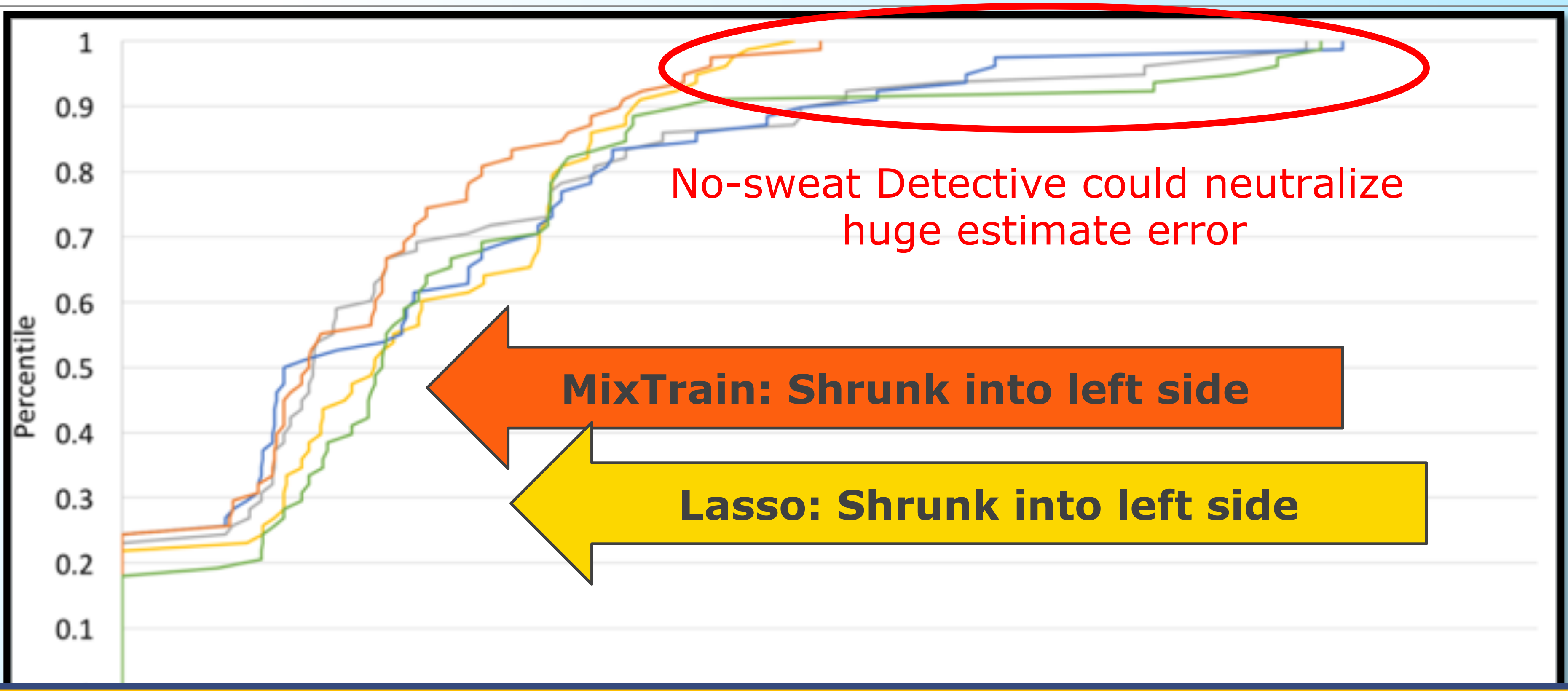
Cumulative distribution of estimate error

Results – Underground Dataset 3/3



Cumulative distribution of estimate error

Results – Underground Dataset 3/3



No-sweat Detective can be applied to existing transfer learning methods to maintain higher accuracy over long time operation

Cumulative distribution of estimate error

Summary and Outlook

Fingerprinting localization mode deteriorates over time

Transfer learning employed majorly

- But additional dataset are randomly sampled

Propose No-sweat Detective

- To identify reference points where environmentally changed
- With no effort using unlabeled fingerprint from user of location services
- Could detect environmental distortion
- Achieved higher recovery with same amount of additional dataset
  - With five-month-long observation

Outlook

- Validate performance recursively
- Substitute unlabeled fingerprint as labeled fingerprint to let No-sweat Detective be autonomous