RF-Sauron: Enabling Contact-Free Interaction on Eyeglass Using Conformal RFID Tag

Baizhou Yang, Ling Chen, Xiaopeng Peng, Jiashen Chen, Yani Tang, Wei Wang, Dingyi Fang, *Member, IEEE*, and Chao Feng

***Abstract*—Smart eyeglasses are emerging as a new medium for human–computer interaction. Existing solutions typically rely on cameras or touchpads, raising privacy invasion concerns or requiring users to physically interact with the glass frames. Here, we present RF-Sauron, a novel mid-air gesture interaction system for eyeglasses, based on radio frequency identification (RFID). The design of our RF-Sauron system involves embedding a conformal RFID tag into the eyeglass frame, where the received signals change with user gestures. We optimize the radiation gain of the tag to avoid view blockage while preserving a long working range. To discriminate different gestures, we propose an adaptively weighted multichannel fusion network to extract their respective distinctive features. To allow efficient adaptation of the pretrained network to new users, we also introduce a novel diagonal dot-product attention in our contrastive learning framework to uncover the feature similarities of different users. The proposed RF-Sauron system was evaluated through extensive experiments, demonstrating an average recognition accuracy of 98.86% across 20 users and a cross-user accuracy of 98.75%.**

***Index Terms*—Contrastive learning, gesture recognition, radio frequency identification (RFID), smart glasses, wireless sensing.**

I. Introduction

SMART eyeglasses for human–computer interaction have received increasing attention in recent years. With the participation of major tech companies like Google [1], Apple [2], and Meta [3], the global smart glass market size was valued at U.S. $6.59 billion in 2024 and is projected to grow at a compound annual growth rate of 9.9% from 2024 to 2030 [4]. It is also estimated that half of the global population will wear eyeglasses daily by 2050 [5]. Transforming ordinary eyeglasses into smart devices could potentially unlock a wide range of applications. In smart homes and offices, for instance, smart eyeglasses could enable users to interact seamlessly with household and office appliances, such as lights, thermostats, printers, and entertainment systems, using hand gestures. This hands-free interaction offers a significant advantage over traditional interfaces like buttons or touchscreens, especially when users have their hands occupied or face mobility constraints. In the manufacturing industry, such hands-free interaction and control capabilities may allow workers to multitask efficiently. For example, they could operate machinery while accessing information or communicating with colleagues without physical involvement, thereby minimizing workflow disruptions. This enhanced functionality has the potential to significantly improve productivity in industrial settings.

Most existing solutions for smart glasses are based on cameras and touch panels [6], [7], [8], [9], [10], [11], [12]. FaceSight [13], for example, integrates an infrared camera on the bridge of the eyeglasses and recognizes user gestures using computer vision approaches. RimSense [14] enables touch-based interaction on eyeglass rims using piezoelectric sensors. While these approaches achieved high recognition accuracy, they tend to be highly sensitive to ambient lighting conditions. In smart homes, where lighting varies significantly throughout the day, this sensitivity can compromise the recognition performance. Likewise, in smart factories and warehouses, lighting conditions may also become too challenging for these systems to function properly. Another concern with vision-based approaches is the potential breach of user privacy [15], where sensitive personal data may be captured by cameras inadvertently, raising potential ethical and security issues. Acoustic approaches [16], [17], [18] have also been explored for gesture recognition in smart eyeglasses. However, these systems typically rely on battery power, which are less energy-efficient and may require frequent recharging, posing practical challenges for long-term uses.

Received 21 October 2024; revised 7 December 2024; accepted 29 December 2024. Date of publication 8 January 2025; date of current version 9 May 2025. This work was supported in part by the National Science Foundation of China under Grant 62302392, and in part by the Project of Shaanxi Province International Science and Technology Cooperation Program under Grant 2024GH-YBXM-08, Grant 2024GH-YBXM-09, and Grant 2024GH-ZDXM-50. *(Corresponding author: Chao Feng.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the Northwest University of China.

Baizhou Yang is with the Shaanxi Key Laboratory of Passive Internet of Things and Neural Computing, School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: yangbai_zhou@163.com).

Ling Chen, Jiashen Chen, and Yani Tang are with the Xi'an Key Laboratory of Advanced Computing and Software Security, School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: 2019117174@stumail.nwu.edu.cn; 2022115102@stumail.nwu.edu.cn; 2022117013@stumail.nwu.edu.cn).

Xiaopeng Peng is with the College of Science, Rochester Institute of Technology, Rochester, NY 14623 USA (e-mail: xxp4248@rit.edu).

Wei Wang and Dingyi Fang are with the Xi'an Key Laboratory of Advanced Computing and System Security, School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: wwang@nwu.edu.cn; dyf@nwu.edu.cn).

Chao Feng is with the Shaanxi International Joint Research Centre for the Battery-Free Internet of Things, School of Information Science and Technology, Northwest University, Xi'an 710127, China (e-mail: chaofeng@nwu.edu.cn).

Digital Object Identifier 10.1109/JIOT.2025.3527126

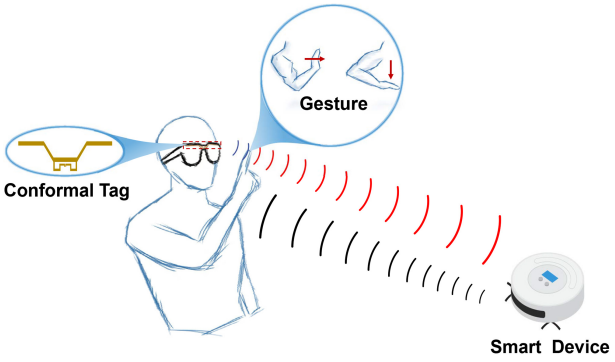

Fig. 1. Application of our RF-Sauron system for contactless interactions in smart eyewear.

Inspired by recent advances in wireless sensing, radio frequency identification (RFID) technology has been widely explored for sensing applications [19], [20], [21], [22]. The RFID tags are ideal for sensing due to their compact size, flexibility, energy efficiency, invariance to lighting conditions, privacy-preserving ability, and low cost (e.g., only a few cents per tag). Unlike acoustic- and vision-based smart glasses that typically rely on battery support, RFID is not limited to power constraints by drawing energy from transmitted signals to enable backscatter transmission. This eliminates the need for onboard batteries, thereby enhancing energy efficiency, reducing the size and weight of the eyewear, and minimizing maintenance demands. RFID tags are also easier to manufacture and deploy. In addition, RFID is invariant to lighting conditions which ensures reliable performance in environments with varying or challenging illumination. Superior recognition accuracy has also been demonstrated in scenarios where user privacy is a priority.

However, applying RFID technologies to eyeglasses for gesture interaction remain challenging. First, traditional commercially available RFID tags are difficult to embed into eyeglass frames because of their complicated topology. Even a slight modification to the structure of tag may result in unpredictable variations in performance. Second, the differentiation of certain types of gestures are particularly difficult. The examples include the symmetrical gestures that exhibit highly similar signal characteristics. Furthermore, training of existing gesture recognition systems typically requires a large amount of data. Repetitive retraining may also required when unseen users are introduced, as a same gesture may differ in characteristics like shape, velocity, and duration when performed by different users.

To address these issues, we introduce RF-Sauron, a low-cost, lightweight, and fine-grained gesture interaction system for smart glasses based on RFID technology. Compared to other smart eyeglass systems that rely on expensive and heavy battery-powered sensors, our RF-Sauron offers a potentially enhanced user experience with more affordable, lightweight, and energy-efficient design. Additionally, improved robustness of our system to environmental changes is also demonstrated. As shown in Fig. 1, the proposed RF-Sauron system consists of a specially designed conformal RFID sensing tag integrated into the eyeglass frame, where the received signals change with user gestures. Our RF-Sauron provides accurate gesture recognition by analyzing these variations using novel neural networks. Several challenges presented in the practical implementation of our system were also addressed.

Challenge 1: Due to the specific structure of the glasses, the shape and design of the sensing tag should fit the frame to avoid obstructing the user's view. Additionally, the tag should be able to receive maximum energy from the reader antenna and be highly sensitive to the user's gestures. To address this, we thoroughly analyze the eyeglasses' structure and select the front edge of the frame as the optimal location for the tag. We then develop an equivalent circuit model to evaluate the mutual coupling between various antenna structures. Following this, we design a multistage antenna workflow that incorporates a T-match structure and an L-shaped feeding mode to achieve impedance matching. Ultimately, the tag meets the spatial and dimensional constraints while maintaining long communication distances.

Challenge 2: Due to the long wavelengths of RFID signals, recognizing similar symmetric gestures is nontrivial. To address the second challenge, we first deploy multiple antennas to capture variations in the reflected signals from different perspectives caused by gesture activity. Then, we introduce a weight-based multichannel fusion network that individually extracts gesture pattern features from both phase and amplitude measurements in each antenna. By assigning different weights to each antenna based on its significance, the network effectively combines these features to achieve accurate and precise gesture recognition.

Challenge 3: Different users exhibit unique behavioral patterns when performing the same gesture, leading to varying signal fluctuations. As a result, the features extracted from measurements become inconsistent across users, making it challenging to apply the pretrained network to new users effectively. Prior methods [23], [24], [25], [26], [27] for extracting user-irrelevant features typically require abundant data and a complex model design. To solve this problem, we propose a novel diagonal dot-product attention (DDPA)-based contrastive learning framework to explore the underlying similarities in data distribution between different users. A key observation is that while certain feature dimensions may differ across users, many similarities persist in other dimensions. This insight inspires us to assign higher weights to similar dimensions and lower weights to dissimilar ones. By doing so, the model concentrates more on the shared patterns across users, improving recognition performance despite variations in user behavior.

We build our system with commodity RFID devices. The effectiveness of our RF-Sauron system is validated through extensive experiments in two different indoor environments. The specific contributions of this work are listed as follows.

1) We introduce RF-Sauron, a cost-effective RFID-based mid-air gesture interaction system for smart eyeglasses. By making use of a low-cost conformal RFID tag, an ordinary eyeglasses is transformed into smart devices capable of gesture sensing.
2) We combine novel designs of both hardware and a neural-network-based gesture recognition in our

RF-Sauron system. Improved performance in tag conformality, communication distance, and discrimination of similar gestures, as well as reduced user dependencies are demonstrated.

3) Extensive experiments demonstrate the effectiveness and robustness of our RF-Sauron system. An average recognition accuracy of 98.86% is demonstrated across 20 users with a cross-user of 98.75%.

II. Related Work

A. Interaction With Smart Glasses

Many systems employ different sensors in smart glasses to perform interactions. For example, some studies [13], [28] use cameras installed on the edge of the glasses to recognize different gestures. While achieving promising processes, such methods are sensitive to lighting conditions and incur privacy invasion. GlassGesture [29] adopts accelerators in the glass to sense head movement. Other works [14], [30] use piezoelectric sensors to convert the edges of the glasses into a touch-sensitive surface, making gesture interaction possible. These solutions require the user to touch the eyeglass frame, which is extremely inconvenient for interaction. Unlike them, RF-Sauron attaches a conformal and battery-free RFID tag in the eyeglass frame, enabling a contactless gesture interaction paradigm. While RF-Sauron is independent of lighting conditions and does not raise privacy concerns. Recently, some works have employed wireless sensing technologies to perform interaction in smart glasses. For example, ReliableEye [31] employs a millimeter-wave radar mounted on the glasses to sense eye blinking and activities. RF-Mic [32] installs a microphone on the glasses to model and analyze facial speech dynamics using acoustic signals. However, these sensing devices cannot conform to the eyeglass frame and add extra weight, leading to user discomfort. In contrast, RF-Sauron integrates a conformal RFID tag within the frame, which is lightweight and does not obstruct the user's view.

B. RFID-Based Sensing

RFID technology has been widely applied in many sensing applications, such as localization [19], [33], activity and gesture recognition [34], [35], [36], object interaction detection [37], [38], and target material identification [39]. For example, GRfid [40] delivers a DTW-based method to achieve precise and stable gesture recognition. RIO [41] leverages the coupling effect of tags to sense touch gestures. RF-CGR [42] transforms received phase information into images for gesture recognition. Cyclops [43] designs a new RFID tag integrated into a contact lens to sense intraocular pressure. Gastag [44] integrates gas-sensitive material and RFID tags to sense different gases. Unlike them, RF-Sauron targets a new application to create a passive, contactless RFID smart glasses system for gesture recognition.

C. Cross-User Gesture Recognition

Deep learning algorithms have been widely investigated in many sensing [45], [46], [47] and recognition tasks [48], [49], including RFID-based gesture recognition [34]. To enable cross-user gesture recognition, many research efforts have been devoted. For example, some studies [50], [51] employ transferring learning networks to adapt to the new user. While effective, these solutions demand extensive data from new users for fine-tuning, making them costly and labor-intensive. Recently, some works have employed meta-learning and contrastive learning to enable models to quickly adapt to new tasks with only a small number of new samples. For instance, MetaSense [52] proposes an adaptive deep mobile sensing system utilizing only a few samples from the target user to fine-turn the model. RF-Net [53] introduces a unified meta-learning framework to enable one-shot human activity recognition. Lai et al. [54] used self-supervised contrastive learning to achieve cross-domain gesture recognition. While promising, the performance of these works significantly degrades when there is a significant difference in the distribution of data from the source domain and the target domain. This is because existing works do not explore the potential similarities in the data distribution under the source and target domains to ensure the performance of the model after migration. Instead, RF-Sauron proposes a novel cross-attention mechanism-based contrastive learning network to mine the potential similarities in data distributions across different users, making the model robust to different cases.

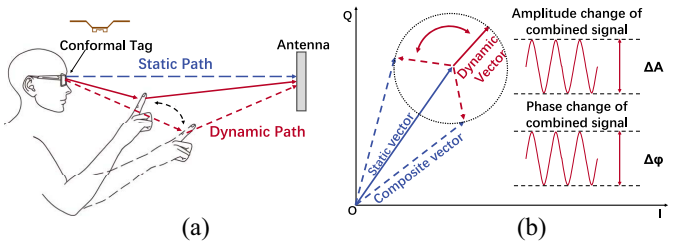

Fig. 2. Phase of the received signal changes as the user's hand moves through the tag and antenna. (a) Examples of user gestures. (b) Signal decomposition model.

III. Preliminaries

A. RFID Basics

An ultrahigh frequency (UHF) RFID system typically consists of a reader and some passive tags. The reader emits a continuous periodic signal to activate nearby tags, and then the tags alter their antenna impedances to reflect the signal back to the reader. Subsequently, the reader can receive the signal, including phase and amplitude information [27], [55].

To understand how a hand gesture affects the received signal, we consider a typical multipath indoor scenario as shown in Fig. 2(a). The signal propagates along three paths, i.e., the Line-of-Sight (LoS) path, the reflection path from the wall, and the reflection path from the user's hand gestures. If we assume there are q reflection paths from the moving user, the received signal can be described as

$$s(t) = A_S e^{j\phi_S} + \sum_q A_q e^{j\left(\frac{2\pi}{\lambda}\int v_q(t)dt + \phi_{dev}\right)} \quad (1)$$

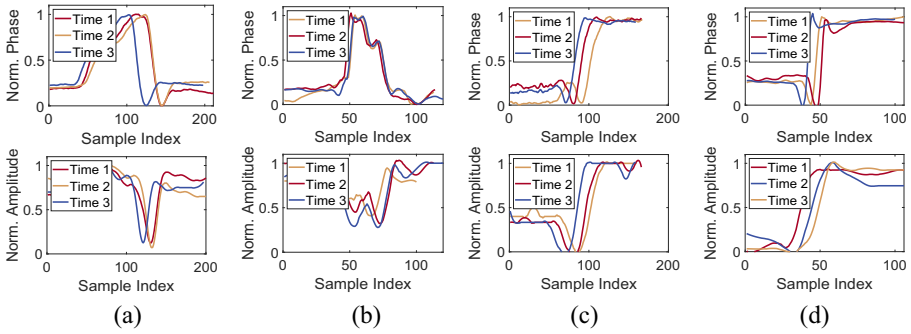

Fig. 3. Three antennas capture phase and amplitude data for one instance of two gestures from a volunteer. (a) Circle clockwise. (b) Triangle. (c) Left Wipe. (d) Right Wipe.

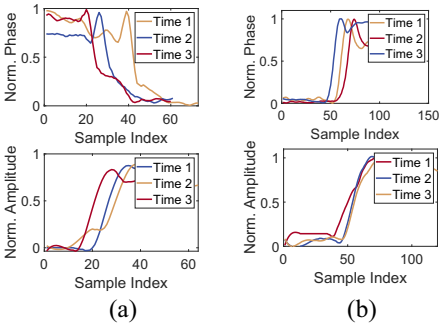

Fig. 4. Antenna reads the phase and amplitude data of the Tick for different user gestures. (a) User1. (b) User2.

where $A_S e^{j\phi_S}$ is the combined complex signal containing the LoS path and static multipath, A_q is the amplitude of the qth reflected from the human body, $v_q(t)$ is the path length change velocity corresponding to the qth path at time, λ denotes the signal wavelength, and ϕ_{dev} is a constant phase offset induced by the tag and reader circuit. According to (1), we can see that when performing a gesture, the dynamic signal varies, which leads to a variation of the phase and amplitude information of the composite signal, as shown in Fig. 2(b). This implies that the phase and amplitude of the received signal can be used to detect the movement of the user, i.e., performing hand gestures in front of the glass.

B. Feasibility Study and Analysis

We conduct a set of benchmark experiments to better understand the correlation between the user gestures and the signal readings. We attach an RFID tag (Alien-9640) in front of the eyeglass frame and let the volunteers wear the eyeglass. Then, we ask each volunteer to stand in front of the reader and perform hand gestures in front of the eyeglass. The distance between the tag and the reader is 1.5 m. In the first experiment, we let one volunteer perform four gestures (Circle clockwise, Triangle, Left Wipe, and Right Wipe) three times. Fig. 3 shows the collected measurements of different gestures. We can see that: 1) different gestures (Circle clockwise and Triangle) can cause different phase and amplitude variations. This indicates that the phase and amplitude variations can be utilized to recognize different gestures and 2) similar gestures (Left Wipe and Right Wipe) induce similar phase and amplitude variations. It means that when recognizing similar gestures, the system could lead to failure.

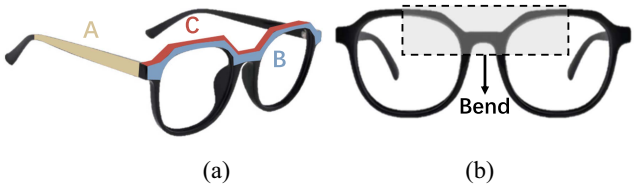

Fig. 5. Design of the conformal tag for smart eyeglasses. (a) Embedding positions. (b) Glass frame.

In the second experiment, we ask two volunteers to perform the same gestures. The collected readings are shown in Fig. 4. We observe that different users cause different phase and amplitude reading changes for the same gesture. This result indicates that the phase and amplitude readings carry adverse user-related information irrelevant to gestures, which could cause a wrong gesture recognition. In the third experiment, we recruit 50 volunteers to wear the glass attached to the Alien-9640 tags and perform a questionnaire that ask them if their view was obstructed. The results show that they all feel that their view is blocked.

In summary, to enable a nonintrusive, accurate, and scalable gesture interaction system using smart eyeglasses, we need to satisfy the following requirements: 1) designing an RFID tag that is conformal with glass frames without obscuring the view of users; 2) Devising a scheme to discriminate similar gestures for accurate gesture recognition; and 3) introducing a scalable approach such that the system can quickly adapt to a new user with minimal human effort.

IV. System Design of RF-Sauron

In this section, we first introduce a novel RFID tag design that conforms to the eyeglass frame. Then, we present a proposed multichannel network to achieve accurate gesture recognition by exploiting antenna diversity in the spatial domain. Finally, we propose a DDPA-based contrastive learning framework to accommodate new users quickly.

A. Conformal Tag Antenna Design

1) How to Integrate the Sensing Antenna: The goal of RF-Sauron's antenna design is to ensure users feel comfortable when wearing eyeglasses to perform hand gestures. To

achieve this, two key factors that significantly impact the user experience need to be solved. First, *view blockage*. Generally, the light enters the eyes through the glass, so any obstruction of the glass lens can impact a person's vision. Thus, our tag cannot obscure the glass lens and only can be deployed around the glass frame, including the temples (A), the front edge of the frame (B), and the upper edge of the frame (C), as shown in Fig. 5(a). The second factor is *tag position*. As mentioned earlier, we have three optional positions to deploy tags. When the tag is placed in position A, the user needs to interact with the gesture on the side of the glasses, which is extremely inconvenient and increases the complexity of the interaction. For position C, due to the orthogonal polarization between the tag antenna and the transmitting antenna, it is impractical to place the antenna in position C, which will result in the inability to receive the signal.

Based on the above considerations, we ultimately select position B for tag deployment. At this location, the user can naturally perform gestures in front of the eyeglasses, aligning with typical user interaction habits. Moreover, this position ensures that the tag receives maximum energy from the reader's antenna and remains sensitive to the user's gestures. Therefore, considering the structure of the eyeglass frame shown in Fig. 5(b), we design the shape of RFID tag to seamlessly fit within the frame.

2) Impedance Matching: Based on the above analysis, we select position B for deploying our sensing tags. However, due to the size constraints of a typical eyeglass frame, such as the antenna being limited to 140 mm in length and 3 mm in width, achieving impedance matching between the designed conformal antenna and the chip becomes a challenge. As a result, the working range sharply decreases from several meters to just 70 cm, significantly limiting gesture recognition scenarios. To enhance the working distance, we need to achieve impedance matching

$$Z_a = Z_c^* \tag{2}$$

where Z_a and Z_c are the impedance of the antenna and the impedance of the chip, respectively.

To achieve impedance matching, a straightforward method is to fine-turn the size of the dipole antenna. However, due to the shape constraint of eyeglasses, the fine-turn scheme is infeasible in our case. To overcome this problem, we borrow the idea of a T-Match ring [56] to facilitate impedance matching between the antenna and the chip. Specifically, we construct an equivalent circuit to illustrate the condition of impedance match, as shown in Fig. 6, which includes a dipole antenna with a T-match circuit and expressed the impedances of the three structures in antenna circuit formulas

$$\begin{cases} Z_a = Z_T + Z_D, & \text{Antenna Circuit} \\ Z_T = j\omega L_t, & \text{T-match Circuit} \\ Z_D = R_d + j\omega L_d - \frac{j\omega}{C_d}, & \text{Dipole Antenna} \end{cases} \tag{3}$$

where Z_T, Z_D are the impedance of the T-match circuit and dipole antenna, respectively. j is the imaginary unit, and ω is the angular frequency. L_t is the inductance of the T-match circuit [57], while L_d and C_d denote the inductance and capacitance of the dipole circuit, respectively.

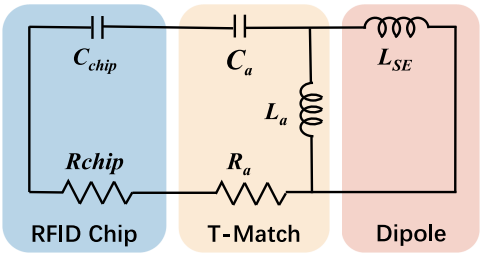

Fig. 6. Equivalent circuit of RFID tag.

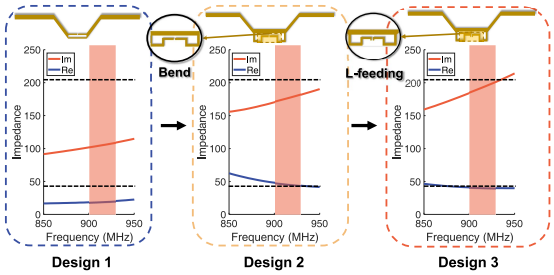

Fig. 7. Optimized impedance matching of tag designs.

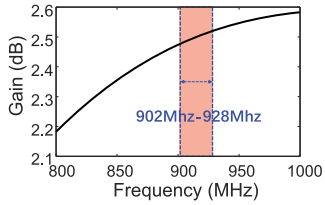

Fig. 8. Gain of the optimized RFID tag.

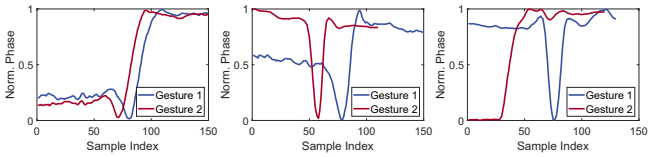

Fig. 9. Three antennas capture phase data for one instance of two gestures from a volunteer. (a) Antenna1. (b) Antenna2. (c) Antenna3.

Based on the equations above, the real part of the antenna impedance is related to resistance, while the imaginary part is associated with inductance and capacitance. To compensate for the degradation introduced by the conformal design in antenna impedance, we need to adapt the form factors of the antenna. However, the length and width are fixed due to the constraint of the form of eyeglasses, we cannot directly fine-tune the dipole structure. To overcome this issue, our basic idea is to modify the T-match structure for the impedance compensation. Specifically, we optimize the size of the T-matching network through odd–even analysis to effectively change the antenna impedance. We then increase the inductance by bending the antenna structure, as shown in Design 2 in Fig. 7, thereby increasing the imaginary part of the antenna impedance. Finally, we introduce a new L-shaped feeding mode into the antenna design, as shown in Design 3 in Fig. 7. Compared to a balanced feed, the vertical and horizontal parts of the L-shaped

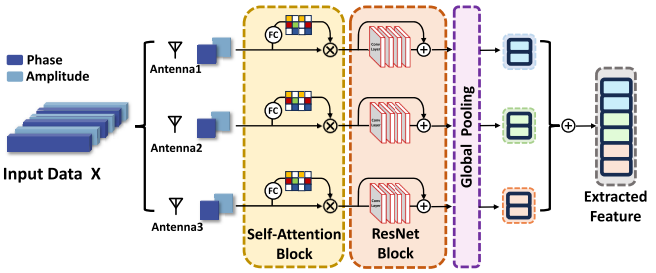

Fig. 10. Feature extraction network in our RF-Sauron.

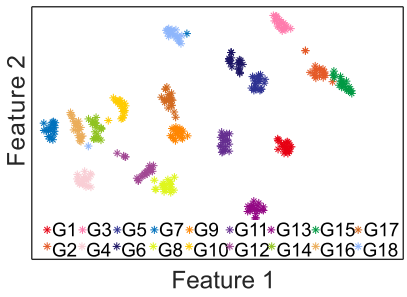

Fig. 11. t-SNE visualization of the combined phase and amplitude features.

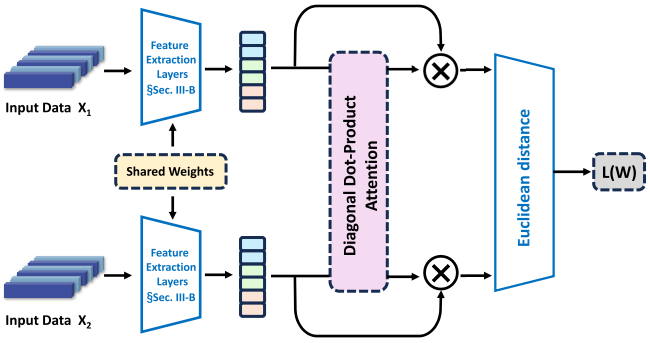

Fig. 12. Architecture of the contrastive network used in RF-Sauron.

feed will generate inductance and capacitive, respectively, which affect the imaginary part of the antenna impedance.

We now leverage the electromagnetic simulation software to verify the effectiveness of our scheme. Fig. 7 shows the impedance of the three antenna design methods during the improvement process, we can see the real and imaginary impedance of the final tag design approach to the impedance of the chip. Additionally, we also simulate the radiation gain in different frequency bands, as shown in Fig. 8, from which we can see that the gain of the optimized antenna is close to 2.5 dB within the range of 902–928-MHz band.

B. Gesture Feature Extraction

As mentioned in Section III-B, similar gestures would cause similar signal patterns for a single receiving antenna, leading to a failure to recognize them. To overcome this issue, our basic idea is to exploit spatial diversity from multiple RFID antennas to extract different signal variations. To illustrate this, we conduct a benchmark by deploying three receiving antennas in three locations, and then ask the same volunteer to perform left wipe and right wipe gestures. The received measurements are plotted in Fig. 9. We can observe that the signal patterns induced by the two gestures are the same in Antenna 1, but are significantly different in Antennas 2 and 3. This result implies that we can combine different data values of the backscatter signal from different antennas to extend the features of different gestures, thus increasing the accuracy of gesture recognition

$$C = Z_p \oplus Z_a. \tag{4}$$

To extract distinguished gesture pattern features, a challenge we face is how to effectively incorporate the measurements from spatially deployed antennas. To address this issue, we propose a weight-based multichannel fusion network to individually extract gesture pattern features and assign different weights to each antenna to combine these features for final gesture recognition. The basic feature extraction network framework is shown in Fig. 10. Specifically, we first design a self-attention block to pay more attention to data from antennas that contain more information about the user's movement. When a user performs a gesture, the block allows the model to assign higher weights to these antennas with pronounced reading changes, thus guiding the model to focus on the most informative features related to the gesture. We take the phase reading data as an example to illustrate the weighting process. Let X_p denote the phase matrix with a size of $K \times M$, where K is the number of antennas and M is the length of the phase reading length. The weighted phase is written as

$$w_p^k = \frac{\exp\left(e_p^k\right)}{\sum_{j=1}^{K} \exp\left(e_p^j\right)} \tag{5}$$

where e_p^k is given by

$$e_p^k = f\left(g_p^k, X_p^k\right) \tag{6}$$

where f is the final fully connected layers, and g_p^k is the weight parameters of the hidden layers. We obtain the weighted antenna phase data Z_p by multiplying the two matrices X_p and W_p together as follows:

$$Z_p = X_p \otimes W_p. \tag{7}$$

In the same way, we can obtain the weighted amplitude Z_a. With the attention mechanism, we can effectively speed up the feature extraction process. Phase and amplitude information are concatenated as follows.

We then pass C into the 4-layer ResNet block to capture unique features from phase and amplitude information. We choose ResNet as the backbone of the feature extraction network due to its identity shortcut connections, which keep all information passing the network while avoiding the problem of gradient explosion or vanishing in deep networks. Specifically, the shortcut connections could be written as

$$V = H(C, w_C) + C \tag{8}$$

where $H(C, w_C)$ represents the residual mapping learning function, w_C denotes the parameters of each layer, C is the

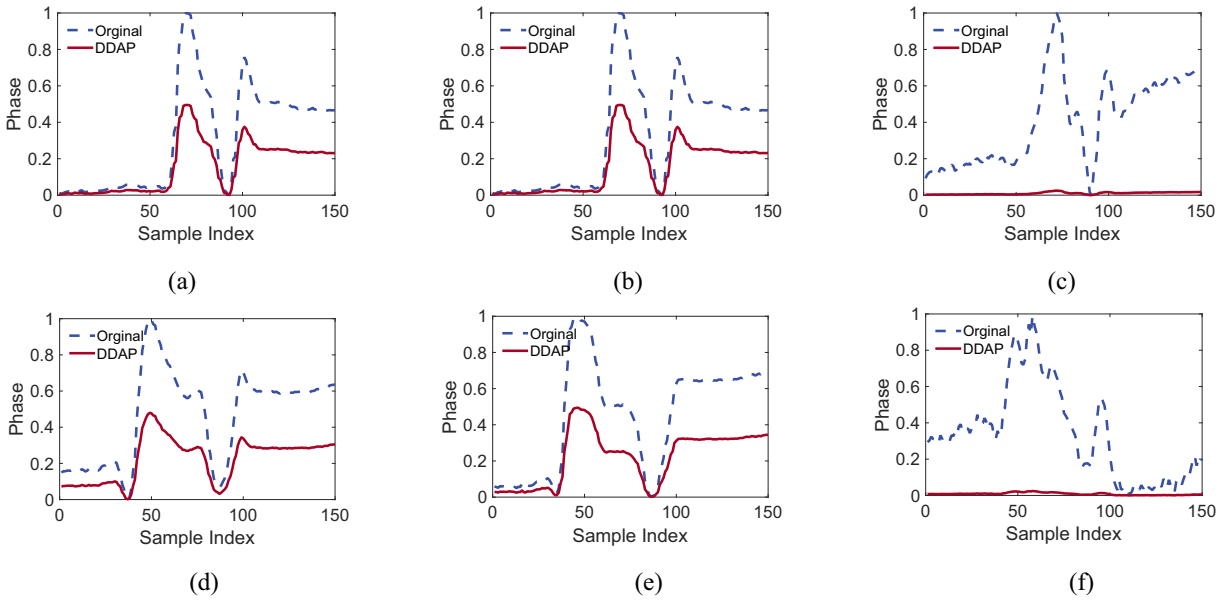

Fig. 13. Normalized phase readings before and after the DDPA process of user1 and user2 performing the gesture "Circle clockwise" from three antennas. (a) Antenna1 of User1. (b) Antenna2 of User1. (c) Antenna3 of User1. (d) Antenna1 of User2. (e) Antenna2 of User2. (f) Antenna3 of User2.

input vector of the ResNet block, and V is the output vector of the ResNet block. Then, the extracted features would be applied to a global pooling layer to scale the magnitude of the data to a uniform scale across different antennas.

To efficiently fuse the data from the individual antennas together while preserving the original representation of the features, we perform a splicing operation on the fused phase and amplitude data from the different antennas

$$Y = (g(V_1), g(V_2), g(V_3), \ldots, g(V_k))^T \tag{9}$$

where $g(\cdot)$ denotes the global pooling function. To visualize the effectiveness of the fusion scheme, we use t-SNE [58] to project the extracted features into a 2-D feature space. As shown in Fig. 11, the learned features of all gestures are independently distributed without any overlap, which showcased the improved performance of our multichannel attention network in extracting distinct features.

C. Accommodating New Users

In this section, our goal is to rapidly adapt the network to new users. As discussed in Section III-B, different users exhibit unique behavioral patterns even when performing the same gesture. Consequently, the features extracted from phase and amplitude measurements are inconsistent across users, rendering the feature extraction network trained on previous users less effective for new ones. To tackle this problem, prior works adopt conservative learning to extract domain-invariant features [27], [59], [60] to eliminate the differences introduced by users' unique habits. However, it requires a large amount of data and very complex model training techniques to achieve a relatively good result, which is inappropriate for the low-cost new user adaptation network we want to achieve.

In our RF-Sauron, we propose a novel DDPA-based contrastive learning framework, as illustrated in Fig. 12. The basic

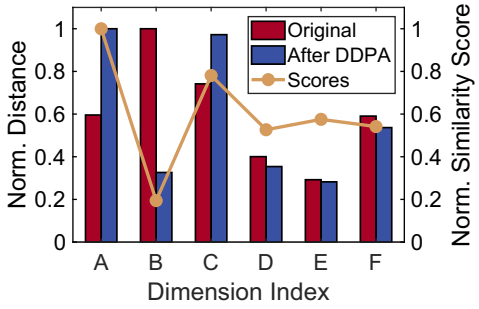

Fig. 14. Normalized distances and similarity scores for data of two users performing "Right Wipe."

idea is to introduce a novel DDPA mechanism, which leverages the observation that although some feature dimensions may differ for different users, many similarities remain in other dimensions. This inspires us to assign weights based on the similarity of the input feature pair (e.g., Y_1 and Y_2), where higher weights are assigned to similar dimensions and lower weights to dissimilar ones. By minimizing the influence of dissimilar dimensions in the loss calculation, the result becomes predominantly determined by the similar feature dimensions. Consequently, this mechanism allows the model to focus on shared patterns across different users, improving recognition performance despite variations in user behavior. Next, we will detail the DDPA mechanism.

1) Diagonal Dot-Product Attention: We first feed the input data pair Y_1 and Y_2, which are extracted features as described in Section IV-B, into a fully connected layer with shared weights

$$\overline{Y}_1 = Y_1 \times W^T + b \tag{10}$$

$$\overline{Y}_2 = Y_2 \times W^T + b \tag{11}$$

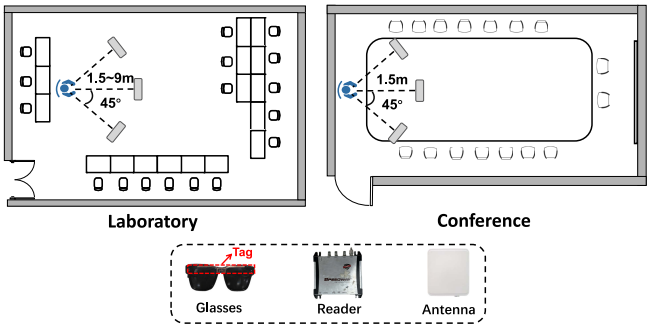

Fig. 15. Experimental setups.

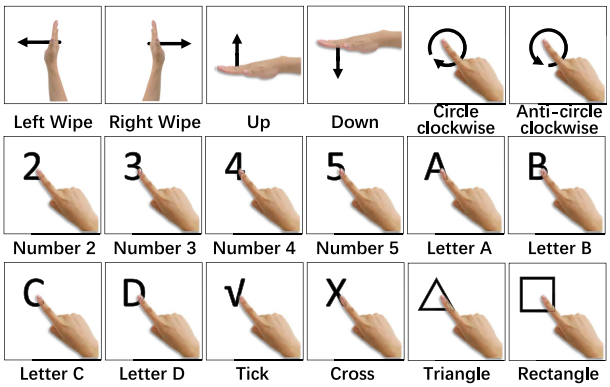

Fig. 16. Schematic of 18 gestures.

where W^T represents the parameters of the fully connected layer and b is the bias. The purpose of using a weight-sharing fully connected layer is to map the features of sample pairs into the same feature space, ensuring that the computed distances between positive and negative sample pairs become more physically interpretable.

Here, we define similar and dissimilar dimensions, respectively, as feature dimensions that exhibit and lack similar patterns. For each of the two pairwise users who perform the same gesture, we consider a phase or an amplitude measurement from one antenna as one dimension, and an input data of six dimensions in total is measured for both the phase and amplitude from three antennas. As illustrated in Fig. 14, the three phase and amplitude dimensions are indexed by ABC and DEF, respectively. Dot-product is employed to calculate the similarity of the features that corresponds to the different antennas for each input feature pair as

$$S = \left(\overline{Y}_{1,1}^T \overline{Y}_{2,1}, \ldots \overline{Y}_{1,i}^T \overline{Y}_{2,i} \ldots, \overline{Y}_{1,K}^T \overline{Y}_{2,K}\right)^T \tag{12}$$

where i is the index of the K antennas, $\overline{Y}_{1,i}$ and $\overline{Y}_{2,i}$ are the subsequence of the feature array split by the corresponding antennae index i, and S is the final similarity matrix containing all the score calculated at the corresponding antenna index.

The normalized similarity score [see (12)] is employed in our DDPA to strengthen similar dimensions and weaken dissimilar dimensions, which allows similar dimensions to carry greater weight when computing the Euclidean distance between a pair of two samples. As shown in Fig. 14, the similarity scores are categorized into three distinct levels based on their values. They include Low level: [0, 0.33), Medium level: [0.33, 0.66), and High level: [0.66, 1.0]. For dimensions with high similarity, their values are enhanced by our proposed DDPA, leading to a greater weight in distance calculations. Conversely, the values of low similarity dimensions diminished, resulting in a reduced weight in the computation of distances. Unlike the traditional attention mechanism, we consider the similarity relationship between the feature subspace from each corresponding antenna pair in each input feature pair. This strategy led to reduced impact of dissimilar features in the loss function, guiding the model to focus on the feature pairs with greater similarities.

Finally, a softmax layer is employed to normalize the resulting similarity matrix and convert the similarity scores into coefficients that should be weighted to the features

$$Y_1' = \frac{\left(e^{S_1}\overline{Y}_{1,1}, e^{S_2}\overline{Y}_{1,2} \ldots, e^{S_K}\overline{Y}_{1,K}\right)^T}{\sum_{j=1}^{K} e^{S_j}} \tag{13}$$

$$Y_2' = \frac{\left(e^{S_1}\overline{Y}_{2,1}, e^{S_2}\overline{Y}_{2,2} \ldots, e^{S_K}\overline{Y}_{2,K}\right)^T}{\sum_{j=1}^{K} e^{S_j}}. \tag{14}$$

2) Contrastive Loss Function: Next, our goal is to construct the contrastive loss function to minimize the feature distance between samples of the same class and maximize the feature distance between samples of different classes. Specifically, for each sample in the dataset, we randomly select another sample from the same class to form a positive sample pair and select a sample from a different class to form a negative sample pair. This approach ensures a balanced number of positive and negative pairs while reducing the total number of sample pairs, which in turn alleviates the computational burden on model training and enhances training efficiency. Then, we employ the following contrastive loss function [61] denoted as $L(W)$ to calculate the distance between samples in one data pair:

$$L(W) = \sum_{i=1}^{N} l \cdot \left[D_W^i(f(g(X_1, X_2)))\right]^2 + (1-l) \cdot \left[\max\left(M - D_W^i(f(g(X_1, X_2))), 0\right)\right]^2 \tag{15}$$

where W denotes the feature extraction network, N is the batch size, $f(\cdot)$ is the DDPA function, $g(\cdot)$ is the feature extraction process, and D_W stands for the MSE distance function [62] of the features out of the two feature extraction networks. Here, we denote the labels as $l = 1$ where the two input samples belong to the same category, and $l = 0$ if X_1 and X_2 belong to different categories.

To verify the effectiveness of our proposed DDPA mechanism, we perform the DDPA operation directly on the original input sample pairs. Fig. 13 illustrates the data waveforms before and after the operation. We can see that the similar features are strengthened while the dissimilar features are weakened, indicating the effectiveness of our method.

3) Model Fine-Tuning: To adapt the pretrained model to new users, we employ fine-tuning to retrain specific model parameters. Unlike approaches that retrain the entire network, we only adjust certain parameters. The rationale is that

while the phase and amplitude characteristics of signals from different users may vary in detail, the overall envelope change remains largely consistent. This allows us to efficiently transfer model parameters from a trained user to a new user with minimal time cost, speeding up the adaptation process.

4) Multiple Gesture Recognition: Traditional contrastive networks function as binary classifiers, determining whether two samples belong to the same class based on feature distance. As such, they are not directly applicable for multiclass gesture classification. To address this, we modify the contrastive network's classification approach to handle multiclass tasks. Instead of simply comparing two samples, we match the test sample against all training samples, sequentially compute the average distance of the test sample to each class, and then assign the gesture category with the smallest average distance as the classification result. The computational process of this classification is given by

$$\hat{y} = \arg \min_{c \in \{1,2,\ldots,C\}} \left(\frac{1}{n_c} \sum_{x \in X_c} d(x_{\text{test}}, x) \right) \tag{16}$$

where c is the category index, ranging from 1 to C, the size of all categories, n_c is the number of date pairs of category c, X_C represents all the data from category c, $d(\cdot)$ is the Euclidean distance calculation process, x_{test} is the test sample waiting to be categorized, and $\hat{y}$ is the final predicated label.

V. Implementation

A. Hardware Implementation

RF-Sauron's system setup is illustrated in Fig. 15, which includes a prototype eyeglass frame embedded with a custom-designed RFID tag antenna, an Impinj Speedway R420 reader, and three directional antennas (each with 9-dBi gain and a 70° beamwidth in both elevation and azimuth). The reader operates at 922.625 MHz with a transmission power of 32 dBm. Antenna 1 is positioned directly in front of the participant, while Antennas 2 and 3 are angled at plus and minus 45°, respectively, to face the participant from different angles.

B. Data Collection

For controlled experiments, we recruit 21 volunteers (12 males and 9 females). Each volunteer wears the eyeglass and repeats 18 different gestures (shown in Fig. 16) 20 times at six different distances. The experiments are conducted extensively in two environments: 1) a laboratory and 2) a conference room (default setup), as shown in Fig. 15. Participants perform the gestures based on their own interpretation. For each environment, we ensure that only one volunteer is present in the sensing area, and no other person is moving nearby. All the experiments have been approved by our Institutional Review Board (IRB). We use accuracy as a metric to evaluate the performance of RF-Sauron.

C. Neural Network Training

We trained the gesture recognition model in our RF-Sauron system for 30 epochs on an NVIDIA GeForce RTX 2080Ti graphics card, with a batch size of 8 and a learning rate of 1e-4. The evaluation is performed on the same GPU.

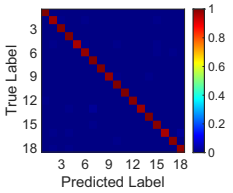

Fig. 17. Confusion matrix of RF-Sauron.

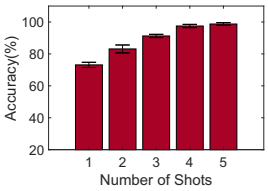

Fig. 18. Performance of cross new users.

VI. Results and Discussion

A. Overall Performance

1) Gesture Recognition Accuracy: To evaluate the gesture recognition performance of RF-Sauron without considering new users, we select data from the conference room as the dataset. We use 80% of each user's measurements for training and 20% for testing. Fig. 17 presents the confusion matrix for 18 gestures, clearly showing that RF-Sauron achieves an average accuracy of 98.86%. This result demonstrates the effectiveness of the proposed method.

2) Cross-User Gesture Recognition Accuracy: To assess RF-Sauron's ability to adapt to new users, we employ leave-one-out cross-validation on the dataset collected from the conference room. Specifically, the data from 20 users are used as the training set, and the remaining user is treated as the new user. This process is repeated for each user. For each new user, we use n samples (n-shot) to fine-tune the pretrained model, and the remaining samples are used to test. Fig. 18 plots the results when varying the number of shots from 1 to 5. As expected, the accuracy gradually improves with the increase in the number of shots. When using five shots, RF-Sauron achieves an average accuracy of 98.75%. These results demonstrate that RF-Sauron can quickly adapt to new users with minimal data, maintaining high performance.

B. Microbenchmarks

1) Verification of the Diagonal Dot-Product Attention Mechanism: To validate the effectiveness of the DDPA mechanism, we conduct an ablation study with the following setups: first, we use the complete model (denoted as DDPA) as the benchmark. We then evaluate three variations: one where the DDPA mechanism is replaced by a traditional attention mechanism, another where it is replaced by a CNN network, and a final version where DDPA is removed entirely, degenerating the model to a basic template matching method. The results are shown in Fig. 19. We observe that RF-Sauron achieves an average recognition accuracy of 71.25%, 76.25%,

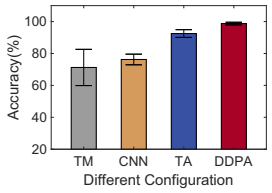

Fig. 19. Performance of DDPA.

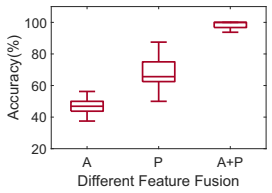

Fig. 20. Performance of feature fusion.

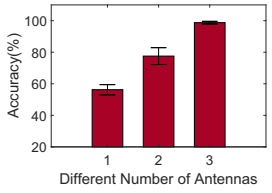

Fig. 21. Performance of different antenna numbers.

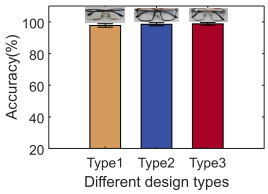

Fig. 22. Performance of different tag designs.

92.50%, and 98.75% for template matching, CNN, traditional attention, and DDPA, respectively. These results highlight the significant improvements provided by the DDPA mechanism, demonstrating its effectiveness in improving the model's gesture recognition performance.

2) Verification of the Feature Fusion: To evaluate the impact of the feature fusion (phase and amplitude readings) on RF-Sauron, we conduct an ablation study with the following scenarios: 1) *Phase Features Only (P):* This scenario involves training and testing the model using only phase features; 2) *Amplitude Features Only (A):* In this case, the model is trained and tested using only amplitude features; and 3) *Combined Features (A + P):* This scenario fuses both phase and amplitude features for training and testing. The results are plotted in Fig. 20. We can see that the performance of the A + P combination achieves the best performance (98.75%), and the performance of phase-based (66.87%) is better than amplitude-based (48.13%). This is because phase readings involve more fine-grained information than amplitude readings. These results demonstrate the proposed fusion scheme can efficiently extract gesture features, and thus achieve a better performance.

3) Verification of the Multiple Antennas: We now evaluate the impact of the number of antennas on RF-Sauron's

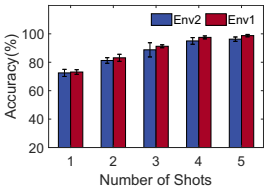

Fig. 23. Performance in different environments.

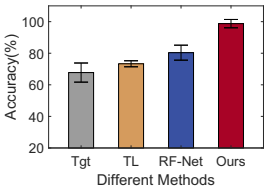

Fig. 24. Comparisons with state of the arts.

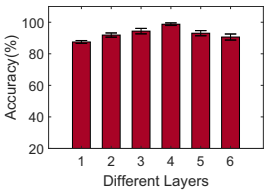

Fig. 25. Performance of different layers.

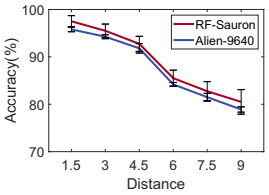

Fig. 26. Performance under different antenna–tag distances.

performance. Specifically, we vary the number of antennas from 1 to 3, and the results are illustrated in Fig. 21. As shown, recognition performance improves as the number of antennas increases. For example, with 1–3 antennas, RF-Sauron achieves an average accuracy of 56.25%, 77.55%, and 98.75%, respectively. The reason is that a small number of antennas can capture only a limited set of gesture features, making it difficult to distinguish between most gestures. Therefore, three antennas are used as the default setup in our RF-Sauron.

4) Generalization of Conformal Tag Design: To assess the generalization of the conformal tag design, we select three different commodity eyeglasses and fabricate three conformal tags based on their respective shapes. Each volunteer is asked to wear each pair of eyeglasses and perform the designated gestures, following the same experimental setup as the default configuration. The results, shown in Fig. 22, indicate that RF-Sauron consistently achieves an average accuracy above 97.75% across all three eyeglass designs. This demonstrates the effectiveness and generalizability of the conformal tag design for various eyeglasses.

C. Impact of Different Factors

1) Environments: To evaluate the robustness of RF-Sauron in various environments, we conduct experiments in two indoor settings: 1) a laboratory and 2) a conference room. The experimental setup remained consistent with our default configuration in both locations. The results are plotted in Fig. 23. Notably, the performance in the office environment shows only slight variation compared to the laboratory, with both environments achieving an average accuracy of approximately 96.37%. This outcome demonstrates that RF-Sauron can maintain stable performance across diverse settings. The underlying reason for this robustness is the proposed contrastive learning framework with DDPA, which effectively extracts common gesture-related features that are invariant to environmental context.

2) Comparison With Prior Works: We compare RF-Sauron's method with three advanced methods: 1) ResNet-based method (Tgt [52]); 2) transfer learning-based method (TL [63]); and 3) meta-learning-based method (RF-Net [53]), where the latter two methods are used to solve the cross-domain problem. Specifically, we calculate the average accuracy in scenarios where instances from one user are used for testing while instances from the remaining users are used for training. Note that TL, RF-Net, and RF-Sauron employ five samples to fine-tune the pretrained model. Fig. 24 presents the result, from which we can observe that RF-Sauron outperforms the other state-of-the-art methods. In particular, TL and RF-Net only achieve an accuracy of 73.33% and 80.35%, respectively. This is because they struggle to resolve the differences in data distribution between the target and source domains caused by variations in user gesture habits, resulting in a degradation that prevents them from achieving high accuracy. In contrast, RF-Sauron introduces a novel conservative learning scheme that effectively mitigates the impact of user-specific gesture patterns, resulting in superior performance.

3) Feature Extraction Network Layers: To evaluate the performance of our system across different layers of the feature extraction network, we conduct experiments under a 5-shot learning condition. The model is trained using between 1 and 6 layers of the ResNet block, which is employed to extract features from both phase and amplitude data. As shown in Fig. 25, the model's performance improves with an increasing number of layers, achieving a peak accuracy of 98.75% at four layers. However, performance declines beyond this point due to overfitting, as excessively deep networks can lead to reduced generalization and increased time and computational costs. Therefore, we selected four layers for the feature extraction network to strike an optimal balance between performance and computational efficiency.

4) Antenna–Tag Distances: To investigate the impact of the antenna–tag distance on RF-Sauron's performance, we vary the distance from 1.5 to 9 m with a step of 1.5 m. At each distance, participants are instructed to perform gestures using our custom-designed conformal tag. As shown in Fig. 26, gesture recognition accuracy decreases with the distance, which is possibly due to the reduced strength of the received signal. Despite this unfavorable trend, the average gesture recognition accuracy of our RF-Sauron appears to be comparable to commercial RFID tags (e.g., Alien-9640), where a remarkable 80% recognition accuracy is achieved at the maximum distance of 9 m. This result affirms the capability of our system for long-range sensing and robustness against deployment distances.

5) Multiple Users: Environment changes can affect the system performance if not properly addressed. Multiuser sensing is particularly challenging in RF sensing due to interference caused by mixed reflected signals from multiple targets at the receiver. Within a certain distance range (e.g., less than 4.5 m), the recognition accuracy decreases notably as the number of interfering users increases. However, this jamming effect diminishes when interferers are more than 4.5 m away from the target user, where recognition accuracies consistently remain above 93% regardless of the number of users or variations in their pairwise distances. It is observed that while the recognition may become challenging when the pairwise distances fall below the threshold, our RF-Sauron has demonstrated promising results for multiuser when they present more than 4.5 m away from each other. Improved robustness to multiple users may benefit from future investigation of multidimensional signal processing techniques [64] for dynamic interference mitigation.

6) Body Movements: Body movements, such as eye blink, mouth motion, and head movement, can potentially affect the accuracy of gesture recognition. In our RFID-based gesture recognition system, the impact of eye blink and mouth motion on gesture sensing is relatively low. These movements are typically subtle, rapid, and involuntary, causing only brief fluctuations in the RFID signal. Since the signals of these movements have a higher frequency compared to those of intentional hand gestures, they are less likely to disrupt the overall gesture recognition process. To further minimize potential impacts, signal processing techniques or machine learning models can be employed to filter out these fluctuations, preserving only signals that are relevant to gestures. In contrast, head movements (e.g., tilting, turning, and nodding) may have greater impacts to RFID signals, particularly because the head is often located near or directly in the line of sight of RFID tags embedded in eyeglass frames. These movements may alter the relative angle between the RFID tag and the antenna, thereby changing the reflection patterns and potentially leading to false detections or misinterpretation of gesture signals. One solution to address this challenge in the future is the use of a low-pass filter to filter out the high-frequency noise caused by head movements. Alternatively, head movements may be labeled and included in the recognition model training.

VII. Conclusion

In summary, we introduce RF-Sauron, an RFID-based mid-air gesture recognition system for smart glasses. Through combined hardware and neural network designs, our RF-Sauron is capable of discriminating similar gestures and adapting to new users efficiently. State-of-the-art performance of our RF-Sauron system is demonstrated through a series of extensive real-world experiments. The presented RF-Sauron is currently limited to a relatively static environment. The robustness of our system to dynamic factors may be improved

by addressing multiple adjacent moving interferences as well as significant head movements through additional data labeling and training of our model.

REFERENCES

[1] A. Klein, C. Sørensen, A. S. de Freitas, C. D. Pedron, and S. Elaluf-Calderwood, "Understanding controversies in digital platform innovation processes: The Google glass case," *Technol. Forecast. Soc. Change*, vol. 152, Mar. 2020, Art. no. 119883.

[2] X. Borah, A. Thangam, and N. Kumari, "Smart glasses with sensors," in *Handbook of Artificial Intelligence and Wearables*. Boca Raton, FL, USA: CRC Press, 2024, pp. 236–245.

[3] E. Waisberg et al., "Meta smart glasses—Large language models and the future for assistive glasses for individuals with vision impairments," *Eye*, vol. 38, no. 6, pp. 1036–1038, 2024.

[4] "Smart glass market size, share and trends analysis report: From 2024 to 2030." Accessed: Dec. 4, 2024. [Online]. Available: https://www.grandviewresearch.com/industry-analysis/smart-glass-market

[5] B. A. Holden et al., "Global prevalence of myopia and high myopia and temporal trends from 2000 through 2050," *Ophthalmology*, vol. 123, no. 5, pp. 1036–1042, 2016.

[6] Z. Zhang, Z. Tian, and M. Zhou, "HandSense: Smart multimodal hand gesture recognition based on deep neural networks," *J. Ambient Intell. Humaniz. Comput.*, vol. 15, pp. 1557–1572, Feb. 2024.

[7] E. Bamani, E. Nissinman, I. Meir, L. Koenigsberg, and A. Sintov, "Ultra-range gesture recognition using a web-camera in human–robot interaction," *Eng. Appl. Artif. Intell.*, vol. 132, Jun. 2024, Art. no. 108443.

[8] P. Bhattacharyya et al., "Helios: An extremely low power event-based gesture recognition for always-on smart eyewear," 2024, *arXiv:2407.05206*.

[9] H. Liu et al., "Ultra-stretchable triboelectric touch pad with sandpaper micro-surfaces for transformer-assisted gesture recognition," *Nano Energy*, vol. 130, Nov. 2024, Art. no. 110110.

[10] D. Jung, C. Gu, J. Park, and J. Cheong, "Touch gesture recognition-based physical human-robot interaction for collaborative tasks," *IEEE Trans. Cogn. Dev. Syst.*, early access, Sep. 24, 2024, doi: 10.1109/TCDS.2024.3466553.

[11] X. Mou, X. Peng, and T. Mou, "38-2: Invited paper: Evaluating optical performance and image quality in augmented reality eyewear: Standardization, challenges, and measurement methods," in *SID Symp. Tech. Dig.*, 2024, pp. 324–326.

[12] X. Mou, X. Peng, and J. Wang, "P-48: Evaluation of field of view in optical see-through near eye displays," in *SID Symp. Tech. Dig.*, 2024, pp. 1548–1550.

[13] Y. Weng, C. Yu, Y. Shi, Y. Zhao, Y. Yan, and Y. Shi, "FaceSight: Enabling hand-to-face gesture interaction on ar glasses with a downward-facing camera vision," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2021, pp. 1–14.

[14] W. Xie, H. Chen, J. Wei, J. Zhang, and Q. Zhang, "RimSense: Enabling touch-based interaction on eyeglass rim using piezoelectric sensors," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 4, pp. 1–24, 2024.

[15] M. Z. Iqbal and A. G. Campbell, "Adopting smart glasses responsibly: Potential benefits, ethical, and privacy concerns with ray-ban stories," *AI Ethics*, vol. 3, no. 1, pp. 325–327, 2023.

[16] M. Tian et al., "RemoteGesture: Room-scale acoustic gesture recognition for multiple users," in *Proc. 20th Annu. IEEE Int. Conf. Sens., Commun., Netw. (SECON)*, 2023, pp. 231–239.

[17] Y. Wang et al., "EchoGest: A highly scalable unseen gesture recognition system based on feature-wise transformation," *IEEE Internet Things J.*, vol. 11, no. 18, pp. 29709–29727, Sep. 2024.

[18] S. Mahmud et al., "ActSonic: Recognizing everyday activities from inaudible acoustic wave around the body," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 8, no. 4, pp. 1–32, 2024.

[19] L. Dodds, I. Perper, A. Eid, and F. Adib, "A handheld fine-grained RFID localization system with complex-controlled polarization," in *Proc. 29th Annu. Int. Conf. Mobile Comput. Netw.*, 2023, pp. 1–15.

[20] M. I. Ahmed, A. Bansal, K. Yuan, S. Kumar, and P. Steenkiste, "Battery-free wideband spectrum mapping using commodity RFID tags," in *Proc. 29th Annu. Int. Conf. Mobile Comput. Netw.*, 2023, pp. 1–16.

[21] J. Wang, J. Xiong, X. Chen, H. Jiang, R. K. Balan, and D. Fang, "TagScan: Simultaneous target imaging and material identification with commodity RFID devices," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 288–300.

[22] C. Feng, J. Xiong, L. Chang, F. Wang, J. Wang, and D. Fang, "RF-identity: Non-intrusive person identification based on commodity RFID devices," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 1, pp. 1–23, 2021.

[23] C. Dian, D. Wang, Q. Zhang, R. Zhao, and Y. Yu, "Towards domain-independent complex and fine-grained gesture recognition with RFID," *Proc. ACM Human-Comput. Interact.*, vol. 4, pp. 1–22, Nov. 2020.

[24] W. Jiang et al., "Towards environment independent device free human activity recognition," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 289–304.

[25] C. Li, M. Liu, and Z. Cao, "WiHF: Enable user identified gesture recognition with WiFi," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, 2020, pp. 586–595.

[26] C. Shi, J. Liu, N. Borodinov, B. Leao, and Y. Chen, "Towards environment-independent behavior-based user authentication using WiFi," in *Proc. IEEE 17th Int. Conf. Mobile Ad Hoc Sens. Syst. (MASS)*, 2020, pp. 666–674.

[27] X. Li et al., "CrossGR: Accurate and low-cost cross-target gesture recognition using Wi-Fi," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 5, no. 1, pp. 1–23, 2021.

[28] Y. Zhang, C. Cao, J. Cheng, and H. Lu, "EgoGesture: A new dataset and benchmark for egocentric hand gesture recognition," *IEEE Trans. Multimedia*, vol. 20, no. 5, pp. 1038–1050, May 2018.

[29] S. Yi, Z. Qin, E. Novak, Y. Yin, and Q. Li, "GlassGesture: Exploring head gesture interface of smart glasses," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, 2016, pp. 1–9.

[30] W. Xie, J. Zhang, and Q. Zhang, "Transforming eyeglass rim into touch panel using piezoelectric sensors," in *Proc. 28th Annu. Int. Conf. Mobile Comput. Netw.*, 2022, pp. 838–840.

[31] W. Zhang and H. Liu, "Toward a reliable collection of eye-tracking data for image quality research: Challenges, solutions, and applications," *IEEE Trans. Image Process.*, vol. 26, pp. 2424–2437, 2017.

[32] Y. Chen, J. Yu, L. Kong, H. Kong, Y. Zhu, and Y.-C. Chen, "RF-Mic: Live voice eavesdropping via capturing subtle facial speech dynamics leveraging RFID," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 2, pp. 1–25, 2023.

[33] B. Liang et al., "{RF-Chord}: Towards deployable {RFID} localization system for logistic networks," in *Proc. 20th USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2023, pp. 1783–1799.

[34] C. Zhao, L. Wang, F. Xiong, S. Chen, J. Su, and H. Xu, "RFID-based human action recognition through spatiotemporal graph convolutional neural network," *IEEE Internet Things J.*, vol. 10, no. 22, pp. 19898–19912, Nov. 2023.

[35] H. Zhang, L. Wang, J. Pei, F. Lyu, M. Li, and C. Liu, "RF-sign: Position-independent sign language recognition using passive RFID tags," *IEEE Internet Things J.*, vol. 11, no. 5, pp. 9056–9071, Mar. 2024.

[36] Q. Qiu, T. Wang, F. Chen, and C. Wang, "LD-recognition: Classroom action recognition based on passive RFID," *IEEE Trans. Comput. Soc. Syst.*, vol. 11, no. 1, pp. 1182–1191, Feb. 2024.

[37] Y. Zhang, Z. Zhan, W. Jin, Y. Li, and D. Shi, "RF-Keypad: A battery-free keypad based on cots RFID tag array," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 6761–6775, Feb. 2204.

[38] B. Zhu et al., "MFD: Multi-object frequency feature recognition and state detection based on RFID-single tag," *ACM Trans. Internet Things*, vol. 4, no. 4, pp. 1–26, 2023.

[39] L. Chang, X. Yang, R. Liu, G. Xie, F. Wang, and J. Wang, "FSS-Tag: High accuracy material identification system based on frequency selective surface tag," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 7, no. 4, pp. 1–24, 2024.

[40] Y. Zou, J. Xiao, J. Han, K. Wu, Y. Li, and L. M. Ni, "GRfid: A device-free RFID-based gesture recognition system," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 381–393, Feb. 2017.

[41] S. Pradhan, E. Chai, K. Sundaresan, L. Qiu, M. A. Khojastepour, and S. Rangarajan, "RIO: A pervasive RFID-based touch gesture interface," in *Proc. 23rd Annu. Int. Conf. Mobile Comput. Netw.*, 2017, pp. 261–274.

[42] Z. Yang, Z. Zhen, Z. Li, X. Liu, B. Yuan, and Y. Zhang, "RF-CGR: Enable Chinese character gesture recognition with RFID," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–16, Oct. 2023.

[43] L. Li, B. Shang, Y. Wu, J. Xiong, X. Chen, and Y. Xie, "Cyclops: A nanomaterial-based, battery-free intraocular pressure (IOP) monitoring system inside contact lens," in *Proc. 21st USENIX Symp. Netw. Syst. Design Implement. (NSDI)*, 2024, pp. 1659–1675. [Online]. Available: https://www.usenix.org/conference/nsdi24/presentation/li-liyao

[44] X. Sun et al., "Gastag: A gas sensing paradigm using Graphene-based tags," in *Proc. 30th Annu. Int. Conf. Mobile Comput. Netw.*, 2024, pp. 342–356.
[45] X. Peng, P. R. Srivastava, and G. A. Swartzlander, "CNN-based real-time image restoration in laser suppression imaging," in *Proc. Imag. Sens. Congr.*, 2021, Art. no. JTh6A–10.
[46] X. Peng, E. F. Fleet, A. T. Watnik, and G. A. Swartzlander, "Learning to see through dazzle," 2024, *arXiv:2402.15919*.
[47] X. Peng, G. J. Ruane, M. B. Quadrelli, and G. A. Swartzlander Jr., "Randomized apertures: High resolution imaging in far field," *Opt. Express*, vol. 25, no. 15, pp. 18296–18313, 2017.
[48] N. Wang, Y. Xiao, X. Peng, X. Chang, X. Wang, and D. Fang, "ContextDet: Temporal action detection with adaptive context aggregation," 2024, *arXiv:2410.15279*.
[49] Y. Xiao et al., "Multi-source eeg emotion recognition via dynamic contrastive domain adaptation," 2024, *arXiv:2408.10235*.
[50] J. Zhang, Z. Tang, M. Li, D. Fang, P. Nurmi, and Z. Wang, "CrossSense: Towards cross-site and large-scale WiFi sensing," in *Proc. 24th Annu. Int. Conf. Mobile Comput. Netw.*, 2018, pp. 305–320.
[51] Z. Ma et al., "RF-Siamese: Approaching accurate RFID gesture recognition with one sample," *IEEE Trans. Mobile Comput.*, vol. 23, no. 1, pp. 797–811, Jan. 2024.
[52] T. Gong, Y. Kim, J. Shin, and S.-J. Lee, "MetaSense: Few-shot adaptation to untrained conditions in deep mobile sensing," in *Proc. 17th Conf. Embed. Netw. Sens. Syst.*, 2019, pp. 110–123.
[53] X. Shen et al., "RF-Net: An end-to-end image matching network based on receptive field," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 8124–8140.
[54] Z. Lai, X. Kang, H. Wang, X. Zhang, W. Zhang, and F. Wang, "Contrastive domain adaptation: A self-supervised learning framework for sEMG-based gesture recognition," in *Proc. IEEE Int. Joint Conf. Biom. (IJCB)*, 2022, pp. 1–7.
[55] J. J. Adams and J. T. Bernhard, "Broadband equivalent circuit models for antenna impedances and fields using characteristic modes," *IEEE Trans. Antennas Propag.*, vol. 61, no. 8, pp. 3985–3994, Aug. 2013.
[56] G. Zamora, S. Zuffanelli, F. Paredes, F. Martı, and J. Bonache, "Design and synthesis methodology for UHF-RFID tags based on the t-match network," *IEEE Trans. Microw. Theory Tech.*, vol. 61, no. 12, pp. 4090–4098, Dec. 2013.
[57] C. A. Balanis, *Antenna Theory: Analysis and Design*. Hoboken, NJ, USA: Wiley, 2016.
[58] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, no. 86, pp. 2579–2605, 2008. [Online]. Available: http://jmlr.org/papers/v9/vandermaaten08a.html
[59] Y. Zheng et al., "Zero-effort cross-domain gesture recognition with Wi-Fi," in *Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services*, 2019, pp. 313–325.
[60] C. Feng et al., "Wi-Learner: Towards one-shot learning for cross-domain Wi-Fi based gesture recognition," *Proc. ACM Interact., Mobile, Wearable Ubiquitous Technol.*, vol. 6, no. 3, pp. 1–27, 2022.
[61] J. Yang, H. Zou, Y. Zhou, and L. Xie, "Learning gestures from WiFi: A Siamese recurrent convolutional architecture," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 10763–10772, Dec. 2019.
[62] G. Palubinskas, "Mystery behind similarity measures MSE and SSIM," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2014, pp. 575–579.
[63] S. A. Rokni, M. Nourollahi, and H. Ghasemzadeh, "Personalized human activity recognition using convolutional neural networks," in *Proc. AAAI Conf. Artif. Intell.*, 2018, pp. 8143–8144.
[64] Y. Xie, J. Xiong, M. Li, and K. Jamieson, "mD-Track: Leveraging multi-dimensionality for passive indoor Wi-Fi tracking," in *Proc. 25th Annu. Int. Conf. Mobile Comput. Netw.*, 2019, pp. 1–16.

Baizhou Yang is currently pursuing the master's degree majoring in software engineering with the School of Information Science and Technology, Northwest University, Xi'an, China.

His current research interest is intelligent wireless sensing.

Ling Chen received the B.E. degree in computer science and technology from Northwest University, Xi'an, China, in 2023, where he is currently pursuing the M.S. degree in computer science and technology with the School of Information Science and Technology.

His research interests include metasurface, wireless sensing, and mobile computing.

Xiaopeng Peng received the M.S. degree from Shanghai Jiao Tong University, Shanghai, China, in 2017, and the Ph.D. degree from Rochester Institute of Technology, Rochester, NY, USA, in 2022.

Her research spans machine learning, deep learning, artificial intelligence, and their applications.

Jiashen Chen is currently pursuing the B.E. degree majoring in Internet of Things engineering with the School of Information Science and Technology, Northwest University, Xi'an, China.

His current research interest is artificial intelligence for wireless sensing.

Yani Tang is currently pursuing the B.E. degree majoring in electronic and information engineering with the School of Information Science and Technology, Northwest University, Xi'an, China.

Her current research interest is wireless sensing and integrated circuit design.

Wei Wang received the Ph.D. degree in information and communication engineering from Northwestern Polytechnical University, Xi'an, China, in 2017.

She is an Associate Professor with the School of Information Science and Technology, Northwest University, Xi'an. Her current research interests include the technology and application of Internet of Things and information security.

Dingyi Fang (Member, IEEE) received the Ph.D. degree in computer science from Northwestern Polytechnical University, Xi'an, China, in 2001.

He is a Professor with the School of Information Science and Technology, Northwest University, Xi'an. His current research interests include Internet of Things, and mobile and wireless computing.

Chao Feng received the Ph.D. degree in computer software and theory from Northwest University, Xi'an, China, in 2022.

He is an Associate Professor with the School of Information Science and Technology, Northwest University. His current research interests include ubiquitous computing and wireless sensing.