Enhancing RFID Observations Using Masked Autoencoders for Improved Indoor Localization

Pritom Dutta[§], Jian Zhang[§], Bernard Amoah[†], Xiangyu Wang[‡], *Kanchon Kanti Podder, Shiwen Mao[†]
[†]Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201, USA
[‡]RFID Lab, Auburn University, Auburn, AL 36849, USA
[§]Department of Information Technology, Kennesaw State University, Marietta, GA 30144, USA
*Department of Electrical and Computer Engineering, Kennesaw State University, Marietta, GA 30060, USA
Email: {pdutta2, kpodder}@students.kennesaw.edu, {jianzhang, smao}@ieee.org, {bza0066, xzw0042}@auburn.edu

***Abstract*—This paper presents the MAE model that uses a Masked AutoEncoder (MAE) to enhance the observations from commercial passive Radio-Frequency Identification (RFID) devices. It is crucial to address the common issue of RFID readers failing to collect observations from all their hop channels and antennas due to environmental effects and device limitations. The proposed method examines the inner rationale among observations from various channels and antennas to reconstruct the missing observations, which can significantly improve the performance of downstream applications. The experiment results show that when we collect more than 70% observation in all antennas at all channels, our MAE model can restore 90% of the missing phase with an error of less than 0.1 radians, which is even less than the error caused by thermal noise in an RFID system. Our MAE model's accuracy in restoring missing data provides a promising future to improve the performance of various RFID applications like localization and motion tracking by providing more complete observations.**

***Index Terms*—Radio-frequency identification (RFID), Ultra-high frequency (UHF), Masked AutoEncoder (MAE), Observation restoration.**

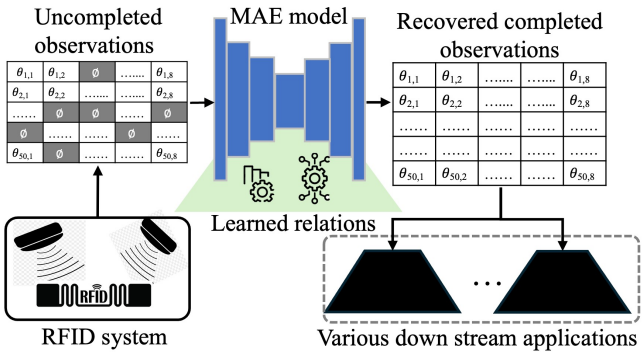

Fig. 1. The architecture of the proposed MAE-based RFID observation enhancement. It illustrates the encoder-decoder structure, where the input is a matrix of RFID phase observations from different antennas and hopping channels. Missing observations, shown as grey blocks, are caused by environmental noise, channel hopping, etc. It facilitates learning the latent relationships among antennas and channels in a self-supervised manner. The model then restores missing values by leveraging these learned relationships and outputs a completed observation matrix to advance the downstream applications.

I. Introduction

Over the past few years, the Radio-frequency identification (RFID) technique, especially the Ultra-high Frequency (UHF) RFID, has been studied and deployed as RFID sensing from various applications [1]–[4], such as ambient environmental sensing, real-time tag localization, and even vital sign monitoring. Unlike the promising results in a lab setting, we saw an obvious performance decrease when they were deployed in real environments. One key reason for this decrease is the loss of RF observations: a UHF RFID reader works with multiple antennas and hops on multiple channels (i.e., 50 hopping channels are mandatory in the USA). However, the dynamic environments and hardware limitations make a reader fail to collect observations from all its antennas at all hopping channels, a common phenomenon of observation sparsity in RFID systems. We must address this problem effectively and efficiently to materialize the promising future of various RFID sensing applications. Several existing works [5], [6] leverage tensor completion for channel information recovery to mitigate the effect of RFID observation sparsity; however, they fail to harness the latent relations among different antennas.

This paper proposes a method to systematically explore the latent relations of hopping channels and antennas in RFID observations. Then, an MAE [7] will be developed to capture these relations and utilize them to recover the missing observations at each channel and antenna. First, we will collect RFID observations and then randomly mask out values at channels and antennas; then, we use MAE to train our model in an unsupervised manner to reconstruct masked-out items, which allows the model to learn the aforementioned latent relations. Later, the well-trained model can be deployed to recover all missing items in an RFID observation to complete it (e.g., the phase value for all 50 channels at all connected antennas). We also evaluate our method by integrating it with an RF Hologram RFID localization, which is introduced in work [8]. Extensive experiment results demonstrate the superior performance of the proposed method in RFID observation restoration and advance the performance of downstream models.

II. Proposed Approach

A. overview of the system

Fig 1 depicts the architecture of the proposed MAE-based RFID observation enhancement. A well-trained MAE model will retain the latent relations among different channels and antennas. Then, it utilizes these relations and the original observation from a commercial RFID reader to recover the

missing items that improve the performance of downstream applications, such as tracking, localization, and other RF sensing. In the following sections, we focus on phase observation, but the proposed method is compatible with other observations, such as the Received Signal Strength Indicator (RSSI).

B. Data processing & MAE model

1) Data Preparation: A typical RFID system comprises $A \geq 1$ antennas to interrogate tags in N hop channels. Our work is considering an RFID reader with $A = 8$ antennas and works on $N = 50$ channels, which is regulated by the FCC standard in the USA. To capture the relationship between the antennas and channels, we represent the phase observations using a matrix, Θ, which captures the phase shift of RFID signals as they are transmitted and reflected back to the antennas. In this work, Θ would be a 50×8 matrix, where each row corresponds to the readings across all antennas for a given channel, and each column corresponds to readings across all channels for a given antenna. This matrix structure allows us to explore how observations are related not only within a single channel or antenna but also across different channels and antennas, revealing hidden patterns in the data. With this understanding, we now represent the completed phase observations as:

$$\mathbf{\Theta} = \begin{bmatrix} \theta_{1,1} & \theta_{1,2} & \cdots & \theta_{1,A} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{N,1} & \theta_{N,2} & \cdots & \theta_{N,A} \end{bmatrix} \tag{1}$$

Here, $\theta_{i,j} \in [-\pi, \pi]$ denotes the phase observation from channel i and antenna j. Due to the experimental effect and the channel hopping mechanism(i.e., according to FCC standards, a reader can take up to 20 seconds to go through all channels), we usually can not collect all these observations at a short interval, especially for motion-tracking applications.

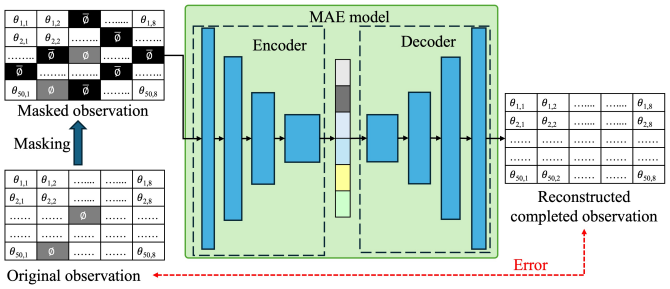

Fig. 2. The proposed MAE model and its training process.

2) MAE Model: The proposed MAE model deploys an encoder-decoder architecture, which is illustrated as the green block in Fig. 2. It comprises an encoder and a decoder with similar but mirroring layers. Due to our small data dimension $N \times A$ (e.g., 50×8), the encoder and decoder can be implemented with a relatively shallow architecture, like $N < 10$ layers. The encoder captures the phase relation among different channels and antennas, creating a compressed and efficient latent representation. All RFID phases constitute the structured matrix $\mathbf{\Theta}$, in which the relative position of phases offers meaningful relation according to channels and antenna displacement. Thus, the encoder should be able to retain these positional features. To this end, the encoder employs Convolutional Neural Network (CNN) layers that have been proven to be fundamental in efficiently capturing both spatial hierarchies and dependencies [9]. Then, the decoder will reconstruct the observation matrix $\hat{\Theta}$ based on the latent representation from the encoder. The decoder also employs CNN layers that are similar but mirror of the structure of the encoder.

C. Model training

Our encoder-decoder architecture facilitates a self-supervised training process: We mask out parts of the original observation $\mathbf{\Theta}$ and allow the MAE model to reconstruct the completed observation $\hat{\mathbf{\Theta}}$. To minimize the error between the $\mathbf{\Theta}$ and $\hat{\mathbf{\Theta}}$, we can empower the MAE model to learn the latent relations in RFID observation without manual labeling. We illustrate such a training process in Fig. 2.

1) Training Dataset: Our training dataset is collected for a testbed where we placed an RFID tag in $\mathcal{P}$ number of positions to collect as many phase values as possible at each position with 2-minute intervals. Due to the environments, reader, and tag characteristics, we will miss some phase values for some antennas at several channels. We mark those phase values as undetected value ϕ in a given observation $\mathbf{\Theta}$, which is highlighted as grey blocks of observations in Fig. 2. During the training process, a masked observation $\mathbf{\Theta}_m$ will be created from $\mathbf{\Theta}$ by randomly masking out a set of $\theta_{i,j}$ and mark them as $\overline{\phi}$, which is highlighted as black blocks of masked observation in Fig. 2. Then, we will input $\mathbf{\Theta}_m$ to reconstruct the $\hat{\mathbf{\Theta}}$ by minimizing the error between the $\hat{\mathbf{\Theta}}$ and $\mathbf{\Theta}$ to enable the MAE model to learn the relations among the channels and antennas.

2) Custom Loss Function: To enable our MAE model to recover the missing phase ϕ in the original observation $\mathbf{\Theta}$, we design a custom loss function to handle masked and unmasked phases separately. We first define the *contributed error*,$\nabla\theta_{i,j}$, at position i, j between original $\mathbf{\Theta}$ and reconstructed observation $\hat{\mathbf{\Theta}}$ to rule out the effect of original missing phases ϕ as:

$$\nabla\theta_{i,j} = \begin{cases} 0, & \text{if } \theta_{i,j} = \phi \\ |\theta_{i,j} - \hat{\theta}_{i,j}|, & \text{otherwise} \end{cases} \tag{2}$$

here, $|\cdot|$ denotes the absolute operation, and $\hat{\theta}_{i,j}$ is the reconstructed phase at position (i, j) of $\hat{\mathbf{\Theta}}$. We define the masked loss function $\mathcal{L}^m$ as:

$$\mathcal{L}^m = \sum_{\substack{i,j \\ \theta^m_{i,j} \neq \overline{\phi}}} \nabla\theta_{i,j} \tag{3}$$

Here, $\theta^m_{i,j}$ is the phase at position i, j of the masked observation $\mathbf{\Theta}_m$. Thus $\mathcal{L}^m$ sums up all the contributed errors at all positions that are masked out. Then, the unmasked loss function $\mathcal{L}^u$ is defined as :

$$\mathcal{L}^u = \sum_{\substack{i,j \\ \theta^m_{i,j} = \overline{\phi}}} \nabla\theta_{i,j} \tag{4}$$

Finally, the final loss function is calculated as a weighted average of the masked and unmasked losses :

$$\mathcal{L} = \alpha \mathcal{L}^m + (1 - \alpha)\mathcal{L}^u \tag{5}$$

This allows the loss function $\mathcal{L}$ to effectively combine the information from both masked and unmasked regions, balancing their contributions to the overall loss based on the parameter $0 < \alpha < 1$.

TABLE I
HYPERPARAMETERS

Hyperparameter	Value
Batch Size	128
Initial Learning Rate	0.001
Learning Rate Drop Factor	0.1
Learning Rate Drop Patience	50
Number of Epochs	100
Early Stopping Patience	100
Early Stopping Monitor	Validation Loss
Early Stopping Mode	min
Optimizer	Adam

III. Experimental Study

A. Experiment Setup

a) Testbed and data collection: In this work, we set up a testbed at our Lab to collect a training dataset and evaluate our system's performance. We deployed a Zebra FX9600 RFID reader and 8 Zebra AN720 S9025PR antennas; an AD-387 U9 UHF RFID tag will be placed in an area of $1.5m \times 1.5m$. During the data collection for training, the area was divided into grids with side lengths of 10cm. Then, we acquired continuously for 2 minutes at each position to ensure the reader operated at all 50 channels, with multiple repetitions conducted to ensure consistency and reliability of the data. Note that even with 2 minutes of interrogation, a non-trivial portion of samples missed several channels for different antennas. This resulted in a dense dataset $\mathcal{D}_d$ with 110 k samples. These samples were collected from 110 distinct positions that were randomly selected in the aforementioned grids. The majority of these samples consisted of more than 90% of phases for all the 8 antennas at 50 channels. While at the evaluation phase, we randomly placed the tag within the area and interrogated it with a randomly $t < 2s$ interval. This usually ended with sparse observations that missed many phases for each antenna, many of them with 60% or more of 8×50 phases. These observations formed our sparse dataset $\mathcal{D}_s$.

b) MAE configuration: We implemented the proposed MAE model using the PyTorch Lightning package and Python 3.8.0. TABLE I summarizes the hyperparameters for our MAE model. The model was trained, and inference was performed on a Dell workstation with an Nvidia RTX 3090 (24GB) and 128GB of RAM. The system takes approximately 7 ms to restore a completed observation $\hat{\mathbf{\Theta}}$.

B. Experiment Results and Analysis

1) Performance of phase restoration at a single position: In this benchmark experiment, we evaluated the performance of our proposed MAE model by comparing it with an existing method, tensor completion, which is introduced in [10]. Tensor completion was originally designed to restore missing data for temperature measurements and only consider a statistical scenario at a single point. To ensure the fairness of comparison, we first trained our MAE model with the data collected at the same position. Then, we evaluated the reconstructed error at observations with 90%, 70%, and 50% phase values for the 8 antennas at 50 channels. We tested each scenario with 200 distinctive samples. Then, we compared them with Tensor completion in terms of their average and maximum error, which is shown in TABLE II. These results show that our MAE model can provide a slightly better average restoration errors at 90% and 70% scenarios with the same level at 50% scenarios. Considering that the thermal or other noise in the RFID system can cause a 0.1 rad error, our average errors in all scenarios can accurately restore the missing phase values. Compared to tensor completion, our most significant improvement lies in the max error, especially for 90% and 70% scenarios, even if the maximum error is less than the thermal error. This demonstrated the superior robustness of the proposed MEA model in stationary RFID sensing applications, such as ambient temperature monitoring, where the tags and reader are both stationary but need to probe with short intervals to ensure the monitoring sensitivity and real-time performance.

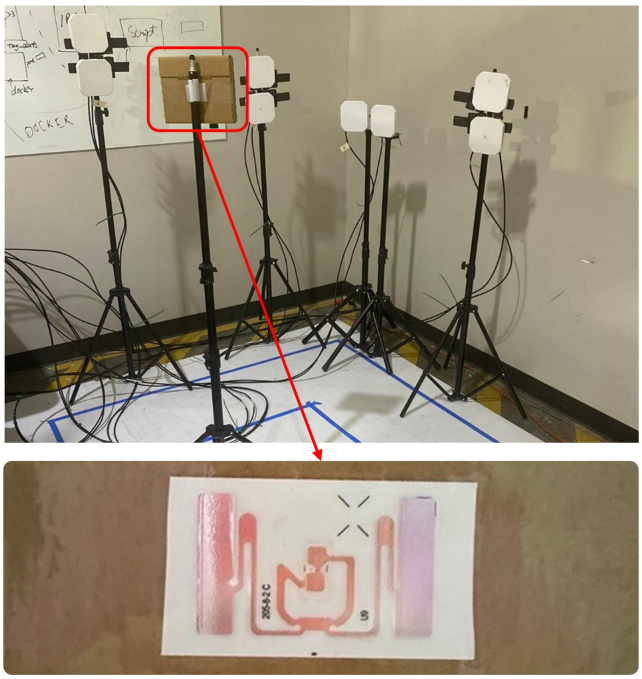

Fig. 3. Our testbed for experiments. The top is the layout of the RFID antennas, and the high adjustable tripod in the middle holds the RFID tag. The bottom is the zoom-out of the RFID tag used in experiments.

2) Performance of phase restoration in an open space: Our MAE model is designed to restore the missing RFID observation for tags that are placed at any position of the reader's reading range. In this experiment, we evaluated its accuracy in phase restoration in the entire testbed, a space with $1.5m \times 1.5m$. To this end, we divided the dense dataset

TABLE II
COMPARISON OF RESTORATION ERRORS AT SINGLE POSITION

	Tensor Completion		Proposed MAE Model	
	Average error	Max error	Average error	Max error
90%	0.016 rad	5.109 rad	0.003 rad	0.005 rad
70%	0.018 rad	5.055 rad	0.011 rad	0.016 rad
50%	0.022 rad	6.135 rad	0.025 rad	0.405 rad

$\mathcal{D}_d$ into three subsets: $\mathcal{D}_t$, $\mathcal{D}_v$, and $\mathcal{D}_e$. Here, $\mathcal{D}_t$ was the training dataset with 70% of samples in $\mathcal{D}_d$ to train our MAE model, and selected $\mathcal{D}_v$ 10% of samples to validate the training process for tuning hyperparameters, detecting overfitting, and guiding improvements to the MAE model. $\mathcal{D}_e$ was the evaluation dataset, which with 20% of samples in $\mathcal{D}_d$. We select the samples in $\mathcal{D}_e$ that were totally different from those in $\mathcal{D}_t$ to ensure the trained MAE model was not seen as an observation in $\mathcal{D}_e$. We deleted the phase values in every sample of $\mathcal{D}_e$ to make them only remain 90%, 80%, 70%, and 60% of observations. These scenarios were selected based on the fact that most RFID observations consisted of more than 60% values for many RFID sensing applications, such as slow motion capture. Then, we inputted all the above observations into our well-trained MAE model and compared the recovered phase values with the phases in original $\mathcal{D}_e$. Fig. 4 presents the cumulative distribution function (CDF) of restoration errors between the true phases in original $\mathcal{D}_e$ and the recovered phases.

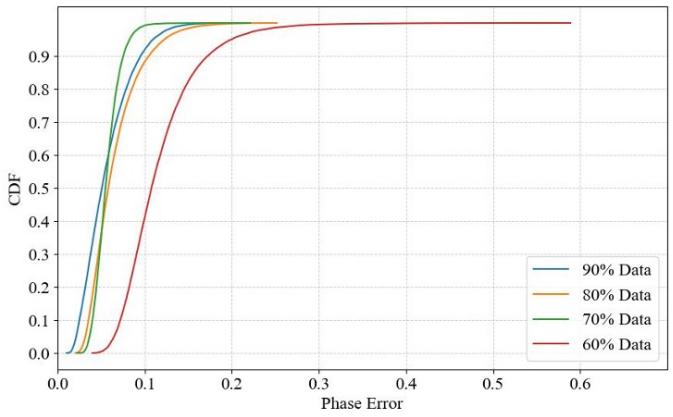

Fig. 4. CDF of phase errors between the truth and recovered phases provided by the proposed MAE model.

In Fig. 4, for scenarios with 70%, 80%, and 90% available data, more than 60% of restored error is less than 0.05 rad. In all of these scenarios, 90% of the restoration errors are less than 0.1 radians, which is the same thermal noise level. Moreover, in the scenario with 60% of the data, 90% of the restoration errors are less than 0.17 radians, only slightly higher than the thermal noise level. Considering many downstream RFID applications [8], [11] designed to handle thermal level noises, our MAE model can smoothly be integrated with them to offer restored observations and significantly improve their performance by completing observations.

IV. Conclusions

In this study, we presented a self-supervised MAE model to capture the latent relations among channels and antennas in RFID observations. The proposed method efficiently learns these relations without any manual labeling or manipulation of the data by providing raw samples from an RFID system. Subsequently, a trained model can restore missing values to form a completed observation from all antennas at all channels. Our proposed MAE model is a task-agnostic method, which only depends on the RFID systems and does not require any information on the downstream applications. Such characteristic enables it to be smoothly integrated with many existing or emerging RFID applications to improve their performance through comprehensive observations. Extensive experiments demonstrate that our method could restore missing phases up to 40%(that is our 60% data in Fig. 4) and offer an error within the thermal noise, demonstrating its effectiveness and superior accuracy. Our future work plans to integrate it with downstream applications, such as localization, motion tracking, and environmental monitoring. We will explore the captured latent RFID relations with downstream task-dependent data to further advance the performance of these applications.

Acknowledgment

This work is supported in part by the NSF under Grants CCSS-2245607 and CCSS-2245608.

References

[1] J. Zhang, Z. Yu, X. Wang, Y. Lyu, S. Mao, S. C. Periaswamy, J. Patton, and X. Wang, "RFHUI: An intuitive and easy-to-operate human-UAV interaction system for controlling a UAV in a 3D space," in *Proc. EAI MobiQuitous 2018*, New York City, NY, Nov. 2018, pp. 1–8.

[2] X. Wang, J. Zhang, Z. Yu, S. Mao, S. Periaswamy, and J. Patton, "On remote temperature sensing using commercial UHF RFID tags," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 715–10 727, Dec. 2019.

[3] X. Wang, J. Zhang, S. Mao, S. C. Periaswamy, and J. Patton, "A framework for locating multiple rfid tags using rf hologram tensors," *Digital Communications and Networks*, Dec. 2023.

[4] L. A. G. Shivaprasad Nageswaran, R. F. Sesek, and G. J. A. Davis, "Using rfid to quantify school bus evacuation training times."

[5] X. Wang, J. Zhang, Z. Yu, E. Mao, S. C. Periaswamy, and J. Patton, "RFThermometer: A temperature estimation system with commercial UHF RFID tags," in *Proc. IEEE ICC'19*, Shanghai, China, May 2019, pp. 1–6.

[6] C. Yang, X. Wang, and S. Mao, "Respiration monitoring with RFID in driving environments," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 2, pp. 500–512, Feb. 2021.

[7] K. He, X. Chen, S. Xie, Y. Li, P. Dollár, and R. Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 16 000–16 009.

[8] X. Wang, J. Zhang, S. Mao, S. C. Periaswamy, and J. Patton, "A framework for locating multiple rfid tags using rf hologram tensors," *Digital Communications and Networks*, 2023.

[9] A. Khan, A. Sohail, U. Zahoora, and A. S. Qureshi, "A survey of the recent architectures of deep convolutional neural networks," *Artificial intelligence review*, vol. 53, pp. 5455–5516, 2020.

[10] X. Wang, J. Zhang, Z. Yu, S. Mao, S. C. Periaswamy, and J. Patton, "On remote temperature sensing using commercial uhf rfid tags," *IEEE Internet of Things Journal*, vol. 6, no. 6, pp. 10 715–10 727, 2019.

[11] L. Yang, Y. Chen, X.-Y. Li, C. Xiao, M. Li, and Y. Liu, "Tagoram: Real-time tracking of mobile RFID tags to high precision using COTS devices," in *Proc. ACM Mobicom'14*, Maui, HI, Sept. 2014, pp. 237–248.