Distributed ESP-Now CSI Human Activity Recognition for Few-Shot Energy Saving in Building Automation

Adolfo Bauchspiess*, Tanmay Mane, Iris Fürst-Walter and Jürgen Becker

***Abstract*— Device-free sensing with wifi signals has gained great attention from the community to enhance building automation systems. In such systems, energy efficiency can be improved through predictive forecasting of occupancy as well as human activities carried out in the room. In this context, we propose a wifi-network of eight ESP32 modules, where we measure Channel State Information (CSI) and couple the sensor data with few-shot bundled majority vote models to perform Human Activity Recognition (HAR). We achieve an accuracy of 74% for HAR. Those are promising results for HAR as a basis for low-power building automation solutions using COTS in single antenna sensor setup.**

I. Introduction

Building automation enhances the comfort of office and residential spaces, where the personal experience is influenced in particular by suitable climate control. However, conventional systems often rely on air quality measures, e.g., temperature and CO_2-density, which are by nature slow and therefore inefficient. In contrast to direct measures, heat load forecasting could be used to reduce delays and prevent overdrive. Such a forecasting can efficiently be realized by Channel State Information (CSI) WiFi sensing. This represents a device-free sensing system which even comes with the benefit of enhanced privacy when compared to camera-based systems.

Device-free sensing with WiFi signals has gained great attention from the community, e.g., [1], [2]. 802.11 b/g/n with OFDM is nowadays available in cell-phones, tablets, notebooks, and many other devices. They provide a multi-carrier CSI that is more versatile than RSSI.

When CSI-based sensing is coupled with transfer learning, knowledge gained from environments with different room layouts, furniture, and usage patterns can be reused. Training a neural network for each environment is out of question, so that transfer-learning, as few-shot learning, e.g., [3] is a feasible approach to automatize buildings. A large data acquisition for one typical room should be few-shot trained for another environment of the building, which is called smart cross-environment automation transfer.

Human activities in an office could vary a lot. Aiming energy-saving we evaluate the proposed approach with male and female subjects at different locations, in different environment scenarios for {sit, stand, walk, empty}, as in [4]. They give us four levels of occupants fine thermal-load estimation, cf. Fig. 1. See Table I for some choices of HAR activities,

Our contribution is as following:

- Network of eight synchronous ESP32 models for CSI wifi signals
- Application of low-cost, readily available COTS
- Bundling of TX-RX lines with majority voting to perform HAR more efficiently and robustly

II. Occupancy-based energy-saving

Occupancy estimation is relevant for energy-saving in building automation, [5], [6]. A video-based occupancy counting system for a feed-forward HVAC (Heating, Ventilation and Air Conditioning) controller was implemented in [5] and 26% energy saving was reported. Thermal load changes are taken into account earlier than by a conventional feedback loop. The PI HVAC controller in Fig. 1 is designed for a nominal load (empty room, first-order model, K, T, with transport delay). In a first automation level, each person entering/leaving the room is accounted (n estimated occupants) by an average thermal load K_{av} per occupant. On a second level the individual activity estimation, Ai, is used to fine tune the thermal load estimation. Low energy activities, like reading, decrease

TABLE I

Typical HAR classes from the literature.

ref.	1st Author	Year	Classes (Activities)
[1]	Yang	2021	bend, clap, **walk**
[3]	Bahadori	2022	jump, **stand, walk, empty**
[2]	Strohmayer	2024	**empty, walk**, walk+arm-waving
[4]	Natarajan	2024	**sit, stand, walk, empty**

*Corresponding author, ENE/FT, University of Brasília, Brasil, adolfobs@ene.unb.br. Tanmay Mane, Iris Fürst-Walter and Jürgen Becker are with ITIV/KIT, Karlsruhe, Germany.

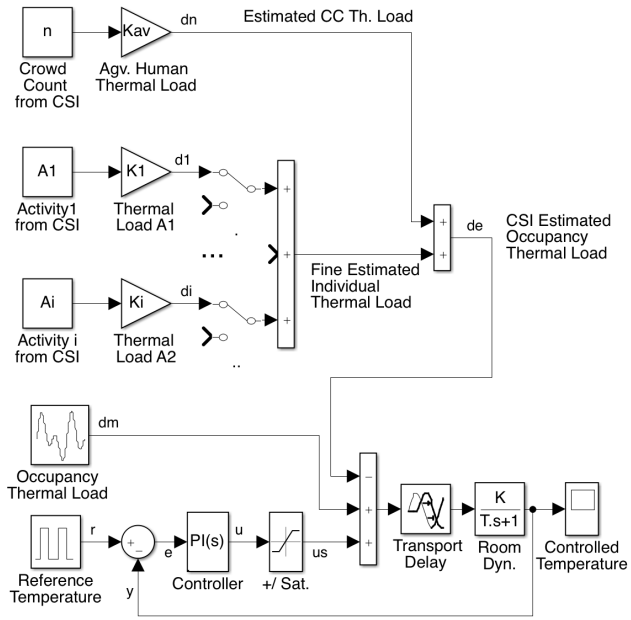

Fig. 1. Feed-forward (disturbance rejection model) thermal control with occupants thermal load estimate. Adapted from [5].

the nominal estimated thermal load, while energy-demanding activities, like sport, would increase the corresponding thermal load estimate. The estimated occupancy thermal load would be:

$$d_e(t) = \underbrace{n(t) * K_{av}}_{\text{count-based}} + \underbrace{\sum_{i=1}^{n} A_i(t) \quad \underbrace{K_i}_{\text{individual's}}}_{\text{HAR-based estimate}} , \qquad (1)$$

where the CSI-estimated values $n(t)$ and $A_i(t)$ are relevant for energy saving. For the illumination, with natural light harvesting, localization of the individuals in the room is also an important information. A good estimation of $d_e(t)$, close to the real occupancy thermal load d_m, favors the anticipatory disturbance rejection. Other real-world disturbances (external temperature, solar radiation, neighbor rooms etc.) are omitted here for clarity (they are taken care by the "normal" feedback loop).

In the conventional feedback control strategy, disturbances $d_m(t)$ hit the process, deviating the controlled variable $c(t)$ from its set point. The negative feedback will bring $c(t)$ back to the desired value using the error information $e(t)$, Fig. 1. Each time the number of occupants, n, changes, the corresponding thermal load (disturbance estimation) is updated. Each occupant's activity (A1,...Ai) is accounted, adding or subtracting from the average.

Typical in-door energy-saving automation approaches, see Fig. 1, that can be implemented with CSI-sensing are:

- Movement/Presence detection, $d_e(t)$
- Counting people at a door/passage, $n(t)$
- Counting people by AoI/zones, $n(t)$
- HAR at Area of Interest (AoI)/zones, $A_i(t)$
- Identify individual's thermal load classes, K_i

In the present work we will focus on HAR at an AoI, aiming building automation. As a proof-of-concept of ESP32 sensing devices for low-cost transfer-learning of HVAC automation between rooms of a building.

III. Related Work

Aftab et al. 2017, [7], shows the importance of using model predictive control with real-time occupancy recognition for HVAC. Video was used with machine learning. Natarajan et al. 2024, [4], implemented a smart LED Lighting control based on ESP32 CSI HAR. Choi et al. 2020, [8], uses only features extracted from CSI in machine learning system. Bahadori 2022, [3], developed a Few-Shot for Multi-Antenna Multi-Receiver CSI learning system. Her few-shot code was the basis to develop the present work on single antenna ESP-32 with ESP-NOW protocol. The great success of prototypical Few-Shot, see [9], has inspired many researchers to try new ways with prototypes.

IV. Distributed ESP-Now CSI Sensing

An overview of the proposed occupancy-based smart building automation can be seen in Fig. 2. Crowd-Counting and HAR are relevant for HVAC and Illumination. a) A typical Area of Interest, AoI, is chosen in the building to automate. A Sensor Network is assembled to cover AoI with good overlapping of Tx-Rx Links. b) CSI is acquired using a routine with typical activities, at different locations, and different persons. c) Classification is learned by three CNN models used in a majority vote. Each model (m1, m2, m3) is a bundle of CSI links. Different bundles, covering specific areas, are trained to obtain the fittest classifier. d) Sensor Networks are built in the test/validation environments. Few-shot adaptation of the prototypes (p1, p2, p3). e) Once validated, new prototypes (pn1, pn2, pn3) are few-shot adapted to all environments of the building to automate with similar architecture.

Using CSI sensors distributed over the environment to cover the AoI, a larger area can be covered (spatial

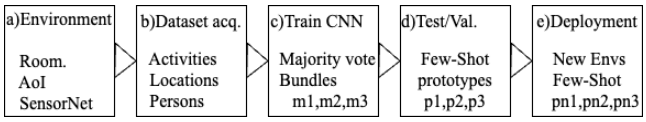

Fig. 2. Overview of occupancy-based smart building automation.

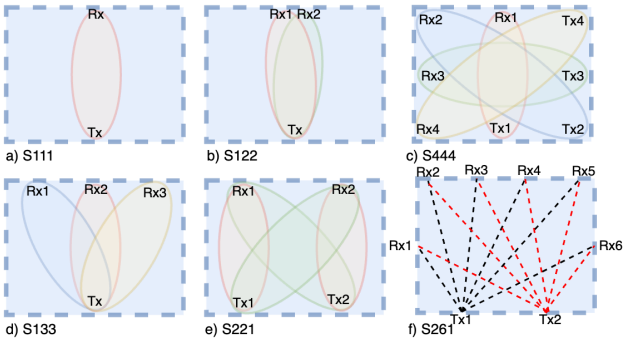

Fig. 3. Tx-Rx 2^{nd} Fresnel CSI ellipsoids cover a given area of interest, as seen in the literature. a) Single link, e.g., [12]. b) Two links with small displacement. "Curtain" used to count passage, e.g., [13]. c) 4 links arrangement used by [8]. d) One Tx - Three Rx (each with 4 antennas), [3]. e) S221 and f) S261, proposed in the present work. Where S261 stands for Two Tx, 6 Rx, and 1 recording module. S261 Fresnel zones are shown in Fig. II).

coverage), as seen in Fig. 3. A challenge is to synchronize the acquisitions. With rapid HAR activities, such as walking or jumping, it is crucial that the CSI being processed describes the same activity.

There are many good results about machine learning, however, for the practical use in building automation, it is necessary that a deep trained model be quickly adapted to different environments, e.g., [10]. In this Cross-Domain context, few shot is a very effective approach, see [11]. Domain diversity considers different environment, location or person. The configuration of the wireless sensors and the localization of the activities have a great impact on the captured CSI signals.

A. S261 Training and Tests Scenarios

Fig. 5 shows the S261 Sensor layout of our training and test scenarios, labeled GDH and ITIV.(GDH: a university's guest room and ITIV: a work office at ITIV/KIT. Engesserstr 3 and 5 in Karlsruhe, respectively). In GDH we see that the red bundle (ellipsoidal links [M1:S12,S13,S15]) gives good coverage to activity "Stand" at location L4. The blue bundle (links [M1-S9:S11]), on the other hand, is better suited to the given "Walk" activity.

Device-free CSI-Sensing is dependent on the spatial distribution of transmitters, receivers, and the

TABLE II
COMPARISON: S2611 ('OURS') AND S133, [3]

	S261	S133
Links	12	48
Antenas	1	4
Subcarriers	52	242
Image	34x34 (Amp/Ph)	242*242 (Amp)
Bandwidth	40 MHz	80 MHz
Hardware	8xESP32	1xR78006, 3xRT-AC86U

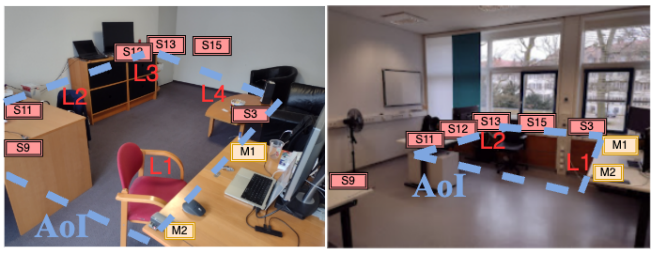

Fig. 4. S261 at Training Scenario, GDH, Room7 (aprox. $6x4m^2$), and Test Scenario, ITIV, Room 1.26, (aprox. $6x9m^2$). Masters M1, M2 close to the collecting notebook. Sensors cover the AoI.

location/orientation of the sensed subject. In [3] one access point and three stations carry four antennas each. With ESP distributed CSI links (single antenna) we have less links but a better spatial coverage, as can be seen in Fig. II.

Table II compares S261, distributed single antenna modules, with flexible bundling, "ours", and sensor network used by [3], S133, which, considering 4 antenna devices should be called more properly, S4.12.3 (four sending antennas, 12 receiving antennas, 3 recording devices, with fixed bundling B444).

Bundles describe how links are grouped for majority vote models (m1,m2,m3), e.g., B321 stands here for a Bundle of 3 red links, m1=(S12.S13.S15), 2 green links, m2=(S9.S11), and 1 blue link, m3=(S3), as shown in Fig. 5. Bundles B321 could be assembled

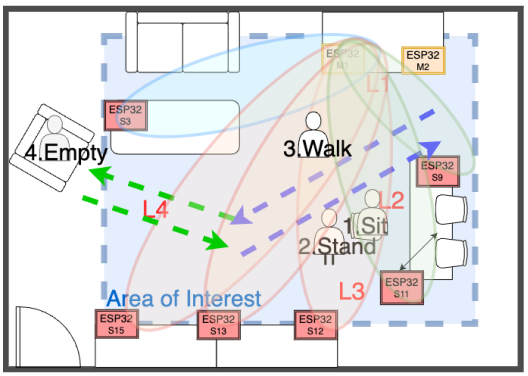

Fig. 5. Train Environment GDH, L2, S261, B321.

with other links, and an additional letter need to be used for experiment identification, e.g. B321a. The six links to M1 are shown, the six links to M2 are omitted, for clarity. The idea of using bundles is to allow location dependent segmentation of the CSI data, useful in a future zone-based automation.

V. S261 ESP-Now Sensor Network

ESP-Now, a Espressif developed MAC connection protocol uses the PHY and MAC layer from 802.11n simplifying the setup when compared with the standard WiFi connection, [14]. With ESP-Now the connection is immediately available after power-on. The ESP-Now protocol allows the transmission of 250 useful bytes and so the idea behind S261 was to use two masters sending a broadcast signal in a short time interval. The CSI from each master will be stored in each sensor node. In the CSI261 configuration (2 Masters, 6 Sensors, 1 Collecting node) 12 CSI links are, almost synchronously, captured. Then each sensor transmits its CSI package (CSI from M1, M2) to the collecting master M1, which saves each package in a .csv file.

In the realization it was noted that using a telecommunication device as a sensing device is not ideal. For sensing synchronous acquisition is critical for correct activity recognition. ESP-32 modules do not expect ACK for broadcast signals. In the MAC layer events occur with different priorities. Some ESP-Now tasks try ”guarantee data delivery”, re-sending data. It is good for a communication but troubles synchronous sensing. Without collisions, a 20 ms acquisition Cycle was expected with ESP-Now (6 x 2.2 ms). Due to missing/duplicate packages we adopted an ”save” acquisition cycle of $> 750ms$). Figure 7 shows, in a simplified way, how bundled data is generated from the CSI_file.csv. For a given sensor network, different bundles can associate links, that are somehow related, expecting so a higher classification accuracy.

The training dataset had 423 CSI images. For the cross-environment tests, 84 images were used. CSI magnitude from S3 to M1 is depicted in Fig. 8 for a full acquisition cycle, 200 s (sit, stand,walk, empty). To enhance the readability the first 5 carriers amplitude and phase are shown in separate. It is interesting to note that the first 5 carriers present a quite similar behavior. Considering all 52 carriers the amplitude shows a considerable spread over the carriers. The corresponding phase spread is significantly smaller.

In Fig. 8 we recognize intuitive rules relating CSI and Human Activity, e.g.: ”sit” has higher CSI values; ”stand” with lower CSI values (WiFi blocked); ”walking” has higher CSI variance; ”empty” with highest ”free sight” CSI, and so on. We understand, however, that a neural network is better suited to ”learn” from CSI data, given the great variability of the signals.

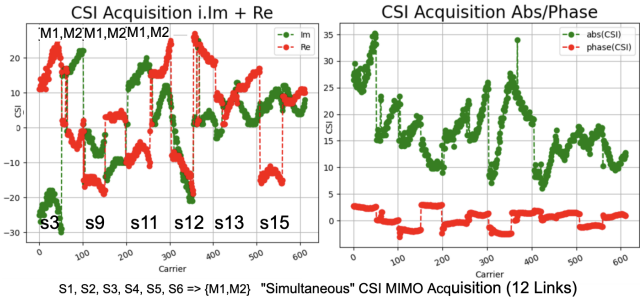

Fig. 6. A Cycle CSI Acquisition S261 (produces one CSI-Image line). CSI Real/Img and Abs./Phase.

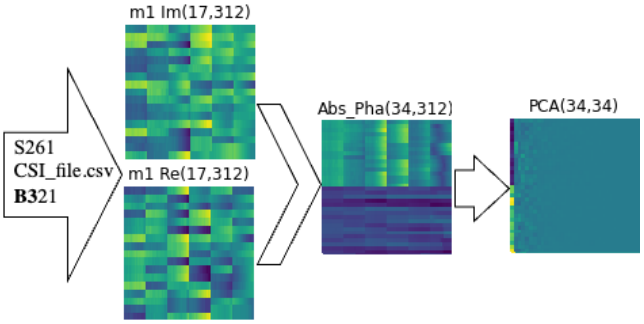

Fig. 7. CSI bundling. For S261, each Broadcast from (M1,M2), produce 2 x 6 x 52 useful CSI subcarrier samples. A Bundle B321, with e.g., m1=(S12.S13,S15) would build images with size (17,2*3*52) for 10 s activity. Absolute value and phase are concatenated and then PCA compressed.All models used in majority vote are PCA compressed to the same (34,34) size.

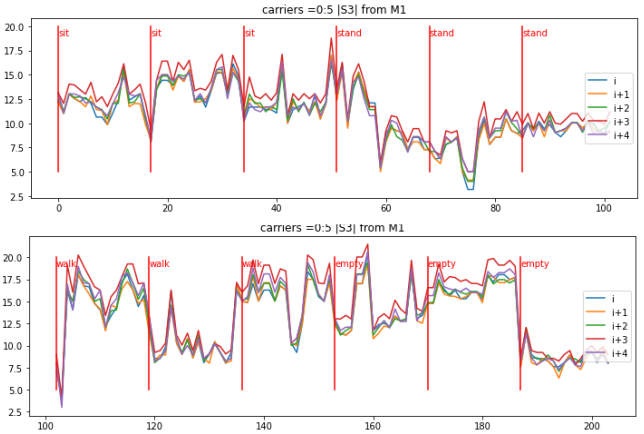

Fig. 8. CSI for link M1-S3 3 min. First 4 carriers of 52 useful carriers. Amplitude. Activity collected over 10 s produce a frame. 30 s for each HAR activity (sit, stand, walk, empty,...)

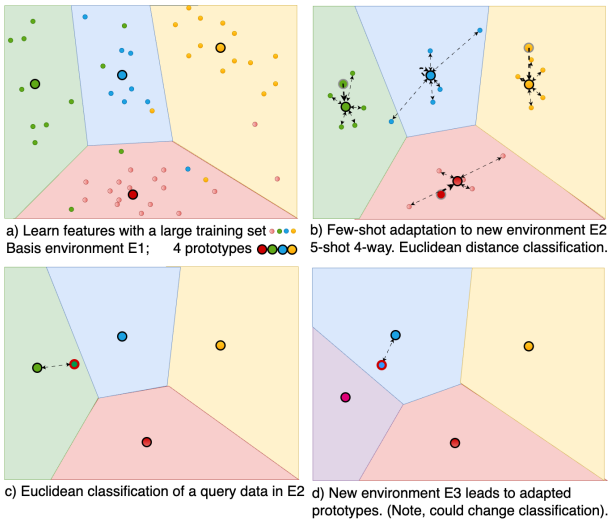

Fig. 9. Few-Shot Prototypical Domain Adaptation.

VI. Bundled Few-Shot Majority Vote

With 34x34 images, a standard CNN classifier with 4 convolutional layers, we obtained 31,54% cross-validation accuracy. With 68x68 and 136x136 images the were even worse.

To implement a cross-environment algorithm that is suited for building automation, we choose Prototypical Few-Shot. We followed the ideas given in [3] for Multi-Antenna Multi-Receiver CSI. Prototypical CNN are trained in one environment (features) and then the classification is adapted with few shots to different environments. As can be seen in Fig. II, we used low cost, distributed, single antenna modules, which are normally sufficient to provide buildings energy-saving.

Fig. 9 illustrates the prototypical few-shot methodology. A machine learning model, e.g. CNN, is used to capture the non-linear feature map using a large data set (small labeled/colored circles), as illustrated in a). The prototypes (larger cicles) represent the n-way resulting classifier. b) For a new Environemt E2, few-shots (5 labeled circles for each class here), are used to adapt the prototypes to the new environment. c) Queries are classified by the shortest distance to the available prototypes. d) A new environment E3, brings renewed prototypes and could give even a different classification. In figure d) the different color used for the left most prototype, instead of green, should illustrate a completely new few-shot class, never trained, 'adopted' by transfer learning mechanism. With bundled majority vote, Fig. 10, we could reach accuracy values, that still need enhancement, but that can already bring considerable

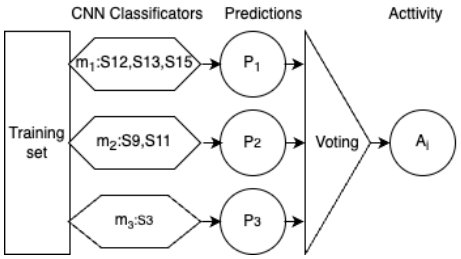

Fig. 10. Majority Vote of bundled B321 models (m1,m2,m3).

energy-saving for real buildings. The three CNN models (m1,m2,m3) used bundled CSI data, that can be freely assembled in an attempt to capture spatial activity dependencies. Overlapping bundles are also possible, as with B432 = m1:(S11.S12.S13.S15), m2:(S11.S12.S13), m3:(S3.S9).

VII. Results

The validation of our few-shot HAR was performed in different environments, locations, persons, and bundles. Fast/long tests employed 100/300 epochs, Table III. All results are averaged over 3 runs.

Diversity:
Cross-Environment (GDH, ITIV)
Cross-Location (1,2,3,4, IITV_AB/IT, ITIV_2AB),
Cross-Person (a,e,i),

Sensor Configuration:
CSI S261: 2xTx-6xRx, TDMA: 100ms, Cycle: 750ms.

Bundles:
B6, B43, B121, B222, B411, B321, B332, B432
E.g. B321 = (S12.S13.S15)(S9.S11)(S3)

TABLE III
Few-shot cross accuracy. a)Fast and b)Long Train.

a) Fast train 100 epochs, 4-Way, 5-Support, 1-Shot Accuracy

	GDH					ITIV		
Bundle	L1Pa	L2Pa	L3Pa	L4Pa	L4Pe	L1Pi	L1Pa	L2Pa
B121	0,567	0,483	0,572	0,567	0,538	0,213	0,406	0,579
B321	0,607	0,493	**0,648**	0,667	0,508	0,310	0,358	**0,641**
B222	0,535	0,591	0,578	0,661	0,610	0,229	0,345	0,573
B411	**0,631**	0,572	0,567	**0,668**	0,620	0,249	0,293	0,607
B332	0,573	0,576	0,609	0,628	0,658	**0,410**	0,400	0,572
B333	0,538	**0,599**	0,631	0,622	0,637	0,390	**0,423**	0,586
B432	0,513	0,584	0,609	0,658	**0,693**	0,382	0,408	0,607

b) Long train 1000 epochs, 4-Way, 5-Support, 1-Shot Accuracy

	GDH			ITIV	
Bundle	L1Pa	L4Pa	L4Pe	L1Pi	L2Pa
B6	0,619	0,659	0,706	0,405	0,605
B43	0,553	0,674	0,740	**0,428**	0,610
B121	0,610	0,624	0,669	0,375	0,662
B321	0,626	**0,748**	0,669	0,396	**0,722**
B332	0,626	0,710	0,730	0,404	0,660
B432	**0,630**	0,656	**0,743**	0,394	0,619
best	0,630	**0,748**	0,743	0,428	0,722

Three complete few-shot majority vote runs have been averaged for Locations GDH (1a,4a,4e) and ITIV (1i,2a) with Bundles B6, B43, B121, B321, B332. (Train and Validation took 151m32.7s on a MacBook Pro M1, 16GB RAM). From Table III some interesting conclusions can be reached:

- **Epochs** - Fast training, with only 100 epochs, already brings useful accuracy. Taking into account that the prototypes still need to be adapted to new environments show that it is not necessary stress the training in building automation tasks.
- **Bundle** - B321 at GDH, Location L4, Person a, brings the highest overall accuracy. (Training was done at L2). A Bundle with superposition, B432, however, brings the best results for locations L1 and L4. From Table III, colored values, we see that each location gives better results with a different bundle. B411, B333, B321 and B432 are even the best for two locations. Environment, Location and Person affected bundle accuracy. Guidance is to be established for best sensor positioning and bundle formation.
- **Training Environment** - Training with Room 7 of GDH, L2, Pa, we got almost similar cross-validation errors for GDH and ITIV. But, it is expected for a real building automation that an "average room" should be chosen for training. A training room in the building to be automatized is, probably, the best strategy.
- **Location** - ITIV is a "very different" test environment. Different furniture in different positions. Sensors in a different arrangement. Person Pa at L2 gives very good results with B321.
- **Person** Person Pi at ITIV L1 (female, lower height than male Pa) brings low accuracy for any bundle. Expanding the group of test persons should fix this.

As B321 performed best at the cross environment ITIV, it is better suited for the automation of that building . But it is clear, that enhanced CSI acquisition and enlargement/augmentation of the data basis is also necessary.

VIII. Conclusions

As expected few-shot transfer learning is a powerful tool for cross-domain applications, as smart building automation. Almost all recent papers dealing with CSI HAR use connection-based WiFi Access Point and Stations. We decided to use ESP-Now, a Espressif connectionless protocol has lower latency than 802.11n. However, acquiring CSI signals with ESP-Now causes many packet re-transmissions, that forces cleaning the CSI-file from duplicate and missing CSI lines. We replaced missing CSI lines from a given sensor with last valid CSI packet from that sensor. This produce a valid CSI image to the CNN, but is a "time-glitch" in particular for the Walking activity. A better procedure, to be tested, is to use a Kalman Filter to interpolate the missing CSI line. Another interesting way to handle it would be with a LSTM, with short memory that can also consider the past states. The implemented proof-of-concept with 10s CSI images classification would be a "CSI-broadcast classification" at, e.g., 200ms, reducing the "sensor-delay" in the feedback loop of Fig. 1, in the sense of a real-time automation. ESP32 is built, in first place, to guarantee communication, not real-time sensing. So, changes in the MAC layer may be needed to improve regular CSI sampling (avoiding redundant retries).

A promising field for occupancy-based automation is to merge CSI sensing networks with in-door already existing Wi-Fi signals from computers, access points, mobile phones, etc. Flexible real-time adaptive learning algorithms need to be developed for these intermittent signals.

Acknowledgment

We are grateful to N. Bahadori, [3], for sharing ReWiS on github.

References

[1] J. Yang, Y. Liu, Z. Liu, Y. Wu, T. Li, and Y. Yang, "A Framework for Human Activity Recognition Based on WiFi CSI Signal Enhancement," *IJAP*, 2021.

[2] J. Strohmayer and M. Kampel, "Data augmentation techniques for cross-domain wifi csi-based human activity recognition," *arXiv*, Jan. 2024.

[3] N. Bahadori, J. Ashdown, and F. Restuccia, "Rewis: Reliable wi-fi sensing through few-shot multi-antenna multi-receiver csi learning," *arXiv*, Apr. 2022.

[4] J. Natarajan, V. Krishnasamy, and M. Singh, "Design of a low-cost and device-free human activity recognition model for smart led lighting control," *IEEE IoT*, vol. 11.4, 2024.

[5] M. P. M. Silva, A. S. Rodrigues, and A. Bauchspiess, "Controle antecipativo por estimativa de carga térmica em vídeo (in portuguese)," in *SBAI*, 2019.

[6] W. Wang, F. Nikseresht, V. G. Rajan, J. Gao, and B. Campbell, "Enabling ubiquitous occupancy detection in smart buildings: A wifi ftm-based approach," in *IEEE 19th DCOSS-IoT*, 2023.

[7] M. Aftab, C. Chen, C. K. Chau, and T. Rahwan, "Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system," *Energy and Buildings*, vol. 154, 2017.

[8] H. Choi, M. Fujimoto, T. Matsui, S. Misaki, and K. Yasumoto, "Wi-cal: Wifi sensing and machine learning based device-free crowd counting and localization," *IEEE Access*, vol. 10, 2022.

[9] X. Li, X. Yang, Z. Ma, and J. H. Xue, "Deep metric learning for few-shot image classification: A review of recent developments," *Pattern Recognition*, vol. 138, 6 2023.

[10] K. Nweye and Z. Nagy, "Martini: Smart meter driven estimation of hvac schedules and energy savings based on wifi sensing and clustering," *Applied Energy*, 10 2021.

[11] W. Jiang, C. Miao, F. Ma, S. Yao, Y. Wang, Y. Yuan, H. Xue, C. Song, X. Ma, D. Koutsonikolas, W. Xu, and L. Su, "Towards environment independent device free human activity recognition." ACM, 10 2018.

[12] W. Zhuang, Y. Shen, C. Gao, L. Li, H. Sang, and F. Qian, "Adaptive scheme for crowd counting using off-the-shelf wireless routers," *CSSE*, vol. 41, 2022.

[13] Y. Jiang, X. Zheng, and C. Feng, "Towards multi-area contactless museum visitor counting with commodity wifi," *Journal on Computing and Cultural Heritage*, 7 2023.

[14] J. Cujilema, G. Hidalgo, D. Hernández-Rojas, and J. Cartuche, "Secure home automation system based on esp-now mesh network, mqtt and home assistant platform," *IEEE Latin America Tr.*, vol. 21, July 2023.