Cross-Domain Inference for Human Localization: Applying Wi-Fi RSSI Data to CSI-Trained Models

Ariel Duschanek-Myers, Thomas Welsh, and Helmut Neukirchen
Department of Computer Science, University of Iceland, Reykjavík, Iceland
{aad10, tomwelsh, helmut}@hi.is

***Abstract*—Wi-Fi signal data can be used to compromise the privacy of individuals. While many existing approaches rely on Channel State Information (CSI), collecting this data on typical IoT devices often requires elevated operating system permissions and specialized drivers. Consequently, this paper investigates the feasibility of utilizing Received Signal Strength Indicator (RSSI) data to predict human locations. RSSI was selected because it is accessible even on devices with limited user permissions, and therefore is more applicable to a wider array of IoT devices. To bypass the tedious process of obtaining training data needed to train an RSSI-based model, an existing Wi-Fi pose prediction project was used in this research. However, that project assumed CSI data as input. Therefore, we investigate the feasibility of cross-domain inference, i.e., feeding RSSI data into that existing CSI-based model. We collected an RSSI dataset, synchronized with video ground-truth of a person moving within a room, to evaluate the model's performance. This evaluation confirmed that RSSI data can predict locations with approximately 80% confidence when human movement is present. This demonstrates that a model trained on CSI data can be used to evaluate low-granularity RSSI data consisting of decibel-milliwatt (dBm) values to roughly locate people in the collection space. These results imply that a wide range of IoT devices can be used for privacy invasion in Wi-Fi-dense environments.**

***Index Terms*—IoT, cross-domain inference, location privacy, Received Signal Strength Indicator (RSSI)**

I. Introduction

The internet has changed the definition of what it means to be connected in the modern world, and everyday smart devices have only further reshaped what 'connected' means. The wireless communications that the Internet of Things (IoT) ecosystem relies on are often forgotten when evaluating what information is exposed by an IoT device. Recent research has identified a wide range of personal data incidentally exposed by these devices, such as location or vital signs [1]. The popular adoption of IoT devices substantially increased the number of interactions between radio frequency (RF) signals and physical objects, including people. Several works have shown that fluctuations in RF signal quality can be used to identify objects and their movements, with more advanced approaches able to identify human pose, accurate location, and even heartbeat [2]–[6].

These works indicate that collecting and analyzing RF has strong implications for privacy invasion, particularly within the context of an increasing number of deployed IoT devices. However, the practical implications of these approaches are unclear due to testing and development under laboratory conditions. For example, these approaches may require access to data (e.g. Channel State Information (CSI) [7]) and model training and inference capabilities unobtainable by resource-constrained and commodity IoT devices.

The contribution of this paper is two-fold: first, investigating the feasibility of using basic Wi-Fi signal data, in the form of Received Signal Strength Indicator (RSSI), for human localization. Such data can be more easily collected via a smart device's Software Development Kit (SDK) than more complex data such as CSI which requires lower-level hardware access permissions, typically inaccessible to developers. An example would be Android-based smart TVs that are ubiquitous and could therefore be used to surveil people. The second contribution of this paper is that to understand if RSSI data can be used for this purpose, we made use of a CSI-trained model for human pose detection [3] with RSSI data collected in a similar laboratory setup. We evaluated this *cross-domain inference* [8] approach and determine that RSSI data can be used to detect human presence and movement in this context.

The rest of this paper is structured as follows. Section II covers background and related work, while Section III details our cross-domain inference methodology and experimental setup. Section IV and V present results and discuss their implications, respectively. Finally, Section VI concludes this paper.

II. Background and Related Work

In wireless sensing, the Received Signal Strength Indicator (RSSI) provides a coarse, easily accessible scalar value representing overall signal power at the Medium Access Control (MAC) layer. While RSSI reflects the aggregate power of the entire channel, CSI provides a granular breakdown of that signal. Channel State Information (CSI) is a physical-layer metric that captures amplitude and phase shifts across multiple subcarriers [1]. While CSI's high resolution makes it ideal for advanced perception tasks, it requires specialized network hardware and elevated operating system permissions. In contrast, RSSI can be easily obtained by an application program.

The work by Wang et al. [2] was the first that inspired the continued investigation into the use of Wi-Fi for locating people in a space. It focuses on using CSI data collected between three pairs of transmitting and receiving antennas in conjunction with the combined ground truths of a Mask Region-based Convolutional Neural Network (R-CNN) and OpenPose [9] models to validate the predictions made by their

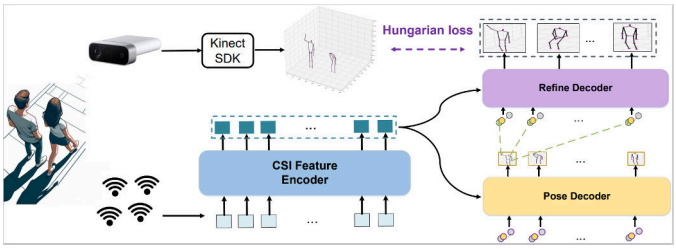


Fig. 1: Training environment for the *Person-in-WiFi 3D* model end-to-end *Multi-Person 3D* Pose Estimation with Wi-Fi. (© 2024 IEEE. Reprinted, with permission, from Fig. 4 of K. Yan et al., "Person-in-WiFi 3D: End-to-End Multi-Person 3D Pose Estimation with Wi-Fi", 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) [3].)

novel *Person-in-WiFi* model. Their research acknowledges the existing use of Wi-Fi-based perception for persons in a space, but focuses on making this perception more fine-grained with the application of CSI data and Computer Vision (CV) models.

Yan et al. [3] continued work on *Person-in-WiFi* with the goal of performing perception in three dimensions. By increasing the number of receivers, the three-dimensional nature of Wi-Fi's broadcast capabilities can be taken advantage of, permitting better estimation where people or limbs overlap. They implemented a new method of generating ground truths by utilizing Microsoft Azure Kinect-generated [10] meshes and the Hungarian Loss function to provide three-dimensional representations for the validation of *Person-in-WiFi 3D* predictions. This increases the flexibility of the ground-truth model when it comes to validating the predictions of the *Person-in-WiFi 3D* model by allowing the re-orientation of the mesh in three dimensions as the validation is performed. It continues using CV models for both parts of the *teacher-student model* for knowledge transfer while decreasing brittleness.

As depicted in Fig. 1 as the *Pose Decoder*, the student model is trained with CSI signal data collected simultaneously as the Kinect captures and labels pose estimations. The pose estimations are used as ground truths during the validation step of the student model's training workflow. The Kinect provides highly reliable pose estimations by using depth readings for objects in the space. These estimates allow the student model to tune out background noise and reflected signals.

The CSI feature encoder takes the collected data and uses the timestamp to group twenty samples of the 30 channels from each of the 3 antennas for each Access Point (AP) together for a frame. A sliding window is used for the frame creation to increase the fluidity of the predicted keypoint movement. The matrix is then passed to a tokenizer that converts the 3 links, 3 antennas, and 20 timepoints into 180 tokens, with each token including a vector of the 30 channel values. This is up-scaled further by 256 by incorporating spatial-temporal embeddings; thus, the model receives a 180 $\times$ 256-dimensional object for evaluation.

Li et al. [4] provided supporting evidence for the use of RSSI data for pose prediction. They focused primarily on the labeling of human poses to provide a minimally invasive method for monitoring elderly occupants of an assisted living space, without the use of video. They performed a comparison between CSI and RSSI data prediction accuracy for the same number of features (Fig. 8 in [4]). Though the CSI data resulted in roughly a 20 percent increase in accuracy, the RSSI data predictions increased in accuracy at the same rate as CSI data did when the number of features was increased. This indicates that predictions made based on a high number of features should still provide a reasonably accurate estimate of a generalized position.

Zhu et al. [5] utilize a Wi-Fi receiver, a router, and a Received Signal Strength-fit model to detect human movement through a wall. This is done by measuring the Doppler effect that moving objects cause in the broadcast Wi-Fi radio frequency's Received Signal Strength. First, the transmitter is placed in a stationary position in a room and the receiver is moved along a defined path to confirm that the signal strength can be used to locate the transmitter based on changes in the measured signal strength along the path. After the model is fit to the RSSI data of the stationary transmitter, the change in signal strength between a stationary transmitter and receiver is correlated with object movement – in this case, people. The receiver relies on a highly directional antenna and Software Defined Radio (SDR) to collect RF data across the Wi-Fi spectrum more broadly than a standard adapter would permit.

The discussed related work focuses on CSI data collected with a specialized implementation in a desktop Operating System (OS), relying on the host's network interface to handle requests for information from the system's Wi-Fi adapter. This is normally not available in consumer IoT devices without using specialized tools to gain root access. Though Li et al. [4] perform a comparison with RSSI data, the effect of increased features is documented but not investigated. If an existing solution with a higher number of features is adapted for use with RSSI data, it is possible the prediction confidence levels will be competitive with that of models based on CSI data.

Besides being CSI-based, the related work also focuses on poses as the output of the model's evaluation process. Poses are dependent on the model recognizing smaller volumes of mass than an entire person, something that is useful but resource-intensive. By shifting the focus to location instead of pose, one can rely on the average location of the entire person's mass. This permits the use of sparse data, such as RSSI signal strength for evaluation and predictions. Moreover, if the predictions are well correlated and have a small amount of standard deviation, a lower score than is permitted for detecting a pose may still be valid. Therefore the rest of this paper seeks to address these gaps in literature by assessing if RSSI data can be used for cross-domain inference with a CSI trained model to permit human localization.

III. Cross-domain Inference CSI Model with RSSI Input

To confirm that the RSSI data exposed in IoT devices is effective with an existing pose prediction model, this section presents a method for cross-domain inference of a CSI-based model using RSSI input. The methodology for evaluating this approach is as follows. First, we configure a laboratory aligned with the methodology in [2] and use it to collect RSSI signal data affected by human poses and movement. Next, we pre-process the collected data through normalization and noise reduction. Finally, we input this data into the existing CSI-based model so that the output can be evaluated. This evaluation considers the model's confidence scores against the ground truth of the human's movements recorded via video.

A. Laboratory Setup

The *Person-in-WiFi 3D* model [3] that we re-use was trained on a specific lab setup; a roughly 3.5 m by 4 m space with three Wi-Fi routers receiving transmission from one Wi-Fi-enabled laptop. Therefore, we replicated this setup in a campus study space with chairs and tables removed. We placed ESP32-C3 boards acting as APs and a Raspberry Pi 4 in the space, along with the necessary power strips (Fig. 2).

The Raspberry Pi 4 was used in place of the Wi-Fi-enabled laptop. After creating an Android application to investigate whether the required RSSI data collection was feasible, the implementation for the receiving device was moved to Linux for simplicity (avoiding lengthy Android app installation procedures) and to ensure empirical consistency. As this work is focusing upon privacy violations in smart environments, the Raspberry Pi was chosen for its similarity to a smart device (such as a smart TV that would, e.g., run Android).

The ESP32-C3 boards serve as simple APs transmitting on a single channel, which is sufficient since signal strength is largely independent of direction in this setup. While this approach may reduce the ability to classify specific joints (i.e. pose) of a human body due to the loss of multi-channel data, the positional predictions in the 3D space should be maintained. As depicted in Fig. 2, the access points and the receiving device are within broadcast range of each other; we maintained the scale expected by the model. Three access points are necessary for 3D positional predictions. A single access point provides only one-dimensional information about material obstruction between it and the receiver. With three access points, we can make three-dimensional predictions based on signal strength variations caused by changes in materials between each transmitter and receiver.

To collect the data, a shell script scans the RSSI value on the WiFi interface of the device[1] and stores them as comma-separated values to be used for model prediction. This script is then implemented as a systemd service which has been configured to run after the network stack is online. The process is then initiated after the device is left stationary in the room.

B. Data Collection

RSSI data collection begins in an empty room, with roughly 2 minutes recorded to establish a baseline for correcting environmental noise in the lab. Although the existing CSI training data[2] from [3] does not control for interference, the teaching model would reinforce inattentiveness to the background noise during the validation step of training. Ground truths created by the Kinect-labeled meshes are based on infrared and visible light data and therefore cannot include radio interference from the training data collection space, see Fig. 1.

A shell script synchronized systems' timestamps via SSH, started the Wi-Fi data collection service on the Raspberry Pi, and began video recording on the laptop. Collection continued until the Enter key was pressed, stopping the Pi service and ending the video recording. This semi-automated workflow allowed us to start collection and leave the room, capturing background noise that was subsequently subtracted to control for changes in the radio frequency environment.

The baseline RSSI data was captured first (empty room with no objects or humans, see the left part of Fig. 3, i.e. before the vertical Empty Room End line), followed by a second capture period with a person in the space (right part of Fig. 3, i.e. after the vertical Occupied Room Start line).

[1]The device operating system was Raspbian 5.6 which is a distribution of Debian Linux with drivers and configuration specific to the Raspberry Pi. The Linux kernel version for this Raspbian build was 6.12.25 and was selected to avoid Wi-Fi driver latency present in later, recent versions.

[2]https://aiotgroup.github.io/Person-in-WiFi-3D/

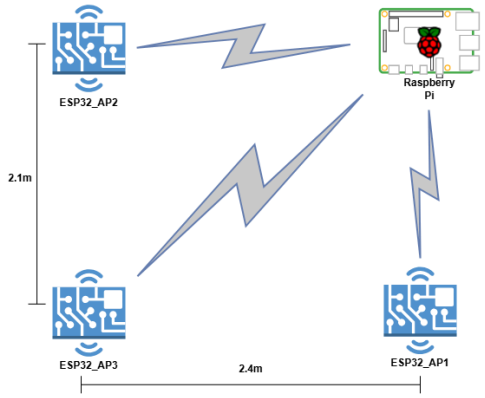


Fig. 2: Lab layout

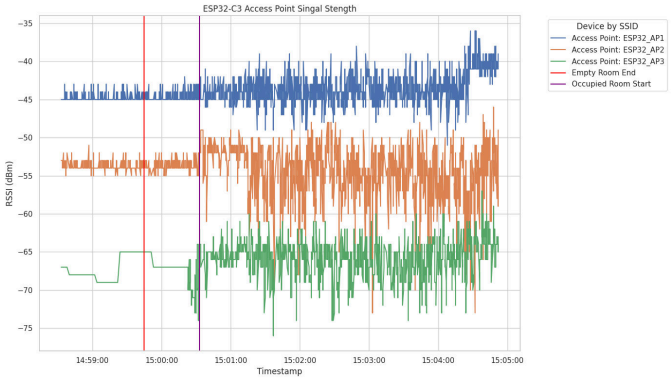


Fig. 3: Pre-processing sample RSSI signal strength data

Listing 1: Normalization Python code

```
empty_linear = 10 ** (np.array(empty_list) / 20.0)
baseline_amp = np.mean(empty_linear)
if baseline_amp == 0:
    baseline_amp = 1e-9
current_linear = 10 ** (current_rssi / 20.0)
normalized_amp = current_linear - baseline_amp
final_amp = normalized_amp * target_baseline
final_amp = np.power(final_amp, dampen)
return final_amp
```

C. RSSI Data Preprocessing

We first focused on signal behavior observed in the empty room. We originally performed testing with two types of USB-C cables and two types of USB wall power adapters. The initial data collection suggested that differences between ESP32_AP3 and the other two ESP32 APs were caused by the combination of the wall adapter and shielded USB cable. To address this, we collected new data using three identical USB-C cables and wall adapter pairs, such that the noise would become similar for all APs. This required adding surge-protected power strips to power each adapter which standardized the noise from the power sources. The Wi-Fi Protected Access (WPA) Supplicant was disabled to prevent conflicting access to the Wi-Fi adapter by multiple services. A Windows laptop was connected to a USB webcam and the Raspberry Pi via Ethernet cable for remote SSH control. The laptop's Wi-Fi adapter was disabled during collection to prevent possible interference from probe requests.

We convert the RSSI decibel-milliwatt (dBm) signal to float linear values to mimic CSI data before the value was scaled and dampened. This allows us to control for differences in room size of our lab and the lab that was in [3] used for training the model that we re-use, see Listing 1. This linear value is computed as $10^{\frac{\text{RSSI}}{2\times 10}}$. To ensure the pre-processing of RSSI data collected in our version of the *Person-in-Wifi 3D* lab setup was not artificially improving predictions, we first evaluated the impact of different combinations of normalization (see Table I) on the resulting confidence scores that the model provides as part of its output. A second evaluation was performed using only linearized data to verify that normalization has not artificially reduced confidence, since RSSI data likely includes some normalization when provided by the Wi-Fi OS driver.

Predictions were produced by the evaluation pipeline and a comparison was done between the maximum confidence scores, see Table I. Based on the difference in the transmitter and room scale, dampening and scaling were applied to the RSSI signal data after linear conversion was performed. Different methods of removing the baseline were tested to determine the best method of removing our specific RF background.

To confirm the need for a filter threshold before data is passed to the model for evaluation, linearized empty room data with and without normalization were evaluated as well. This confirmed noisy spaces with only the median value of the empty room subtracted still provide enough information for the model's attention mechanism to make overly confident predictions, with a maximum confidence of 0.6529. All noise suppression methods besides Per-AP Baseline removal resulted in the empty room data having higher confidence predictions than the normalized, live test data. The visual representation of the data, Fig. 3, makes it apparent that the length and amplitude of changes in signal strength is a useful method for filtering evaluation candidates.

TABLE I: Absolute Maximum Confidence Scores Across Different Data Normalizations. See the footnotes for the normalization applied.

Data Type	Each AP Baseline[1]	Each AP Baseline Z-Score[2]	Tilt Gradient[3]	Global Mean Subtraction[4]
Noisy Data[5]	0.5774	0.5405	0.7025	0.7922
Dampen Data[6]	0.5492	0.5112	0.7370	0.7904
Scale Data[7]	0.6355	0.3633	0.6252	0.7237
Norm. Data[8]	0.4333	0.3501	0.4997	0.6089
Empty Data[9]	0.3996	0.3978	0.5716	0.6529
Empty Norm. Data[10]	0.3356	0.3319	0.2750	0.3264

[1] `current_linear_for_ap / baseline_amp_for_ap`
[2] `((current_linear_for_ap-g_mean)/g_std)+1.0`
[3] `current_linear_for_ap*(0.9,1.1,num_subcars)`
[4] `current_linear_for_ap - baseline_amp`
[5] `target_baseline=1.0, dampen=1.0`
[6] `target_baseline=1.0, dampen=0.8`
[7] `target_baseline=0.6, dampen=1.0`
[8] `target_baseline=0.6, dampen=0.8`
[9] `target_baseline=1.0, dampen=1.0`
[10] `target_baseline=0.6, dampen=0.8`

The maximum confidence score increased from 0.6089 to 0.7922 when the data was left 'noisy' without scale and dampening. If the model were retrained with visual and signal data in the lab space, the reflection and background noise captured in the RSSI signal data would be trained out of relevance and the median and mean confidence of the data with noise would be increased.

A baseline to account for possible changes in the RF environment of the lab before the evaluation process was repeated with the *Person-in-WiFi 3D* model, such that the linearized data had the average of the empty room subtracted. We also tested subtracting each AP's average linearized value before applying a noise gate, sigmoid, and angular embedding.

D. Model Integration

The collected, preprocessed RSSI data is fed to the trained model, which produces 100 predictions per frame, each with a confidence score. These scores reflect either successful positive predictions or confident false predictions.

The *Person-in-WiFi 3D* model [3] depends on specific versions of CUDA and PyTorch, which in turn depend on older package versions required by MMDetection, MMCV, and Object Perception & Application (OPERA). Upgrading these libraries would necessitate changes to imports and implementation throughout the codebase. To avoid extensive refactoring, we leveraged compatibility options for older versions of Nvidia's CUDA toolkit in combination with the used NVIDIA A100 Tensor Core GPUs.

The output of the evaluation stage of the model is controlled by a configuration parameter, `max_per_img`. We set this to 100 to ensure all 100 query results from the PETRHead stage of the *Person-in-WiFi 3D* model are recorded as output. We also pass the eval_options 'out' parameter a `.pkl` file name for storing the results. The result consists of 100 predictions generated for each frame (due to the 100 queries made by the model's evaluation of a frame). A prediction consists of 14 keypoints that represent joints in a humanoid pose and of a score indicating the model's confidence on the accuracy of the prediction of that specific pose.

Since the model outputs 100 different predictions for each analyzed frame, we filter for the highest confidence score to identify the most likely pose for a frame. To validate that this prediction is not an outlier, we plot all 1400 keypoints per frame from the raw output. This visualization reveals whether the prediction confidence score correlates with the model's actual perception of reality as triggered by the evaluated RSSI data.

We generate statistics and visualizations for human validation: a graph of prediction score deviation indicates the range of prediction scores. Comparing the maximum score per 100 predictions per frame between the normalized and noisy data will test the theory that RSSI data can produce meaningful predictions regardless of the controlled nature of the collection space. We expected keypoint (=joint) predictions to have low confidence, but the keypoints should be clustered together.

To determine if the model's confidence scores are accurately predicting real world events, the collected data can be correlated with the video based on the time point recorded with each scan result. Though the scans are meant to be performed twenty times per second, this is an optimistic frequency. As a result of this sampling schedule, frames of pre-processed data are roughly representative of one second. We can therefore plot the prediction frames and compare the video at the likely time code.

IV. Results and Evaluation

In this section we present the results and evaluation of our cross-domain approach presented in Section III. To correlate the RSSI data with the captured video (i.e. our ground truth), Table II lists the human movement events, where some example video frames are given in Fig. 4.

A. Event Ordering

Using the timestamps from Table II, the RSSI data was annotated, as shown in Fig. 5. The colors and labels assigned to each event in Table II and Fig. 5 are used consistently throughout the remaining figures.

B. Prediction Correlation Analysis

As the *Person-in-WiFi 3D* model makes 100 predictions per frame for each skeletal keypoint we perform statistical analysis on each keypoint. The standard deviation of each keypoint, representing a joint in the pose model, remains in the same standard deviation range when the room is empty, as depicted in Fig. 6. The lower the standard deviation value, the more closely the keypoints from all the 100 predictions are in each frame. A lower value indicates a higher level of consensus between the location of the predictions. By combining this with the confidence scores, we can determine the predictions

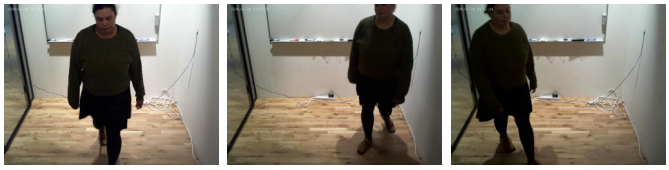
TABLE II: Event Timestamps

Event	Time Stamp
1: Room Empty	15:14:16.696-15:16:29
2: Door Open	15:16:29-15:16:32
3: Raising Alternating Arms	15:16:35-15:16:47
4: Walking Front to Back	15:16:47-15:17:09
5: Walking ESP32_AP1 to ESP32_AP2	15:17:09-15:17:32
6: Walking Pi to ESP32_AP3	15:17:33-15:17:57
7: Arms Out Turning	15:17:58-15:18:05
8: Arms Bent Turning	15:18:05-15:18:13
9: Raising/Lowering Both Arms	15:18:15-15:18:23

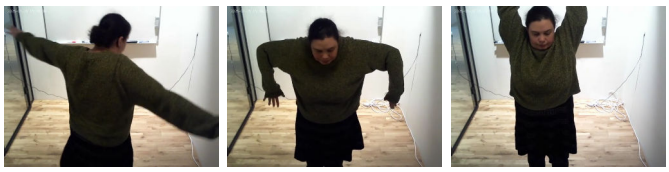
TABLE III: Skeletal Keypoint Connections and Body Regions

Region	**Connection** (A, B)	**Description**
Head	(12, 13)	Head to Neck
Torso	(10, 6)	Upper Torso to Left Hip
	(6, 8)	Left Hip to Right Hip
	(8, 11)	Right Hip to Upper Torso
Arms	(13, 10)	Neck to Left Shoulder
	(10, 7)	Left Shoulder to Left Elbow
	(7, 3)	Left Elbow to Left Wrist
	(13, 11)	Neck to Right Shoulder
	(11, 9)	Right Shoulder to Right Elbow
	(9, 5)	Right Elbow to Right Wrist
Legs	(6, 2)	Left Hip to Left Knee
	(2, 0)	Left Knee to Left Ankle
	(8, 4)	Right Hip to Right Knee
	(4, 1)	Right Knee to Right Ankle

Raising alternating arms



Walking front to back and both diagonals



Rotating with arms straight out | Rotating with bent arms | Raising both arms

Fig. 4: Example images from video of each event in Table II

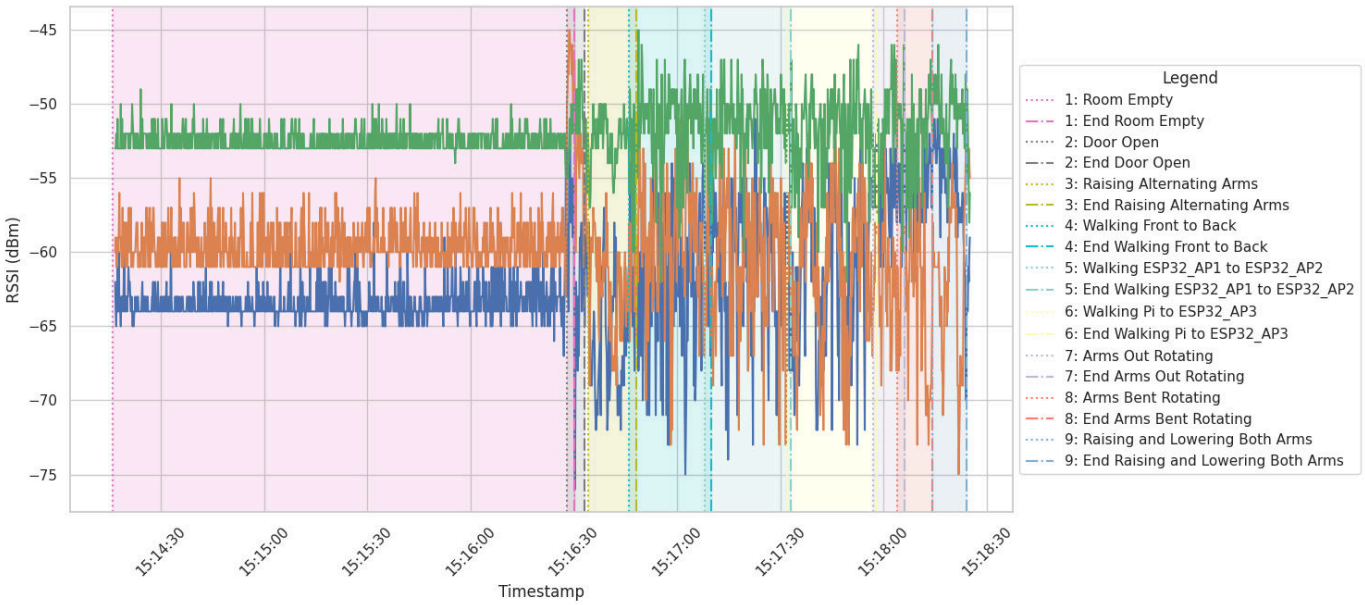


Fig. 5: RSSI signal strength with the time range of actions labeled.

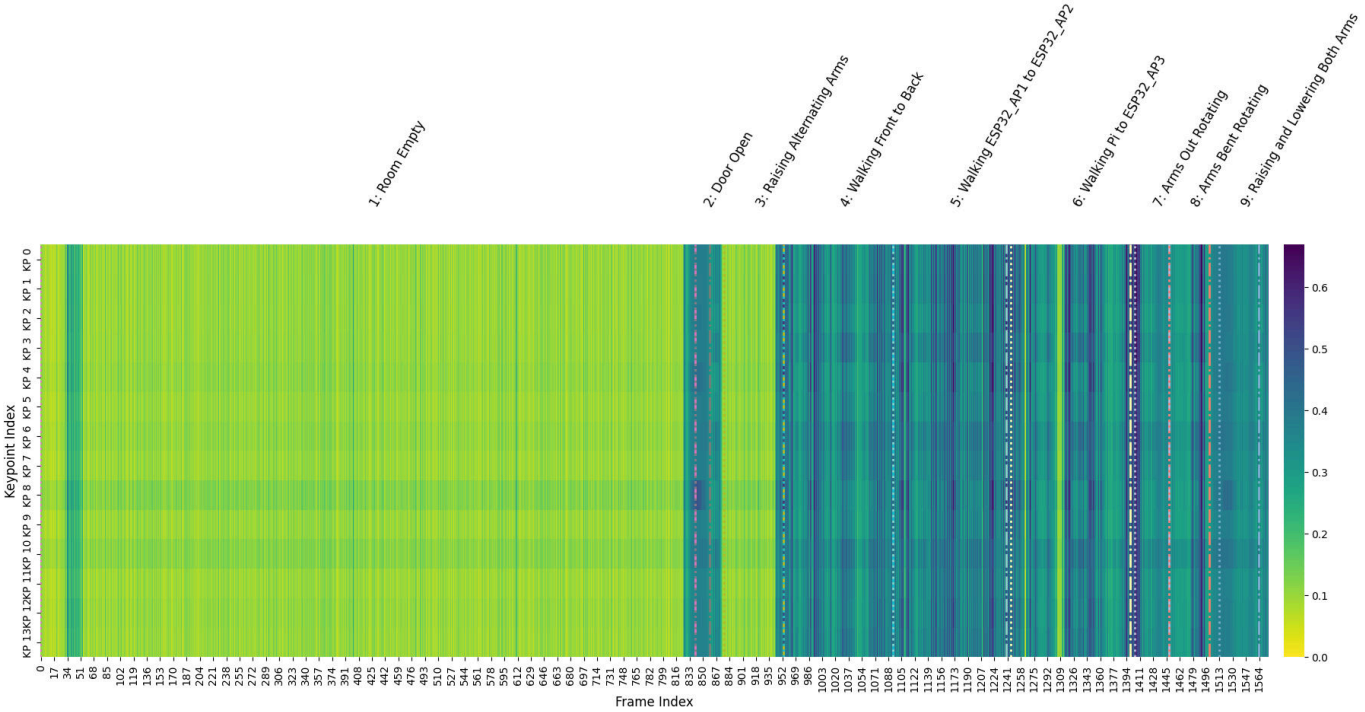


Fig. 6: Standard deviation between each frame's 100 predictions.

were precise based on the model's understanding of CSI data, even if the location is inaccurate compared to reality.

Figure 6 in combination with Table III shows that the keypoints with the highest standard deviation throughout the predictions are 6, 8, and 10 (i.e. the visible horizontal bars in Fig. 6); all points are part of the torso region, i.e. where the human body has most of its mass that influences Wi-Fi signals, e.g. via Doppler effect. There is also an observable vertical bar of low standard deviation frames when the door was opened to allow the person into the lab space. We would expect the model to be confused by the change in distance between the ESP32_AP3 transmitter and the receiver raising the signal strength. It may also perceive the change as evidence an object has entered the space and changed the reflectance of broadcast packets.

To try and improve the delineation between these two states, we applied a minimum threshold and sigmoid value to remove noise and smooth the data in preparation for scaling to mimic the CSI *phase denoising* performed in the *Person-in-WiFi 3D* project. The angular embedding was then applied by subtracting 0.5 to center the sigmoid value and the multiplying it by π^2, see Listing 2. The embedding is based on a unit circle with a point traveling around its perimeter, a distance of less than π from the circle's center indicates the signal strength was

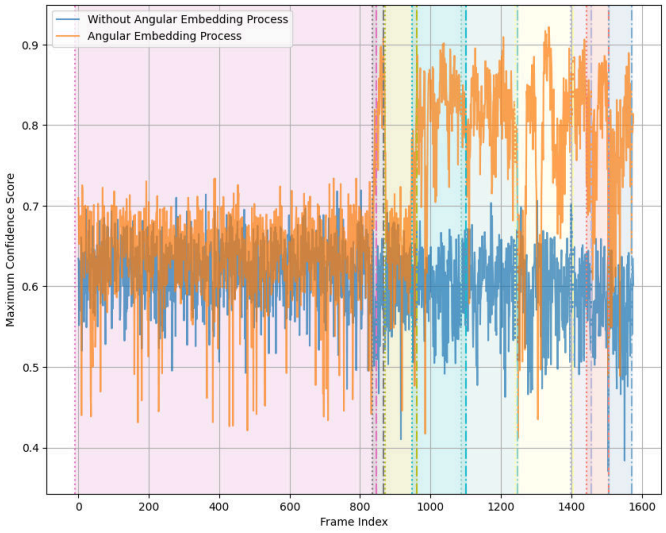


Fig. 7: Change in maximum confidence for each frame when a noise gate, smoothing, and angular embedding are performed.

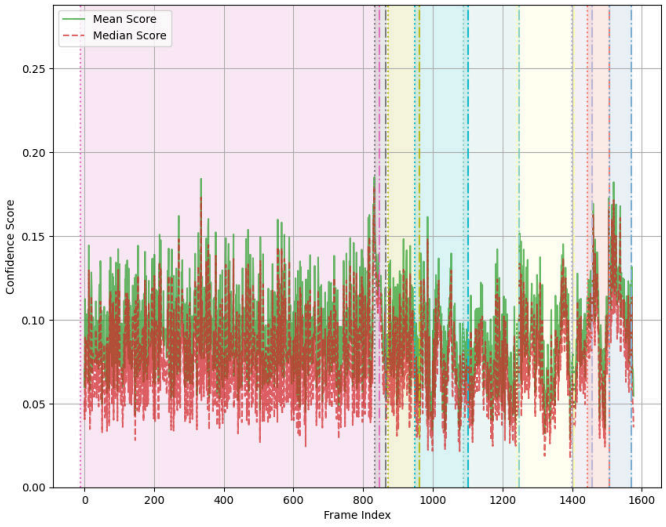


Fig. 8: Mean and median of the results dataset with angular embedding.

lower than expected. A noise threshold of 0.00095 and K value of 4000 were selected to preserve the model's responsiveness to fluctuations while quieting the signal. After this, values will fall within $\pm\pi$ as expected for phase data and represents the voltage observed at the antenna.

Fig. 7 plots the maximum confidence score for pre-processed data with and without angular embedding. The figure shows that the confidence of the predictions in the empty room are nearly the same as for the occupied room. Comparing the two, we can see that the maximum confidence values for the empty and occupied room are clearly separated into ranges with limited overlap. The mean and median of the predictions, Fig. 8 and 9, are also lower on average with the increased amplitude of the angular embedded data, indicating the prediction confidence scores are based on a feature of the

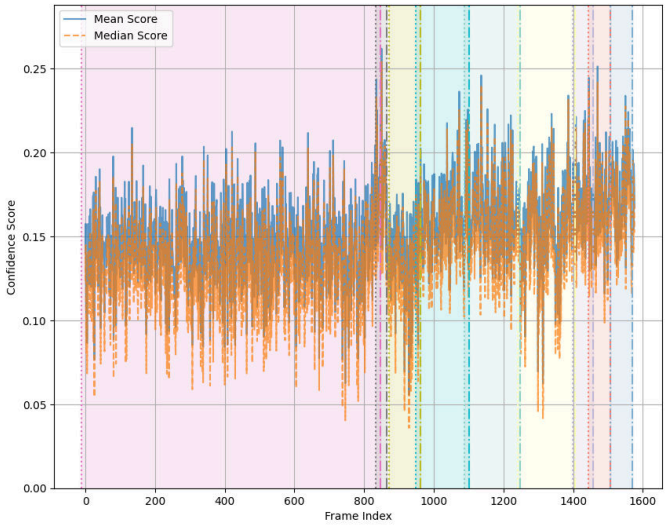


Fig. 9: Mean and median of the results dataset with subtraction of global mean only.

Listing 2: Angular embedding Python code

```
filtered = np.abs(diff) - noise_threshold
diff_centered = np.sign(diff) * np.maximum(0,filtered)
sig = 1 / (1 + np.exp(-K * diff_centered))
return (np.pi ** 2) * (sig - 0.5)
```

data, not of the amplitude of the input. This suggests that the noise of the RSSI data has incorrectly been perceived as an object moving in the collection space. Therefore, the angular embedding had the desired effect of increasing the occupied room's prediction confidence, as shown by the orange line in Fig. 7. However, the mean and median scores were lowered as a result as seen in Fig. 8. This is in contrast to the predictions based on the pre-processing method shown as the blue maximum confidence score line in Fig. 7 and the mean and median graph of Fig. 9. This is to be expected as the original training method was a bipartite matching loss detection transformer. The normal training pipeline creates the 100 raw predictions exactly as our implementation does, but the prediction that matches the ground truth is given a higher score by the teaching model. The student pose decoder model produces a higher score for predictions it perceives as matching the pose being made in the collection space as a result of this reinforcement.

Overall, the maximum prediction confidence score reached 0.71894246, as depicted by the blue maximum confidence score line in Fig. 7. This is higher than random guessing and demonstrates the model's evaluation process found the RSSI signal data to be generally as meaningful as CSI. With the use of angular embedding, the maximum prediction confidence score, as shown by the orange maximum confidence score line in Fig. 7, reached 0.92170554 with roughly 0.8, or 80%, average confidence when the room is occupied.

Therefore, the correlation of labeled events with the RSSI and predictions (Table II and Fig. 5), indicate the possible use

of RSSI data for predicting the location of novel objects and people. This is further supported by the high confidence scores (Fig. 7), and the decrease in standard deviation (Fig. 6) when the room is occupied. Together this supports predication via a model from an existing Wi-Fi positioning or pose prediction project without additional training.

V. Discussion

In combination, the results have shown that the statistical behavior of the raw predictions supports the possible use of RSSI data for predicting human presence and movement.

Implications for Privacy. The results of this research suggest individual privacy is at risk of possible location tracking in IoT-rich environments using consumer hardware for RSSI data collection. This is in contrast to CSI data collection methods that require lower-level permissions in the operating system to access. Related works focus on CSI data makes the use of consumer Wi-Fi adapters less feasible as many do not expose Physical (PHY)-layer information. While legacy drivers in Linux can sometimes expose such functionality, manufacturers do not provide access in consumer equivalent drivers. As an alternative to the single-core ESP32-C3, the use of ESP32 Xtensa devices would be an accessible alternative to specialized Wi-Fi adapters for CSI collection. These devices are representative of a wide range of IoT devices on the market and provide multi-core processors with higher processing speeds.

Threats to Validity. Internal threats to validity focus on the physical experimental setup. The space used in the *Person-in-WiFi 3D* lab environment was set up in an open floor-plan space; this is in strong contrast to the small concrete room used for this project's lab. The open area reduces general reflectance of the space and leads to less background noise. It is possible the echo effect of the Wi-Fi signal bouncing off of objects and walls repeatedly makes predictions less accurate and the lack of granular information in RSSI data compounds this. We accounted for this by subtracting away the noise from the empty room. External threats to validity include the lack of alignment with the *Person-in-WiFi 3D* model's focus on CSI data during training in combination with the unavailability of a Kinect Azure or equivalent device. To address the limitation of access to a Kinect Azure for ground truth generation, a secondhand version of the Kinect could be used and the depth or field for the infrared vision could be supplemented with an additional LED array and sensor(s). This would likely further increase the need for compute power during collection as the Kinect Azure development kit appears to synchronize its sensor data before forwarding the captured content.

VI. Conclusion

With the increased use of Wi-Fi and other wireless communications in everyday devices, such as IoT devices, the ambient surveillance vectors grow as well. Traditional methods of surveillance rely on cameras and microphones, features that are well understood and controlled by the average user. In contrast, wireless signals are ephemeral and invisible, making them far less apparent as potential tools for monitoring. This dichotomy of ubiquity and ambiance makes the design and auditing of smart device SDKs a public cybersecurity obligation.

By investigating the feasibility of using RSSI data as input for the CSI-focused *Person-in-WiFi 3D* model, this research has quantified the confidence and precision that ambient signal data can elicit from a model trained with CSI data. This demonstrates low-density signal data can still trigger high-certainty predictions with a low degree of deviation, proving the obscurity of the vector does not protect the user, but instead necessitates a focus on secure design that prioritizes privacy.

Acknowledgment

We would like to thank the authors of [3] for sharing their model and training data with us. This project has received co-funding from the Icelandic government as well as from the European Union's Digital Europe Programme and the European Cybersecurity Competence Centre under grant agreement no. 1011226821 Eyvör National Coordination Centre for Cybersecurity Iceland and 101127307 Defend Iceland: Nationwide bug bounty platform.

References

[1] F. Wang, T. Zhang, W. Xi, H. Ding, G. Wang, D. Zhang, Y. Cui, F. Liu, J. Han, J. Xu, and T. X. Han, "A survey on Wi-Fi sensing generalizability: Taxonomy, techniques, datasets, and future research prospects," *IEEE Communications Surveys & Tutorials*, vol. 28, 2026, https://doi.org/10.1109/COMST.2026.3670854.

[2] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-WiFi: Fine-grained person perception using WiFi," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, https://doi.org/10.1109/ICCV.2019.00555.

[3] K. Yan, F. Wang, B. Qian, H. Ding, J. Han, and X. Wei, "Person-in-WiFi 3D: End-to-end multi-person 3D pose estimation with Wi-Fi," in *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, https://doi.org/10.1109/CVPR52733.2024.00098.

[4] F. Li, M. A. A. Al-qaness, Y. Zhang, B. Zhao, and X. Luan, "A robust and device-free system for the recognition and classification of elderly activities," *Sensors*, 2016, https://doi.org/10.3390/s16122043.

[5] J. Zhu, P. Zhang, J. Wang, and Z. Chen, "Through-wall human sensing with WiFi passive radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 4, 2021, https://doi.org/10.1109/TAES.2021.3069767.

[6] X. Wang, C. Yang, and S. Mao, "On CSI-based vital sign monitoring using commodity WiFi," *ACM Trans. Comput. Healthcare*, vol. 1, no. 3, May 2020, https://doi.org/10.1145/3377165.

[7] Z. Wang, J. A. Zhang, H. Zhang, M. Xu, and J. Guo, "Passive human tracking with WiFi point clouds," *IEEE Internet of Things Journal*, vol. 12, no. 5, 2025, https://doi.org/10.1109/JIOT.2024.3487193.

[8] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," *Proceedings of the IEEE*, vol. 109, no. 1, pp. 43–76, 2021, https://doi.org/10.1109/JPROC.2020.3004555.

[9] Z. Cao, G. Hidalgo, T. Simon, S.-E. Wei, and Y. Sheikh, "OpenPose: Realtime multi-person 2D pose estimation using part affinity fields," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 1, 2019, https://doi.org/10.1109/TPAMI.2019.2929257.

[10] J. Bertram, T. Krüger, H. M. Röhling, A. Jelusic, S. Mansow-Model, R. Schniepp, M. Wuehr, and K. Otte, "Accuracy and repeatability of the Microsoft Azure Kinect for clinical measurement of motor function," *PLOS ONE*, vol. 18, 2023, https://doi.org/10.1371/journal.pone.0279697.