Feature Attribution for Human Sensing with Radio Signals

Shuokang Huang, Julie McCann

Department of Computing, Imperial College London, London SW7 2AZ, United Kingdom
s.huang21@imperial.ac.uk, j.mccann@imperial.ac.uk

Abstract

Human sensing with radio signals has emerged as a non-intrusive and occlusion-robust alternative to vision-based approaches, and WiFi signals further support device-free sensing. However, current approaches deeply rely on neural networks, whose black-box nature hinders model transparency and explainability, limiting the use of WiFi-based human sensing in critical fields. For model explainability, recent works have studied saliency methods which attribute model outputs to important features, but they mostly bias in favor of common modalities (*e.g.*, images, time series). This paper proposes a Matryoshka-like saliency method, MatryMask, an initial exploration of feature attribution for human sensing with radio signals. Compared to existing methods that require empirical knowledge about the sparsity of important features, MatryMask regularizes multiple masks to highlight salient areas at different scales, adapting to the uncertain and varying sparsity of important features in radio signals. To effectively perturb radio signals, we devise a novel frequency-removal perturbation beyond existing spatial/time-domain perturbations. Experimentally, MatryMask outperforms state-of-the-art saliency methods and significantly improves the attribution performance by up to 38.1∼70.6% for three tasks.

Code — https://github.com/huangshk/MatryMask

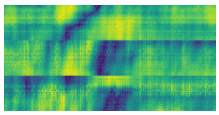
1 Introduction

Human sensing is foundational to user behavior analysis for various applications, such as healthcare (Yang, Zhang, and Zhang 2022), smart homes (Ding et al. 2022b), security monitoring (Zhu et al. 2022), and augmented/virtual reality (Tan et al. 2022). Recently, using radio signals has demonstrated remarkable potential for non-intrusive human sensing (Nirmal et al. 2021), since it does not require on-body sensors or cameras to collect data, enhancing user privacy and convenience. Compared to camera-based sensing, the use of radio signals is more robust to adverse conditions, involving smoke, darkness, and non-line-of-sight situations (Kong et al. 2022). Specifically, radio signals can be collected from ubiquitous existing WiFi devices, which further enable device-free human sensing without additional deployments (Hussain, Sheng, and Zhang 2020).

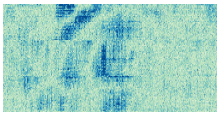
Theoretically, human bodies interfere with radio signals and lead to signal fluctuations (Wei et al. 2015). For WiFi devices, such fluctuations are recorded in Channel State Information (CSI), which thereby contains implicit features for human sensing (Wang et al. 2015). In practice, the implicit features in WiFi CSI are intertwined with excessive noise and not observable for human perception. To this end, various neural networks have been used to extract implicit human features from WiFi CSI, such as Long Short-Term Memory (LSTM) (Yousefi et al. 2017), Convolutional Neural Networks (CNNs) (Wang et al. 2019), CNN-LSTM hybrids (Kong et al. 2022), and Transformer variants (Li et al. 2021). These models have made considerable contributions to WiFi-based human sensing, especially for three fundamental tasks (Chen, Zhou, and Lin 2023): human identification, localization, and activity recognition.

Despite the promising performance of neural networks, their black-box nature hinders model transparency and explainability, limiting the practical use of WiFi-based human sensing. Such transparency and explainability are crucial for users to understand, trust, and improve human sensing. Specifically, it is worth studying whether these models have really captured the signal fluctuations caused by human interference in radio signals. Hence, it necessitates the attribution of model outputs to important features, highlighting significant *channels* and *time* for human sensing. The highlighted temporal channel features help us *understand* how models distinguish human-related patterns from noise in radio signals, enhancing stakeholder *trust* to abate potential ethical and legal concerns, especially in critical fields such as healthcare (Amann et al. 2020) and public security (Hassija et al. 2024). Moreover, understanding how models recognize human-related patterns can help analyze the strengths and weaknesses of models to develop better human sensing. Such an obvious and essential need motivates us to explore feature attribution for WiFi-based human sensing.

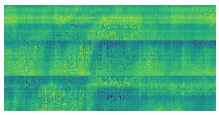
For feature attribution, *saliency methods* aim to highlight input features that are responsible for model outputs, gaining much interest in computer vision (Fong, Patrick, and Vedaldi 2019), time-series analysis (Tonekaboni et al. 2020), *etc*. Diverse saliency methods have been presented, such as surrogate-based methods (Ribeiro, Singh, and Guestrin 2016), gradient-based methods (Sundararajan, Taly, and Yan 2017), and attention-based methods (Choi



(a) Original radio signal.



(b) Learned attribution mask.



(c) Perturbed radio signal.

Figure 1: The attribution mask (b) highlights the importance of features in a radio signal sample (a) for human activity recognition (HAR). For an HAR model, THAT (Li et al. 2021), perturbing important features (c) significantly diminishes the predicted probability of activity "Fall", proving that the attribution mask can outline the human-related pattern recognized by the model.

et al. 2016). Recently, perturbation-based methods (Crabbé and Van Der Schaar 2021; Enguehard 2023) have become popular because they require less intervention in black-box models, using the variations in model outputs caused by input perturbations to produce *post-hoc* attributions. However, most saliency methods bias in favor of common modalities (*e.g.*, images, time series) (Ghassemi, Oakden-Rayner, and Beam 2021), while it is more challenging to highlight feature importance based on radio signals because: (1) Human-related patterns in radio signals are intertwined excessive noise. Existing methods depend on the empirical knowledge (*e.g.*, sparsity) of common modalities for attribution, but such knowledge cannot be intuitively observed in radio signals. Across radio signal samples, the sparsity of important features is uncertain and varying, making existing methods less effective, as discussed in Section 2; (2) Radio signals are initially designed for wireless communication in the frequency domain, and thus the spatial/time-domain perturbations for other modalities are less meaningful.

In this paper, we propose a Matryoshka-like saliency method, MatryMask, to explore feature attribution for human sensing with radio signals, as shown in Figure 1. To address the uncertain and varying sparsity of important features, MatryMask includes multiple masks to highlight salient areas at different scales. We regularize the beginning masks to capture large salient areas, while the subsequent masks focus on increasingly smaller salient areas. After applying mask-associated perturbations to inputs in a nest fashion, MatryMask highlights important features according to the variations in model outputs, adapting to the uncertainty and variability in radio signals. In line with the frequency-domain nature of radio signals, we devise a frequency-removal perturbation to remove principle frequency components from inputs, effectively perturbing human features that are essentially the frequency response of wireless channels (Wang et al. 2015). Finally, we aggregate all learned attribution masks into an optimal mask, which pinpoints significant *temporal channel elements* for human sensing and outlines human-related patterns recognized by models. Our main contributions are as follows:

- We propose MatryMask, a novel perturbation-based saliency method for human sensing with radio signals. To the best of our knowledge, this is the first attempt to study the explainability of WiFi-based human sensing.
- We build multiple masks in MatryMask to learn salient areas at different scales and perturb inputs in a nested fashion. We design a frequency-removal perturbation beyond existing spatial/time-domain perturbations.
- We conduct extensive experiments to demonstrate the attribution performance of MatryMask for three human sensing tasks. We also illustrate the effectiveness of MatryMask via quantitative analysis and ablation study.
- We further adopt a widely-used synthetic dataset (Enguehard 2023) based on a hidden Markov model (HMM) to prove the superiority of MatryMask in feature attribution for general time-series models (Appendix B).

2 Preliminary

Problem Formulation. Given a radio signal sample $\mathbf{X} \in \mathbb{R}^{C\times T}$ with C channels over T time steps, a human sensing model $f(\cdot)$ takes $\mathbf{X}$ as an input to yield an output $f(\mathbf{X})$. We regard $f(\cdot)$ as a *black-box* model whose intermediate layers and gradients are inaccessible. To highlight feature importance for $f(\cdot)$, a perturbation-based saliency method learns an attribution mask $\mathbf{M} \in \mathbb{R}^{C\times T}$ of the same dimensions as $\mathbf{X}$. Each element $m_{c,t} \in \mathbf{M}$ indicates the importance of feature $x_{c,t} \in \mathbf{X}$ for $f(\mathbf{X})$, where $c \in [1, 2, ..., C]$ is the channel index, and $t \in [1, 2, ..., T]$ denotes the time step. With $m_{c,t}$ in the range of $[0, 1]$, $m_{c,t}$ close to 1 means the feature at channel c and time t is important, while $m_{c,t}$ close to 0 means the feature at channel c and time t is unimportant.

Typically, perturbation-based saliency methods include two steps: (1) applying perturbations $\mu(\cdot)$ to the input based on the mask $\mathbf{M}$; (2) optimizing $\mathbf{M}$ according to the variations in the model output. Without loss of generality (Crabbé and Van Der Schaar 2021; Enguehard 2023), the perturbed input can be presented as $\Phi_{\mathbf{M}}(\mathbf{X}) = \mathbf{M}\mathbf{X} + (1-\mathbf{M})\mu(\mathbf{X})$. To optimize $\mathbf{M}$, the objective function normally includes: an *error* term $\mathcal{L}_e(f(\mathbf{X}), f(\Phi_{\mathbf{M}}(\mathbf{X})))$ and a *regularization* term $\mathcal{L}_r(\mathbf{M})$. The *error* term represents variations in the output $f(\Phi_{\mathbf{M}}(\mathbf{X}))$ with respect to the original output $f(\mathbf{X})$. The *regularization* term regulates the sparsity of $\mathbf{M}$. Overall, the objective of perturbation-based saliency method can be defined as: $\arg\min_{\mathbf{M}} \mathcal{L}_e(f(\mathbf{X}), f(\Phi_{\mathbf{M}}(\mathbf{X}))) + \mathcal{L}_r(\mathbf{M})$. We formulate this objective in the *preservation* mode and discuss the *deletion* mode in Appendix A.8.

Mask Sparsity. In perturbation-based saliency methods, a pivotal issue is how to ensure that *only* important features are highlighted, which is equivalent to regularizing mask sparsity in $\mathcal{L}_r(\mathbf{M})$. To account for this, DynaMask (Crabbé

and Van Der Schaar 2021) introduces a term $\mathcal{L}_a(\mathbf{M}) = \|\text{vecsort}(\mathbf{M}) - \mathbf{r}_a\|^2$ to highlight a proportion a of input features as salient areas. However, we do not have prior knowledge of the optimal proportion, since the human-related patterns in radio signals cannot be intuitively observed. ExtrMask (Enguehard 2023) adopts the l^1-norm to regularize mask sparsity as $\lambda_1\|\mathbf{M}\|_1$, but the choice of fixed λ_1 is challenged by the varying sparsity of important features across radio signal samples. Therefore, MatryMask builds multiple masks to learn salient areas at different scales adaptively.

Meaningful Perturbations. During mask optimization, the associated perturbations $\mu(\cdot)$ are applied to model inputs to yield perturbed outputs. Therefore, meaningful perturbations are key to high-quality attribution masks, such as Gaussian blurs for images (Fong and Vedaldi 2017; Fong, Patrick, and Vedaldi 2019) and moving averages for time series (Crabbé and Van Der Schaar 2021). ExtrMask (Enguehard 2023) further introduces a neural network $\text{NN}(\cdot)$ to generate perturbations, but it is vague about how to define the architecture of $\text{NN}(\cdot)$. The learning of $\text{NN}(\cdot)$ also leads to increased computational overhead. More importantly, there is a gap between the above spatial/time-domain perturbations and the frequency-domain nature of radio signals. To cope with these, we devise a frequency-removal perturbation to simply yet effectively perturb radio signals.

3 Methodology

3.1 MatryMask

We establish N masks $\mathbf{M}^+ = \{\mathbf{M}^n \in \mathbb{R}^{C\times T} | n \in [1, ..., N]\}$, where each $\mathbf{M}^n$ has the same dimensions as the input $\mathbf{X}$, as shown in Figure 2. $m^n_{c,t}$ in the range of $[0, 1]$ denotes an element of $\mathbf{M}^n$. In a nested fashion, we apply the perturbations associated with masks to generate a Matryoshka-like perturbed input as:

$$\Phi_{\mathbf{M}^n}(\mathbf{X}) = \begin{cases} \mathbf{M}^n\mathbf{X} + (1-\mathbf{M}^n)\,\mu(\mathbf{X}) & \text{if } n = 1, \\ \mathbf{M}^n\Phi_{\mathbf{M}^{n-1}}(\mathbf{X}) + & \\ \quad (1-\mathbf{M}^n)\,\mu(\Phi_{\mathbf{M}^{n-1}}(\mathbf{X})) & \text{if } n > 1. \end{cases} \quad (1)$$

Such formulations can be regarded as a generalization of DynaMask (Crabbé and Van Der Schaar 2021) and ExtrMask (Enguehard 2023), which are equivalent to MatryMask when $N = 1$. However, the single mask in DynaMask and ExtrMask can hardly determine feature importance of different degrees, because the elements in each mask are enforced to be near 0 or 1. Therefore, the idea behind using multiple masks is that *the features of higher importance will be highlighted by more masks* compared to less important features. By aggregating multiple masks, we can highlight the importance of features to varying degrees and thereby achieve better adaptability and attribution performance.

Mask Sparsity. On each mask $\mathbf{M}^n$, we use a regularization term $\mathcal{L}_{a^n}(\mathbf{M}^n)$ to highlight a proportion a^n of features as salient areas:

$$\mathcal{L}_{a^n}(\mathbf{M}^n) = \|\text{vecsort}(\mathbf{M}^n) - \mathbf{r}_{a^n}\|^2. \quad (2)$$

$\text{vecsort}(\mathbf{M}^n)$ sorts the elements in $\mathbf{M}^n$ in an ascending order, and $\mathbf{r}_{a^n} = [0, 0, ..., 1, 1]$ is a vector with $(1-a^n)\cdot C\cdot T$ zeros followed by $a^n \cdot C \cdot T$ ones. One of the primary novelties in MatryMask is that we alleviate the rigorous tuning of $a^n \in \mathbb{R}$ by setting $a^n \in [a^{\min}, a^{\max}]$. Accordingly, we can *simply* adopt a large range for a^n, such as $a^{\min} = 0.1$ and $a^{\max} = 0.9$, to cope with the uncertain sparsity of important features. Meanwhile, we define $a^{n-1} > a^n > a^{n+1}$ to make the initial masks cover large salient areas and the following masks capture increasingly focused salient areas, enabling MatryMask to adapt to varying sparsity of important features. For all masks in $\mathbf{M}^+$, the regularization term of MatryMask can be summarized as:

$$\mathcal{L}_r(\mathbf{M}^+) = \sum_{n=1}^{N} \mathcal{L}_{a^n}(\mathbf{M}^n). \quad (3)$$

After the model $f(\cdot)$ takes $\Phi_{\mathbf{M}^N}(\mathbf{X})$ as input to yield a perturbed output $f(\Phi_{\mathbf{M}^N}(\mathbf{X}))$, an error term $\mathcal{L}_e(f(\mathbf{X}), f(\Phi_{\mathbf{M}^N}(\mathbf{X})))$ encourages $\mathbf{M}^+$ to highlight important features that minimize the distance between original outputs and perturbed outputs. Typically (Crabbé and Van Der Schaar 2021; Enguehard 2023), $\mathcal{L}_e$ can be implemented using the cross-entropy loss for classification tasks or the squared error for regression tasks.

Mask Optimization. The overall objective of MatryMask is to optimize $\mathbf{M}^+$ by minimizing the loss function including the error term and the regularization term:

$$\arg\min_{\mathbf{M}^+} \mathcal{L}_e(f(\mathbf{X}), f(\Phi_{\mathbf{M}^N}(\mathbf{X}))) + \mathcal{L}_r(\mathbf{M}^+), \quad (4)$$

where $\mathcal{L}_e$ urges $\mathbf{M}^+$ to highlight important features that contribute to the original output, and $\mathcal{L}_r$ makes $\mathbf{M}^+$ adapt to uncertain and varying sparsity of important features. In Equation (4), we do not rely on any extra hyperparameters, such as λ_1 in ExtrMask (Enguehard 2023), to balance $\mathcal{L}_e$ and $\mathcal{L}_r$, because we can strengthen the adaptability of MatryMask by simply using *more* masks in $\mathbf{M}^+$ along with a *larger* range of $a^n \in [a^{\min}, a^{\max}]$. Compared to previous works, such formulations mitigate the laborious tuning of hyperparameters in mask optimization. For example, the attribution performance of ExtrMask (Enguehard 2023) is highly sensitive to differing hyperparameters, as shown in its experiments. We will further discuss the benefits of using more masks (*i.e.*, an increasing N) and a larger range of $a^n \in [a^{\min}, a^{\max}]$ via quantitative analysis in Section 4.2.

Mask Aggregation. After the optimization of $\mathbf{M}^+$, there are two strategies to aggregate all masks into an optimal mask $\mathbf{M}^*$. (1) The multiplicative strategy calculates the product of all masks in $\mathbf{M}^+$ as: $\mathbf{M}^* = \prod_{n=1}^{N} \mathbf{M}^n$. (2) The additive strategy calculates the summation of all masks in $\mathbf{M}^+$ as: $\mathbf{M}^* = \frac{1}{N}\sum_{n=1}^{N} \mathbf{M}^n$. Since the masks in $\mathbf{M}^+$ have learned salient areas at different scales, such aggregations integrate all masks to create an adaptive optimal attribution mask. We will further compare these two aggregation strategies via ablation study in Appendix A.7.

The optimal attribution mask $\mathbf{M}^*$ highlights the important features to pinpoint the significant temporal channel elements that contribute to human sensing, helping us understand *how models recognize human-related patterns from multiple radio channels over a time period*.

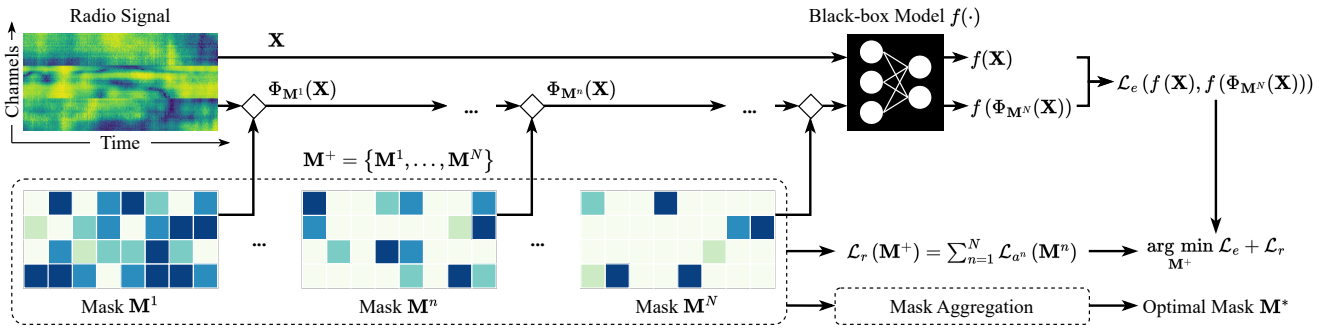


Figure 2: MatryMask constructs N masks to highlight the feature importance of an input $\mathbf{X}$ for the black-box model $f(\cdot)$ to produce an output $f(\mathbf{X})$. Each mask $\mathbf{M}^n$ is regularized to highlight a proportion a^n of features as salient areas.

3.2 Frequency-removal Perturbation

Radio signals are designed and modulated for wireless communication in the frequency domain, where useful implicit features are actually the frequency response to human interference (Wang et al. 2015). Therefore, for perturbation-based saliency methods based on radio signals, it necessitates perturbations in the frequency domain as a substitute for spatial/time-domain perturbations (Fong, Patrick, and Vedaldi 2019; Enguehard 2023). Recent studies (Zhou et al. 2022b,a; Yi et al. 2023) have also proved that frequency-domain analysis helps improve the performance of various models, while perturbation-based saliency methods in the frequency domain are still unexplored.

In response to this, we propose a frequency-removal perturbation to effectively perturb radio signals by removing principle frequency components from inputs. Despite the frequency-domain nature of radio signals, they are typically recorded in the time domain for convenience, and thus we use the discrete Fourier transform $\mathcal{F}(\cdot)$ to convert each time-domain input into the frequency domain. Given an input $\mathbf{X} \in \mathbb{R}^{C\times T}$ with C channels and T time steps, each channel is individually converted into the frequency domain, deriving $\mathbf{Z} = \mathcal{F}(\mathbf{X}) \in \mathbb{R}^{C\times K}$ with C channels and K frequency components, where K usually equals to T. Each $z_{c,k} \in \mathbf{Z}$ is transformed from $x_{c,t} \in \mathbf{X}$ by:

$$z_{c,k} = \sum_{t=1}^{T} x_{c,t} \cdot e^{\frac{-j\cdot 2\pi\cdot(t-1)}{T}(k-1)}, \tag{5}$$

where $k \in [1, 2, ..., K]$, and j is the imaginary unit. In each channel $\mathbf{Z}_c \in \mathbb{R}^K$ of $\mathbf{Z}$, we filter out $v \cdot K$ components of the largest amplitudes as $\text{top}_v(|\mathbf{Z}_c|) \in \mathbb{R}^{v\cdot K}$, where v is the removal ratio. According to $\text{top}_v(|\mathbf{Z}_c|)$, principle frequency components in $\mathbf{Z}$ are removed to yield $\mathbf{Z}' = \rho_v(\mathbf{Z}) \in \mathbb{R}^{C\times K}$, where each element $z'_{c,k} \in \mathbf{Z}'$ can be described as:

$$z'_{c,k} = \rho_v(z_{c,k}) = \begin{cases} 0 & |z_{c,k}| \in \text{top}_v(|\mathbf{Z}_c|) \text{ and } k \neq 1, \\ z_{c,k} & \text{otherwise.} \end{cases} \tag{6}$$

We *do not* remove $z_{c,1}$ because the purpose of frequency removal is to perturb useful implicit features in signal fluctuations, while $z_{c,1}$ represents the static component of each channel and does not involve signal fluctuations. Essentially, $z_{c,1}$ disentangles feature-agnostic information from each channel to build meaningful perturbations. After frequency removal, $\mathbf{Z}'$ represents the perturbations in the frequency domain, while in Equation (1) we need to apply perturbations in the input domain. Therefore, we use inverse discrete Fourier transform $\mathcal{F}^{-1}(\cdot)$ to convert it back into the time domain as $\mathbf{X}' = \mathcal{F}^{-1}(\mathbf{Z}') \in \mathbb{R}^{C\times T}$. Each element $x'_{c,t} \in \mathbf{X}'$ can be calculated by:

$$x'_{c,t} = \frac{1}{K}\sum_{k=1}^{K} z'_{c,k} \cdot e^{\frac{j\cdot 2\pi\cdot(k-1)}{K}(t-1)}. \tag{7}$$

Overall, we outline the frequency-removal perturbation as:

$$\mu(\mathbf{X}) = \mathbf{X}' = \mathcal{F}^{-1}(\rho_v(\mathcal{F}(\mathbf{X}))). \tag{8}$$

This method not only perturbs radio signals in the frequency domain effectively but also is compatible with any input domains, extending frequency-removal perturbations as plug-and-play modules to other perturbation-based saliency methods (Enguehard 2023; Liu et al. 2024; Lu et al. 2024).

4 Experiments

We use MatryMask to highlight feature importance for three WiFi-based human sensing tasks: human identification, localization, and activity recognition (HAR).

Datasets and Models. (1) For human identification, we implement a CNN-1D model (Wang et al. 2019) on the MM-Fi dataset (Yang et al. 2024). (2) For human localization, we employ an attention-based bidirectional LSTM (ABLSTM) (Chen et al. 2018) model on the ARIL dataset (Wang et al. 2019). (3) For HAR, we establish a THAT (Li et al. 2021) model on the Office dataset (Yousefi et al. 2017). Following previous works (Enguehard 2023), we shuffle the samples in each dataset and split them into 5 folds for cross-validation. More details of datasets and models are in Appendix A.1.

Attribution Baselines. We compare MatryMask with 8 saliency methods. (1) For surrogate-based methods, we adopt Lime (Ribeiro, Singh, and Guestrin 2016). (2) For gradient-based methods, we use Integral Gradient (IG) (Sundararajan, Taly, and Yan 2017) and GradSHAP (GS) (Lundberg et al. 2018). (3) For attention-based methods, we employ Retain (Choi et al. 2016). (4) For perturbation-based

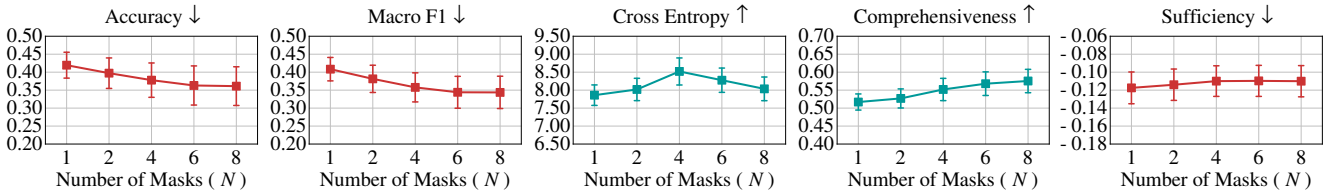
Task	Method	Acc. ↓	F1 ↓	CE ↑	Comp. ↑	Suff. ↓
Identification	Random	0.8250±0.0161	0.8212±0.0131	0.3742±0.0477	0.0079±0.0049	0.0689±0.0283
	Lime (Ribeiro, Singh, and Guestrin 2016)	0.8167±0.0167	0.8121±0.0178	0.4182±0.0460	0.0264±0.0051	0.0646±0.0244
	Retain (Choi et al. 2016)	0.8028±0.0264	0.7985±0.0263	0.4491±0.0515	0.0328±0.0131	0.0733±0.0220
	IG (Sundararajan, Taly, and Yan 2017)	0.7269±0.0250	0.7138±0.0165	0.5765±0.0346	0.1613±0.0057	-0.0163±0.0231
	GS (Lundberg et al. 2018)	0.7407±0.0217	0.7289±0.0219	0.5550±0.0390	0.1409±0.0039	-0.0071±0.0230
	FO (Suresh et al. 2017)	0.7019±0.0249	0.6861±0.0279	0.6798±0.0365	0.1981±0.0094	-0.0254±0.0263
	AFO (Tonekaboni et al. 2020)	0.6389±0.0234	0.6274±0.0283	0.9870±0.0445	0.2775±0.0098	-0.0628±0.0109
	DynaMask (Crabbé and Van Der Schaar 2021)	0.5861±0.0350	0.5724±0.0310	5.1746±0.5478	0.3130±0.0136	-0.1082±0.0163
	ExtrMask (Enguehard 2023)	0.5907±0.0343	0.5606±0.0436	2.0649±0.4456	0.3332±0.0443	-0.0580±0.0176
	MatryMask (Ours)	**0.3630**±0.0545	**0.3440**±0.0444	**8.2763**±0.3383	**0.5679**±0.0328	**-0.1097**±0.0173
Localization	Random	0.9849±0.0053	0.9835±0.0064	0.0698±0.0106	0.0131±0.0022	0.2927±0.0202
	Lime (Ribeiro, Singh, and Guestrin 2016)	0.9512±0.0092	0.9476±0.0123	0.1611±0.0347	0.0496±0.0088	0.3011±0.0160
	Retain (Choi et al. 2016)	0.9075±0.0354	0.8967±0.0467	0.3392±0.1654	0.0865±0.0345	0.2574±0.0252
	IG (Sundararajan, Taly, and Yan 2017)	0.8888±0.0122	0.8811±0.0143	0.4128±0.0598	0.1206±0.0127	0.2298±0.0215
	GS (Lundberg et al. 2018)	0.9125±0.0124	0.9094±0.0193	0.2966±0.0355	0.0926±0.0133	0.2512±0.0163
	FO (Suresh et al. 2017)	0.8759±0.0158	0.8651±0.0223	0.4463±0.0584	0.1256±0.0136	0.2335±0.0157
	AFO (Tonekaboni et al. 2020)	0.8329±0.0223	0.8154±0.0260	0.7923±0.1479	0.1699±0.0223	0.1939±0.0210
	DynaMask (Crabbé and Van Der Schaar 2021)	0.7855±0.0145	0.7631±0.0229	0.7888±0.0465	0.2109±0.0131	0.1819±0.0289
	ExtrMask (Enguehard 2023)	0.7733±0.0234	0.7489±0.0301	0.9608±0.1240	0.2243±0.0186	0.1418±0.0254
	MatryMask (Ours)	**0.6908**±0.0174	**0.6754**±0.0188	**1.3074**±0.1206	**0.3084**±0.0090	**0.0463**±0.0092
Activity Recognition	Random	0.9118±0.0074	0.8756±0.0163	0.0994±0.0130	0.0113±0.0024	0.1356±0.0130
	Lime (Ribeiro, Singh, and Guestrin 2016)	0.8846±0.0180	0.8349±0.0414	0.2262±0.0408	0.0470±0.0104	0.3711±0.0642
	Retain (Choi et al. 2016)	0.9068±0.0080	0.8649±0.0213	0.1401±0.0200	0.0200±0.0047	0.2208±0.0254
	IG (Sundararajan, Taly, and Yan 2017)	0.8821±0.0142	0.8316±0.0298	0.2181±0.0090	0.0651±0.0052	-0.0053±0.0055
	GS (Lundberg et al. 2018)	0.8402±0.0099	0.7940±0.0219	0.3812±0.0318	0.1100±0.0042	**-0.0060**±0.0044
	FO (Suresh et al. 2017)	0.8523±0.0113	0.8089±0.0285	0.3625±0.0207	0.1024±0.0029	0.0172±0.0071
	AFO (Tonekaboni et al. 2020)	0.7732±0.0182	0.7243±0.0332	0.7827±0.0721	0.1902±0.0145	-0.0022±0.0047
	DynaMask (Crabbé and Van Der Schaar 2021)	0.7218±0.0461	0.6557±0.0527	1.6118±0.2630	0.2445±0.0445	0.0873±0.0395
	ExtrMask (Enguehard 2023)	0.6305±0.0292	0.5499±0.0305	3.1275±0.1971	0.3717±0.0250	0.0680±0.0203
	MatryMask (Ours)	**0.5993**±0.0423	**0.5351**±0.0287	**9.0896**±0.9062	**0.4075**±0.0450	0.1228±0.0362

Table 1: Performance of saliency methods for three tasks based on radio signals. "Random" generates attribution masks from a uniform distribution over $[0, 1)$. ↑ indicates that the higher results are the better; ↓ indicates that the lower results are the better.

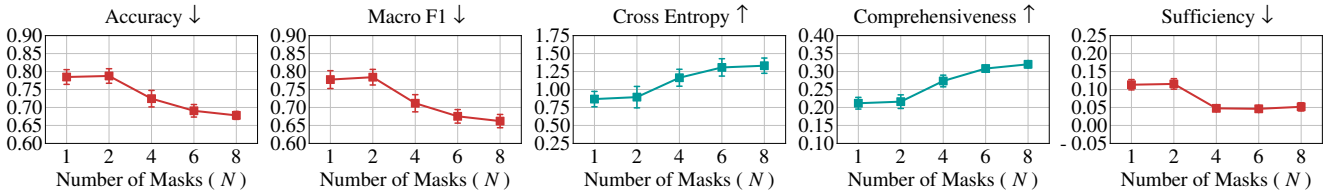
methods, we discuss Feature Occlusion (FO) (Suresh et al. 2017), Augmented FO (AFO) (Tonekaboni et al. 2020), DynaMask (Crabbé and Van Der Schaar 2021), and ExtrMask (Enguehard 2023), which have shown state-of-the-art attribution performance for time-series models. More details of baselines are introduced in Appendix A.2.

Evaluation Metrics. We follow previous works (Crabbé and Van Der Schaar 2021; Enguehard 2023; Lu et al. 2024) to measure the performance of saliency methods in two steps: (1) each method generates attribution masks; (2) a certain ratio of input features are perturbed to evaluate *if the masks have highlighted the most important features*. In step (2), all perturbed features are replaced with average values $\overline{x}_{c,t} = \frac{1}{T}\sum_{t=1}^{T} x_{c,t}$ for a fair comparison between different methods. For example, MatryMask only applies frequency-removal perturbations in step (1), while in step (2) its optimal attribution masks are evaluated *in the same way* as other methods. Specifically, we perturb 20% of most important features to measure Accuracy (Acc.), Marco F1, Cross Entropy (CE), and Comprehensiveness (Comp.), while we preserve 20% of most important features and perturb the others to measure Sufficiency (Suff.). Different perturbed/preserved ratios are discussed in Appendix A.5. We report the means and standard deviations of results via 5-fold cross-validation, following previous works (Enguehard 2023). More details of metrics are in Appendix A.3.

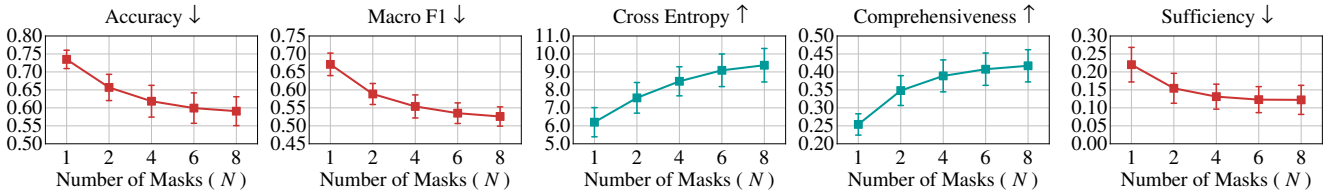
Implementation Details. Each black-box model is trained for 200 epochs, and more details of model training are presented in Appendix A.4. To compare with baselines (Section 4.1), we establish 6 masks ($N = 6$) in MatryMask, opt for $\left[a^{\min}, a^{\max}\right] = [0.2, 0.8]$ as the range of a^n, and use $v = 0.5$ as the removal ratio for frequency-removal perturbations. We schedule $\{a^n\}_{n=1}^{N}$ linearly from $a^1 = a^{\max}$ to $a^N = a^{\min}$. The quantitative analysis (Section 4.2) discusses a series of N values and different ranges of $a^n \in \left[a^{\min}, a^{\max}\right]$. We vary the removal ratio v for further evaluation in Appendix A.6. We initialize mask elements in MatryMask as 0.5 and use Adam (Kingma and Ba 2014) to train MatryMask for 200 epochs with a learning rate of 10^{-3}. For the classification tasks in human sensing, we use the cross-entropy loss to implement the error term $\mathcal{L}_e\left(f\left(\mathbf{X}\right), f\left(\Phi_{\mathbf{M}^N}(\mathbf{X})\right)\right)$. By default, we use the multiplicative strategy to aggregate masks, and the additive strategy is discussed in ablation study (Appendix A.7). We apply Equation (4) to optimize MatryMask in the preservation mode and discuss the deletion mode in Appendix A.8.



(a) Attribution for human identification using the CNN-1D model (Wang et al. 2019) on the MM-Fi dataset (Yang et al. 2024).



(b) Attribution for human localization using the ABLSTM model (Chen et al. 2018) on the ARIL dataset (Wang et al. 2019).



(c) Attribution for human activity recognition using the THAT model (Li et al. 2021) on the Office dataset (Yousefi et al. 2017).

Figure 3: Performance of MatryMask with respect to different numbers of masks (N). Red indicates that the lower results are the better; Cyan indicates that the higher results are the better.

4.1 Comparison Results

Table 1 shows that MatryMask achieves state-of-the-art attribution performance for three human sensing tasks based on radio signals. For human identification and localization, MatryMask consistently outperforms the second best methods by 10.6∼38.1% in Accuracy, 9.9∼38.7% in Macro F1, 36.0∼59.9% in Cross Entropy, 37.5∼70.6% in Comprehensiveness, and 1.8∼67.6% in Sufficiency. For human activity recognition (HAR), MatryMask is also the best attribution method on $4/5$ metrics except for Sufficiency. We ascribe this to the gap between using only 20% of important features to measure Sufficiency and the fact that there are indeed more important features for HAR. Due to the adaptability of MatryMask, it will highlight more than 20% of important features, and thus only preserving 20% of important features cannot reflect the advantages of MatryMask. This can be observed from the results in Supplemental Figure 1, where MatryMask attains increasingly better performance than other methods when more important features are preserved. Specifically, when we preserve 60% of most important features, MatryMask is superior to all other methods with the lowest Sufficiency of -0.021±0.008.

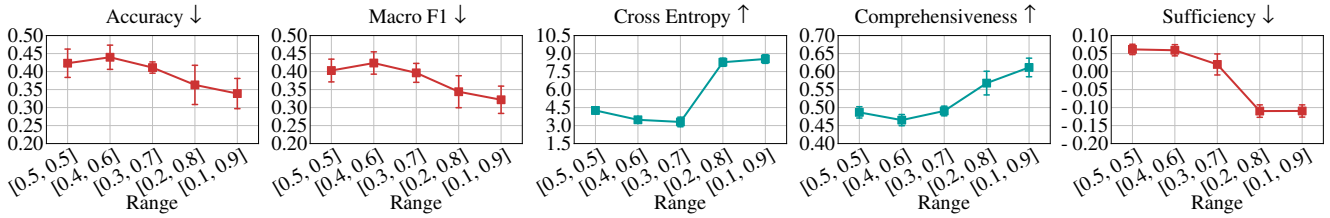
4.2 Quantitative Analysis

The Number of Masks (N). MatryMask aims to tackle the uncertain sparsity of important features by using multiple masks which highlight salient areas at different scales. Herein, we analyze the effect of using different numbers of masks $N \in [1, 2, 4, 6, 8]$ in MatryMask, as shown in Figure 3. We clearly see that MatryMask can achieve better performance when more masks (*i.e.*, an increasing N) are used. Meanwhile, when there are already multiple masks (*e.g.*, $N = 6$) in MatryMask, further increasing masks still improves the performance to a smaller extent. These results illustrate the effectiveness of multiple masks in MatryMask.

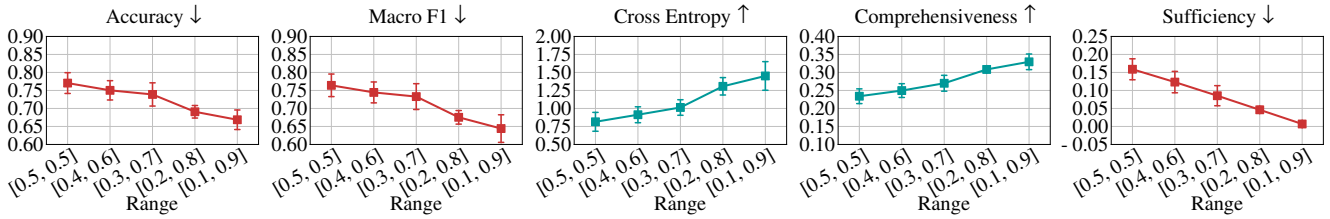
The Range of $a^n \in [a^{\min}, a^{\max}]$. To adapt to the varying sparsity of important features across instances, MatryMask adopts a range of $a^n \in [a^{\min}, a^{\max}]$. In Figure 4, we discuss the performance of MatryMask with respect to different ranges of $a^n \in [a^{\min}, a^{\max}]$ increasing from $[0.5, 0.5]$ to $[0.1, 0.9]$. Obviously, we can observe the increasing performance of MatryMask along with larger ranges, which enable MatryMask to learn salient areas at different scales without the rigorous tuning of a^n, proving the adaptability of MatryMask using larger ranges of $a^n \in [a^{\min}, a^{\max}]$.

5 Related Work

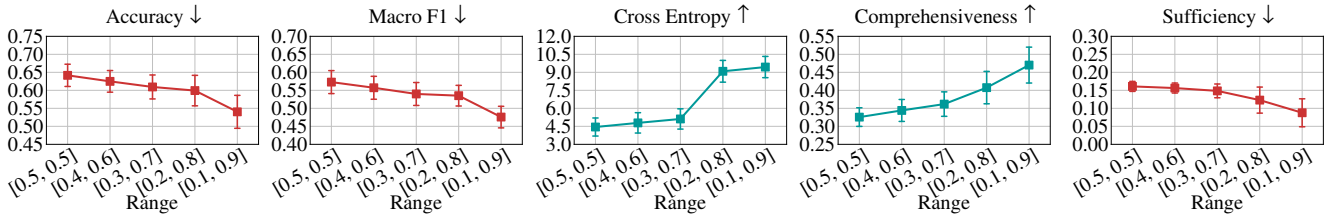
WiFi-based Human Sensing. Using the human features from WiFi CSI, three primary sensing tasks thrive, including human identification, localization, and activity recognition (HAR). (1) For human identification, initial methods have explored Multilayer Perceptrons (MLPs) (Yang et al. 2023), LSTM (Ding, Wang, and Fu 2020), and CNNs (Wang et al. 2022a; Yang et al. 2022a,b). Recently, CNN-LSTM hybrids (Mo and Kim 2021; Zhang et al. 2021; Kong et al. 2022) have outperformed other methods by exploiting the strengths of both LSTM and CNNs. (2) For human localization, the use of Sparse Auto-encoder (Gao et al. 2017),



(a) Attribution for human identification using the CNN-1D model (Wang et al. 2019) on the MM-Fi dataset (Yang et al. 2024).



(b) Attribution for human localization using the ABLSTM model (Chen et al. 2018) on the ARIL dataset (Wang et al. 2019).



(c) Attribution for human activity recognition using the THAT model (Li et al. 2021) on the Office dataset (Yousefi et al. 2017).

Figure 4: Performance of MatryMask with respect to different ranges of $a^n \in [a^{\min}, a^{\max}]$.

LSTM (Ding et al. 2022a), and CNNs (Wang et al. 2019) has made remarkable progress in localizing users based on WiFi CSI. (3) HAR is one of the most challenging tasks in WiFi-based human sensing. WiFi-based HAR models have employed MLPs (Zhang et al. 2018), LSTM (Yousefi et al. 2017), CNNs (Moshiri et al. 2021; Zhang et al. 2022, 2023), ABLSTM (Chen et al. 2018), *etc*. Combining attention layers with multi-scale CNNs, THAT (Li et al. 2021) builds a two-stream convolution augmented transformer to establish state-of-the-art WiFi-based HAR. However, WiFi-based human sensing deeply relies on black-box neural networks that lack model transparency and explainability.

Saliency Methods for Feature Attribution. Deep neural networks empower the rapid growth of Artificial Intelligence (AI), while their black-box nature catalyzes the research on explainable AI (XAI) for model transparency and explainability. Recent studies have delved into various saliency methods to analyze feature importance (Lundberg and Lee 2017) with respect to model outputs. (1) Surrogate-based methods approximate black-box models with transparent surrogates to estimate feature importance, such as Lime (Ribeiro, Singh, and Guestrin 2016) and TimeX (Queen et al. 2024), but it is challenging to choose appropriate transparent models as surrogates. (2) Gradient-based methods obtain feature importance from model gradients based on back-propagation (Ancona et al. 2018), such as Layer-wise Relevance Propagation (Bach et al. 2015), DeepLIFT (Shrikumar, Greenside, and Kundaje 2017), IG (Sundararajan, Taly, and Yan 2017), and GS (Lundberg et al. 2018). However, the intermediate layers and gradients of black-box models may be inaccessible in practice. (3) Attention-based methods use the inherent matrices in attention layers (Vaswani et al. 2017) to establish explainable models, such as Retain (Choi et al. 2016) and DSAN (Lin et al. 2020a), while their feasibility is limited since they depend on predefined attention layers which may be absent in black-box models. (4) Perturbation-based methods have gained increasing popularity since they can acquire model-agnostic and post-hoc attributions, such as FO, AFO (Tonekaboni et al. 2020), DynaMask (Crabbé and Van Der Schaar 2021), ExtrMask (Enguehard 2023), ContraLSP (Liu et al. 2024), and CGS-Mask (Lu et al. 2024). However, feature attribution remains unexplored in the field of human sensing with radio signals.

6 Conclusion

This paper presents MatryMask which highlights important features to help us understand how models recognize signal fluctuations caused by human interference. To tackle the uncertain and varying sparsity of important features, multiple masks in MatryMask are regularized to capture salient areas at different scales. Given the frequency-domain nature of radio signals, we formulate a frequency-removal perturbation to effectively perturb inputs. Experiments prove that MatryMask achieves state-of-the-art attribution performance in human identification, localization, and activity recognition.

References

Amann, J.; Blasimme, A.; Vayena, E.; Frey, D.; Madai, V. I.; and Consortium, P. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20: 1–9.

Ancona, M.; Ceolini, E.; Öztireli, C.; and Gross, M. 2018. Towards better understanding of gradient-based attribution methods for Deep Neural Networks. In *International Conference on Learning Representations*.

Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; and Samek, W. 2015. On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS ONE*, 10(7): e0130140.

Chen, C.; Zhou, G.; and Lin, Y. 2023. Cross-Domain WiFi Sensing with Channel State Information: A Survey. *ACM Computing Surveys*, 55(11): 1–37.

Chen, Z.; Zhang, L.; Jiang, C.; Cao, Z.; and Cui, W. 2018. WiFi CSI based passive human activity recognition using attention based BLSTM. *IEEE Transactions on Mobile Computing*, 18(11): 2714–2724.

Choi, E.; Bahadori, M. T.; Sun, J.; Kulas, J.; Schuetz, A.; and Stewart, W. 2016. Retain: An interpretable predictive model for healthcare using reverse time attention mechanism. *Advances in Neural Information Processing Systems*, 29.

Crabbé, J.; and Van Der Schaar, M. 2021. Explaining time series predictions with dynamic masks. In *International Conference on Machine Learning*, 2166–2177. PMLR.

Ding, J.; Wang, Y.; and Fu, X. 2020. Wihi: WiFi based human identity identification using deep learning. *IEEE Access*, 8: 129246–129262.

Ding, J.; Wang, Y.; Si, H.; Gao, S.; and Xing, J. 2022a. Three-Dimensional Indoor Localization and Tracking for Mobile Target Based on WiFi Sensing. *IEEE Internet of Things Journal*, 9(21): 21687–21701.

Ding, J.; Wang, Y.; Si, H.; Ma, J.; He, J.; Liang, K.; and Fu, S. 2022b. Multimodal Fusion-GMM based Gesture Recognition for Smart Home by WiFi Sensing. In *2022 IEEE 95th Vehicular Technology Conference*, 1–6. IEEE.

Enguehard, J. 2023. Learning perturbations to explain time series predictions. In *International Conference on Machine Learning*, 9329–9342. PMLR.

Fong, R.; Patrick, M.; and Vedaldi, A. 2019. Understanding deep networks via extremal perturbations and smooth masks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2950–2958.

Fong, R. C.; and Vedaldi, A. 2017. Interpretable explanations of black boxes by meaningful perturbation. In *Proceedings of the IEEE International Conference on Computer Vision*, 3429–3437.

Gao, Q.; Wang, J.; Ma, X.; Feng, X.; and Wang, H. 2017. CSI-based device-free wireless localization and activity recognition using radio image features. *IEEE Transactions on Vehicular Technology*, 66(11): 10346–10356.

Ge, Y.; Taha, A.; Shah, S. A.; Dashtipour, K.; Zhu, S.; Cooper, J.; Abbasi, Q. H.; and Imran, M. A. 2022. Contactless WiFi Sensing and Monitoring for Future Healthcare-Emerging Trends, Challenges, and Opportunities. *IEEE Reviews in Biomedical Engineering*, 16: 171–191.

Ghassemi, M.; Oakden-Rayner, L.; and Beam, A. L. 2021. The false hope of current approaches to explainable artificial intelligence in health care. *The Lancet Digital Health*, 3(11): e745–e750.

Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; and Hussain, A. 2024. Interpreting black-box models: a review on explainable artificial intelligence. *Cognitive Computation*, 16(1): 45–74.

Huang, Q.; Chen, H.; and Zhang, Q. 2020. Joint design of sensing and communication systems for smart homes. *IEEE Network*, 34(6): 191–197.

Hussain, Z.; Sheng, Q. Z.; and Zhang, W. E. 2020. A review and categorization of techniques on device-free human activity recognition. *Journal of Network and Computer Applications*, 167: 102738.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kong, H.; Lu, L.; Yu, J.; Zhu, Y.; Tang, F.; Chen, Y.-C.; Kong, L.; and Lyu, F. 2022. Push the limit of wifi-based user authentication towards undefined gestures. In *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*, 410–419. IEEE.

Li, B.; Cui, W.; Wang, W.; Zhang, L.; Chen, Z.; and Wu, M. 2021. Two-stream convolution augmented transformer for human activity recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 286–293.

Lin, H.; Bai, R.; Jia, W.; Yang, X.; and You, Y. 2020a. Preserving dynamic attention for long-term spatial-temporal prediction. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 36–46.

Lin, Y.; Gao, Y.; Li, B.; and Dong, W. 2020b. Revisiting indoor intrusion detection with WiFi signals: do not panic over a pet! *IEEE Internet of Things Journal*, 7(10): 10437–10449.

Liu, Z.; Zhang, Y.; Wang, T.; Wang, Z.; Luo, D.; Du, M.; Wu, M.; Wang, Y.; Chen, C.; Fan, L.; and Wen, Q. 2024. Explaining Time Series via Contrastive and Locally Sparse Perturbations. In *Proceedings of the 12th International Conference on Learning Representations*, 1–21.

Lu, F.; Li, W.; Sun, Y.; Song, C.; Ren, Y.; and Zomaya, A. Y. 2024. CGS-Mask: Making Time Series Predictions Intuitive for All. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 14149–14157.

Lundberg, S. M.; and Lee, S.-I. 2017. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30.

Lundberg, S. M.; Nair, B.; Vavilala, M. S.; Horibe, M.; Eisses, M. J.; Adams, T.; Liston, D. E.; Low, D. K.-W.; Newman, S.-F.; Kim, J.; et al. 2018. Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering*, 2(10): 749–760.

Mo, H.; and Kim, S. 2021. A deep learning-based human identification system with wi-fi csi data augmentation. *IEEE Access*, 9: 91913–91920.

Moshiri, P. F.; Nabati, M.; Shahbazian, R.; and Ghorashi, S. A. 2021. CSI-Based Human Activity Recognition using Convolutional Neural Networks. In *2021 11th International Conference on Computer Engineering and Knowledge*, 7–12. IEEE.

Nirmal, I.; Khamis, A.; Hassan, M.; Hu, W.; and Zhu, X. 2021. Deep learning for radio-based human sensing: Recent advances and future directions. *IEEE Communications Surveys & Tutorials*, 23(2): 995–1019.

Queen, O.; Hartvigsen, T.; Koker, T.; He, H.; Tsiligkaridis, T.; and Zitnik, M. 2024. Encoding time-series explanations through self-supervised model behavior consistency. *Advances in Neural Information Processing Systems*, 36.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

Shrikumar, A.; Greenside, P.; and Kundaje, A. 2017. Learning important features through propagating activation differences. In *International Conference on Machine Learning*, 3145–3153. PMLR.

Sundararajan, M.; Taly, A.; and Yan, Q. 2017. Axiomatic attribution for deep networks. In *International Conference on Machine Learning*, 3319–3328. PMLR.

Suresh, H.; Hunt, N.; Johnson, A.; Celi, L. A.; Szolovits, P.; and Ghassemi, M. 2017. Clinical intervention prediction and understanding using deep networks. *arXiv preprint arXiv:1705.08498*.

Tan, S.; Ren, Y.; Yang, J.; and Chen, Y. 2022. Commodity WiFi Sensing in Ten Years: Status, Challenges, and Opportunities. *IEEE Internet of Things Journal*, 9(18): 17832–17843.

Tian, Z.; Li, Y.; Zhou, M.; and Li, Z. 2018. WiFi-based adaptive indoor passive intrusion detection. In *2018 IEEE 23rd International Conference on Digital Signal Processing*, 1–5. IEEE.

Tonekaboni, S.; Joshi, S.; Campbell, K.; Duvenaud, D. K.; and Goldenberg, A. 2020. What went wrong and when? Instance-wise feature importance for time-series black-box models. *Advances in Neural Information Processing Systems*, 33: 799–809.

Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, Ł.; and Polosukhin, I. 2017. Attention is all you need. *Advances in Neural Information Processing Systems*, 30.

Wang, D.; Yang, J.; Cui, W.; Xie, L.; and Sun, S. 2022a. CAUTION: A Robust WiFi-Based Human Authentication System via Few-Shot Open-Set Recognition. *IEEE Internet of Things Journal*, 9(18): 17323–17333.

Wang, F.; Feng, J.; Zhao, Y.; Zhang, X.; Zhang, S.; and Han, J. 2019. Joint activity recognition and indoor localization with WiFi fingerprints. *IEEE Access*, 7: 80058–80068.

Wang, J.; Varshney, N.; Gentile, C.; Blandino, S.; Chuang, J.; and Golmie, N. 2022b. Integrated sensing and communication: Enabling techniques, applications, tools and data sets, standardization, and future directions. *IEEE Internet of Things Journal*, 9(23): 23416–23440.

Wang, W.; Liu, A. X.; Shahzad, M.; Ling, K.; and Lu, S. 2015. Understanding and modeling of wifi signal based human activity recognition. In *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, 65–76.

Wei, B.; Hu, W.; Yang, M.; and Chou, C. T. 2015. Radio-based device-free activity recognition with radio frequency interference. In *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*, 154–165.

Yang, J.; Chen, X.; Zou, H.; Lu, C. X.; Wang, D.; Sun, S.; and Xie, L. 2023. SenseFi: A library and benchmark on deep-learning-empowered WiFi human sensing. *Patterns*, 4(3).

Yang, J.; Chen, X.; Zou, H.; Wang, D.; and Xie, L. 2022a. AutoFi: Towards automatic WiFi human sensing via geometric self-supervised learning. *IEEE Internet of Things Journal*.

Yang, J.; Chen, X.; Zou, H.; Wang, D.; Xu, Q.; and Xie, L. 2022b. EfficientFi: Toward large-scale lightweight WiFi sensing via CSI compression. *IEEE Internet of Things Journal*, 9(15): 13086–13095.

Yang, J.; Huang, H.; Zhou, Y.; Chen, X.; Xu, Y.; Yuan, S.; Zou, H.; Lu, C. X.; and Xie, L. 2024. Mm-fi: Multi-modal non-intrusive 4d human dataset for versatile wireless sensing. *Advances in Neural Information Processing Systems*, 36.

Yang, Z.; Zhang, Y.; and Zhang, Q. 2022. Rethinking Fall Detection With Wi-Fi. *IEEE Transactions on Mobile Computing*.

Yi, K.; Zhang, Q.; Fan, W.; Wang, S.; Wang, P.; He, H.; An, N.; Lian, D.; Cao, L.; and Niu, Z. 2023. Frequency-domain MLPs are More Effective Learners in Time Series Forecasting. In *Advances in Neural Information Processing Systems*, volume 36, 76656–76679.

Yousefi, S.; Narui, H.; Dayal, S.; Ermon, S.; and Valaee, S. 2017. A survey on behavior recognition using WiFi channel state information. *IEEE Communications Magazine*, 55(10): 98–104.

Zhang, J.; Tang, Z.; Li, M.; Fang, D.; Nurmi, P.; and Wang, Z. 2018. CrossSense: Towards cross-site and large-scale WiFi sensing. In *Proceedings of the 24th Annual International Conference on Mobile Computing and Networking*, 305–320.

Zhang, R.; Jiang, C.; Wu, S.; Zhou, Q.; Jing, X.; and Mu, J. 2022. Wi-fi sensing for joint gesture recognition and human identification from few samples in human-computer interaction. *IEEE Journal on Selected Areas in Communications*, 40(7): 2193–2205.

Zhang, R.; Wu, S.; Jiang, C.; Cui, Y.; and Jing, X. 2021. WirelessID: Device-free human identification using gesture signatures in CSI. In *2021 IEEE 94th Vehicular Technology Conference*, 1–4. IEEE.

Zhang, Y.; Yin, Y.; Wang, Y.; Ai, J.; and Wu, D. 2023. CSI-based location-independent Human Activity Recognition with parallel convolutional networks. *Computer Communications*, 197: 87–95.

Zhou, T.; Ma, Z.; Wen, Q.; Sun, L.; Yao, T.; Yin, W.; Jin, R.; et al. 2022a. Film: Frequency improved legendre memory model for long-term time series forecasting. *Advances in Neural Information Processing Systems*, 35: 12677–12690.

Zhou, T.; Ma, Z.; Wen, Q.; Wang, X.; Sun, L.; and Jin, R. 2022b. Fedformer: Frequency enhanced decomposed transformer for long-term series forecasting. In *International Conference on Machine Learning*, 27268–27286. PMLR.

Zhou, Y.; Huang, H.; Yuan, S.; Zou, H.; Xie, L.; and Yang, J. 2023. MetaFi++: WiFi-Enabled Transformer-based Human Pose Estimation for Metaverse Avatar Simulation. *IEEE Internet of Things Journal*.

Zhu, G.; Wu, C.; Zeng, X.; Wang, B.; and Liu, K. R. 2022. Who Moved My Cheese? Human and Non-human Motion Recognition with WiFi. In *2022 IEEE 19th International Conference on Mobile Ad Hoc and Smart Systems*, 476–484. IEEE.