Wi-Fi See It All: Generative Adversarial Network-augmented Versatile Wi-Fi Imaging

Chenning Li[1], Zheng Liu[1], Yuguang Yao[1], Zhichao Cao[1], Mi Zhang[1], Yunhao Liu[1,2]
[1]Michigan State University [2]Tsinghua University

ABSTRACT

Wi-Fi imaging has attracted significant interests due to the ubiquitous availability of Wi-Fi devices today. In this paper, we present *Wi-Fi See It All* (WiSIA), a versatile Wi-Fi imaging system built upon commercial off-the-shelf (COTS) Wi-Fi devices, which is able to simultaneously detect objects and humans, segment their boundaries, and identify them within the image plane. To achieve this, WiSIA utilizes three techniques. First, instead of constructing the image plane at the receiver side using a high-cost antenna array and complex parameter estimation, WiSIA pushes the image plane to the object side with two pairs of transceivers and 2D-IFFT. Second, WiSIA extracts the specific physical signature of the signals reflected from multiple objects to segment their boundaries. Third, WiSIA incorporates a cGAN (conditional Generative Adversarial Network) to enhance the boundary of different objects. We have implemented WiSIA using COTS Wi-Fi devices and evaluated it using a rich set of experiments. Our results demonstrate the efficacy of WiSIA. It outperforms the state-of-the-art vision-based method in dark and occlusion scenarios, demonstrating its superiority in such challenge scenarios.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Networks** → *Mobile networks.*

KEYWORDS

Wireless Sensing, Wi-Fi Imaging, Deep Learning

ACM Reference Format:
Chenning Li, Zheng Liu, Yuguang Yao, Zhichao Cao, Mi Zhang, and Yunhao Liu. 2020. Wi-Fi See It All: Generative Adversarial Network-augmented Versatile Wi-Fi Imaging. In *The 18th ACM Conference on Embedded Networked Sensor Systems (SenSys '20), November 16–19, 2020, Virtual Event, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3384419.3430725

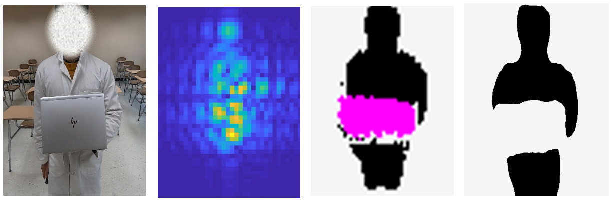

(a) RGB image (b) Wi-Fi image (c) WiSIA mask (d) R-CNN mask

Figure 1: (a) The RGB image of a person holding a laptop. (b) Wi-Fi image recovered from the wave front of Wi-Fi signals. (c) WiSIA segmentation masks: black and pink color represent the human body and the laptop, respectively. (d) Mask R-CNN segmentation masks: the laptop is not correctly segmented as its color is similar to the background.

1 INTRODUCTION

Although light is the most commonly used media for creating images, the concept of imaging itself is applicable to any kind of coherent light (i.e., electromagnetic wave) [14]. In recent years, Wi-Fi imaging – the use of Wi-Fi signals to create images of objects and humans – has attracted significant interests [8, 14, 15, 20, 26, 38, 39] due to the ubiquity of Wi-Fi signals today. In this paper, we move beyond the general Wi-Fi imaging and explore the possibility of using Wi-Fi signals radiated from commercial off-the-shelf (COTS) Wi-Fi devices to achieve *versatile* Wi-Fi imaging which is able to *detect* objects and humans, *segment* their boundaries, and *identify* them within the image plane.

In comparison to visible light, Wi-Fi signals have two unique advantages that benefit imaging. First, different from RGB pixel-wise representation of visible light imaging, Wi-Fi imaging provides a distinctive dimension for object segmentation. Specifically, state-of-the-art Wi-Fi techniques (e.g., 802.11n and later version) leverage Orthogonal Frequency Division Multiplexing (OFDM) to modulate data [21]. In OFDM, the band of a channel is divided into multiple orthogonal sub-carriers. The Channel State Information (CSI) extracted from these sub-carriers contains the changes of both amplitude and phase. This indicates how signals traverse from transmitter's antenna to the receiver's antenna through the line-of-sight (LOS) path and several non-line-of-sight (NLOS) paths reflected, scattered or refracted by surrounding objects [40, 44, 45, 47]. The amplitude and phase of the pixel-wise wave front are sensitive to the location [46], texture [49] and reflected area [42] of each object and can be used to distinguish the boundaries. Second, Wi-Fi signals are not visible so that they can be recorded in a dark environment. Wi-Fi signals are able to penetrate obstacles which visible light can not (e.g., wall, cloth, bag, luggage) [15, 18, 38]. Visible light imaging also suffers from quality degradation under poor lighting conditions and blockage of obstacles.

As an example, Figure 1(a) shows a RGB image in which a man is holding a laptop while Figure 1(b) shows the corresponding Wi-Fi image of the same scene constructed by our wave front recovery method (§4). As shown, in the RGB image, the boundary of the

SenSys '20, November 16–19, 2020, Virtual Event, Japan

ACM ISBN 978-1-4503-7590-0/20/11…$15.00
https://doi.org/10.1145/3384419.3430725

laptop is vague. This is because its color is relatively similar with the background. In such a case, as shown in Figure 1(d), the vision-based image segmentation method (e.g., Mask R-CNN [13]) cannot correctly segment the laptop. In contrast, in the Wi-Fi image, the boundary of the yellow area in the middle is much more obvious since the attenuation and locations are different between the reflection areas of the laptop and the human body. Taking this input, we can further enhance its visibility and generate a fine-grained image (§6) shown in Figure 1(c) where black and pink colors represent the human body and the laptop.

Such versatile Wi-Fi imaging has the potential to enable a wide range of applications. For example, for security checking in public areas (e.g., theater, cinema, airport, stadium), it can detect, localize, and identify the carried metallic objects [38] which, however, cannot be detected and recognized by RGB camera due to occlusion (e.g., hidden under clothes) [13]. In a further example, autonomous driving would also benefit from such versatile Wi-Fi imaging under complicated road situations in which the RGB representation of objects and background are hard to distinguish.

Design Challenges: the design of such a versatile Wi-Fi imaging system involves the following challenges.

- **High-cost Imaging.** In a camera imaging system, the light reflected or scattered by surrounding objects is projected on an image plane through the optical lens of a camera. This then renders each pixel of the image. In a Wi-Fi imaging system, however, a receiver antenna records the superposed wave fronts of Wi-Fi signals traversing along different paths so that it cannot directly obtain the amplitude and phase of the pixel-wise wave front on the image plane. Some works [14, 15, 20] need a high-cost antenna array and complex algorithm to render the image plane.
- **Binary Object Tagging.** In practice, several different kinds of objects exist simultaneously in many cases. For example, a person may bring a smartphone, wallet and drink bottles when entering a stadium. We need to distinguish all these items using one system. Due to the underutilized feature space of Wi-Fi signals, most of existing works target binary object tagging, namely they assume all of objects in a scene belong to the same category.
- **Coarse-grained Segmentation.** Due to the constraint of Wi-Fi channel bandwidth, the pixel resolution of Wi-Fi imaging is coarse-grained. For example, the Wi-Fi image shown in Figure 1(b) is far blurred than the RGB image shown in Figure 1(a). With the low quality Wi-Fi imaging, some works [26, 38] fail to support fine-grained object segmentation.

In this paper, we propose *Wi-Fi See It All* (WiSIA), a generative adversarial network-augmented versatile Wi-Fi imaging system. WiSIA offers efficient countermeasures to solve the challenges mentioned above. First, by utilizing the principle of ray tracing and the relative motion between Wi-Fi antennas and the objects, WiSIA leverages a pair of Wi-Fi transceivers to model the image plane on the object side instead of the antenna side with computation-efficient 2D IFFT. Second, WiSIA exploits the diverse polarization properties of the signals reflected by different objects to enlarge the feature space to segment the boundaries of humans and objects detected in the image plane. Third, WiSIA incorporates a conditional Generative Adversarial Networ) model to refine the boundaries.

We have implemented WiSIA using COTS Wi-Fi devices and have conducted experiments to evaluate its performance across various scenarios (e.g., clothing, environments, locations, poor light, occlusion and multi-objects). Our results show the high efficiency and accuracy of WiSIA, achieving 90% accuracy for the object profiling and tagging classification. It outperforms the state-of-the-art vision-based method [13] in dark and occlusion scenarios, demonstrating its superiority in such challenge scenarios.

In summary, our contributions are as follows:

- To the best of our knowledge, WiSIA is the first versatile Wi-Fi imaging system that is able to simultaneously detect objects and humans, segment their boundaries, and identify them within the image plane.
- To design WiSIA, We have developed novel techniques related to ray tracing, wave polarization and deep learning to enable real-time Wi-Fi imaging, multi-object detection and identification, and fine-grained segmentation enhancement.
- We implemented a prototype of WiSIA using COTS Wi-Fi devices, and evaluated its performance in various scenarios. The experimental results demonstrate the efficacy of WiSIA in comparison with vision-based approaches [13].

The rest of the paper is organized as follows. §2 describes the related work. §3 provides an overview of WiSIA. The details of the design of §2 are presented in §4, §5, and §6. §7 describes the implementation and our evaluation results. We discuss the limitation and open issues in §8. We conclude our work in §9.

2 RELATED WORK

In this section, we categorize the existing literature into general-purpose Wi-Fi imaging, and application-specific Wi-Fi imaging. We summarize the recent works within each category, and compare the most related ones to ours in Table 1.

General-Purpose Wi-Fi Imaging. To evaluate the feasibility and sensibility of computational imaging, Wision [15] first emulates the Synthetic Aperture Radar (SAR) [33] for the imaging radar system. This requires a $(8, 8)$ stationary antenna array with multiple different vantage view points. It then adopts the beamforming to extract the depth information by detecting the maximum intensity for the same direction. Due to the limited wavelength (approx. 6 cm) of Wi-Fi signals and antenna array length, Wision can only detect the target without the detailed information (e.g., shape, type). With a similar idea built upon SAR, Karanam et al. [8, 20] associate the Wi-Fi power measurement (Received Signal Strength Indicator, RSSI) with each voxel in the Markov Random Field model of the discrete imaging space. Thus rendering the 3D binary imaging using loopy belief propagation [48]. It does not require a high-cost massive antenna array, however it demands antenna scanning to formulate a virtual antenna array equivalent to 150×150. Beyond the existence of targets, it profiles the shape of the single-type object more accurately. To enhance the contrast of Wi-Fi imaging for holography [14], Holl et al. [14] extract the wave front using antenna scanning for data collection. It also employs dark-field propagation to suppress the multi-path reflection. This verifies the feasibility of holography for a single metallic cross-shaped phantom object with COTS Wi-Fi. It presents the 3D hologram of

Table 1: A comparison of state-of-the-art works on Wi-Fi Imaging.

Reference	COTS Device	#(Tx,Rx) Ant[a]	Real-time	Multi-Object Segmentation	Contrast Enhancement
Wision [15]	✗	(1,8×8)	✓	✗	✗
C. Karanam [20]	✓	(1,150×150)[b]	✗	✗	✗
P. Holl [14]	✓	(1,50×40)[c]	✗	✗	Dark-field Propagation
C. Wang [38]	✓	(2,2)	✓	✗	✗
P. Proffitt [26]	✗	(180, 180)[d]	✗	✗	✗
WiPose [18]	✓	(1,9)	✗	✗	DNN with Forward Kinetics [1]
F. Wang [39]	✓	(3,3)	✓	✗	DNN with U-net [30]
RF_Avatar [51]	✗	(4,16)	✓	✗	DNN with Attention [37]
WiSIA	✓	(1,6)	✓	✓	cGAN for Image Translation

[a] Equivalent static antenna array for antenna scanning. [b] The RX drone measures the RSSI every 2 cm for the 3m×3m area.

[c] Achieve a resolution of 4.0 cm × 7.2 cm for the 2m×3m area.

[d] The receive beam sweeps from +90° to −90° for one transmit beam angle between +90° and −90°.

the floor map in simulation. In comparison with WiSIA, the antenna scanning and antenna array significantly increase the deployment cost in practice.

To detect "suspicious" objects in bags via Wi-Fi imaging, Wang et al. [38] utilize recorded CSI measurements from COTS Wi-Fi and employ machine learning based techniques (k-NN and SVM) for material detection (e.g., metal and liquid). Similarly, Proffitt et al. [26] propose to steer the antenna automatically and further incorporate the Mask R-CNN [13] for object detection. In contrast, instead of only object detection, WiSIA targets the more challenging object segmentation tasks which classify the types of the detected objects deriving their contours as segmentation masks.

Application-Specific Wi-Fi Imaging. To segment objects via Wi-Fi imaging, effective contrast enhancement methods are required. Human-centering segmentation attracts much interest as a single-type object, such as pose estimation [2, 18, 50, 52] and profiling recovery [39, 51]. Most of these rely on deep learning techniques for skeleton and mesh recovery. RF-Pose [50] and RF-Pose3D [52] leverage the teacher-student network for cross-modality learning achieving 14 key joint estimation in 2-D and 3-D scenarios. To accurately profile the joint motion of a person, WiPose [18] expands the Body-coordinate Velocity Profile (BVP) [54] to 3D space and incorporates the forward kinematic module [1] into a Deep Neural Network (DNN) for in-situ pose estimation. Wang et al. [39] develop another DNN (e.g., U-Net) with specially designed loss functions for body segmentation and joint estimation. RF-Avatar [51] further leverages the attention-based [37] DNN to recover 3D meshes of the human body. Different from them, beyond the human, our system aims for a versatile Wi-Fi imaging system which can enable multi-objects detection, segmentation and identification.

In contrast to prior works, we focus on the task of Wi-Fi based image segmentation. The goal is to not only to capture objects and humans but also to localize their boundaries within the image plane. Moreover, WiSIA is the only versatile one that simultaneously achieves low-cost (e.g. COTS device, no antenna array) and real-time Wi-Fi imaging, multi-object segmentation and fine-grained contrast enhancement which are the must to design a practical system in many applications.

3 WISIA OVERVIEW

WiSIA is designed to achieve the following goals: 1) achieve computation-efficient Wi-Fi imaging with COTS devices; 2) support detection and segmentation of multiple types of objects in the same scene; 3) enable fine-grained segmentation masks. Figure 2 illustrates an overview of the system architecture of WiSIA. WiSIA utilizes two pairs of transceivers with two receivers (e.g., three antennas pointing to three orthogonal directions) sharing the same transmitter antenna to record Wi-Fi CSI. This imitates light and camera imaging systems. The output of WiSIA is the segmentation masks and identities of the objects in the scene. In the middle segment of the figure, WiSIA consists of a cascade of three core components: *wave front construction* module, *pixel-wise illumination* module, and *segmentation refinement* module. We briefly describe the challenges and countermeasures of each module as follows.

Wave Front Construction. Resembling a light and camera imaging system, WiSIA should enable the encoding of pixel-wise spatial information from the Wi-Fi radiation bouncing from humans and objects, namely the wave front [14]. Wi-Fi radiation, however, retains the intensity and direction information inherently in CSI measurements while eliminating their pixel-wise spatial distribution by superposing at the Wi-Fi receiver [15]. The multi-path effect further aggravates its entanglements since radiation scattered from surrounding objects can distort direct reflected ones. It is challenging to reconstruct the wave front of the image plane efficiently by avoiding the cumbersome computation of Lagrange multiplier estimation while alleviating the multi-path effect of surroundings.

Given the raw CSI measurements collected from the Wi-Fi devices, WiSIA extracts the dominant Wi-Fi radiation which traverses the direct reflected path of the surrounding objects. It then constructs the wave front by tracing Wi-Fi radiation from different directions back to an image plane at the object side. This is achieved by associating continuous temporal snapshots of channels captured by multiple Wi-Fi sub-carriers to the pixel-wise wave front of Wi-Fi radiation using 2D IFFT. Compared with SAR-based approaches in Table 1, WiSIA only needs 1×6 antenna pairs and can be operational in real-time.

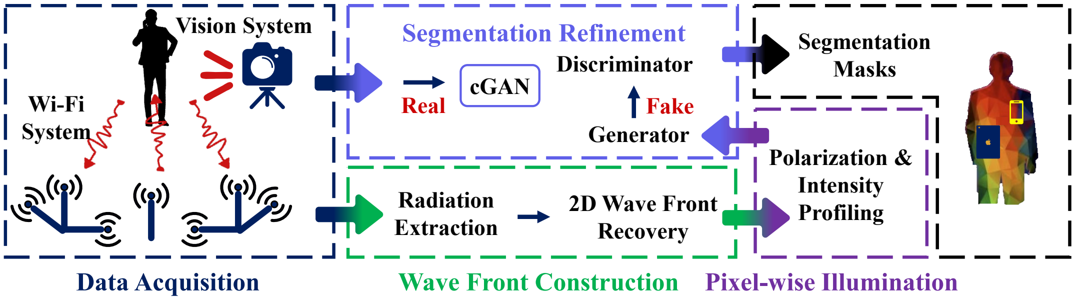


Figure 2: An overview of the system architecture of WiSIA.

Pixel-wise Illumination. Once the wave front is constructed, it is challenging to make each pixel sufficiently distinct for object tagging when multiple kinds of objects simultaneously exist. WiSIA designs physical signatures that encode the properties of intensity and polarization of Wi-Fi radiation to "illuminate" each pixel. WiSIA also incorporates a content-aware scheme borrowed from the computer graphics literature [5] to improve the contrast of physical signatures in the wave front. This enhances the reliability of the weakly reflected Wi-Fi radiation.

Segmentation Refinement. The constructed image plane only achieves the coarse-grained multi-object segmentation with low spatial resolution due to the narrow bandwidth of Wi-Fi. Enhancing the object boundaries of Wi-Fi imaging is critical. WiSIA incorporates the cGAN into the process improving the spatial resolution with the limited vantage views of Wi-Fi receivers. This refines the constructed wave front with pixel-wise physical signature. With cGAN, WiSIA induces the segmentation mask generator to fool the adversarial discriminator while keeping close to the ground truth image with designed loss function. This generates high-quality segmentation masks as the system output.

4 WAVE FRONT CONSTRUCTION

In this section, we present our design for constructing the wave front (e.g., amplitude and phase) of each pixel on an image plane which is put at the object side. This module consists of two tasks. First, we extract the dominant Wi-Fi radiation that is directly reflected by the surrounding objects from the raw CSI measurements. Then, we utilize the ray tracing and 2D IFFT to calculate the wave front for each pixel on the image plane. For both tasks, we require a slight relative motion between the antennas of Wi-Fi transceivers and the objects in the scene. There are no special requirements for the moving trajectory and velocity. Hence, we can easily implement the relative motion in practice. For example, the antennas of Wi-Fi transceivers can be fixed on a sliding-table which is continuously moving front and back when the objects are static. Moreover, the chest fluctuation of human breath, body sway and regular walking can be also counted when the antennas are static.

4.1 Dominant Wi-Fi Radiation Extraction

Upon receiving the raw CSI measurements, the first step is to extract the dominant Wi-Fi radiation bouncing off the surrounding objects. Besides the dominant Wi-Fi radiation we are interested in, the raw CSI measurements contain the signals traversing from the LOS path along the transmitter antenna to receiver antennas and burst noise brought by low-cost COTS devices [28, 29]. Given the relative motion between Wi-Fi transceivers and the objects, the observed frequency of the dominant Wi-Fi radiation will be continuously changing due to the Doppler Effect. As a result, the CSI power (i.e., conjugate multiplication of CSI) of the dominant Wi-Fi radiation is changing as well. Regarding the frequency of the CSI energy changing, in comparison with the dominant Wi-Fi radiation, burst noises bring generally higher frequency while the LOS signals incur generally lower frequency. Thus we can suppress the interference signals by adopting a band-pass filter. Specifically, we apply a band-pass Butter-worth filter to the CSI power series of all subcarriers and keep the CSI phase components are not distorted. We empirically set its cutting off frequency as 0.5 Hz and 80 Hz [28] respectively, delivering the dominant sanitized Wi-Fi radiations.

4.2 Wave Front Construction via Radiation Tracing

With the extracted information of the dominant Wi-Fi radiation, resembling the pixel-wise photograph of the light field, WiSIA constructs the wave front of the image plane by tracing Wi-Fi radiations superposed at the receiver back to the image plane, retrieving the intensity and phase information for each pixel.

Existing solutions can be generally categorized into two approaches. In the first approach, a massive antenna array [6, 15] or an antenna scanning method [14, 26] is adopted to observe objects from multiple vantage views. It then recovers the spatial diversity using 2D Fourier transform, which is either high-cost or cumbersome in deployment. In the second approach, a numerical fitting method is proposed with limited antennas, rendering it as an optimization problem which is similar with the non-convex 2D Non-Uniform Discrete Fourier transform [36, 49] or a method of Lagrange multipliers [18, 54]. However, the under-constrained question suffers from the local minimum and high computation complexity. To avoid the shortcomings of these two approaches, WiSIA instead leverages the multi-dimensional information of CSI $H(t, f)$ across the continuous packets t and multiple sub-carriers f endowed by OFDM modulation.

As shown in Figure 3(a), we take a scene where a person is standing on the ground as an example to illustrate the coordinate system and the relationship among Wi-Fi transceivers, the image plane and

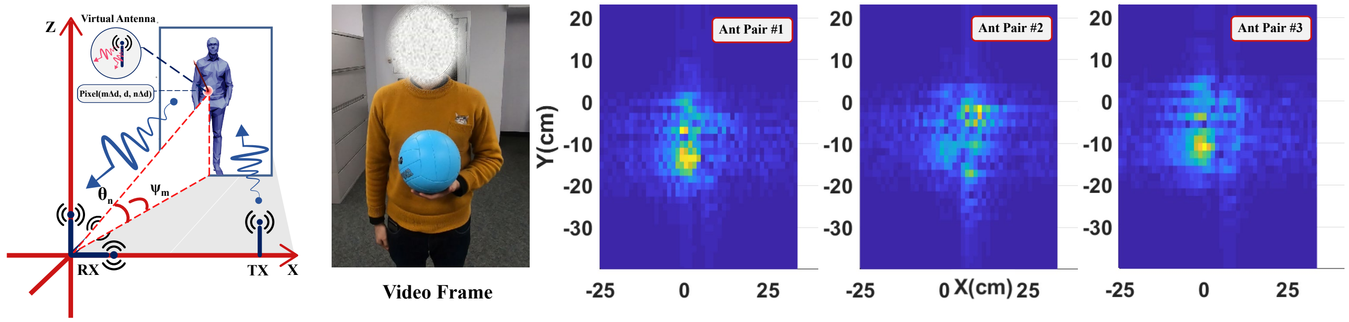


Figure 3: (a) Constructing the wave front by tracing Wi-Fi radiations back to virtual antennas at each pixel on the image plane. (b-e) The coarse-grained wave front observed by three different pairs of transceiver antennas.

the person. We define the 3D coordinate with receiver antennas as the origin $(0, 0, 0)$, in which the transceiver plane (x-z plane) is vertical to the ground. The image plane is parallel with the transceiver plane. $M \times N$ pixels are uniformly distributed with the spacing interval Δd along the x-axis and z-axis, respectively. The person body consists of many small areas (called *reflector*) which have different depth along the y-axis and reflect signals from transmitter's antenna to receiver's antennas. For a reflector of the person denoted as $r_{m,n}$, its coordinate vector is $\vec{s}(r_{m,n}) = (m\Delta d, d(m,n), n\Delta d)$ corresponding to the direction of azimuth angle $\Psi_m = arctan(m\Delta d/d(m,n))$, elevation angle $\theta_n = arctan(n\Delta d/\sqrt{d(m,n)^2 + (m\Delta d)^2})$ and the depth $d(m,n)$. Then, we correlate the reflector $r_{m,n}$ with the pixel whose coordinate is (m,n) on the image plane. Thus the wave front of each pixel corresponds to the wave front of the reflector. In this way, we can keep the property of the reflector as much as possible to increase the pixel discrimination for imaging generation.

To recover the wave front of each pixel, we trace Wi-Fi radiation from the receiver back to each corresponding reflector on the body of the target person. For the reflector $r_{m,n}$, we set a virtual antenna which is equivalent to a new transmitter with intensity and phase information of the corresponding pixel in the wave front indicated as $h_{m,n}$. By doing this, we can compute the phase shift of Wi-Fi signals induced by the propagation path from the reflector to the receiver's antennas whose distance is $|\vec{s}(r_{m,n})|$. Then we can derive the amplitude and phase information of the corresponding wave front that received by the receiver's antennas $(0, 0, 0)$ after signal propagation, denoted as $h^{Rx}_{m,n}$. From the basic physics principle, the $h^{Rx}_{m,n}$ can be formulated as a complex expression related to $|\vec{s}(r_{m,n})|$.

$$\vec{s}(r_{m,n}) = (d(m,n) \cdot tan(\Psi_m), d(m,n), \frac{d(m,n) \cdot tan(\theta_n)}{cos(\Psi_m)}) \tag{1}$$

$$h^{Rx}_{m,n} = \alpha h_{m,n} exp(-j\frac{2\pi|\vec{s}(r_{m,n})|}{\lambda}) \tag{2}$$

Where α and λ indicate the amplitude attenuation factor and the wavelength of the Wi-Fi radiations, respectively.

Our goal is to recover the wave fronts $\mathbb{H}_{M,N} \triangleq [h_{m,n}]_{M,N}$ of all pixels $\forall m \in [1, M]$ and $\forall n \in [1, N]$ on the image plane by tracing Wi-Fi radiations. To represent the phase shift for the wave front of each pixel (m,n) induced by the wave propagation, we first define the basis function $B_{m,n}$ related to each reflector $r_{m,n}$ through the $\vec{s}(r_{m,n})$ using Equation (1) as follows:

$$B_{m,n} = exp(-j2\pi|\vec{s}(r_{m,n})|/\lambda) \tag{3}$$

Then, we correlate the dominant CSI measurement $H(t,f)$, at packet t and subcarrier frequency f for a receiver antenna with the corresponding basis function as Equation (4). Specifically, we have the following summation in terms of $H(t,f)$ due to the superposition of Wi-Fi radiations from the reflectors that corresponding to the pixels on the image plane:

$$H(t,f) = \sum_{m=1}^{M} \sum_{n=1}^{N} h_{m,n} B_{m,n} \tag{4}$$

Note that here we assume the amplitude attenuation α of Equation 1 is approximately consistent for every pixel in the image plane and can be unified in $H(t,f)$.

Now, dominant Wi-Fi radiation $H(t,f)$ is known. To recover $h_{m,n}$ of every pixel, we find the summation computation of Wi-Fi radiations in Equation (4) is similar to a 2D Inverse Fast Fourier Transformation (2D IFFT). Thus we utilize the relative motion along the y-axis between the transceiver and the objects to transform it into the 2D IFFT problem. Mathematically, given the velocity of the relative motion along the y-axis is v_y, the displacement can be denoted as $\vec{d_s}(m,n) = (0, v_y\Delta t, 0)$ for the targeting reflector $r_{m,n}$ with the coordinate vector $(m\Delta d, d(m,n), n\Delta d)$. Note that each pixel has corresponding initial distance $|\vec{s}(r_{m,n})|_{t_0}$ along the direction $\vec{s}(r_{m,n})/||\vec{s}(r_{m,n})||$. For different pixels, the depth $d(m,n)$ in $\vec{s}(r_{m,n})$ is a variable parameter. Thus we can transform the basic function $B_{m,n}$ to represent the phase shift which is changing with time as follows:

$$\begin{aligned} B_{m,n} &= exp(-j2\pi(|\vec{s}(r_{m,n})|_{t_0} + \vec{d_s} \cdot \frac{\vec{s}(r_{m,n})}{||\vec{s}(r_{m,n})||})/\lambda) \\ &= exp(-j2\pi|\vec{s}(r_{m,n})|_{t_0}/\lambda) \cdot exp(-j2\pi v_y cos\theta cos\Psi t/\lambda) \end{aligned} \tag{5}$$

Note that we only consider the variance of t while replacing λ in the second term using λ_c at the central frequency f_c [6, 7]. Therefore, Equation (4) can be transformed as follows:

$$H(t,f) = \sum_{m=1}^{M}\sum_{n=1}^{N} h_{m,n}exp(-j2\pi(f\cdot\frac{|\vec{s}(r_{m,n})|_{t_0}}{c} + t\cdot\frac{v_y cos\theta cos\Psi}{\lambda_c})) \tag{6}$$

Specifically, Equation (7) shows a standard 2-D FFT given a two-dimensional input signal $f(m,n)$ and output signal $F(a,b)$. In our problem, the packet t and the frequency of subcarrie f correspond to variable a and b. By substituting m and n for variables $M|\vec{s}(r_{m,n})|_{t_0}/c$ and $Nv_y cos\theta cos\Psi/\lambda_c$, namely the depth and angle of each reflector, we utilize the 2D IFFT to recover the wave front $\mathbb{H}_{M,N}$ by associating each reflector $r_{m,n}$ with known $H(t,f)$ across packets and subcarriers.

$$F(a,b) = \sum_{m=0}^{M}\sum_{n=0}^{N} f(m,n)exp(-j2\pi(a\cdot\frac{m}{M} + b\cdot\frac{n}{N})) \tag{7}$$

To illustrate the recovered wave fronts with 2D IFFT, we represent the power distribution for reconstructed results in Figure 3(b-e), each of which is plotted using the data from two mirror antennas of the symmetric receivers as shown in our prototype implementation in §5.2. And it can only capture the coarse-grained contour of human and existence of the volleyball, making it impossible to be interpreted in comparison with RGB images.

5 PIXEL-WISE ILLUMINATION

Upon getting the wave front of each pixel in the image plane, rather than only taking the signal intensity to render each pixel, we need to extract representative features so that multiple objects can be naturally distinguished and tagged. WiSIA utilizes signal polarization as a physical signature to illuminate each pixel in the image plane for object tagging. In a conceptual sense, the meaning of the polarization based physical signature to WiSIA imaging is that of the brightness to a RGB photograph. We first introduce the background on polarization of electromagnetic waves (§5.1). Then dedicated approaches are designed to extract the physical signature for each pixel and verify its feasibility to distinguish multiple objects in a scene (§5.2). To bootstrap the pixel-wise resolution for physical signature profiling, we borrow the content-aware scheme from the computer graphics to enhance the profiling contrast, making it more distinguishable (§5.3).

5.1 Background on Polarization

Resembling light, Wi-Fi radiations propagate as the electromagnetic wave while the electric field oscillates perpendicularly to the direction of propagation. For an unpolarized wave, its electric field vectors vibrate in all planes perpendicular to the direction of propagation. Moreover, the unpolarized wave can be converted to a polarized wave if the electric field vectors are restricted to a single plane by filtration of the beam with specialized materials [25]. The polarized wave can be decomposed with two orthogonal linear polarization as p- and s-polarization, in which the p-polarized wave has an electric field polarized parallel to the plane of incidence, while s-polarized light is perpendicular to this plane. Different polarized waves can be produced while bouncing off various surfaces of materials and textures, such as scattering off a smooth non-metallic surface, a smooth metallic surface, and a rough surface [34, 49].

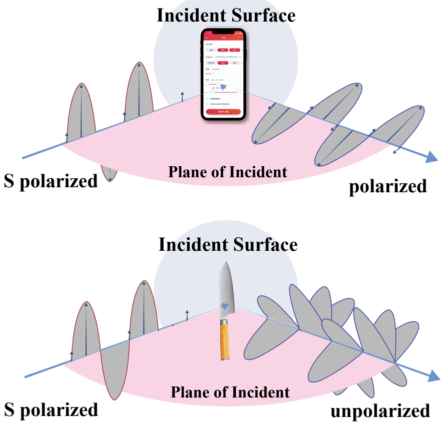


Figure 4: Theoretical analysis on physical signature: differently polarized waves determined by the material and texture of the reflected surface.

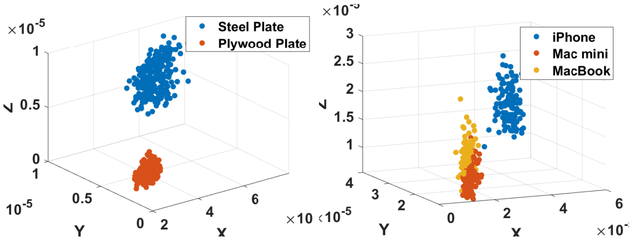
Figure 4 shows two examples of the polarization change when the signal is reflected by different objects. In the top figure, the incident wave is pure s-polarized. When the wave is reflected by the screen glass of a smartphone, the reflected wave is polarized as well, but the polarization direction contains both s- and p- parts. In the other example shown in the bottom figure, the s-polarized incident wave becomes an unpolarized wave after the reflection of the steel material of a knife. Hence, according to the polarization property of the reflected wave, we can distinguish different objects appearing in the same scene.

Mathematically, in Figure 4, the incident wave is s-polarized and can be depicted as $\vec{E}_s^{(in)}(t) = Ae^{j\omega t}\hat{e}_s^{(in)}$, where $\hat{e}_s^{(in)}$ is the unit s-polarization vector along the incident direction. The reflected wave can be further denoted as $\vec{E}^{(ref)}(t) = Ae^{j\omega t}(r_{ss}\hat{e}_s^{(ref)} + r_{sp}\hat{e}_p^{(ref)})$, where $\hat{e}_s^{(ref)}$ and $\hat{e}_p^{(ref)}$ are the unit s-polarization and p-polarization vectors along the reflection direction. Specifically, for a polarized wave, its polarization can be completely converted to the cross direction after reflection with different materials. WiSIA aims to measure the specific parameters r_{ss} and r_{sp}, delivering a physical signature influenced by the material, texture, geometric and areas of reflectors' surfaces.

5.2 Physical Signature Profiling

To associate each pixel with the corresponding object, we design a synthetic physical signature based on the polarization and signal power properties of the Wi-Fi radiations bouncing off different surfaces of objects. The underlying principle is that objects with various materials, textures, and reflected areas to the transceiver plane can induce distinguishable variances in polarization and magnitude, respectively.

To obtain the polarization of the reflected waves, our receiver consists of three mutually perpendicular linearly-polarized antennas which are utilized to monitor the Wi-Fi radiations transmitted



(a) The impact of material and texture (b) The impact of reflected areas

Figure 5: Experimental observations: distinguishable 3D point cloud in the feature space for (a). the metallic and plywood plate of the same size. (b). daily items of the same material and textures with various reflected areas.

by a vertically polarized antenna. Two receivers are symmetrically put at the opposite side of the transmitter, as shown in Figure 7 to increase the detectable reflection area of the objects. Then we collect $100ms$ packets continuously and derive the wave fronts on the image plane with the CSI measurements from each pair of symmetric antennas of the two receivers. Since we have three pairs of symmetric antennas towards different directions, we can recover three different wave fronts of the same image plane that indicate the polarization property at different directions, which form a similar feature space in comparison to the RGB channels of photographs.

To verify the feasibility of material recognition using polarization, we compute the mean of pixel-related weighted amplitudes for each of the three wave fronts of the image plane, delivering a 3D feature point scattered in the feature space. We further render the 3D point cloud by calculating multiple 3D feature points for each testing object (e.g. a metallic plate and a plywood plate of the same size $2' \times 2' \times .0625''$) with continuous temporal segments, shown in Figure 5(a). On the one hand, feature points for each object are clustered and stay consistent across continuous segments. On the other hand, we can distinguish the two clusters of feature points readily, rendering the impact of different material and texture on the polarization of the wave front.

The second physical property is the magnitude of Wi-Fi radiations. Since it is demonstrated the power distribution of spectrograms changes as reflection areas vary across different objects. Given the same area on the image plane, the accumulated power of the corresponding wave fronts increases as the reflections areas of the object are increasing. It can also be applied in the power distribution of the recovered wave fronts of the image plane. To evaluate the impact of the reflected area, we conduct the preliminary using three objects composed of the same material (mainly aluminum) and textures, including Macbook Pro ($13.75'' \times 9.48'' \times .61''$), Mac mini ($7.7'' \times 7.7'' \times 1.4''$) and iPhone 8 ($5.48'' \times 2.65'' \times 0.29''$). Illustrated in Figure 5(b), the 3D point cloud shows each cluster of items stays consistent while keeps away from each other, making it reliable as a physical signature for object tagging. Nevertheless, an overlapping area between the clusters of Mac Mini and MacBook appears in the feature space. This is mainly attributed to the coarse-grained resolution of the wave front and can be resolved by our cGAN model that refines the boundary of object segmentation masks (§6).

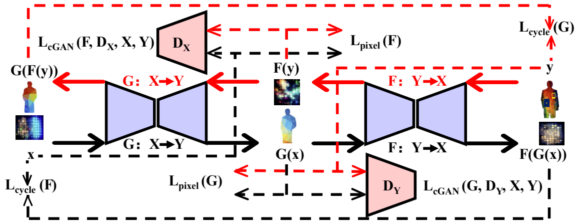


Figure 6: The cGAN to refine the segmentation masks by enhancing the pixel-wise contrast of the constructed wave front, including the loss propagation (Dashed line) for domains of $X \rightarrow G(X) \rightarrow F(G(X))$ (Black) and $Y \rightarrow F(Y) \rightarrow G(F(Y))$ (Red).

5.3 Profiling Enhancement and Optimization

To improve the resolution and sensing sensibility of the wave fronts in the image plane, it is required to enhance the profiling contrast of the physical signature. Referring to the idea of content-aware scheme in computer graphics [5], the profiling contrast of the wave front can be increased directly by computing the derivative of the wave front, rendering distinguishable boundaries of various targets readily. We resort to the differential operation for computation efficiency. Specifically, a 2D Sobel operator G [19] is applied to the recovered wave front of the image plane. Thus the final feature map $\mathbb{F}_{M,N}$ can be derived concerning the wave front $\mathbb{H}_{M,N}$ as follows:

$$\frac{\partial^2 \mathbb{H}_{M,N}}{\partial M\, \partial N} \approx \mathbb{F}_{M,N} = Convolution\ (G, \mathbb{H}_{M,N}) \tag{8}$$

To further enhance the profiling contrast, we employ a Gaussian filter [26] to focus on the central area of the image plane by filtering out the power noise at the marginal area.

6 SEGMENTATION REFINEMENT

In this section, by taking the wave front scratches of the image plane and the segmentation masks of different objects extracted from RGB images as input, we utilize a cGAN model to learn a model that can generate the fine-grained masks for each object in a scene. Basically, it can be formulated as an image translation problem by computing the mapping function from the data domain X of wave front scratches to the domain of Y segmentation masks. Illustrated in Figure 6, we design two pairs of adversarial network with respective generator and discriminator, one is the $G : X \longrightarrow Y$ and D_Y and the other is $F : Y \longrightarrow X$ and D_X. Further, we design three loss functions for these two mapping function learning, including the L_{cGAN}, L_{pixel} and L_{cycle} which are introduced in detail later.

6.1 Adversarial Learning on Wave Front

Upon receiving the derivative of the illuminated wave front in the image plane as the feature map $F_{M,N}$, we can further refine the scratch of the wave front and feed it into a generative adversarial model which can generate much finer-grained segmentation masks. Generally, it can be broadly described as the image-to-image translation, converting an image from one representation of a given domain, X [16, 55], to another, Y. For example, mapping from sketches to photographs [31], or from wave front scratches to segmentation masks here. Our challenge is to learn a mapping function from the extracted feature maps, $\{x_i\}_{i=1}^{T} \triangleq \{F_{M,N}\}_{i=1}^{T}$, in the wave front domain to ground truth segmentation masks extracted from RGB images, $\{y_i\}_{i=1}^{T} \triangleq \{P_{M,N}\}_{i=1}^{T}$, for T continuous samples. Supposing the data distribution $x \sim p_{data}(f) \in X$, $y \sim P_{data}(y) \in Y$, a mapping function $G : X \rightarrow Y$ is required for the translation from the wave front scratches to ground truth segmentation masks. And we build our model on the cGAN [10] since the adversarial loss induced by the discriminator model can bootstrap the learning ability of the generator model for the internal representations of data iteratively, making the generated images more realistic.

Illustrated in Figure 6, upon receiving the derived wave front scratch, we feed it into the generator model, which is composed of the encoder-decoder model. For the mapping from wave front scratches to ground truth segmentation masks $G : X \rightarrow Y$ and its discriminator D_Y, we introduce the adversarial loss as follows [10]

$$\begin{aligned} L_{cGAN}(G, D_Y, X, Y) =& \mathbb{E}_{y \sim p_{data}(y)}[logD_Y(y)] \\ &+ \mathbb{E}_{x \sim p_{data}(x)}[1 - logD_Y(G(x))] \end{aligned} \tag{9}$$

where G tries to fool the D_Y by generating fake segmentation masks that look similar to ones in domain Y, while D_Y is required to distinguish the fake images $G(x)$ and real y, leading to the adversarial learning by optimizing $min_G max_{D_Y} L_{cGAN}(G, D_Y, X, Y)$. For the bi-directional adversarial learning from ground truth segmentation masks to wave front scratches, we have the equivalent optimization problem for $F : Y \rightarrow X$ and its discriminator D_X as $min_F max_{D_X} L_{cGAN}(F, D_X, X, Y)$.

6.2 Network Architecture Design

Nevertheless, we cannot apply the general generator and discriminator directly. Since it is facing two challenges as follows.

- **Coarse-grained resolution of Wi-Fi radiations.** Limited by the narrow bandwidth of CSI measurements and the limited number of subcarriers, the recovered wave front of the image plane is coarse-grained in spatial resolution, making it difficult for recognizing and locating objects with fine-grained segmentation boundary.
- **Cumbersome deployment cost.** It is essential for the data-thirsty deep neural network to collect massive data, especially paired dataset required for supervised learning, where data pairs $x_i, y_{i}{}_{i=1}^{T}$ are available [9, 31]. Thus the rigid time synchronization is required for the wave front scratches and ground truth segmentation masks for the paired dataset. However, to accurately synchronize the Wi-Fi transceiver with the camera is hard to achieve in real deployment.

To alleviate the impact of the non-synchronization of recovered wave front scratches and the ground truth segmentation masks, we explore the significant influence of the unpaired data in coarse-grained time synchronization. And it demonstrates the translation performance drops since the unpaired dataset cannot induce an individual input x to match a sole output y in a meaningful way [55]. Mathematically, it cannot guarantee an efficient bijection between the scratches and ground truth images and often leads to the known problem of mode collapse, where all inputs map to the few outputs and the optimization procedures fails [11]. Thus we borrow the idea of the cycle-gan [55] and design a cycle-consistency loss to relax the rigid requirement of time synchronization. Differently, limited by the coarse-grained resolution of wave front scratches, we adopt the cross-entropy loss to transform the regression problems for the whole image into a pixel-wise classification problem. Given the $L_{cycle}(G)$ and $L_{cycle}(F)$, it can be formulated as follows:

$$\begin{aligned} L_{cycle}(G, F) =& \mathbb{E}_{y \sim p_{data}(y)}[CrossEntropy(G(F(y)), y)] \\ &+ \mathbb{E}_{x \sim p_{data}(x)}[||(F(G(x)) - x)||_1] \end{aligned} \tag{10}$$

Experimental evaluation demonstrates that it is difficult and inefficient to learn the mapping from wave front scratches to ground truth segmentation masks with the enhancement of the L_{cGAN} and L_{cycle}. And the generator should not only fool the discriminator but also learn from the ground truth, especially at the pixel scale. Thus we adopt the pixel-to-pixel classification loss to bootstrap the learning aggressively. Instead of using the L_1 norm for image-scale regression, we apply the cross-entropy loss in the domain Y at the pixel scale:

$$\begin{aligned} L_{pixel}(G, F) =& \mathbb{E}_{x \sim p_{data}(x)}[CrossEntropy(G(x), y)] \\ &+ \mathbb{E}_{y \sim p_{data}(y)}[||(F(y) - x)||_1] \end{aligned} \tag{11}$$

Our final objective is the combination of losses below:

$$\begin{aligned} L(G, F, D_X, D_Y) =& \gamma L_{cGAN}(G, D_Y, X, Y) + \gamma L_{cGAN}(F, D_X, X, Y) \\ &+ \lambda L_{cycle}(G, F) + \beta L_{pixel}(G, F) \end{aligned} \tag{12}$$

where γ, λ and β are the weights for each loss, in which λ can alleviate the impact of the non-synchronized wave front scratches and ground truth images while β is utilized for encouraging the generator to be near the ground truth at the pixel scale.

Note that the whole network, shown in Figure [10], is designed to optimize the problem. And we evaluate the ablations of the full objective to verify the effectiveness of designed losses (Sec. §7.4):

$$G^*, F^* = \underset{G,F}{\operatorname{argmin}} \underset{D_X, D_Y}{\operatorname{argmax}} L(G, F, D_X, D_Y) \tag{13}$$

7 EVALUATION

7.1 Evaluation Methodology

Implementation: We prototype WiSIA using three mini-desktops equipped with commercial linear polarized antennas (Ettus Vert2450). Each mini-desktop has an Intel 5300 wireless NIC, and runs the Linux 802.11n CSI tool [12] to collect CSI measurements. To alleviate interference, we select the channel 165 at 5.825 GHz [53]. The packet rate is set to 1000 packets per second and the available bandwidth is 20 MHz. Note that we set the three receiving antennas of the same receiver orthogonal to each other to capture polarized Wi-Fi radiations [49], shown in Figure 7. We implemented the core components of WiSIA using MATLAB/PyTorch [24] in the local laptop, and ran the evaluation tasks on the remote server. To achieve the image segmentation in real-time, we set the sampling rate for generating segmentation masks at 20 frames per second. The size of the reflected area is denoted as (height × width) for common objects. And we apply the centimeter (cm) as a unit without specific denotation.

Experimental Setup: Unless stated otherwise, our evaluation across experiments are conducted in settings as described below.

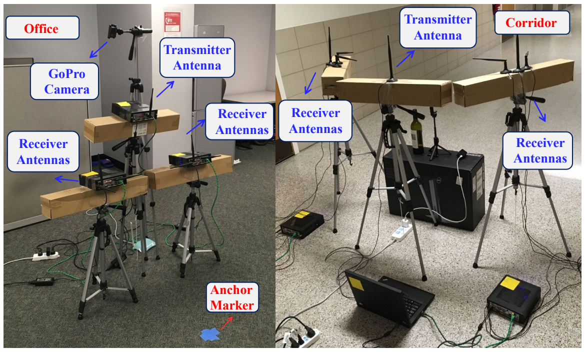


Figure 7: Illustration of our experimental setup.

We have evaluated WiSIA in three different scenarios, including (1) a 25 sq.m classroom, (2) a 15 sq.m office, and (3) a long corridor spanning 4m in width, shown in Figure 9(a). And our evaluation consists of three parts with 16 different setups, including image segmentation of the target person (1) holding 5 various objects for in-domain and cross-domain testing [54], such as the volleyball, vase, laptop, wine bottle, and iPhone, (2) holding objects in 5 different locations, covering both sides of the body and various distances to the transceivers, (3) under 6 special conditions such as poor lights, occlusions, and multi-object/person scenarios. For each setup, we collect 240 images as our dataset. We train our model once with 200 images and test it using the remaining 40 images. We further denote the material, texture and reflected area of the collected daily objects to verify the effectiveness of our feature extraction module, shown in Table 2.

Note that we evaluate the cross-domain performance of our system across instances, clothes and environments, in which we train our model using the dataset from one domain (e.g., the office) while testing its performance in new domains (e.g., classroom and corridor) directly. And the cross-domain ability explores the resistance of our system to the dynamics of wireless signals, reducing the deployment cost (e.g., model retraining and data recollection) significantly while being deployed in new domains.

For experimental convenience, all imaging objects are moving to provide a relative motion towards transceivers at the distance of 70cm (§4.2). For example, the target person is instructed to take a deep breath and we put multiple objects on the cart which can be moved relatively towards the transceivers for the multi-object testing. Note that all experiments are approved by our IRB. We further truncate the collected Wi-Fi signals using a time window of 50ms at the sampling rate 1000, delivering the raw CSI measurement as the dimension of 50×30 with 30 subcarriers. Taking as input the denoised CSI measurements of three antennas, the Wi-Fi radiation tracing module generates the wave front with the size of $64 \times 48 \times 3$ on the image plane, analogous to the RGB image. As a result, we render the segmentation mask by feeding the wave front into our segmentation refinement module.

Ground Truth Acquisition: To teach our cGAN for the image segmentation, we leverage a computer vision architecture [13] for the ground truth acquisition and comparison study. For simplicity, one GoPro camera is required to capture photos in-situ steadily, illustrated in Figure 7. To improve the performance of our system, we further annotate the segmentation masks of Mask R-CNN manually. Since Mask R-CNN suffers from center conditions, such as the interference of the background and poor light conditions.

Table 2: Properties of experimental objects

Object	Material	Texture[a]	Reflected Area (cm^2)
#1 volleyball	leather	1	20.70cm[b]
#2 vase	ceramic	5	17×13
#3 laptop	plastic	2	38.47×25.4
#4 wine Bottle	glass	3	307×9
#5 iPhone	aluminium	4	14×7

[a] The value drops with a rougher surface.
[b] The sphere diameter indicates the reflected area of a volleyball.

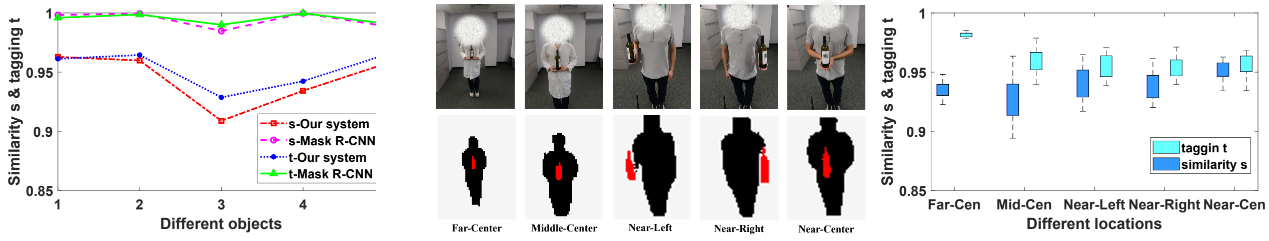
7.2 System Benchmark

Metrics: To measure the quality of the segmentation mask generated by WiSIA, we consider the following metrics:

- *Szymkiewicz-Simpson similarity* $s = S_c/\sqrt{S_i S_g}$ measures the accuracy of object profiling, in which S_i and S_g indicate the object tagging area (e.g. person and object) in the generated and ground truth images, respectively [22] while S_c is the area of their intersection. The tagging area is calculated pixel-by-pixel, rendering the similarity value equal to 1 as the perfect matching.
- *Tagging accuracy* t measures the pixel-wise tagging accuracy by comparing the pixel-wise classification results. The ratio is derived as the metric t between the number of the matched pixels in the generated segmentation mask and that of the ground truth, rendering the perfect tagging accuracy equal to 1.

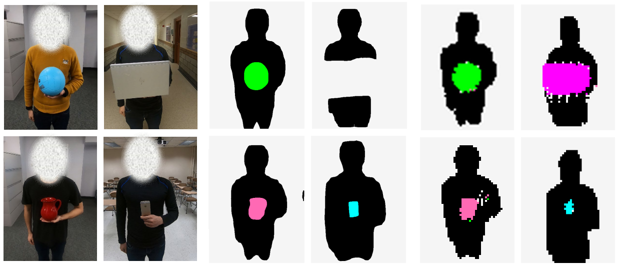
Performance on Different Objects: First, we evaluate WiSIA in the general cases, in which the target person carries different objects standing at the same location. Illustrated in Figure 9, visualized examples demonstrate the feasibility and accuracy of WiSIA in profiling and tagging at the pixel scale. Specifically, WiSIA can reconstruct the segmentation mask at a finer-grained spatial resolution, such as the left elbow and the right shoulder in the top row, delivering a comparable result with the state-of-the-art in the computer vision field. Besides, it can profile and tag the volleyball and vase accurately as Mask-R-CNN does. Note that Mask R-CNN cannot recognize the laptop due to its vantage view in the image while our system does. We can however observe the rough border of the segmented part in the segmentation mask, especially the Wi-Fi segmentation mask of our system. And some pixels are even tagged as other colors due to the noise signals from COTS Wi-Fi devices. We can further refine it by increasing the pixel intensity of the wave front or leverage more antennas to provide more vantage views for WiSIA.

To further demonstrate the feasibility of our system quantitatively, we plot the similarity s and tagging accuracy t in Figure 8(a). It shows that WiSIA can reach 0.9 in s and t for those objects for all 5 tested objects, which is comparable to the state-of-the-art in computer vision [13] and acoustics imaging [22]. Note that the performance of WiSIA drops for object#3 with the similarity s of 0.90, which indicates the profiling for the laptop is relatively less



(a) Human with different objects (b) Visual examples for objects in different locations (c) Object in different locations

Figure 8: Overall performance of WiSIA for the target person with different objects in various locations.



(a) RGB image (b) Mask R-CNN (c) Wi-Fi mask

Figure 9: Examples of the reconstructed segmentation mask of the person with various objects across environments.

accurate. The accuracy degradation appears when using Mask R-CNN method as well. The reason behind this is that the laptop has a rougher surface, which scatters the Wi-Fi radiations and thus reduces the spatial resolution. Another reason is the low-quality ground truth of the laptop affected by its color and limited vantage view, making it less accurate and robust for training cGAN.

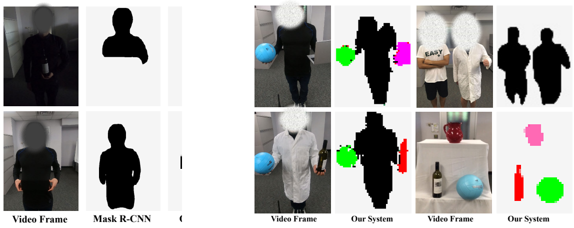
Cross-domain Evaluation: To evaluate the cross-domain ability of WiSIA, we train the model using the dataset of the experiment on all objects in the office while testing it in other domains, including various instances, clothes, and environments. Table 3 lists our statistical evaluation results. We can see the performance of the similarity s and the tagging accuracy t are above 0.8 for most setups, comparable with the state-of-the-art technique in computer vision (Mask R-CNN). Note that WiSIA suffers more across clothes and environments. Since the dynamic setups introduce noticeable interference into the narrow-band Wi-Fi signals. Nevertheless, we can still. It demonstrates the stability of the designed feature and the consistency of its cross-domain performance, which is unaffected by the variances of different sampling instances, clothes of the target person or the performing scenarios. Figure 9 however shows the effectiveness of our system visually in the corridor and classroom scenarios. we know the performance of Mask R-CNN can suffer in some situations, like the missing laptop in Figure 1(d) and 9. Thus WiSIA is also complementary to existing works for the instance segmentation task in the field of computer vision.

Table 3: Performance metrics for cross-domain evaluation.

Similarity s[a]	#1	#2	#3	#4	#5
Instance	0.9952	0.9997	0.9793	0.9992	0.9952
	0.9656	0.9636	0.9145	0.9329	0.9594
Clothes	1	0.9723	0.9915	1	1
	0.807	0.7938	0.7713	0.6719	0.8130
Corridor	1	1	0.7877	0.9996	1
	0.7847	0.7964	0.6974	0.9308	0.7237
Classroom	1	1	0.7884	0.9992	1
	0.8355	0.8505	0.8238	0.9349	0.7760
Tagging t	#1	#2	#3	#4	#5
Instance	0.9812	0.9997	0.9866	0.9994	0.9962
	0.964	0.968	0.9327	0.9412	0.9659
Clothes	1	0.967	0.9942	1	1
	0.7835	0.7983	0.7706	0.7415	0.8491
Corridor	1	1	0.8538	0.9997	1
	0.7894	0.8121	0.7605	0.9385	0.7841
Classroom	1	1	0.8501	0.9993	1
	0.8359	0.8657	0.8218	0.9421	0.8113

[a] For each cell, two rows denote measurements for Mask R-CNN (upper) and WiSIA (lower), respectively

Performance at Different Locations: To evaluate the effectiveness of our wave front construction module on the spatial resolution, we let a person hold the same object with different locations toward the body and transceivers, for example, on the left, center and right of the human body as well as at the far (2.1m), middle (1.4m), near distance (0.7m) to the transceivers. Figure 8(b) shows that we can estimate the position of the arms and body shape as well as the location of the wine bottle from the left to the right. The arm holding the wine bottle can be pinpointed by observing the density profile steepening and hollowing. Note that profiling of the wine bottle is not so smooth due to two reasons. First, the narrow bandwidth of Wi-Fi radiations limits the spatial resolution from the source signals. Second, it suffers from the property of the reflector, such as the near-far problem [3], which can be alleviated with more receiver antennas. Figure 8(c) further shows WiSIA can profile and tag the human and the wine bottle accurately, delivering high visual performance. Specifically, the similarity s and tagging accuracy t can reach 0.9 and 0.95 for all 5 locations, rendering the effective range up to 2.1m.



(a) Poor light and occlusions

(b) Multi-object segmentation

Figure 10: Visual examples of WiSIA for special cases.

Poor Lighting Evaluation: We also evaluate the performance of WiSIA under poor lighting conditions, where the vision-based solutions usually have poor performance. Illustrated in the top of Figure 10(a), the image on the left shows the targeting person holding a wine bottle, which is very blurry due to the poor lighting. The center and right images are the recovered image by Mask R-CNN and WiSIA, respectively. This result indicates Mask R-CNN suffers from poor lighting, whereas WiSIA is robust to the poor lighting condition and delivers consistent performance. It demonstrates the advantages of WiSIA over vision-based approaches in poor lighting condition. We further evaluate our system against the vision-based method [13] qualitatively, Table 4 shows both achieve a comparable similarity for the object profiling against the background. Mask R-CNN performs a little worse since it sometimes only profiles the partial body part as the Figure 10(a) shows. To measure the classification of the object, we compute the tagging t against the target object. It shows Mask R-CNN cannot recognize the small-size object for most cases, rendering the tagging t as 0.0078 while our system works normally with the pixel-wise classification accuracy of 0.7277.

Occlusion Evaluation: Furthermore, we evaluate the performance of WiSIA under occlusion, where the vision-based solutions no longer have any effect. Illustrated in the bottom of Figure 10(a), the image on the left shows the targeting person holding a keyboard, which is occluded under the clothes. The center and right images are the recovered image by Mask R-CNN and WiSIA, respectively. This result demonstrates Mask R-CNN no longer takes any effects while WiSIA delivers a consistent performance without suffering. Table 4 further computes the performance metrics for our system against the vision-based method, in which our system achieves the classification accuracy of the tagging t 0.6504 while Mask R-CNN cannot work at all. Therefore, WiSIA can be utilized to augment the vision based entrance checking while the malicious people may occlude illegal objects under their clothes or bags.

Multi-Object Evaluation: Finally, we validate the feasibility of WiSIA in the scenarios that multiple people and objects simultaneously appear. Figure 10(b) demonstrates its ability for image segmentation of multiple objects in three scenarios including a person holding two objects, two persons standing side by side and three objects putting on a sliding table, respectively. For the scenario where no person exists, the table is sliding forth and back to construct the relative motion. The results show that WiSIA cannot only successfully generate the segmentation masks of the multiple objects, but also accurately recognize each kind of object at the pixel level. While the borders of recovered objects are blurred, which can be improved by increasing the pixel resolution of wave front.

Table 4: Performance metrics against the vision based method.

Setup	Similarity s		Tagging t*[a]	
	Ours	Vision [13]	Ours	Vision [13]
Poor light	0.7763	0.7151	0.7277	0.0078
Occlusion	0.9357	0.9053	0.6504	0

[a] We compute the tagging t for classifying objects.

Table 5: Time consumption comparison between the whole process for WiSIA and Lagrange multiplier on a single NVIDIA Titan V GPU.

Module	Wave front Recovery	Feature Design	cGAN	Lagrange Multiplier
Time(s)	0.0027	0.001	0.001	29.51

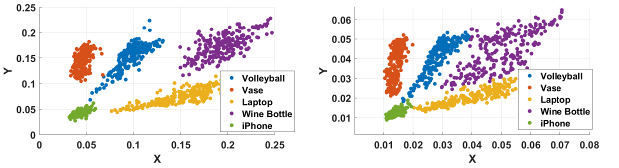
7.3 Running Time Analysis

To demonstrate the computation efficiency of our system, we measure the processing time of major components with multiple instances in WiSIA. For 40ms CSI measurements, the total processing time is less than 3ms in average, rendering a real-time system for stream processing. We also measured the time consumption of the numerical fitting method using the Lagrange multiplier [18, 54]. The average total processing time is 29.51s. Note that conditions for the comparison keep consistent except that we leverage the Lagrange multiplier to recover a 32×24 image plane which is less than our 64×48. And it demonstrates the ineffectiveness of the numerical fitting in recovering the wave front.

7.4 Ablation Study

To verify the effectiveness of our designed modules, we do the ablation study against the profiling enhancement in §5.3 and the cGAN design in §6.2. We first illustrate the 3D feature point cloud in Figure 11(a) with the enhancement optimization, rendering distinct physical signatures for all objects in Table 2. While the distinctness can be reduced by removing the enhancement optimization. For example, Figure 11(b) shows the physical signature of the wine bottle can be more scattered in the feature space, making it more difficult to be classified against the other objects.

We further evaluate our segmentation refinement module in Figure [10], especially the weights of various loss functions, including γ, λ and β for loss L_{cGAN}, L_{cycle} and L_{pixel} (Sec. §6.2). As listed in Table 6, it has a low similarity s with the loss β of the L_{pixel} equal to 0 or 2. The rationale lies that a zero β means no utilization of pixel loss so it cannot make the generated image mask near the ground truth. While a larger β can reduce the impact of the L_{cycle}, inducing the mode collapse problem [11] with unpaired data, where all inputs map to the few outputs and the optimization procedures fail. In the experiments for different locations of the target person



(a) With enhancement optimization (b) Without enhancement optimization

Figure 11: 3D point cloud representing distinct physical signatures for various objects.

Table 6: Performance with various weights of loss functions given the fixed weight $\gamma = 2$ for the L_{cGAN}

$\lambda \setminus \beta$	0	1	2
0	(0.244, 0.937)	(0.749, 0.951)	(0, 0.913)
1	(0, 0.917)	(0, 0.918)	(0.241, 0.935)
2	(0.262, 0.933)	**(0.859, 0.975)**	(0.261, 0.931)

^{a}The data pair for each cell denotes the (s,t) for WiSIA

and object, the mode collapse can remove the ability of the generator to distinguish various locations of the object, leading to a terrible profiling performance with a low similarity s. We can see that it achieves the best performance with the λ of L_{cycle} and β of L_{pixel} equal to 2 and 1, rendering the necessity of each type of loss function.

8 DISCUSSION & LIMITATIONS

Although WiSIA demonstrates its feasibility and effectiveness in image segmentation of multiple objects located in various locations, it has some limitations and future potentials.

Deployment cost: To extract the wave front without the massive antenna array, a relative motion is required between the imaging targets and the transceivers in §4.2, such as making a deep breath by the person or moving the transceivers' antennas to simulate a massive antenna. Although it increases the deployment cost, the requirements can be easily satisfied in many scenarios such as the security surveillance [38] and drone based monitoring [49].

Surface properties and effective range: The preliminary demonstrates that our system can extract features to represent various materials, textures and reflected areas of objects in Figure 5. Figure 11(a) also verifies the distinct physical signal across 5 evaluated objects in Table 2, which is consistent with the existing works [42, 49]. However, we also observe different scatters for each type of object. For example, physical signatures of the laptop and the wine bottle can be more scattered compared with others, rendering the degraded performance for segmentation in Figure 8(a). It can be attributed to properties of the plastic and glass, making it not so distinct with other materials with respect to the polarization as the metal does [49]. Further, the effective distance to transceivers is up to 2.1m in the current setting. We expect to enlarge the segmentation range by using more transceivers and advanced deep learning techniques [39].

Continuous movement and generalization: Currently, we only evaluate our system for targets on the spot and have to re-train our cGAN model for different setups. On one hand, it can be improved to deal with the continuous movement of a user walking since the moving objects can provide more resilience to the background interference by enhancing the Wi-Fi radiations bouncing off the target person. To alleviate the dynamic variances introduced by the motion of various body parts (e.g., the moving arm, torso and leg), more signal processing and deep learning techniques can be adopted for noise reduction and high-level feature extraction, such as the multi-level interference cancellation [4, 22, 49] and RNN-based network (e.g., LSTM and GRU) to extract the temporal feature from the sequential data [18, 23, 54]. On the other hand, our system can be generalized and extended with high-quality wireless signals, which can provide a larger feature space and spatial resolution. For example, the 60GHz transmission of IEEE 802.11ad with higher bandwidth and frequency will improve the spatial resolution of the recovered 2D wave front by our wave front construction module, to the 5mm scale [14].

Future trends: The implications of our evaluation are manifold. It verifies the feasibility of understanding wireless signals from the view of the light field, toward through-the-wall holography for the security surveillance and floor tomography. Besides its fundamental interest to understand the untraceable wireless signals resembling the light, it can also bootstrap the pervasive sensing, ranging from object tracking [4, 36], activity recognition [17, 27, 43], human identification [32, 35, 41]. We leave designing the 3D imaging segmentation counterpart as our future work.

9 CONCLUSION

In this paper, we presented the design, implementation and evaluation of WiSIA, a generative adversarial network-augmented versatile Wi-Fi imaging system based on COTS Wi-Fi devices. WiSIA involves a number of novel techniques as follows. First, it can trace the wave front of Wi-Fi radiation back to the image plane with efficient 2D IFFT. Second, by fine-grained profiling and pixel-wise tagging, it encodes the prior knowledge of objects into the image segmentation and explores the physical signature of Wi-Fi radiations beyond the intensity, which can enhance the spatial resolution by pixel-wise illumination. Lastly, it employs the cGAN model and multiple loss functions for the translation from the wave front of Wi-Fi signals to object segmentation masks. Experimental evaluation based on the COTS Wi-Fi devices demonstrates the feasibility and effectiveness of WiSIA in image segmentation with high precision. Inspired by the promising results we have achieved, we plan to focus on designing the 3D image segmentation system as our future work.

10 ACKNOWLEDGMENT

We sincerely thank the anonymous shepherd and reviewers for their helpful comments on the paper. This work was partially supported by National Science Foundation (NSF) Awards CNS-1909177, CNS-1617627, and PFI:BIC-1632051.

REFERENCES

[1] Karim Abdel-Malek and Jasbir Arora. 2013. *Human Motion Simulation: Predictive Dynamics*.

[2] Fadel Adib, Chen-Yu Hsu, Hongzi Mao, Dina Katabi, and Fredo Durand. 2015. Capturing the Human Figure through a Wall. (2015).

[3] Fadel Adib, Zachary Kabelac, and Dina Katabi. 2015. Multi-Person Localization via RF Body Reflections. In *Proceedings of USENIX NSDI*.

[4] Fadel Adib, Zachary Kabelac, Dina Katabi, and Robert C. Miller. 2014. 3D Tracking via Body Radio Reflections. In *Proceedings of USENIX NSDI*.

[5] Shai Avidan and Ariel Shamir. 2007. Seam Carving for Content-Aware Image Resizing. In *Proceedings of ACM SIGGRAPH*.

[6] Roshan Ayyalasomayajula, Aditya Arun, Chenfeng Wu, Sanatan Sharma, Abhishek Sethi Deepak Vasisht, and Dinesh Bharadia. 2020. Deep Learning based Wireless Localization for Indoor Navigation. In *Proceedings of ACM MobiCom*.

[7] Roshan Ayyalasomayajula, Deepak Vasisht, and Dinesh Bharadia. 2018. BLoc: CSI-Based Accurate Localization for BLE Tags. In *Proceedings of ACM CoNEXT*.

[8] S. Depatla, C. R. Karanam, and Y. Mostofi. 2017. Robotic Through-Wall Imaging: Radio-Frequency Imaging Possibilities with Unmanned Vehicles. *IEEE Antennas and Propagation Magazine* (2017).

[9] David Eigen and Rob Fergus. 2015. Predicting Depth, Surface Normals and Semantic Labels with a Common Multi-Scale Convolutional Architecture. In *Proceedings of IEEE/CVF ICCV*.

[10] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative Adversarial Networks. In *Proceedings of NIPS*.

[11] Ian J. Goodfellow. 2017. NIPS 2016 Tutorial: Generative Adversarial Networks. (2017).

[12] Daniel Halperin, Wenjun Hu, Anmol Sheth, and David Wetherall. [n.d.]. Predictable 802.11 packet delivery from wireless channel measurements. ([n. d.]).

[13] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. In *Proceedings of IEEE/CVF ICCV*.

[14] Philipp M. Holl and Friedemann Reinhard. 2017. Holography of Wi-fi Radiation. *Phys. Rev. Lett.* (2017).

[15] Donny Huang, Rajalakshmi Nandakumar, and Shyamnath Gollakota. 2014. Feasibility and Limits of Wi-fi Imaging. In *Proceedings of ACM SenSys*.

[16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-Image Translation with Conditional Adversarial Networks. (2017).

[17] Wenjun Jiang, Chenglin Miao, Fenglong Ma, Shuochao Yao, Yaqing Wang, Ye Yuan, Hongfei Xue, Chen Song, Xin Ma, Dimitrios Koutsonikolas, et al. 2018. Towards environment independent device free human activity recognition. In *Proceedings of ACM MobiCom*.

[18] Wenjun Jiang, Hongfei Xue, Chenglin Miao, Shiyang Wang, Sen Lin, Chong Tian, Srinivasan Murali, Haochen Hu, Zhi Sun Sun, and Lu Su. 2020. Towards 3D Human Pose Construction Using WiFi. In *Proceedings of ACM MobiCom*.

[19] Nick Kanopoulos, Nagesh Vasanthavada, and Robert L Baker. 1988. Design of an image edge detection filter using the Sobel operator. *IEEE Journal of solid-state circuits* (1988).

[20] Chitra R. Karanam and Yasamin Mostofi. 2017. 3D Through-wall Imaging with Unmanned Aerial Vehicles Using Wifi. In *Proceedings of ACM IPSN*.

[21] Yongsen Ma, Gang Zhou, and Shuangquan Wang. 2019. WiFi sensing with channel state information: A survey. (2019).

[22] Wenguang Mao, Mei Wang, and Lili Qiu. 2018. AIM: Acoustic Imaging on a Mobile. In *Proceedings of ACM MobiSys*.

[23] Wenguang Mao, Mei Wang, Wei Sun, Lili Qiu, Swadhin Pradhan, and Yi-Chao Chen. 2019. RNN-Based Room Scale Hand Motion Tracking. In *The 25th Annual International Conference on Mobile Computing and Networking*. ACM.

[24] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in PyTorch. (2017).

[25] Hari Prakash and Naresh Chandra. 1971. Density Operator of Unpolarized Radiation. *Phys. Rev. A* (1971).

[26] P. C. Proffitt and H. Wang. 2018. Static Object Wi-Fi Imaging and Classifier. In *Proceedings of IEEE International Symposium on Technologies for Homeland Security*.

[27] Qifan Pu, Sidhant Gupta, Shyamnath Gollakota, and Shwetak Patel. 2013. Whole-home gesture recognition using wireless signals. In *Proceedings of ACM MobiCom*.

[28] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, and Kyle Jamieson. 2017. Widar: Decimeter-Level Passive Tracking via Velocity Monitoring with Commodity Wi-Fi. In *Proceedings of ACM Mobihoc*.

[29] Kun Qian, Chenshu Wu, Yi Zhang, Guidong Zhang, Zheng Yang, and Yunhao Liu. 2018. Widar2.0: Passive Human Tracking with a Single Wi-Fi Link. In *Proceedings of ACM MobiSys*.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. 2015. U-net: Convolutional networks for biomedical image segmentation. In *Proceedings of Springer MICCAI*.

[31] Patsorn Sangkloy, Jingwan Lu, Chen Fang, Fisher Yu, and James Hays. 2016. Scribbler: Controlling Deep Image Synthesis with Sketch and Color. (2016).

[32] Cong Shi, Jian Liu, Hongbo Liu, and Yingying Chen. 2017. Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In *Proceedings of ACM Mobihoc*.

[33] Paul Thompson, Daniel E. Wahl, Paul H. Eichel, Dennis C. Ghiglia, and Charles V. Jakowatz. 1996. *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*. Kluwer Academic Publishers.

[34] Shoji Tominaga and Akira Kimachi. 2008. Polarization imaging for material classification. *Optical Engineering* (2008).

[35] Deepak Vasisht, Anubhav Jain, Chen-Yu Hsu, Zachary Kabelac, and Dina Katabi. 2018. Duet: Estimating User Position and Identity in Smart Homes Using Intermittent and Incomplete RF-Data. (2018).

[36] Deepak Vasisht, Swarun Kumar, and Dina Katabi. 2016. Decimeter-Level Localization with a Single WiFi Access Point. In *Proceedings of USENIX NSDI*.

[37] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, undefinedukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Proceedings of NIPS*.

[38] C. Wang, J. Liu, Y. Chen, H. Liu, and Y. Wang. 2018. Towards In-baggage Suspicious Object Detection Using Commodity WiFi. In *Proceedings of IEEE CNS*.

[39] F. Wang, S. Zhou, S. Panev, J. Han, and D. Huang. 2019. Person-in-WiFi: Fine-Grained Person Perception Using WiFi. In *Proceedings of IEEE/CVF ICCV*.

[40] Ju Wang, Hongbo Jiang, Jie Xiong, Kyle Jamieson, Xiaojiang Chen, Dingyi Fang, and Binbin Xie. 2016. LiFS: Low Human-Effort, Device-Free Localization with Fine-Grained Subcarrier Information. In *Proceedings of ACM MobiCom*.

[41] Wei Wang, Alex X Liu, and Muhammad Shahzad. 2016. Gait recognition using wifi signals. In *Proceedings of ACM UbiComp*.

[42] Wei Wang, Alex X Liu, Muhammad Shahzad, Kang Ling, and Sanglu Lu. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of ACM MobiCom*.

[43] Yan Wang, Jian Liu, Yingying Chen, Marco Gruteser, Jie Yang, and Hongbo Liu. 2014. E-eyes: device-free location-oriented activity identification using fine-grained wifi signatures. In *Proceedings of ACM MobiCom*.

[44] Bo Wei, Wen Hu, Mingrui Yang, and Chun Tung Chou. 2019. From Real to Complex: Enhancing Radio-Based Activity Recognition Using Complex-Valued CSI. *ACM Transactions on Sensor Networks* (2019).

[45] Dan Wu, Daqing Zhang, Chenren Xu, Yasha Wang, and Hao Wang. 2016. WiDir: walking direction estimation using wireless signals. In *Proceedings of ACM UbiComp*.

[46] C. Xu, B. Firner, Y. Zhang, and R. E. Howard. 2016. The Case for Efficient and Robust RF-Based Device-Free Localization. *IEEE Transactions on Mobile Computing* (2016).

[47] Zheng Yang, Zimu Zhou, and Yunhao Liu. 2013. From RSSI to CSI: Indoor localization via channel response. *ACM Comput. Surv* (2013).

[48] Jonathan S. Yedidia, William T. Freeman, and Yair Weiss. 2000. Generalized Belief Propagation. In *Proceedings of NIPS*.

[49] Diana Zhang, Jingxian Wang, Junsu Jang, Junbo Zhang, and Swarun Kumar. 2019. On the Feasibility of Wi-Fi Based Material Sensing. In *Proceedings of ACM MobiCom*.

[50] M. Zhao, T. Li, M. A. Alsheikh, Y. Tian, H. Zhao, A. Torralba, and D. Katabi. 2018. Through-Wall Human Pose Estimation Using Radio Signals. In *Proceedings of IEEE CVPR*.

[51] Mingmin Zhao, Yingcheng Liu, Aniruddh Raghu, Tianhong Li, Hang Zhao, Antonio Torralba, and Dina Katabi. 2019. Through-Wall Human Mesh Recovery Using Radio Signals. In *Proceedings of IEEE/CVF ICCV*.

[52] Mingmin Zhao, Yonglong Tian, Hang Zhao, Mohammad Abu Alsheikh, Tianhong Li, Rumen Hristov, Zachary Kabelac, Dina Katabi, and Antonio Torralba. 2018. RF-based 3D Skeletons. In *Proceedings of ACM SIGCOMM*.

[53] Y. Zheng, C. Wu, K. Qian, Z. Yang, and Y. Liu. 2017. Detecting radio frequency interference for CSI measurements on COTS WiFi devices. In *Proceedings of IEEE ICC*.

[54] Yue Zheng, Yi Zhang, Kun Qian, Guidong Zhang, Yunhao Liu, Chenshu Wu, and Zheng Yang. 2019. Zero-Effort Cross-Domain Gesture Recognition with Wi-Fi. In *Proceedings of ACM MobiSys*.

[55] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proceedings of IEEE/CVF ICCV*.