Multi-Band Wi-Fi Sensing with Matched Feature Granularity

Yu, Jianyuan; Pu, Wang; Koike-Akino, Toshiaki; Wang, Ye; Orlik, Philip V.; Buehrer, R. Michael

TR2022-105   September 17, 2022

Abstract

Complementary to the fine-grained channel state information (CSI) and coarse-grained received signal strength indicator (RSSI) measurements, the mid-grained spatial beam attributes (i.e., beam SNR) during the millimeter-wave (mmWave) beam training phase were recently repurposed for Wi-Fi sensing applications such as human activity recognition and indoor localization. This paper proposes a multi-band Wi-Fi sensing framework to fuse features from both CSI from 5 GHz bands and the mid-grained beam SNR at 60 GHz with feature granularity matching that pairs feature maps from the CSI and beam SNR at different granularity levels with learnable weights. To address the issue of limited labeled training data, we propose to pretrain an autoencoderbased multi-band Wi-Fi fusion network in an unsupervised fashion. For specific sensing tasks, separate sensing heads can be attached to the pretrained fusion network with fine-tuning. The proposed framework is thoroughly validated for three sensing applications using in-house experimental datasets: 1) pose recognition; 2) occupancy sensing; and 3) indoor localization. Comparison to a list of baseline methods demonstrates the effectiveness of granularity matching. An ablation study is performed as a function of the amount of labeled data, the latent space dimension, and learning rates

Multi-Band Wi-Fi Sensing
with Matched Feature Granularity

Jianyuan Yu*, Pu Wang*, Toshiaki Koike-Akino, Ye Wang, Philip V. Orlik, and R. Michael Buehrer

Abstract—Complementary to the fine-grained channel state information (CSI) and coarse-grained received signal strength indicator (RSSI) measurements, the mid-grained spatial beam attributes (i.e., beam SNR) during the millimeter-wave (mmWave) beam training phase were recently repurposed for Wi-Fi sensing applications such as human activity recognition and indoor localization. This paper proposes a multi-band Wi-Fi sensing framework to fuse features from both CSI from 5 GHz bands and the mid-grained beam SNR at 60 GHz with *feature granularity matching* that pairs feature maps from the CSI and beam SNR at different granularity levels with learnable weights. To address the issue of limited labeled training data, we propose to pre-train an autoencoder-based multi-band Wi-Fi fusion network in an unsupervised fashion. For specific sensing tasks, separate sensing heads can be attached to the pre-trained fusion network with fine-tuning. The proposed framework is thoroughly validated for three sensing applications using in-house experimental datasets: 1) pose recognition; 2) occupancy sensing; and 3) indoor localization. Comparison to a list of baseline methods demonstrates the effectiveness of granularity matching. An ablation study is performed as a function of the amount of labeled data, the latent space dimension, and learning rates.

Index Terms—Wi-Fi sensing, WLAN sensing, channel state information, beam attributes, fusion, autoencoder, millimeter wave, fingerprinting, deep learning.

I. Introduction

Wi-Fi sensing (or WLAN sensing) has received tremendous attention over the past decade. More recently (September 2020), IEEE 802.11 Standards Association established a new task group for WLAN sensing for making greater use of 802.11 technologies to empower new industrial and commercial applications in home security, entertainment, energy management (HVAC, light, device power savings), elderly care, and assisted living [1], [2].

Most Wi-Fi sensing frameworks use either fine-grained channel state information (CSI) [3]–[15] or coarse-grained RSSI measurements [16]–[32]; see the more detailed literature review in the next section. The conventional RSSI measurement suffers from measurement instability and coarse granularity of the channel information, leading to limited accuracy for localization. CSI measurements are more fine-grained but require access to physical-layer interfaces and high computational power to process the large amount of sub-carrier data, particularly for 802.11ac/ax (Wi-Fi 5/6/6E) and upcoming 802.11be (Wi-Fi 7) due to the large bandwidth, high throughput from MIMO, and high frame rate [33].

These limitations motivate the recent adoption of mid-grained intermediate channel measurements which are more informative (e.g., in the spatial domain) than RSSI measurements and easier to access than lower-level CSI measurements [34]–[39]. Specifically, spatial beam SNRs that are inherently available (with zero overhead) from beam training for IEEE 802.11ad/ay standards operating at millimeter-wave (mmWave) bands [40]–[42] can be used to construct a fingerprinting/training dataset. The use of mid-grained channel measurements was made possible by earlier efforts in [43]–[45] which enabled easy access to beam SNR measurements from commercial off-the-shelf (COTS) 802.11ad Wi-Fi routers.

Fusion-based approaches have been considered before to improve robustness and accuracy. Most fusion-based localization methods make use of heterogeneous sensor modalities such as on-device or wearable sensors, e.g., inertial navigation system (INS), magnetic sensors, accelerometers, ultrasound, and other sensors such as cameras, Bluetooth, and ZigBee. These heterogeneous sensor signals are then fused with Wi-Fi signals (either RSSI or CSI) in the traditional maximum likelihood framework or more advanced deep learning context. For instance, continuous localization of indoor pedestrians was achieved using INS with tracking errors adjusted by the Wi-Fi [46]. DeepFusion combines heterogeneous wearable (e.g., smartphone and smartwatch) and wireless (Wi-Fi and acoustic) sensors for human activity recognition [47]. A recent survey paper can be found in [48].

When only wireless radio frequency (RF) sensors (e.g., Wi-Fi, Bluetooth, ZigBee, LTE) are considered, the fusion of the coarse-grained RSSI, Bluetooth, and ZigBee measurements was considered in [49] using the k nearest neighbor (kNN) method. In [50], fine-grained CSI was used first to extract Angle-of-Arrival (AoA) measurements and then fused with RSSI measurements from Bluetooth in the context of the traditional maximum-likelihood estimation (MLE) framework.

If we further narrow down the scope to Wi-Fi-only measurements, the fusion between the CSI and RSSI can be done in a way similar to [50] or in a straightforward manner by concatenating the scalar RSSI measurements to the high-dimensional CSI measurements. For CSI Wi-Fi measurements, [51] proposed to fuse the phase and amplitude of the fine-grained CSI, as opposed to the magnitude-only CSI, for the purpose of localization. When multiple access points (APs) are available, the fusion of multi-view CSI measurements over APs was considered in [52] using generalized inter-view and intra-view discriminant correlation analysis. [53] proposed to fuse the probability maps from multiple APs and multiple transmit antennas. Moreover, [54] proposed the Maximum Mean Discrepancy (MMD) as a transfer learning technique
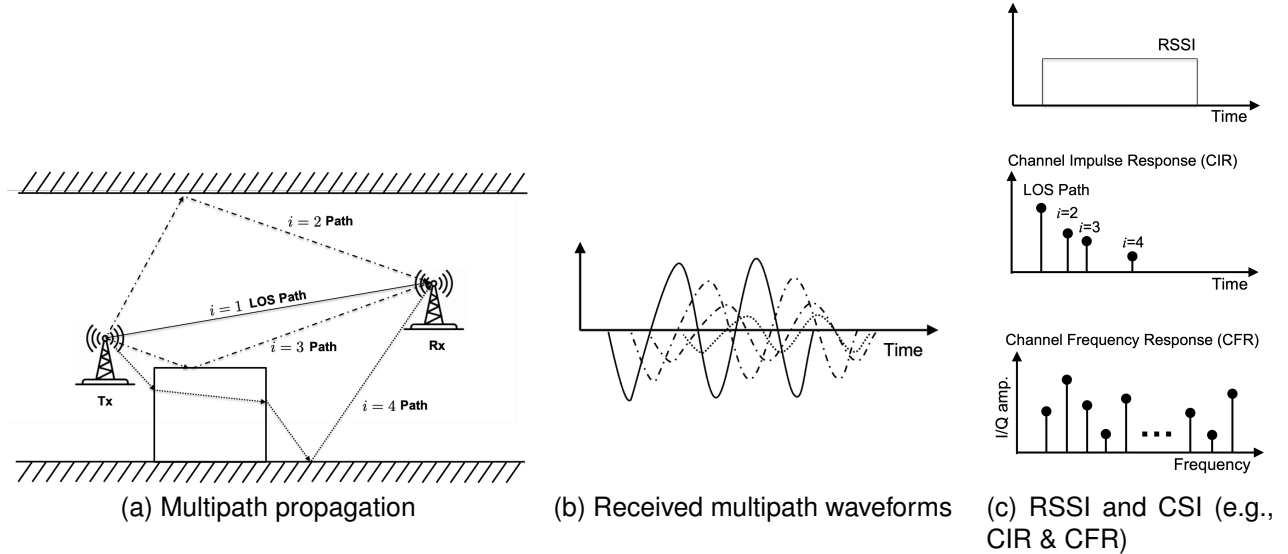
Fig. 1. The coarse-grained RSSI and fine-grained CSI channel measurements at sub-7 GHz frequency bands, e.g., 2.4 GHz and 5 GHz in 802.11ac.

for indoor localization with different placements of the APs.

In this paper, we propose to fuse the fine-grained CSI measurements at 5 GHz from multiple spatial streams and the mid-grained beam SNR measurements from the mmWave band at 60 GHz. Such a fusion is motivated by the observation that these two distinct kinds of Wi-Fi channel measurements probe the scene with different lenses (i.e., different physical space (time/delay versus space) and antenna domains (omni-directional antennas versus directional beampatterns) and, hence, provide rich yet complementary features in different physical domains: 5-GHz CSI reflects the time/delay-domain profile while the 60-GHz beam SNR possesses the spatial-domain features. In particular, our contributions and results are summarized as follows:

- Ours is the first work to fuse the two distinct kinds of Wi-Fi channel measurements at different frequency bands. The mid-grained beam SNRs are richer and more stable channel measurements than the coarse-grained RSSI and can lead to more meaningful fusion with the fine-grained CSI.
- We introduce a multi-band Wi-Fi fusion method that accounts for the granularity matching between the CSI and beam SNR measurements. The granularity matching is realized via a learnable fusion block that pairs any two feature maps from CSI and beam SNR measurements at different feature levels and assigns adaptive weights to the paired feature maps.
- We propose an autoencoder-based multi-band fusion network that can be pre-trained in an unsupervised fashion. With the pre-trained fusion network, we propose attaching multi-task heads to the fused features by fine-tuning the fusion block and training the multi-task heads from the scratch.
- We implement an in-house multi-band Wi-Fi testbed consisting of commercial 802.11ac- and 802.11ad-compliant Wi-Fi routers and collect real-world measurements in standard room environments (e.g., apartment and lab)

for three Wi-Fi sensing tasks: 1) pose recognition, 2) occupancy sensing, and 3) indoor localization.
- Finally, we conduct a comprehensive performance analysis by including four baseline methods and evaluating the performance as a function of the amount of labeled training data, the latent space, and the learning rate.

II. WI-FI CHANNEL MEASUREMENTS AND SENSING TECHNIQUES: AN OVERVIEW

In the following, we provide a literature review on Wi-Fi channel measurements and their application to indoor localization, human activity recognition, and wireless sensing.

A. RSSI: Received Signal Strength Indicator

In a typical indoor environment, the baseband signal voltage at an RF receiver at a given time is measured as [55]

$$V = \sum_{i=1}^{N} \|a_i\| e^{-j\theta_i}, \tag{1}$$

where a_i and θ_i are the amplitude and phase of the i-th multipath component, and N is the total number of multipath components.; see Fig. 1 (a) and (b) for an illustration. Then RSSI is the received power in decibels (dB) as

$$\text{RSSI} = 10 \log_2 \left(\|V\|^2 \right). \tag{2}$$

Early Wi-Fi based indoor localization systems used RSSI measurements to estimate indoor location in a direct local-ization fashion [16]–[19]. For fingerprinting-based methods, RSSI was used directly as fingerprinting data in systems such as Radar [20], Compass [21], and Horus [22] due to easy access to 802.11ac- and 802.11n-compliant devices. Classical machine learning methods, e.g., the k-nearest neighbor (kNN) and support vector machine (SVM), and probabilistic Bayesian methods have been applied to RSSI fingerprinting measurements [20], [22]–[29]. Advanced deep learning methods such
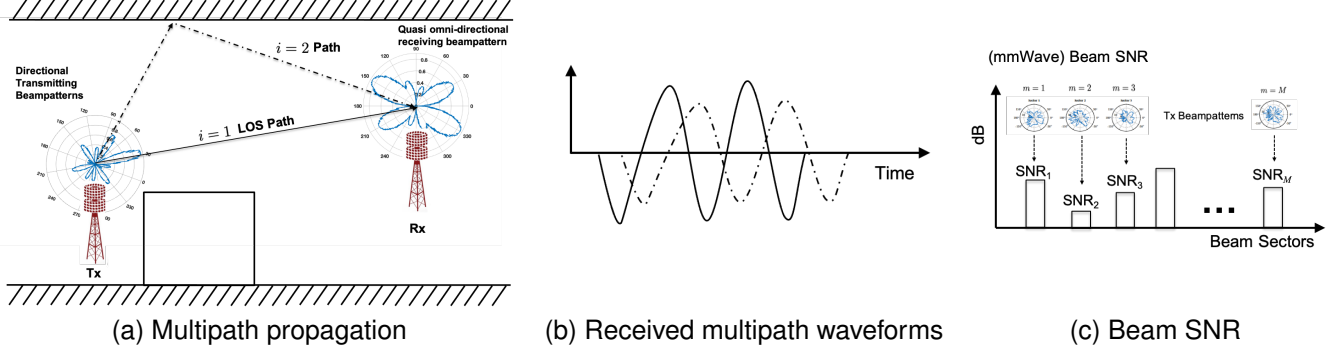
Fig. 2. The mid-grained beam SNR measurement at mmWave frequency bands, i.e., 60 GHz in 802.11ad.

as multi-layer neural networks [31] and recurrent neural networks (RNNs) [32] showed improved localization performance over classical machine learning approaches. Recent work in [56] proposed an extreme learning machine framework that locates floor and then predicts finer position using RSSI fingerprints in a multiple floor building.

Nevertheless, RSSI measurements, as a simple superposition of multipath voltages (1), fluctuate over time at a given location and are only a scalar measurement as shown in (2). RSSI is considered a *coarse-grained* channel measurement since it results in a single value per channel.

B. CSI: Channel State Information

By modeling the wireless channel as a temporal linear filter, one can measure the channel impulse response (CIR) $h(\tau)$ as

$$h(\tau) = \sum_{i=1}^{N} a_i e^{-j\theta_i} \delta\left(\tau - \tau_i\right), \qquad (3)$$

where τ_i is the delay of the i-th multipath, and $\delta(\cdot)$ is the Dirac delta function. In other words, the channel can be represented by the multipath amplitude and phase distribution at corresponding delays; see the middle figure of Fig. 1 (c) for an illustration. Given the CIR $h(t)$, the received signal $r(t)$ is the temporal convolution of the preamble $s(t)$ and $h(t)$: $r(t) = s(t) \otimes h(t)$. The channel frequency response (CFR) or, equivalently, the Fourier transform of the CIR, is often measured as

$$H(f) = S^*(f)R(f), \qquad (4)$$

where $R(f)$ is the Fourier transform of $r(t)$ and $S^*(f)$ is the complex conjugate of the Fourier transform of $s(t)$. In commercial Wi-Fi devices, a group of sampled CFRs at a list of subcarriers are measured as

$$H(f_k) = \|H(f_k)\|e^{j\angle H(f_k)}, k = 1, 2, \cdots, K, \qquad (5)$$

where f_k is the k-th subcarrier frequency; see the bottom figure of Fig. 1 (c) for an illustration. These complex-valued CFRs are reported to upper network layers in the format of CSI [57]. Compared with the coarse-grained RSSI, the CSI provides a *fine-grained* channel measurement with better capability to resolve multipath in the time or frequency domain. While still suffering from temporal fluctuation, the

CSI provides a set of complex-valued random variables (i.e., multipath components) compared with the scalar RSSI.

Efforts on CSI extraction from COTS Wi-Fi devices, e.g., Intel Wi-Fi Link 5300 radio [58], Atheros 802.11n chipsets [59], and Cypress 802.11ac chipsets [60], have enabled access to the CSI over a bandwidth of up to 80 MHz, and, more recently, 160 MHz from a Broadcom 802.11ax Wi-Fi chipset [61], at sub-7 GHz (e.g., 2.4 and 5 GHz bands). This has sparked a large swath of learning-based Wi-Fi sensing applications [3], [8], [62]–[72]. For instance, ConFi [9] used convolutional neural networks (CNNs) to train CSI measurements from three antennas, to classify the location and estimate location coordinates. [10] fingerprinted full CSI over multiple time instants, calibrated their phases, and fitted one autoencoder for one location. CSI measurements can also be directly trained to regress the coordinate [11], [12]. More recently, [70] used annotations from camera images to train fine-grained CSI measurements for pose and human tracking. Extracted environment-independent Doppler profile [73] from the CSI phase information can represent human activity, and such phase information can be also used for indoor localization.

C. Beam SNR: mmWave Beam Training Measurement

At mmWave bands, another type of channel measurements from the mandatory beam training phase provides a *mid-grained* measurement in the beam angle domain for Wi-Fi sensing. During the mandatory beam training phase defined by the 802.11 ad/ay standards, directional probing beampatterns are used to determine desired directions for subsequent data communication to compensate large path loss at mmWave bands [40]–[42]. For each probing beampattern (also referred to as beam sectors), beam SNR is computed as a measure of beam quality.

For a given pair of transmitting and receiving beampatterns, corresponding beam SNR can be defined as

$$h_m = \text{BeamSNR}_m = \frac{1}{\sigma^2} \sum_{i=1}^{N} \gamma_m(\theta_i)\zeta_m(\psi_i)P_i, \qquad (6)$$

where m is the index of beampattern, θ_i and ψ_i are the transmitting and receiving azimuth angles for the ith path, respectively, P_i is the signal power at the ith path, $\gamma_m(\theta_i)$ and $\zeta_m(\psi_i)$ are the transmitting and receiving beampattern
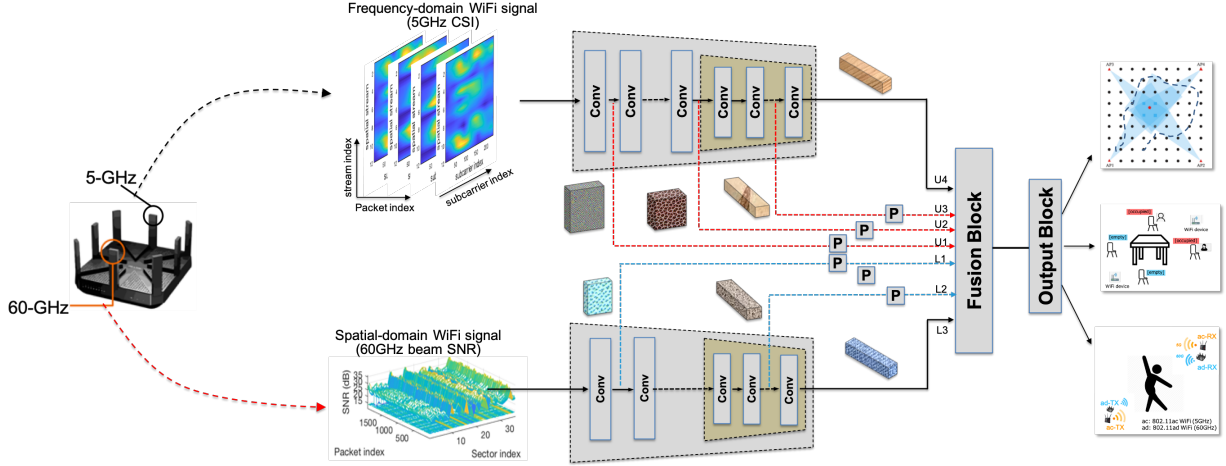
Fig. 3. Multi-band Wi-Fi fusion with granularity matching.

gains at the ith path for the mth beampattern, respectively, and σ^2 is the noise variance. Fig. 2 (a) shows an example of $I = 2$ paths between the transmitting side that probes the spatial domain using a directional beampattern and the receiving side which is in a listening mode with a quasi-omni-directional beampattern, while Fig. 2 (c) shows a number of pre-determined beampatterns used at the transmitting side to compute corresponding beam SNRs. Due to antenna housing and hardware impairments, these beampatterns exhibit fairly irregular shapes at the 60 GHz band [43], [44].

Earlier efforts in [43]–[45] formulated a direct localization using the beam SNR measurement as a constrained optimization and required dedicated chamber measurements of beam patterns. Our previous work in [34] and [35] considered model-based signal processing and classical machine learning approaches to learn the nonlinear mapping from the beam SNR to the location-and-orientation information. More advanced deep learning approach was applied in LOS [36] and NLOS scenarios [37].

D. CSI versus Beam SNR

By comparing Fig. 1 (a) with Fig. 2 (a), several observations can be made between the CSI and beam SNR. First, multipath propagation is richer at sub-7 GHz bands (i.e., 4 paths) than 60 GHz bands (i.e., 2 paths). Second, the through-wall path (denoted as dotted lines) in Fig. 1 (a) does not survive in the 60 GHz link due to its much shorter wavelength and less capability to penetrate obstacles (e.g., wall). Third, mmWave Wi-Fi devices are equipped with a phased array that enables highly directional beampatterns while antenna elements of sub-7 GHz Wi-Fi devices are mostly in an omni-directional mode. Although multiple antenna elements at sub-7 GHz can be used for beamforming, the beampatterns are less directional than the mmWave Wi-Fi device due to the smaller number of antennas and the relatively large inter-element spacing. Consequently, these two distinct kinds of Wi-Fi channel measurements probe the scene with different lenses and may exhibit complementary channel features in different

TABLE I
SYMBOLS

Symbols	Meaning
$H(f)$	Channel Frequency Response
h_m	beam SNR
$\mathbf{C}$	set of CSI
$\mathbf{h}$	set of beamSNR
$\mathbf{C_r}$	set of labeled CSI
$\mathbf{h_r}$	set of labeled beamSNR
M_s	number of subcarriers (CSI)
N_s	number of spatial stream (CSI)
M	number of beam SNR
d	latent space size
λ	fusion weight

physical space (time/frequency versus space) and domains (omni-directional antennas versus directional beampatterns).

III. MULTI-BAND WI-FI SENSING WITH SUPERVISED LEARNING

In this section, we introduce the multi-band Wi-Fi fusion framework that combines the features from both the fine-grained CSI measurements and the mid-grained beam SNRs. Particularly, we account for the feature granularity by designing a learnable granularity matching network that pairs multi-scale features from the CSI and beam SNR and assigns linear weights to the paired features. Specifically, Fig. 3 shows the multi-band Wi-Fi sensing network that consists of three blocks: 1) feature extraction, 2) feature fusion with granularity matching, and 3) separate output blocks for individual sensing tasks. Such a framework is then trained end-to-end for human activity recognition, occupancy detection, and localization tasks, respectively, in a supervised fashion. In the following, we introduce each block in more detail along with symbols defined in Table I.

A. Preliminary

Like traditional fingerprinting-based Wi-Fi sensing methods, we follow the standard procedure by collecting both CSI and beam SNR measurements corresponding to each class
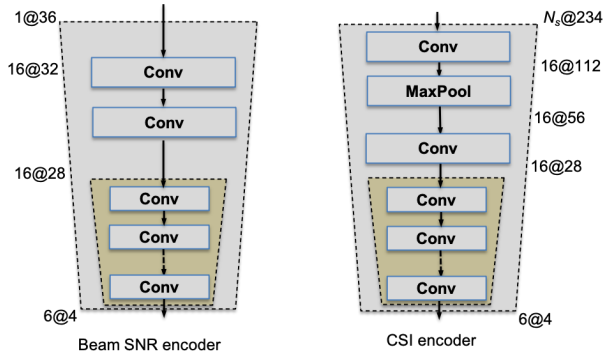
Fig. 4. Separate CSI and beam SNR encoder architectures.

Fig. 5. Feature fusion with granularity matching.

(e.g., pose, occupancy pattern, and location/orientation) as the fingerprinting data. Specifically, we use $\mathbf{C} \in \mathbb{C}^{N_s \times M_s}$ and $\mathbf{h} \in \mathbb{R}^{M \times 1}$ to denote the CSI measurements from N_s spatial streams over M_s subcarriers and the beam SNRs from M beampatterns. For a given class l, R fingerprinting snapshots, $\mathbf{C}_1(l), \ldots, \mathbf{C}_R(l)$ and $\mathbf{h}_1(l), \ldots, \mathbf{h}_R(l)$, are collected to construct the offline training dataset. By collecting many realizations of both Wi-Fi channel measurements at L classes, we will have L sets of training data in the training dataset. For the labeled data, the label class l is automatically attached to both $\mathbf{C}_r(l)$ and $\mathbf{h}_r(l)$.

B. Feature Extraction

In Fig. 3, the feature extraction block employs two separate convolution networks to encode the feature maps for CSI-only and beam SNR-only measurements. Specifically,

$$\mathbf{y}_\ell^c = f_\ell^c(\mathbf{y}_{\ell-1}^c, \boldsymbol{\theta}_\ell^c), \quad \ell = 1, 2, \ldots, N_c$$
$$\mathbf{y}_\ell^h = f_\ell^h(\mathbf{y}_{\ell-1}^h, \boldsymbol{\theta}_\ell^h), \quad \ell = 1, 2, \ldots, N_h, \quad (7)$$

where $\mathbf{y}_0^c = \mathbf{C}_r(l)$, $\mathbf{y}_0^h = \mathbf{h}_r(l)$, and $f_\ell^{c/h}$ denotes the convolution operation (including batch normalization, pooling and activation functions) for CSI-only and beam SNR-only feature extraction with kernel parameters $\boldsymbol{\theta}_\ell^{c/h}$ at the ℓ-th layer. Overall, we use N_c and N_h convolution layers for, respectively, the CSI and beam SNRs to gradually shrink the dimension of the feature maps while increasing the number of feature maps. The ℓ-th feature map is denoted as $\mathbf{y}_l^{c/h} \in \mathbb{R}^{N_\ell^{c/h} \times H_\ell^{c/h} \times W_\ell^{c/h}}$, where $\{N_\ell^{c/h}, H_\ell^{c/h}, W_\ell^{c/h}\}$ are, respectively, the number of channels, height and width of the ℓ-th feature map. Note that the number of layers for CSI can be different from that for beam SNRs, i.e., $N_c \neq N_h$.

To be more specific, we use 5 convolutional blocks, respectively, for the CSI and beam SNR encoders, as shown in Fig. 4. Each convolution block consists of a standard convolution layer followed by a batch normalization layer and a rectified linear unit (ReLU) activation layer. For the CSI encoder, we use the first 2 convolution blocks and 1 maxpooling layers to increase the number of feature maps to 16 and downsize the dimension to 28. Similarly, the beam SNR encoder uses 2 convolution blocks to downsize the input to the same feature space. Then the 3 subsequent convolution blocks share the same structure as highlighted in tan colored blocks of Fig. 4.
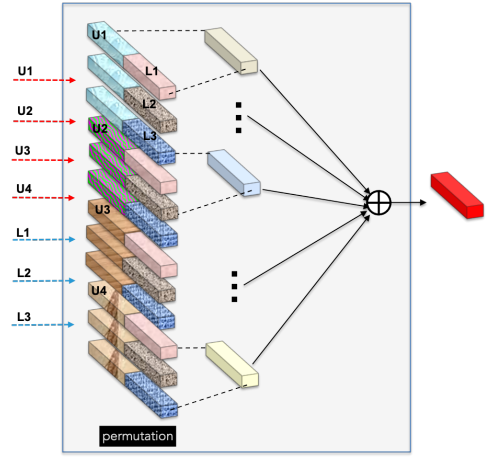
C. Feature Fusion

Since the CSI and beam SNRs are two distinct Wi-Fi channel measurements with different channel feature granularities, one needs to take into account the granularity correspondence for the multi-band Wi-Fi fusion. For instance, CSI features at later convolution layers may have a better correspondence to beam SNR features at earlier layers as the CSI is likely to have finer granularity than the beam SNR. To this end, we extract multi-scale feature maps at selected convolution layers as two sets of threads

$$\{\mathbf{y}_{\ell \in \mathbb{L}_c}^c\} \quad \& \quad \{\mathbf{y}_{\ell \in \mathbb{L}_h}^h\}, \quad (8)$$

where $\mathbb{L}_c$ and $\mathbb{L}_h$ denote the selected layer indices of the CSI and, respectively, beam SNR encoders. For instance of Fig. 3, we have 4 layer indices in $\mathbb{L}_c = \{1, 3, 5, 6\}$ and 3 elements in $\mathbb{L}_h = \{1, 4, 5\}$.

With an identity mapping or a linear projection to a smaller dimension, the features from the upper encoder branch are mapped to $\{\mathbf{u}_\ell\}_{\ell=1}^{|\mathbb{L}_c|}$ and the lower ones as $\{\mathbf{l}_\ell\}_{\ell=1}^{|\mathbb{L}_h|}$ in Fig. 3, where $|\mathbb{L}|$ denoting the cardinality of the set $\mathbb{L}$. To find correspondences between the extracted multi-scale features from the CSI and beam SNR, the proposed fusion scheme concatenates two selected feature maps from the two encoders: one is from the CSI encoder and the other from the beam SNR encoder.

$$\mathbf{y}_p = [\mathbf{u}_\ell; \mathbf{l}_{\ell'}], \ell \in \{1, \cdots, |\mathbb{L}_c|\}, \ell' \in \{1, \cdots, |\mathbb{L}_h|\}, \quad (9)$$

where $p = 1, \cdots, P$ denotes the fusion pair index with $P = |\mathbb{L}_c||\mathbb{L}_h|$. For instance, we have $P = 12$ pairs in Fig. 5 by pairing one of four CSI features $\mathbf{u}_l$ and one of three beam SNR features $\mathbf{l}_{\ell'}$.

Then, each concatenated feature is fused with a fully connected layer into the same latent dimension d,

$$\mathbf{f}_p = \mathbf{W}_p \mathbf{y}_p + \mathbf{b}_p, \quad p = 1, \cdots, P, \quad (10)$$

where $\mathbf{W}_p$ is the projection weight matrix for the p-th pair with $\mathbf{b}_p$ denoting the bias term. The final fusion layer is a

linear combination of the fused feature map as

$$\mathbf{f} = \sum_{p=1}^{P} a_p \mathbf{f}_p, \tag{11}$$

where a_p is the fusion weight that can be learned from the training data. The fusion weights act as a soft selection vector on all concatenated feature vectors. A higher weight implies that the corresponding pair of features has a better granularity matching level. In the case of $\{a_p\}_{p<P} = 0$ and $a_P = 1$, the fused feature map (11) only combines the features at the final layers of the two encoders. In contrast, with non-zero values of $\{a_p\}_{p<P}$, the proposed fusion scheme allows multi-scale features from earlier layers to combine with their corresponding values from the other branch and contribute to the final fused features.

D. Output Block

In Fig. 3, the fused feature map $\mathbf{f}$ is fed into the output block to generate the output vector $\mathbf{u}$. The output block consists of a number of fully-connected layers:

$$\mathbf{z}_\ell = g_\ell(\mathbf{z}_{\ell-1}, \boldsymbol{\gamma}_\ell), \quad \ell = 1, 2, \ldots, N_g, \tag{12}$$

where $\mathbf{z}_0 = \mathbf{f}$, $\mathbf{u} = \mathbf{z}_{N_g}$, and $g_\ell = \sigma_\ell(\mathbf{W}_\ell \mathbf{z}_\ell + \mathbf{b}_\ell)$ with $\boldsymbol{\gamma}_\ell = \{\mathbf{W}_\ell, \mathbf{b}_\ell\}$ and σ_ℓ denoting the activation function such as ReLU for $\ell < N_g$ and an identity mapping when $\ell = N_g$ at the final output layer. Note that, for different sensing applications, the dimension of the output vector $\mathbf{u}$ may be different, depending on the number of classes. For instance, $N = 8$ for the pose recognition and occupancy sensing and $N = 16$ for the indoor localization.

E. Loss Function

To train the multi-band Wi-Fi sensing network with labeled training data $\{\mathbf{C}_r(l), \mathbf{h}_r(l)\}$ with the label l, corresponding output of the last layer $\mathbf{u}$ is first normalized with the softmax operation as

$$s_n = \exp(u_n) / \sum_{i=1}^{N} \exp(u_i), \quad n \in \{1, 2, \ldots, N\}. \tag{13}$$

where s_n is the nth element of the normalized output u_n. Then, the cross-entropy loss function is computed over the score vector $\mathbf{s} = [s_1, s_2, \ldots, s_N]$ and the corresponding one-hot label vector $\mathbf{c} = [c_1, c_2, \ldots, c_N]$ as

$$L_{\text{classification}} = -\sum_n c_n \log(s_n). \tag{14}$$

where the one-hot label vector $\mathbf{c}$ is 1 at the l-th element and 0 elsewhere.

IV. MULTI-BAND WI-FI SENSING WITH TRANSFER LEARNING

Although it is simple to train the above network separately for each sensing task, the offline fingerprinting phase is time- and manpower-consuming [74], [75]. To label the data, one has to associate both channel measurements with the ground-truth labels, in the form of pose gesture, occupancy pattern,

or user location. This issue is severe when one needs to build up the training dataset when the number of classes is large. Even worse, after a Wi-Fi sensing system is deployed, the ground-truth information is hard to obtain, while the Wi-Fi devices can still gather or "listen" the environment and access to multi-band Wi-Fi channel measurements.

To meet the challenges of limited labeled training data, we consider an unsupervised fusion approach using a multi-stream autoencoder to train the encoder and fusion blocks with unlabeled training data $\mathbf{C}_r$ and $\mathbf{h}_r$ with $r > R$, possibly pooled from multiple sensing tasks. With the unsupervised trained encoder and fusion blocks, a transfer learning approach is then applied to fine-tune the fusion block and retrain a new attached output block for individual sensing task with limited labeled data $\mathbf{C}_r(l)$ and $\mathbf{h}_r(l), r = 1, \cdots, R$.

A. Unsupervised Multi-Stream Autoencoder with Feature Granularity Matching

Fig. 6 shows a multi-stream autoencoder architecture that uses the same two-branch encoder as in the previous section to extract the features of the two Wi-Fi channel measurements, the same fusion block taking multi-scale features from the two encoder branches and matching the feature granularity, and a new two-branch decoder to synthesize the two input Wi-Fi channel measurements from the same fused feature at the output of the fusion block.

The multi-stream autoencoder can be trained end-to-end using unlabeled ("listening") multi-band Wi-Fi data that can be pooled from multiple sensing tasks. The motivation here is that, without any class label, one can still train or update the feature extractor and fusion blocks continuously at local Wi-Fi devices or edge units. Whenever new labeled data are available (e.g., users share their location sporadically), the limited amount of labeled data can be used to train the output block for the task-specific sensing task while leaving the training of feature extraction and fusion blocks to the unlabeled data. In the following, we elaborate on each building block of Fig. 6.

1) Encoder: At the encoder side, we re-use the same two encoder networks to extract the feature maps for CSI and beam SNR measurements in Fig. 4. Multi-scale features from multiple levels of both encoder branches are extracted as $\mathbf{y} = \{\mathbf{y}_{\ell \in \mathbb{L}_c}^c, \mathbf{y}_{\ell \in \mathbb{L}_h}^h\}$ for the feature fusion with granularity matching.

2) Fusion: For the fusion block, we also re-use the feature fusion in (11) that takes the selected feature maps $\mathbf{y}_p$ to generate the fused feature map $\mathbf{f}$ of (11) with the granularity matching scheme in Fig. 5.

3) Decoder: At the decoder side, the fused feature $\mathbf{f}$ is fed to two separate decoders to recover the input CSI and beam SNR measurements simultaneously. Particularly,

$$\begin{aligned} \mathbf{f} &= \mathbf{t}_0^c = \mathbf{t}_0^h, \\ \mathbf{t}_\ell^c &= d_\ell^c(\mathbf{t}_{\ell-1}^c, \boldsymbol{\zeta}_\ell^c), \quad \ell = 1, 2, \ldots, N_c \\ \mathbf{t}_\ell^h &= d_\ell^h(\mathbf{t}_{\ell-1}^h, \boldsymbol{\zeta}_\ell^h), \quad \ell = 1, 2, \ldots, N_h, \end{aligned} \tag{15}$$

where $\mathbf{f}$ is the output of the fusion block in Fig. 6 that is simultaneously fed into the two decoders, $d_\ell^{c/h}$ denotes the
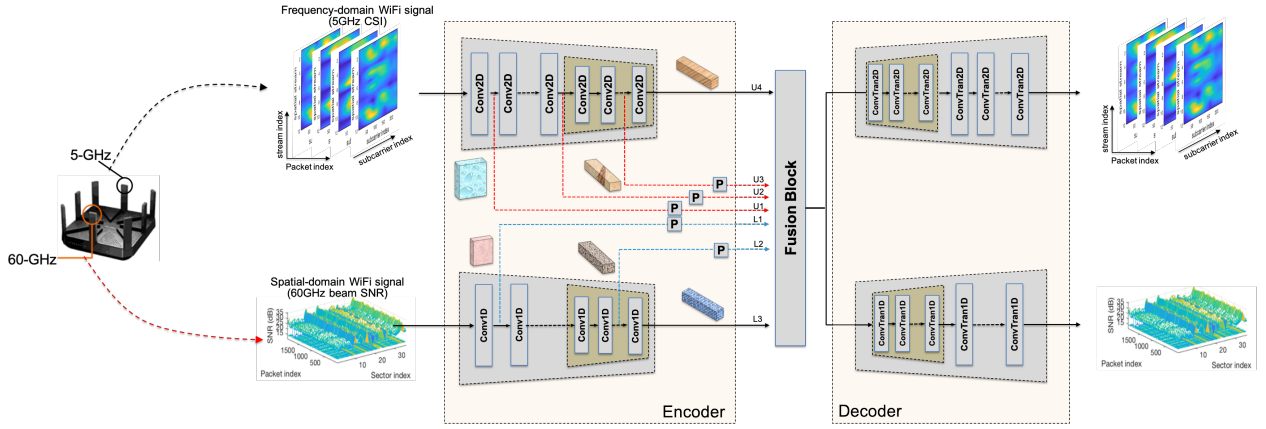
Fig. 6. Unsupervised training of feature extraction and fusion blocks with unlabeled ("listening") multi-band Wi-Fi data pooled from multiple sensing tasks.
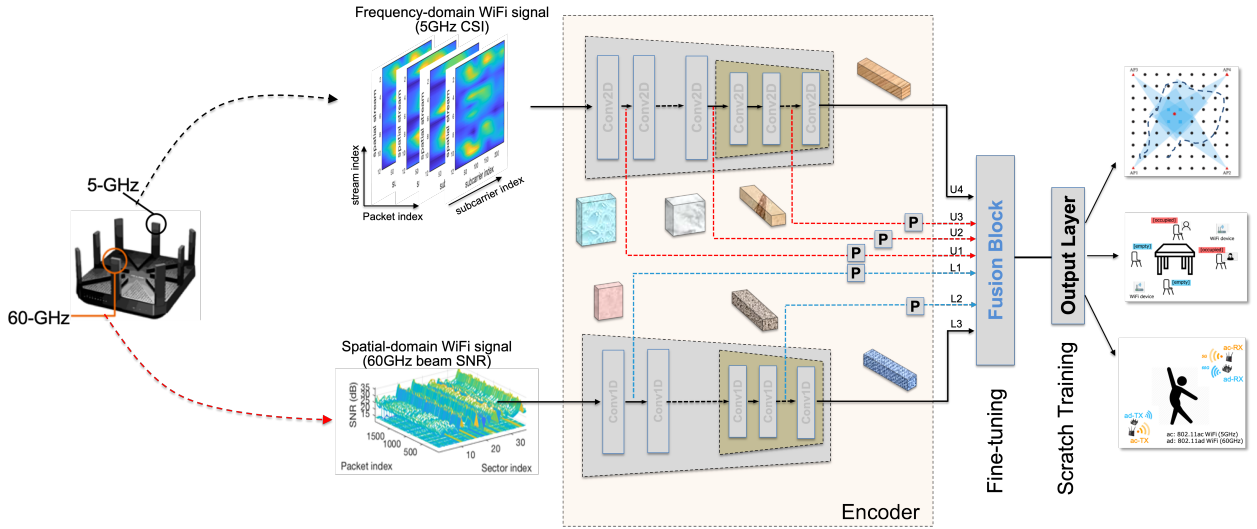

Fig. 7. Multi-band Wi-Fi transfer learning with the pretrained encoder from Fig. 6, fine-tuned fusion block from Fig. 6, and newly added task-specific output block trained using limited labeled data.

transposed convolution or upsampling operation at the ℓ-th layer of the two decoders with corresponding parameters $\boldsymbol{\zeta}_\ell^{c/h}$. Usually, we implement the decode network with a mirrored architecture of its encoder network. Therefore, we keep the same number of layers, N_c and N_h, in the decoders. The decoder architectures for the beam SNR and CSI mirror their corresponding encoder architectures in Fig. 4.

4) Loss Function: To train the autoencoder-based fusion network with unlabeled data, we adopt a weighted mean-squared error (MSE) as the loss function:

$$L_{\text{MSE}}(w_c) = \lambda \sum_r \left(\mathbf{y}_0^c - \mathbf{t}_{N_c}^c\right)^2 + (1-\lambda) \sum_r \left(\mathbf{y}_0^h - \mathbf{t}_{N_h}^h\right)^2 \tag{16}$$

where $\mathbf{y}_0^c = \mathbf{C}_r$ and $\mathbf{y}_0^h = \mathbf{h}_r$ are the CSI and beam SNR training samples, $\mathbf{t}_{N_{c/h}}^{c/h}$ are corresponding outputs of the two decoders, and λ is the hyperparameter to balance the reconstruction error between the CSI and beam SNR branches.

B. Transfer Learning with Limited Labeled Data

Once the autoencoder-based multi-band Wi-Fi fusion network in Fig. 6 is trained, we use the transfer learning to freeze the encoder parameters, fine-tune the fusion block, remove the decoder block, and attach a new output classification block for each sensing task, as shown in Fig. 7.

With the labeled training data for each sensing task, we retrain the output block which takes the same form of (12) and outputs the vector $\mathbf{u}$. Specifically, the transfer learning computes the classification loss function of (14) and back-propagates the gradient of the loss function with respect to the parameters in the output block in a large step size, e.g., a learning rate of 0.01, while fine-tuning the parameters within the fusion block with a small training step size (e.g., a learning rate of 0.001 or smaller) on the linear fusion weights $\{a_p\}_{p=1}^P$ of (11) and an even smaller learning rate on $\{\mathbf{W}_p, \mathbf{b}_p\}$ of (7). It is worth noting that, for each sensing task, the transfer learning needs to apply separately to train separate output blocks. Once the entire network of Fig. 7 is re-trained, we
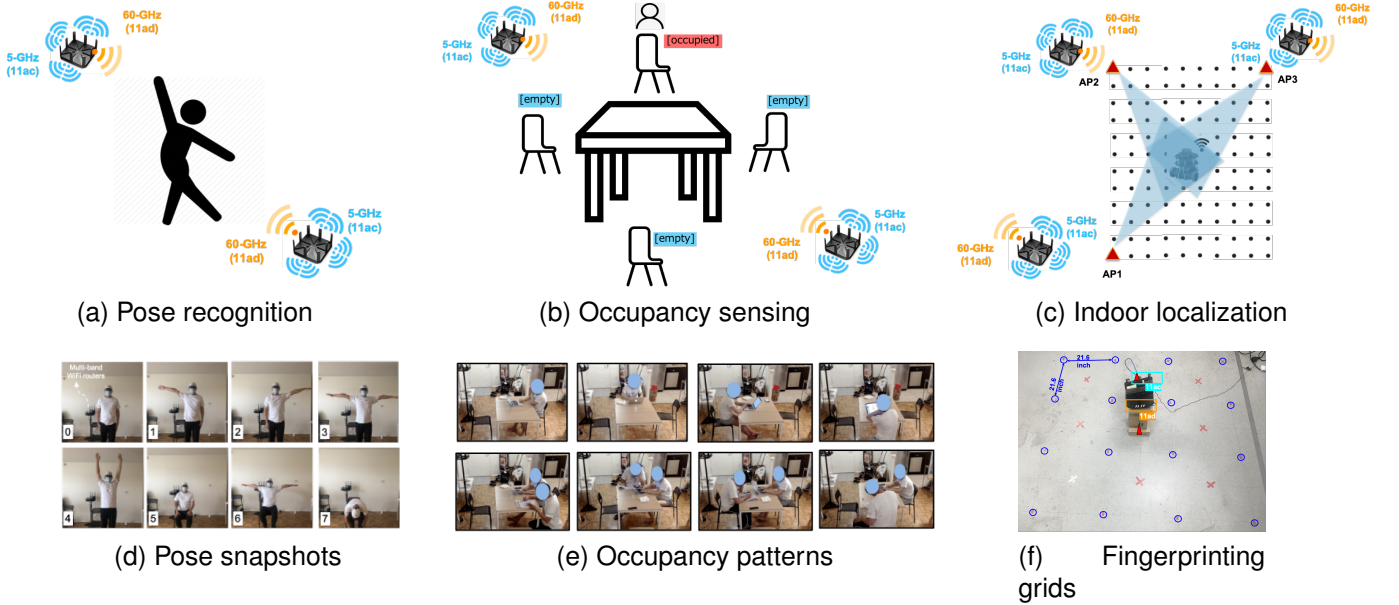
Fig. 8. Multi-band Wi-Fi sensing tasks and experiment settings.

deploy the network with test datasets for different sensing tasks.

V. MULTI-BAND WI-FI TESTBED AND SENSING TASKS

In the following, we provide detailed Wi-Fi testbed setup and experiment configuration for three sensing tasks: 1) pose recognition; 2) occupancy sensing; and 3) indoor localization. We also briefly overview preprocessing steps to calibrate the multi-band Wi-Fi data.

A. Experiment Setup

The data collection system consists of multiple commercial 802.11ac- and 802.11ad-compliant routers and devices in a configuration shown in Fig. 8 (a)-(c) for the three sensing tasks. Both 5 GHz CSI and 60 GHz beam SNR measurements are recorded in the routers and sent to a workstation via Ethernet cables. The data collection system is deployed in standard indoor room settings, as shown in Fig. 8 (d)-(f).

CSI from 802.11ac devices: We use ASUS RT-AC86U routers with 3 external and 1 internal antennas to extract the CSI measurements at 5 GHz and modified its firmware using the Nexmon CSI Extractor Tool of [60]. It allows per-frame CSI extraction for up to 4 spatial streams using all four receive chains on Broadcom and Cypress Wi-Fi chips with up to 80 MHz bandwidth in both the 2.4 and 5 GHz bands. It also supports devices such as the low-cost Raspberry Pi platform and mobile platforms such as Nexus smartphones. Other CSI tools are also available at 2.4 GHz and 5 GHz [58], [59], but none of them supports 80 MHz wide channels within the newer 802.11ac standard (with respect to 802.11n/g). In addition, it supports various Tx-Rx antenna configurations, up to 4×4 MIMO. In our in-house testbed, we use a pair of ASUS routers to record all 16 spatial streams over 242 subcarriers. However, to mimic a standard mobile terminal, we only use 3 spatial streams which are equivalent to a configuration of using 1 ASUS routers with 3 external antennas and 1 mobile user with a single antenna.

Beam SNR from 802.11ad devices: We use TP-Link Talon AD7200 routers to collect beam SNRs at 60 GHz. Complying with IEEE 802.11ad standards, this router implements a Qualcomm QCA9500 transceiver that supports a single stream communication in 60 GHz range using analog beamforming over 32-element planar array. To support beam scanning with multiple beampatterns, a pre-determined codebook is implemented in the firmware to control the RF phase shifters. A total of $M = 36$ pre-determined beampatterns are implemented in the QCA9500 chip with a few measured beampatterns shown in Fig. 2-C. The raw beam SNRs are extracted by using the open-source software package in [76]. By matching the patterns of IEEE 802.11ad beam training frames with the memory inside the chip, one can identify parts of the firmware handling the beam training frames and extract beam SNR measurements from these memory addresses. For Talon AD7200 routers, the beam SNR measurements are further quantized in a stepsize of 0.25 dB. Overall, from one beam training, one AP can collect $M = 36$ beam SNRs for the 36 transmitting beampatterns. It is noted that the resulting beampatterns depart from the theoretical ones and exhibit fairly irregular shapes due to hardware imperfections and housing at 60 GHz.

B. Sensing Tasks and Datasets

With the above multi-band Wi-Fi testbed, we conducted multiple sensing tasks as follows:

1) Pose Recognition: Fig. 8 (a) shows the experimental configuration for the pose recognition by using one Wi-Fi station (including both 802.11ac and 802.11ad routers) in the front of the subject and one station behind the subject. Both

stations are placed on a stand of a height of 1.20 meters with a distance of approximately 2 meters. As shown in Fig. 8 (d), the subject is asked to perform a total of 8 poses including distinct gestures like 'sit', 'stand with left arm lifted', 'stand with both arms lifted', etc. For each pose, we recorded 7 independent sessions with different time durations all taken on the same day and with sufficient time separation between any two consecutive sessions. We grouped the measurements in the first four data sessions into the training data set and the last three sessions into the test data set to create the maximum time separation between training and testing. The numbers of both CSI and beam SNR samples (in brackets) in the training and test datasets are listed in the second and third columns of Table II.

2) Occupancy Sensing: Fig. 8 (b) shows the experimental configuration for occupancy sensing around a table in a small living room environment. Two stations are placed in the top left and bottom right corners in a diagonal configuration. The selected 8 occupancy patterns cover both a single subject sitting on one of the four chairs and two subjects, as shown in Fig. 8 (e). During the data collection, the subject is free to move body parts such as typing on the laptop and reading the paper as long as the subject is fixed at the location. We also recorded 7 independent sessions on the same day with sufficient time separation between any two consecutive sessions. Measurements were grouped into the training and test datasets in the same way as in the pose recognition task. The numbers of both beam SNR and CSI samples (in brackets) in the training and test datasets are listed in the fourth and fifth columns of Table II. Similar data augmentation is performed to align the beam SNR samples with the higher density CSI data in the training dataset.

3) Indoor Localization: Finally, we consider indoor localization in a standard lab environment with multiple AP stations (including one 802.11ac and three 802.11ad routers) and one user on the grid, as denoted by blue dots in Fig. 8 (f). The AP stations are placed around the three corners of the fingerprinting area, denoted as triangles in Fig. 8 (c). The number of grids is $4 \times 4 = 16$ with a grid size of 54.86 cm. For each grid, we place another station with both 802.11ac and 802.11ad devices acting as a single-antenna mobile user and record multiple data sessions for training and independent testing.

C. Wi-Fi Signal Preprocessing

For the fingerprinting dataset, we perform preprocessing steps to calibrate CSI and beam SNR measurements. For the CSI measurements from commercial routers (ASUS RT-AC86U in our case), known hardware issues include [59], [77]–[80]

1) Power control uncertainty: The raw CSI amplitude is compensated by the automatic gain controller (AGC) and mixed with a power amplifier uncertainty error, causing an amplitude offset;
2) Sample frequency offset (SFO): the sampling frequencies between the transmitter and receiver also exhibit an offset due to non-synchronized clocks, causing a time

(a) raw CSI amp. (b) calibrated CSI amp.

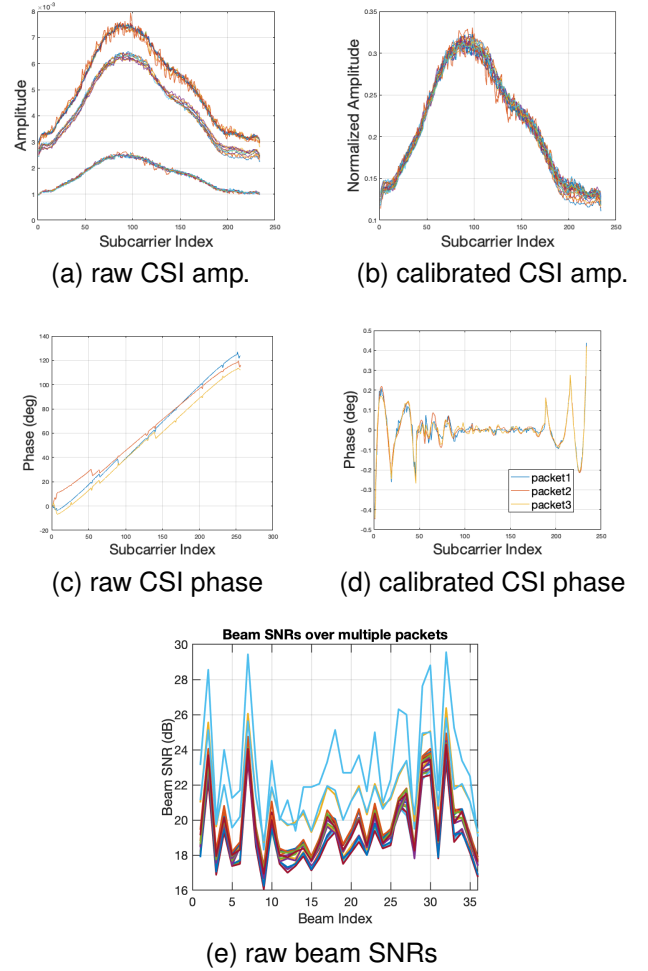(c) raw CSI phase (d) calibrated CSI phase

(e) raw beam SNRs

Fig. 9. Multi-band Wi-Fi measurement calibration and preprocessing.

shift to the digitally sampled signal after the analog-to-digital converter (ADC);
3) Package boundary detection (PBD) error: due to correlator sensitivity of packet detector, the packet detection at the receiver introduces another time shift, resulting in a random phase offset to the raw CSI;
4) Carrier frequency offset (CFO): The carrier frequencies between the transmitter and receiver cannot be perfectly synchronized. The CFO is initially compensated by a CFO corrector at the receiver. However, the compensation is incomplete due to the hardware imperfection, leaving residual CFO and leading to a phase offset over time;
5) Phase offsets between RF chains: The CSI phases at different RF chains are not synchronized perfectly and it results in semi-deterministic phase offsets between RF chains;

result in both amplitude and phase distortions to raw CSI measurements.

For Item 1), the CSI amplitude offsets in individual bands can be removed by an averaging or a standardization operation [59]; see Fig.9 (a) and (b). For Items 2)-4), the phase offset over subcarriers and/or packets can be calibrated by either a linear phase fitting approach used in the SpotFi [77, Algorithm

TABLE II

NUMBER OF BEAM SNR AND CSI (IN THE BRACKET) SAMPLES FOR EACH
CLASS FOR POSE RECOGNITION AND OCCUPANCY SENSING.

Class ID	Training-Pose	Test-Pose	Training-Occ.	Test-Occ.
0	434 (6510)	151	508 (7620)	230
1	499 (7485)	149	717 (10755)	290
2	325 (4875)	173	438 (6570)	189
3	347 (5205)	129	483 (7245)	184
4	238 (3570)	88	602 (9030)	235
5	314 (4710)	96	815 (12225)	217
6	272 (4080)	119	535 (8025)	200
7	432 (6480)	135	608 (9120)	271

TABLE III

TASK-SPECIFIC SUPERVISED TRAINING: AVERAGE PROBABILITY OF
SUCCESSFUL CLASSIFICATION AS A FUNCTION OF THE PERCENTAGE OF
LABELED DATA R TO ALL TRAINING DATA $r + R$ (I.E., $R/(r + R)$)

$R/(r + R)$	CSI	bSNR	IF	FF	GM
All-Pose	74.8%	92.3%	89.5%	90.0%	**94.4%**
40%	64.0%	86.8%	84.2%	83.0%	**90.2%**
20%	57.1%	81.8%	83.2%	79.9%	**89.9%**
10%	48.7%	77.9%	73.0%	77.8%	**80.0%**
All-Occ.	89.6%	88.5%	**96.0%**	93.4%	95.5%
40%	80.2%	86.7%	92.7%	92.1%	**94.4%**
20%	76.4%	86.0%	91.2%	91.2%	**93.2%**
10%	73.1%	76.0%	83.5%	82.6%	**86.8%**
All-Loc.	85.2%	92.5%	**96.4%**	95.2%	95.8%

1] or a multi-subband phase correction approach in [59], [78]. The compensated phases (with respect to the first subcarrier) of multiple packets are stable over time, as shown in Fig.9 (c) and (d). The combined amplitude and phase compensated CSI can serve as an indirect measurement of the time-of-flight (ToF) or delay power profile of the multi-path propagation.

To further extract the angle (e.g., AoA) information across multiple antennas, one may need to compensate the initial phase offsets (of the phased-locked loops (PLLs)) between RF chains of Item 5) by connecting a coaxial cable between the transmitter and receiver [80]. However, due to the semi-deterministic nature of this phase offset and the fixed amount of phase permutations (4 phase permutations for 3 antennas), we skip this antenna calibration and treat the raw CSI across antennas as augmented CSI data. Moreover, we remove guard subcarriers as corresponding amplitudes are always zeros.

On the other hand, we found that the beam SNRs as positive scalar values (so no need for the phase calibration) are much more consistent than the CSI measurement, except for a scaling effect over multiple packets; see Fig. 9 (e). As a result, the raw beam SNRs in the training data are augmented by multiplying them with a random scaling factor in $[0.9, 1.2]$ and then standardized over each beam index. Table II shows the numbers of beam SNR and CSI samples in the training dataset to show the amount of data augmentation performed on the beam SNR measurements to match the CSI values. However, we only show the number of beam SNRs in the test dataset since no data augmentation is used. Instead, in the test dataset we reduce the CSI data to match the number of beam SNRs.

VI. PERFORMANCE EVALUATION

In the following, we present the performance evaluation for the proposed multi-band Wi-Fi fusion framework using real-world experimental datasets collected from the above testbed. For each of these three tasks, we compare the proposed fusion methods with two single-modal (CSI-only and beam SNR-only) methods and two multi-modal methods.

A. Baseline Methods

For a fair comparison, we consider a list of 4 baseline methods with the proposed fusion network with granularity matching (GM) for both supervised learning and transfer learning:

- **CSI-Only** uses only the fine-grained CSI measurements as the input. The neural network follows the same architecture of the upper CSI encoder of Fig. 3 for supervised learning and the upper CSI branch (both encoder and decoder) of Fig. 6 for unsupervised fusion. The only difference is that the input is only the CSI and we only change the first convolutional layer configuration to reflect the change in the input dimension while keeping all other layers the same.
- **bSNR-Only** uses only the mid-grained beam SNR (bSNR) measurements as the input. Similarly, we slightly change the first convolution layer of Fig.3 for supervised learning and the lower beam SNR branch (encoder and decoder) of Fig. 6 for unsupervised fusion while keeping other layers the same.
- **Input-Fusion** (IF) concatenates the CSI and beam SNR at the input. The concatenated input is then fed to the CSI encoder of Fig. 3 with the first convolution layer slightly changed and the CSI branch of Fig. 6 for unsupervised fusion.
- **Feature-Fusion** (FF) follows the architectures of Fig.3 and Fig.6 by only fusing the outputs of the two encoders without the granularity matching.

For performance evaluation, we use the confusion matrix **C** as the criterion:

$$\mathbf{C}(i,j) = \frac{1}{T_j} \sum_{t=1}^{T_j} \mathbb{1}[\hat{l} = i], \qquad (17)$$

where i and j are indices, respectively, for the estimated and true labels, and T_j is the number of samples in the test dataset for the index j. In addition, $\hat{l}$ is the estimate by using the tth sample batch from the test data collected at the jth label.

B. Supervised Classification

In Table III, we first present the results for the task-specific supervised training for the 5 considered methods for pose recognition, occupancy sensing, and indoor localization. When all training data in Table II are labeled (e.g., the two rows of All-Pose and All-Occ.), all methods give reasonably good performance except the CSI-only method for the pose recognition. For the pose recognition, it seems that the CSI-only measurements provide less distinct features compared
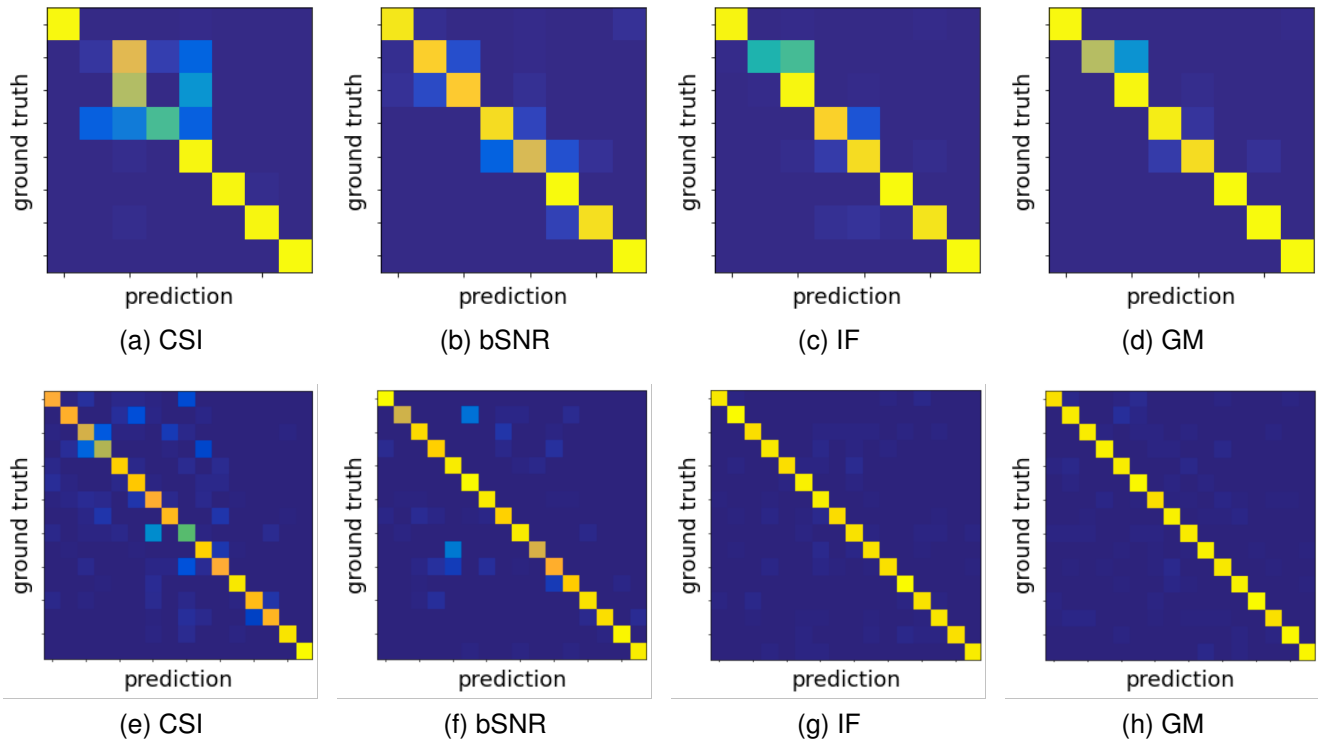
Fig. 10. Confusion matrices for pose recognition (top row) and indoor localization (bottom row) when all training data are labeled.
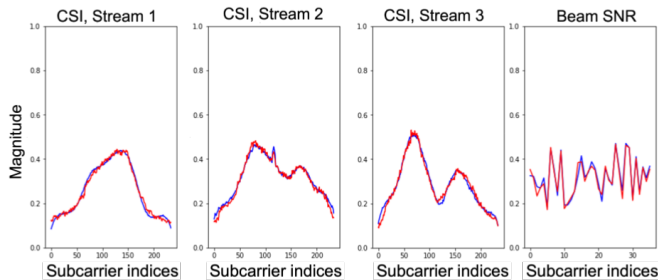

Fig. 11. Waveform comparison between the input to the autoencoder-based unsupervised fusion and its output in Fig. 6. Blue: original input waveforms; Red: reconstructed output waveforms from the same fused latent space.

TABLE IV
UNSUPERVISED FUSION+TRANSFER LEARNING: AVERAGE PROBABILITY OF SUCCESSFUL CLASSIFICATION AS A FUNCTION OF THE PERCENTAGE OF LABELED DATA R TO ALL TRAINING DATA $r + R$ (I.E., $R/(r + R)$)

$R/(r+R)$	CSI	bSNR	IF	FF	GM
All-Pose	75.4%	87.5%	86.0%	89.3%	**91.2%**
40%	72.3%	84.6%	82.6%	86.8%	**87.5%**
20%	74.0%	85.9%	82.1%	85.0%	**86.5%**
10%	75.9%	85.3%	82.0%	86.5%	**88.0%**
All-Occ.	90.1%	75.2%	84.0%	88.5%	**92.5%**
40%	89.6%	71.8%	82.2%	88.0%	**90.3%**
20%	**88.4%**	71.9%	82.9%	87.4%	86.6%
10%	78.2%	76.7%	83.8%	88.0%	**89.2%**

with the beam SNRs as one can see the classification accuracy drops from 92.3% to 74.8%. This makes the fusion between a "good" measurement and a "bad" measurement much more challenging. Indeed, it is noted that, for the pose recognition, a simple input fusion between the CSI and beam SNR reduces the classification accuracy from 92.3% for the beam SNR-only to 89.5%. With the FF, the accuracy is slightly better at 90.0% but still less than that of the beam SNR-only method. Finally, with the GM that takes into account the feature granularity correspondence, it can achieve better accuracy compared with the beam SNR-only method. Corresponding confusion matrices are shown in the top row of Fig. 10.

For the occupancy sensing application, the CSI and beam SNR measurements give similar accuracy levels around 89% which indicates the similar quality of both measurements. In this case, the simple IF gives the highest performance at

96.0% while the FF and GM provide an accuracy of 93.4% and 95.5%, respectively. Overall, it is seen that, when both measurements are of good quality, all three fusion methods can boost the performance of individual measurements by a margin of 5% − 8%.

For indoor localization, all methods give relatively good results. All three fusion methods can further improve the accuracy to more than 95% compared with 92.5% for the beam SNR-only measurement and 82.5% for the CSI-only measurements. Corresponding confusion matrices are shown in the bottom row of Fig. 10.

C. Unsupervised Fusion and Transfer Learning

Meanwhile, we provide the results for unsupervised fusion with transfer learning for the 5 considered methods for both pose and occupancy sensing. The results are summarized in Table IV. Fig. 11 illustrates the comparison between the
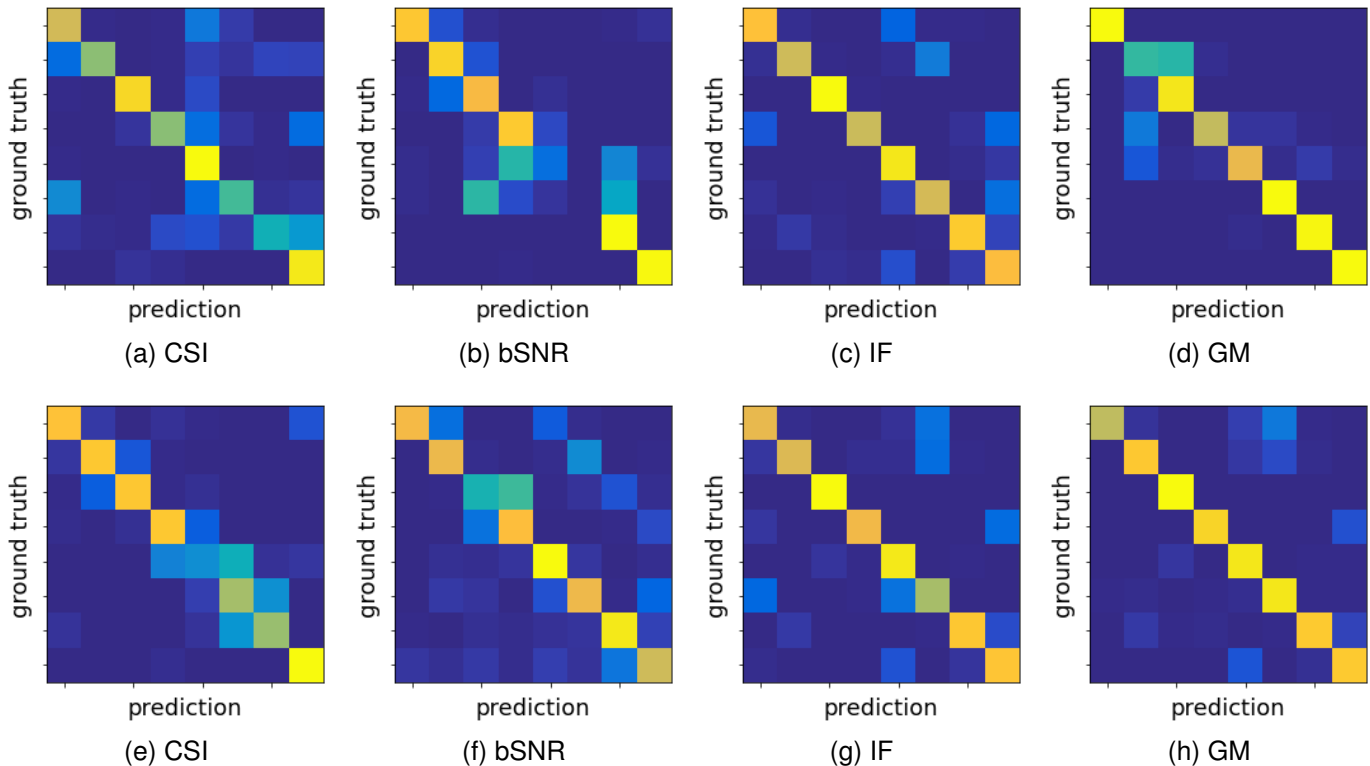
Fig. 12. Confusion matrices for occupancy sensing with supervised training (top row) and unsupervised fusion and transfer learning (bottom row) when only 10% training data are labeled.

original input waveforms, i.e., 3 spatial streams of CSI and 1 set of beam SNRs, and the reconstructed output waveforms from the same fused latent space using the autoencoder-based unsupervised fusion network of Fig. 6.

When all training data are labeled (e.g., the two rows of All-Pose and All-Occ.), the GM can provide the best classification results compared with individual measurement-based methods and the two other fusion methods. It is also interesting to compare the results with their counterparts in Table III. It is expected that the classification accuracy of the unsupervised fusion can be relatively worse than the supervised training when all training data are labeled. The results show that our model can learn well with unlabelled data, and predict better than its corresponding supervised model with limited labeled data.

D. Impact of Labeled Sample Size

We also performed a comprehensive evaluation on the impact of the percentage of the labeled data to the whole training data, namely $R/(r+R)$. In Table III and Table IV, we list the classification accuracy for all considered methods when the percentage of labeled data reduces to 40%, 20%, and 10%. It is worth noting that we draw 40%, 20% and 10% according to a uniform distribution independently. Therefore, the case of 40% may not contain all the 20% or 10% dataset.

For the supervised training, one can notice that the performance gradually decreases as the number of labeled data is less. With only 10% labeled training data, the fusion-based methods give the classification accuracy in between 73%
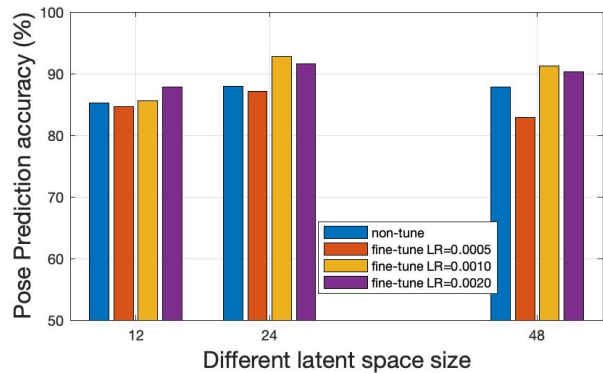

Fig. 13. Impact of latent space dimension d and the learning rate (lr) used for finetuning of the fusion block.

and 86.8% for both pose recognition and occupancy sensing. Corresponding confusion matrices for the occupancy sensing are shown in the top row of Fig. 12.

For the unsupervised fusion and transfer learning in Table IV, one can also conclude that the performance degrades as the number of labeled data reduces. However, compared with the results in Table III, the amount of performance degradation appears to be smaller in Table IV, potentially due to the use of the autoencoder-based unsupervised fusion. More noticeably, the performance can be better with only 10% of the labeled training data than that with more labeled training data for different methods. This observation is in contrast to
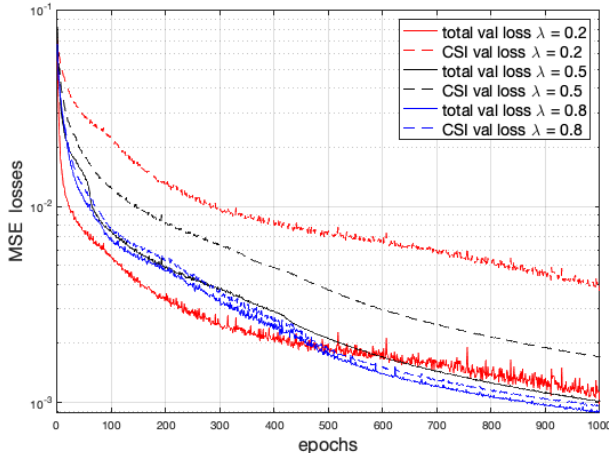
Fig. 14.   Learning error trajectories in terms of MSEs over epochs.

Table III using the supervised training where a steady trend of decreasing performance can be seen for each method when the number of labels is reduced. Our suspicion is that, since we train the fusion block in an unsupervised fusion by pooling all data from all three sensing tasks, the features corresponding to the CSI, beam SNR or combined may be smoothed out. As a result, by finetuning the network with limited labeled data, the variation of sensing performance may be larger than the supervised training in Table III. On the other hand, when the percentage of labeled training data reduces to 10%, the unsupervised fusion and transfer learning can achieve better results than their supervised learning counterparts. For a better visual comparison, we also show the confusion matrices in the bottom row of Fig. 12 by using the unsupervised fusion and transfer learning.

E. Impact of Latent Space Dimension and Finetuning Rate

Following the above discussion on the transfer learning when only 10% labeled training data are available, we further consider the impact of the latent space dimension d and the learning rate (lr) used for fine-tuning of the fusion block. Particularly, we vary the dimension of the fused feature from $d = 12$, $d = 24$ to $d = 48$ and use 4 learning rates, i.e., $lr = \{0, 5 \times 10^{-4}, 10^{-3}, 2 \times 10^{-3}\}$, for fine-tuning the fusion weight in (11). As shown in Fig. 13, the fine-tuning (with $lr > 0$) appears to always improve the classification accuracy for all three cases of latent space dimensions, although the best learning rate varies from one case to another. For instance, $lr = 2 \times 10^{-3}$ gives the best performance when $d = 12$, while $lr = 10^{-3}$ is the best for $d = 24$ and $d = 48$.

Meanwhile, if we compare the results over d, the best average performance is given by the latent space dimension $d = 24$ with $d = 48$ comes the second. $d = 12$ gives the worst performance among the three considered latent space dimensions. One plausible explanation for $d = 24$ being the best is that the latent space dimension of $d = 24$ is in between the input dimension of the beam SNR $M = 36$ and the output dimension of $N = 8$, while $d = 12$ is limited in terms of capacity and $d = 48$ is too large with respect to the input

dimension.

F. Learning Trajectories

Fig. 14 shows the learning error trajectories in terms of the validation loss as a function of epochs for the autoencoder-based unsupervised fusion. More specifically, the validation loss is computed using the validation dataset according to the weighted MSE of (16) with three choices of the weight λ. We also include the corresponding CSI-based MSE component in the figure. It is seen that, by placing larger weights on the CSI terms, the total validation loss converges to smaller values (blue curves versus red curves), while smaller weights (particularly $\lambda = 0.2$) leads to quicker MSE reduction at the beginning, e.g., for the first 400 epochs.

We then used the three pretrained encoder and fusion networks and applied the transfer learning with 10% labeled data. The classification accuracy improves from 88.4% when $\lambda = 0.2$, 89.2% when $\lambda = 0.5$ (also reported in Table IV), to 90.3% when $\lambda = 0.8$.

G. Feature Visualization

Finally, we use the tool of t-distributed Stochastic Neighbor Embedding (t-SNE) [81] to visualize high-dimensional latent space in the autoencoder of Fig. 6 (a). t-SNE converts similarities among data points to joint probabilities and tries to minimize the Kullback-Leibler divergence between the joint probabilities of the low-dimensional embedding and the high-dimensional data.

Fig. 15 shows the t-SNE results when the latent space dimension is $d = 24$. We include a) CSI-only autoencoder, b) beam SNR-only autoencoder, c) input fusion, and d) granularity matching to illustrate the cluster separation in the corresponding latent space. It is clear that the CSI measurements are less separated in the latent space as compared with the beam SNR. The best clustering separation appears to be achieved by granularity matching in Fig. 15 (d).

VII. Conclusion

This paper considered multi-band Wi-Fi sensing tasks (pose recognition, occupancy sensing and indoor localization) by matching the feature granularity between Wi-Fi channel measurements at 5 GHz and 60 GHz via a learning-based fusion block. For the three sensing tasks considered using labeled training data, we can achieve around 5% gain in accuracy on average over the best baseline methods available. To mitigate the requirement of a large amount of training data, we proposed the pre-training of a multi-band fusion network in an unsupervised fashion followed by a fine tuning of each sensing head with limited labels. It is shown that using this approach we can maintain equal performance using only 40% of the labeled data.

Looking forward, we plan to extend the multi-band Wi-Fi fusion network to regression tasks such as continuous activity monitoring and moving object (robot, human, pet) tracking. It is also critical to improve the robustness or invariance of the proposed method with respect to the variation of (human) subjects and environmental changes such as furniture displacement.
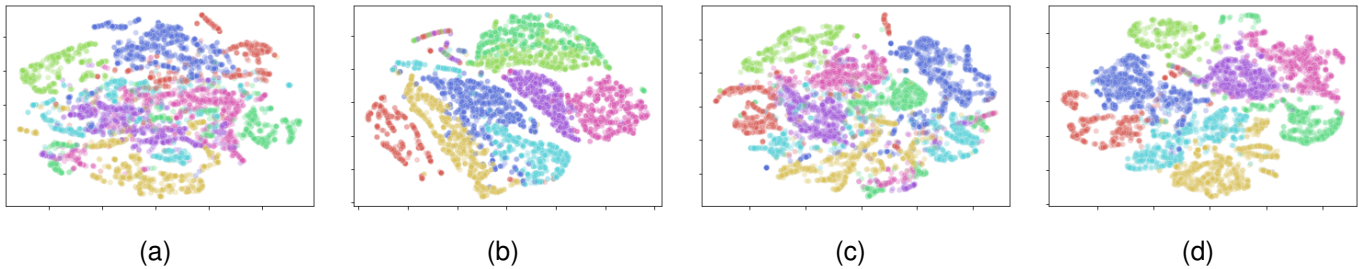
Fig. 15. Visualization of latent space clusters using t-SNE for a) CSI; b) beam SNR; c) input fusion; and d) granularity matching.

REFERENCES

[1] "Part 11: Wireless lan medium access control (MAC) and physical layer (PHY) specifications amendment: Enhancements for wireless local area network (WLAN) sensing," *IEEE Std 802.11bf*, pp. 1–628, Dec 2012.

[2] J. Xiong and K. Jamieson, "ArrayTrack: A Fine-Grained indoor location system," in *NSDI*, Lombard, IL, April 2013, pp. 71–84.

[3] D. Vasisht, S. Kumar, and D. Katabi, "Decimeter-level localization with a single WiFi access point," in *NSDI*, Santa Clara, CA, Mar. 2016, pp. 165–178.

[4] C. Chen, Y. Chen, Y. Han, H. Lai, F. Zhang, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on WiFi platforms: A multi-antenna approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 122–134, Feb 2017.

[5] C. Chen, Y. Chen, Y. Han, H. Lai, and K. J. R. Liu, "Achieving centimeter-accuracy indoor localization on wifi platforms: A frequency hopping approach," *IEEE Internet of Things Journal*, vol. 4, no. 1, pp. 111–121, Feb 2017.

[6] X. Wang, L. Gao, S. Mao, and S. Pandey, "CSI-based fingerprinting for indoor localization: A deep learning approach," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 1, pp. 763–776, Jan 2017.

[7] X. Wang, L. Gao, and S. Mao, "Biloc: Bi-modal deep learning for indoor localization with commodity 5GHz WiFi," *IEEE Access*, vol. 5, pp. 4209–4220, 2017.

[8] C. Hsieh, J. Chen, and B. Nien, "Deep learning-based indoor localization using received signal strength and channel state information," *IEEE Access*, vol. 7, pp. 33 256–33 267, 2019.

[9] H. Chen, Y. Zhang, W. Li, X. Tao, and P. Zhang, "ConFi: Convolutional neural networks based indoor Wi-Fi localization using channel state information," *IEEE Access*, vol. 5, pp. 18 066–18 074, 2017.

[10] X. Wang, L. Gao, and S. Mao, "CSI phase fingerprinting for indoor localization with a deep learning approach," *IEEE Internet of Things Journal*, vol. 3, no. 6, pp. 1113–1123, Dec 2016.

[11] C. Xiang et al., "Robust sub-meter level indoor localization - a logistic regression approach," in *ICC*, May 2019, pp. 1–6.

[12] ——, "Robust sub-meter level indoor localization with a single WiFi access point: Regression versus classification," *IEEE Access*, vol. 7, pp. 146 309–146 321, 2019.

[13] H. Zou, Y. Zhou, J. Yang, H. Jiang, L. Xie, and C. J. Spanos, "Deepsense: Device-free human activity recognition via autoencoder long-term recurrent convolutional network," in *ICC*, 2018, pp. 1–6.

[14] Y. Zhang, Y. Zheng, K. Qian, G. Zhang, Y. Liu, C. Wu, and Z. Yang, "Widar3.0: Zero-effort cross-domain gesture recognition with Wi-Fi," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2021.

[15] J. Yu, P. Wang, T. Koike-Akino, and P. V. Orlik, "Multi-modal recurrent fusion for indoor localization," in *ICASSP*, 2022, pp. 5083–5087.

[16] X. Li, "RSS-based location estimation with unknown pathloss model," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, pp. 3626–3633, December 2006.

[17] S. Mazuelas et al., "Robust indoor positioning provided by real-time RSSI values in unmodified WLAN networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 821–831, Oct 2009.

[18] M. Pajovic et al., "An unsupervised indoor localization method based on received signal strength (RSS) measurements," in *GLOBECOM*, Dec 2015, pp. 1–6.

[19] C. Liu et al., "RSS distribution-based passive localization and its application in sensor networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2883–2895, April 2016.

[20] P. Bahl and V. N. Padmanabhan, "RADAR: an in-building RF-based user location and tracking system," in *INFOCOM*, vol. 2, March 2000, pp. 775–784.

[21] T. King et al., "COMPASS: A probabilistic indoor positioning system based on 802.11 and digital compasses," in *WiNTECH*, Sep. 2006.

[22] M. Youssef and A. Agrawala, "The horus location determination system," *Wirel. Netw.*, vol. 14, no. 3, pp. 357–374, Jun. 2008.

[23] M. Brunato and R. Battiti, "Statistical learning theory for location fingerprinting in wireless LANs," *Computer Networks*, vol. 47, no. 6, pp. 825 – 845, 2005.

[24] Z. Wu et al., "Location estimation via support vector regression," *IEEE Trans. on Mobile Computing*, vol. 6, no. 3, pp. 311–321, March 2007.

[25] D. Li, B. Zhang, Z. Yao, and C. Li, "A feature scaling based k-nearest neighbor algorithm for indoor positioning system," in *GLOBECOM*, Dec 2014, pp. 436–441.

[26] S. He and S.-G. Chan, "Wi-Fi fingerprint-based indoor positioning: Recent advances and comparisons," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 1, pp. 466–490, 2016.

[27] A. Kushki, K. N. Plataniotis, and A. N. Venetsanopoulos, "Kernel-based positioning in wireless local area networks," *IEEE Transactions on Mobile Computing*, vol. 6, no. 6, pp. 689–705, June 2007.

[28] C. Figuera et al., "Nonparametric model comparison and uncertainty evaluation for signal strength indoor location," *IEEE Transactions on Mobile Computing*, vol. 8, no. 9, pp. 1250–1264, Sep. 2009.

[29] S. Fang and T. Lin, "Indoor location system based on discriminant-adaptive neural network in IEEE 802.11 environments," *IEEE Transactions on Neural Networks*, vol. 19, no. 11, pp. 1973–1978, Nov 2008.

[30] X. Lu et al., "Robust extreme learning machine with its application to indoor positioning," *IEEE Transactions on Cybernetics*, vol. 46, no. 1, pp. 194–205, Jan 2016.

[31] H. Dai, W. Ying, and J. Xu, "Multi-layer neural network for received signal strength-based indoor localisation," *IET Communications*, vol. 10, no. 6, pp. 717–723, 2016.

[32] M. T. Hoang, B. Yuen, X. Dong, T. Lu, R. Westendorp, and K. Reddy, "Recurrent neural networks for accurate RSSI indoor localization," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 639–10 651, Dec 2019.

[33] T. Fukushima, T. Murakami, H. Abeysekera, T. Fujihashi, T. Watanabe, and S. Saruwatari, "Feasibility study of practical AoA estimation using compressed CSI on commercial WLAN devices," *IEEE Access*, pp. 1–1, 2022.

[34] M. Pajovic, P. Wang, T. Koike-Akino, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial MMWave WiFi–Part I: RSS and Beam Indices," in *GLOBECOM*, Dec. 2019.

[35] P. Wang, M. Pajovic, T. Koike-Akino, H. Sun, and P. Orlik, "Fingerprinting-based indoor localization with commercial mmwave WiFi–Part II: Spatial beam SNRs," in *GLOBECOM*, Dec 2019.

[36] T. Koike-Akino, P. Wang, M. Pajovic, H. Sun, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmwave WiFi: A deep learning approach," *IEEE Access*, vol. 8, pp. 84 879–84 892, 2020.

[37] P. Wang, T. Koike-Akino, and P. V. Orlik, "Fingerprinting-based indoor localization with commercial mmwave WiFi: NLOS propagation," in *GLOBECOM*, December 2020.

[38] J. Yu, P. Wang, T. Koike-Akino, and P. V. Orlik, "Human pose and seat occupancy classification with commercial mmwave WiFi," in *IEEE GLOBECOM Workshop on Integrated Sensing and Communication (ISAC)*, December 2020.

[39] T. Koike-Akino, P. Wang, and Y. Wang, "Quantum transfer learning for Wi-Fi sensing," in *ICC*, May 2022.

[40] "IEEE standards 802.11ad-2012: Enhancements for very high throughput in the 60 GHz band," in *IEEE Standards Association*, 2012.

[41] T. Nitsche et al., "IEEE 802.11ad: directional 60 GHz communication for multi-Gigabit-per-second Wi-Fi," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 132–141, 2014.

[42] Y. Ghasempour *et al.*, "IEEE 802.11ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi," *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.

[43] D. Steinmetzer *et al.*, "Compressive millimeter-wave sector selection in off-the-shelf IEEE 802.11ad devices," in *CoNEXT*, December 2017.

[44] G. Bielsa *et al.*, "Indoor localization using commercial off-the-shelf 60 GHz access points," in *INFOCOM*, April 2018, pp. 2384–2392.

[45] S. K. Saha *et al.*, "Fast and infuriating: Performance and pitfalls of 60 GHz WLANs based on consumer-grade hardware," in *SECON*, June 2018.

[46] L. Chen, E. H. Wu, M. Jin, and G. Chen, "Intelligent fusion of Wi-Fi and inertial sensor-based positioning systems for indoor pedestrian navigation," *IEEE Sensors Journal*, vol. 14, no. 11, pp. 4034–4042, 2014.

[47] H. Xue *et al.*, "DeepFusion: A deep learning framework for the fusion of heterogeneous sensory data," in *MobiHoc*, 2019, pp. 151–160.

[48] X. Guo, N. Ansari, F. Hu, Y. Shao, N. R. Elikplim, and L. Li, "A survey on fusion-based indoor positioning," *IEEE Communications Surveys & Tutorials*, vol. 22, no. 1, pp. 566–594, 2020.

[49] M. L. Rodrigues, L. F. M. Vieira, and M. F. M. Campos, "Fingerprinting-based radio localization in indoor environments using multiple wireless technologies," in *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, 2011, pp. 1203–1207.

[50] A. U. Ahmed., R. Arablouei., F. de Hoog., B. Kusy., and R. Jurdak., "Multi-radio data fusion for indoor localization using Bluetooth and WiFi," in *PECCS*, 2019, pp. 13–24.

[51] X. Dang, X. Si, Z. Hao, and Y. Huang, "A novel passive indoor localization method by fusion CSI amplitude and phase information," *Sensors*, vol. 19, no. 4, 2019.

[52] T. F. Sanam and H. Godrich, "A multi-view discriminant learning approach for indoor localization using amplitude and phase features of CSI," *IEEE Access*, vol. 8, pp. 59 947–59 959, 2020.

[53] E. Gonultas *et al.*, "CSI-based multi-antenna and multi-point indoor positioning using probability fusion," *arXiv:2009.02798*, 2020.

[54] Z. Gao, Y. Gao, S. Wang, D. Li, and Y. Xu, "Crisloc: Reconstructable csi fingerprinting for indoor smartphone localization," *IEEE Internet of Things Journal*, vol. 8, no. 5, pp. 3422–3437, 2020.

[55] Z. Yang, Z. Zhou, and Y. Liu, "From RSSI to CSI: Indoor localization via channel response," *ACM Comput. Surv.*, vol. 46, no. 2, Dec. 2013.

[56] J. Yan, G. Qi, B. Kang, X. Wu, and H. Liu, "Extreme learning machine for accurate indoor localization using rssi fingerprints in multifloor environments," *IEEE Internet of Things Journal*, vol. 8, no. 19, pp. 14 623–14 637, 2021.

[57] D. Halperin, W. Hu, A. Sheth, and D. Wetherall, "Predictable 802.11 packet delivery from wireless channel measurements," in *SIGCOMM*, 2010, pp. 159–170.

[58] ——, "Tool release: Gathering 802.11n traces with channel state information," *SIGCOMM Comput. Commun. Rev.*, vol. 41, no. 1, Jan. 2011.

[59] Y. Xie, Z. Li, and M. Li, "Precise power delay profiling with commodity WiFi," in *MobiCom*, 2015, pp. 53–64.

[60] F. Gringoli, M. Schulz, J. Link, and M. Hollick, "Free your CSI: A channel state information extraction platform for modern Wi-Fi chipsets," in *WiNTECH*, 2019, pp. 21–28.

[61] F. Gringoli, M. Cominelli, A. Blanco, and J. Widmer, "Ax-csi: Enabling csi extraction on commercial 802.11 ax wi-fi platforms," in *Proceedings of the 15th ACM Workshop on Wireless Network Testbeds, Experimental evaluation & CHaracterization*, 2022, pp. 46–53.

[62] X. Lu *et al.*, "Robust occupancy inference with commodity WiFi," in *2016 IEEE 12th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Oct 2016, pp. 1–8.

[63] Y. Zeng, P. H. Pathak, and P. Mohapatra, "WiWho: WiFi-based person identification in smart spaces," in *2016 15th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, April 2016, pp. 1–12.

[64] D. Wu, D. Zhang, C. Xu, H. Wang, and X. Li, "Device-free WiFi human sensing: From pattern-based to model-based approaches," *IEEE Communications Magazine*, vol. 55, no. 10, pp. 91–97, Oct 2017.

[65] H. Zou *et al.*, "WiFi-based human identification via convex tensor shapelet learning," in *Proceedings of the 2018 AAAI Conference on Artificial Intelligence*, 2018, pp. 1711–1718.

[66] J. Yang, H. Zou, H. Jiang, and L. Xie, "Device-free occupant activity sensing using WiFi-enabled iot devices for smart homes," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3991–4002, Oct 2018.

[67] Y. Gu *et al.*, "Emosense: Data-driven emotion sensing via off-the-shelf WiFi devices," in *ICC*, May 2018, pp. 1–6.

[68] ——, "Besense: Leveraging WiFi channel data and computational intelligence for behavior analysis," *IEEE Computational Intelligence Magazine*, vol. 14, no. 4, pp. 31–41, Nov 2019.

[69] Z. Wang, K. Jiang, Y. Hou, W. Dou, C. Zhang, Z. Huang, and Y. Guo, "A survey on human behavior recognition using channel state information," *IEEE Access*, vol. 7, pp. 155 986–156 024, 2019.

[70] F. Wang *et al.*, "Joint activity recognition and indoor localization with WiFi fingerprints," *IEEE Access*, vol. 7, pp. 80 058–80 068, 2019.

[71] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang, "Person-in-wifi: Fine-grained person perception using WiFi," in *International Conference on Computer Vision (ICCV)*, July 2019.

[72] L. Zhang, X. Ruan, and J. Wang, "WiVi: A ubiquitous violence detection system with commercial WiFi devices," *IEEE Access*, vol. 8, pp. 6662–6672, 2020.

[73] F. Meneghello, D. Garlisi, N. D. Fabbro, I. Tinnirello, and M. Rossi, "Environment and person independent activity recognition with a commodity ieee 802.11 ac access point," *arXiv preprint arXiv:2103.09924*, 2021.

[74] W. Zhao, S. Han, R. Q. Hu, W. Meng, and Z. Jia, "Crowdsourcing and multisource fusion-based fingerprint sensing in smartphone localization," *IEEE Sensors Journal*, vol. 18, no. 8, pp. 3236–3247, April 2018.

[75] Y. Jiang, B. Liu, Z. Wang, and X. Yi, "Start from scratch: A crowdsourcing-based data fusion approach to support location-aware applications," *Sensors*, vol. 19, no. 20, 2019.

[76] D. Steinmetzer, D. Wegemer, and M. Hollick, "Talon tools: The framework for practical IEEE 802.11ad research," in *Available: https://seemoo.de/talon-tools/*, 2018.

[77] M. Kotaru, K. Joshi, D. Bharadia, and S. Katti, "Spotfi: Decimeter level localization using WiFi," in *SIGCOMM '15*, 2015, p. 269–282.

[78] Y. Zhuo, H. Zhu, H. Xue, and S. Chang, "Perceiving accurate CSI phases with commodity WiFi devices," in *INFOCOM*, 2017, pp. 1–9.

[79] H. Zhu, Y. Zhuo, Q. Liu, and S. Chang, "π-Splicer: Perceiving accurate CSI phases with commodity WiFi devices," *IEEE Transactions on Mobile Computing*, vol. 17, no. 9, pp. 2155–2165, 2018.

[80] D. Zhang, Y. Hu, Y. Chen, and B. Zeng, "Calibrating phase offsets for commodity WiFi," *IEEE Systems Journal*, vol. 14, no. 1, pp. 661–664, 2020.

[81] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of Machine Learning Research*, vol. 9, pp. 2579–2605, 2008.
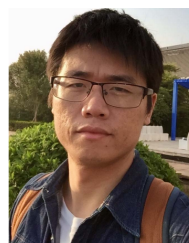
Jianyuan Yu Jianyuan Yu received the B.Sc. degree in electrical engineering from the Beijing Institute of Technology, Beijing, China, in 2014, and the M.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 2017. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Virginia Tech, Blacksburg, VA, USA. He has been a research intern with the Mitsubishi Electric Research Laboratories, Cambridge, MA, USA. His current research interests include wireless sensing, indoor localization, angle of arrival estimation and dynamic spectrum access.

Pu Wang (S'05-M'12-SM'18) received the Ph.D. degree in Electrical Engineering from the Stevens Institute of Technology, Hoboken, NJ, USA, in 2011.

He is now a Senior Principal Research Scientist at the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA, where he was an intern in 2010. His current research interests include signal processing, Bayesian inference, machine learning, and their applications to wireless sensing, wireless communications, and autonomous systems.

Dr. Wang was selected as a Distinguished Speaker of the Society of Petrophysicists and Well Log Analysts (SPWLA) in 2017 for the work "Dipole Shear Anisotropy Using Logging-While-Drilling Sonic Tools". He was a recipient of the Conrad Schlumberger Award for Technical Depth twice in 2017 and 2015 from the Schlumberger (North America) Reservoir Symposium, the IEEE Jack Neubauer Memorial Award from the IEEE Vehicular Technology Society in 2013, and the Outstanding Paper Award from the IEEE AFICON Conference in 2011. He also received the Outstanding Doctoral Thesis in Electrical Engineering Award in 2011, the Edward Peskin Award in 2011, the Francis T. Boesch Award in 2008, and the Outstanding Research Assistant Award in 2007 from the Stevens Institute of Technology. He was an Associate Editor and is now a Senior Area Editor for *IEEE Signal Processing Letters*.

Toshiaki Koike-Akino (M'05-SM'11) received the B.S. degree in electrical and electronics engineering, the M.S. and Ph.D. degrees in communications and computer engineering from Kyoto University, Kyoto, Japan, in 2002, 2003, and 2005, respectively. During 2006–2010, he was a Postdoctoral Researcher with Harvard University, Cambridge, MA, USA, and joined MERL, Cambridge, MA, USA, in 2010. He is currently a Distinguished Research Scientist at MERL, working on signal processing for communications, sensing, and computing. He was the recipient of the YRP Encouragement Award 2005, the 21st TELECOM System Technology Award, the 2008 Ericsson Young Scientist Award, the IEEE GLOBECOM'08 Best Paper Award in Wireless Communications Symposium, the 24th TELECOM System Technology Encouragement Award, and the IEEE GLOBECOM'09 Best Paper Award in Wireless Communications Symposium. He is a Fellow of Optica (formerly OSA).
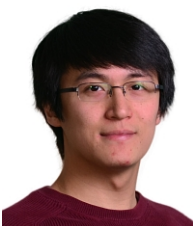
Ye Wang received his B.S. degree in Electrical and Computer Engineering from Worcester Polytechnic Institute in 2005, and his M.S. and Ph.D. degrees in Electrical and Computer Engineering from Boston University in 2009 and 2011. In 2012, he joined Mitsubishi Electric Research Laboratories, where he had also previously completed an internship in 2010. His research interests are information theory, machine learning, signal processing, communications, and data privacy/security.

Philip V. Orlik (M'97-SM'12) was born in New York, NY in 1972. He received the B.E. degree in 1994 and the M.S. degree in 1997 both from the State University of New York at Stony Brook. In 1999 he earned his Ph. D. in electrical engineering also from SUNY Stony Brook.

Since 2000 he has been with Mitsubishi Electric Research Laboratories Inc. located in Cambridge, MA where he is currently Vice President and Research Director responsible for research in the areas of signal processing, data analytics, robotics, and electronic devices. His primary research focus is on advanced wireless and wired communications, sensor/IoT networks. Other research interests include vehicular/car-to-car communications, mobility modeling, performance analysis, and queuing theory.

R. Michael Buehrer (F'16) joined Virginia Tech from Bell Labs as an Assistant Professor with the Bradley Department of Electrical and Computer Engineering in 2001. He is currently a Professor of Electrical Engineering and is the director of Wireless @ Virginia Tech, a comprehensive research group focusing on wireless communications, radar and localization. During 2009 Dr. Buehrer was a visiting researcher at the Laboratory for Telecommunication Sciences (LTS) a federal research lab which focuses on telecommunication challenges for national defense. While at LTS, his research focus was in the area of cognitive radio with a particular emphasis on statistical learning techniques.

Dr. Buehrer was named an IEEE Fellow in 2016 "for contributions to wideband signal processing in communications and geolocation." His current research interests include machine learning for wireless communications and radar, geolocation, position location networks, cognitive radio, cognitive radar, electronic warfare, dynamic spectrum sharing, communication theory, Multiple Input Multiple Output (MIMO) communications, spread spectrum, interference avoidance, and propagation modeling. His work has been funded by the National Science Foundation, the Defense Advanced Research Projects Agency, the Office of Naval Research, the Army Research Office, the Air Force Research Lab and several industrial sponsors.

Dr. Buehrer has authored or co-authored over 80 journal and approximately 250 conference papers and holds 18 patents in the area of wireless communications. In 2021 he was the co-recipient of the Vanu Bose Award for the best paper at MILCOM'21. In 2010 he was co-recipient of the Fred W. Ellersick MILCOM Award for the best paper in the unclassified technical program. He is currently an Area Editor IEEE Wireless Communications. He was formerly an associate editor for IEEE Transactions on Communications, IEEE Transactions on Vehicular Technologies, IEEE Transactions on Wireless Communications, IEEE Transactions on Signal Processing, IEEE Wireless Communications Letters, and IEEE Transactions on Education. He has also served as a guest editor for special issues of The Proceedings of the IEEE, and IEEE Transactions on Special Topics in Signal Processing. In 2003 he was named Outstanding New Assistant Professor by the Virginia Tech College of Engineering and in 2014 he received the Dean's Award for Excellence in Teaching.