IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 12 January 2024, accepted 31 January 2024, date of publication 6 February 2024, date of current version 12 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3363088

RESEARCH ARTICLE

Knowledge Distillation Based Deep Learning Model for User Equipment Positioning in Massive MIMO Systems Using Flying Reconfigurable Intelligent Surfaces

ABDULLAH AL-AHMADI
Department of Electrical Engineering, College of Engineering, Majmaah University, Al Majma'ah 11952, Saudi Arabia
e-mail: a.alahmadi@mu.edu.sa

The author extends the appreciation to the Deanship of Postgraduate Studies and Scientific Research at Majmaah University for funding this research work through the project number (R-2024-968).

ABSTRACT This paper introduces an innovative deep learning-based approach for User Equipment positioning in massive MIMO systems, enhanced by Flying Reconfigurable Intelligent Surfaces. The study emphasizes the transformative impact of Massive MIMO technology, which substantially increases the capacity and efficiency of wireless networks through extensive antenna arrays. We present a knowledge distillation model where a comprehensive teacher model is trained on massive MIMO data to predict User Equipment positions, which in turn guides a more compact student model. This results in a system that not only retains high accuracy with a lower mean absolute error of 3.67m but also benefits from reduced complexity and resource requirements. Our findings reveal that the distilled student model achieves a significant reduction in computational load while maintaining precise positioning capabilities, showcasing the potential for practical deployment in future wireless communication systems.

INDEX TERMS Knowledge distillation, massive MIMO, reconfigurable intelligent surfaces, user equipment positioning.

I. INTRODUCTION

Massive Multiple-Input Multiple-Output (Massive MIMO) technology marks a groundbreaking evolution in wireless communication systems. It involves the integration of an unprecedented number of antennas at both the transmitting and receiving ends. This extensive array of antennas is not merely a quantitative increase but brings about a qualitative leap in how wireless systems operate [1]. Unlike traditional MIMO systems, which utilize a relatively modest number of antennas, Massive MIMO employs a vast array, often numbering in the hundreds. This substantial increase in antennas leads to a dramatic enhancement in network capacity and efficiency.

The principal advantage of Massive MIMO lies in its ability to dramatically increase the spectral and energy efficiency of wireless networks [2]. By deploying a large number of antennas, the system significantly broadens the bandwidth availability within the existing radio spectrum. This enhancement is not just about handling more data; it's about doing so more efficiently, utilizing the same spectral resources [3]. This efficient use of the spectrum is crucial in an era where the demand for data is growing exponentially, driven by the proliferation of mobile devices and the Internet of Things (IoT).

A key feature of Massive MIMO is its capability to harness spatial multiplexing. This technique allows for the simultaneous transmission of multiple data streams to various users in the same frequency band. By doing so, Massive MIMO can significantly amplify the capacity of wireless channels. This capacity increase is instrumental in

The associate editor coordinating the review of this manuscript and approving it for publication was Wence Zhang.

meeting the high data rate demands of modern wireless communication, especially in the context of emerging 5G networks. Massive MIMO also offers enhanced resilience against common signal transmission issues like fading and thermal noise. This robustness results in more reliable and high-quality communication, a crucial requirement for the smooth functioning of modern wireless networks.

The integration of Massive MIMO with meta surfaces or intelligent reflecting surfaces (RISs) represents a significant advancement in wireless technology [4]. RIS are engineered to manipulate electromagnetic waves intelligently, enabling control over the signal propagation environment. This control is achieved through numerous miniature elements on the RIS, each capable of altering the phase shift of incident electromagnetic waves. In Massive MIMO systems, the incorporation of RIS is a game-changer. It allows for unprecedented control over the propagation of signals, leading to improved signal quality and reduced interference. This integration not only enhances the performance of Massive MIMO systems but also addresses critical challenges associated with hardware complexity and power consumption. By optimizing signal paths and reducing the need for power-intensive components, the combined system of Massive MIMO and RIS significantly elevates the spectral and energy efficiency, making it an ideal solution for next-generation 5G communication networks.

In the complex ecosystem of Massive MIMO systems, the positioning of User Equipment (UE) is a critical factor that significantly influences network performance and efficiency [5]. Precise localization of UE is essential for achieving superior signal quality, which in turn leads to higher data transmission rates and an enhanced overall user experience. This precision is particularly vital given the dense deployment of antennas in Massive MIMO systems, which allows for concurrent service to multiple UEs. The technology leverages spatial diversity and multiplexing, elevating network capacity and energy efficiency to meet the rising data demands in wireless networks. Optimal beamforming, a technique enabled by accurate UE positioning, plays a pivotal role in this context. It allows for focused signal transmission directly to the intended UE, minimizing interference and optimizing the use of network resources. This precision in signal delivery is crucial for maintaining high-quality communication, especially in densely populated areas where the risk of interference is high.

Accurate UE positioning extends its benefits to effective interference management, a critical consideration in densely populated network environments [6]. By accurately pinpointing UE locations, the base station can fine-tune its transmission parameters, significantly reducing interference and ensuring signal integrity. This improvement is not just technical; it translates into enhanced service quality and user experience, especially in urban areas where user density and the potential for interference are high.

Moreover, the advanced positioning capabilities inherent in Massive MIMO systems are fundamental for the deployment of various location-based services, which have become integral to a multitude of modern applications and services. These include navigation systems, asset tracking, and even augmented reality applications, all of which rely on precise and reliable location data. In emergency scenarios, the ability to accurately locate UEs can be critical. It enables quicker response times, providing first responders with exact location information, which can be vital in saving lives. This aspect of Massive MIMO systems is not just a technical improvement but a societal benefit, enhancing public safety and emergency response capabilities.

In addition to these practical applications, the precise positioning capabilities of Massive MIMO systems also open the door to a range of innovative services and applications [7]. The improved accuracy in UE localization allows for more personalized and location-aware services, catering to the unique needs and preferences of individual users. This could lead to more targeted marketing, enhanced user experiences in augmented reality environments, and even more efficient urban planning and management, leveraging the data gathered from precise user positioning.

A. RELATED WORK

1) MASSIVE MIMO SYSTEMS AND USER EQUIPMENT POSITIONING

Wen et al. [8] provide a comprehensive examination of user localization techniques in 5G networks, with a particular focus on the role of massive MIMO technology. The study underscores the critical importance of massive antenna arrays in 5G networks, not only for enhancing communication capabilities but also for improving the precision of user localization. These arrays exploit the spatial signatures of users, enabling more accurate localization. The survey delves into the development of advanced channel estimation routines and localization methods specifically designed for massive MIMO environments. This includes a detailed discussion on the potential of leveraging multipath signals, which are typically considered detrimental in traditional communication systems. By utilizing geometric channel models, the authors argue that these multipath signals can be transformed into a valuable asset for enhancing localization accuracy.

Furthermore, they critique the conventional two-step process used in localization, which typically involves extracting distance and/or angle information from signal measurements and then applying triangulation techniques to pinpoint user positions. They highlight the limitations of this approach, particularly its tendency to reduce the rich information contained in waveforms to mere point estimates. This simplification can lead to significant performance losses, especially in scenarios where measurements are ambiguous or complex. The authors suggest that this conventional method may not be well-suited

for the intricate signal environments characteristic of 5G networks, where signals often interact with various objects in the environment, creating complex propagation paths.

The paper also explores the integration of machine learning techniques with massive MIMO systems to further enhance localization accuracy. The authors propose that machine learning algorithms, particularly those based on deep learning, can effectively process the high-dimensional data generated by massive MIMO systems. These algorithms can learn from the environment-specific characteristics of signal propagation, leading to more accurate and robust localization methods. This approach represents a significant shift from traditional model-based localization techniques, offering a more adaptable and efficient solution for the dynamic and heterogeneous environments typical in 5G networks.

2) DEEP LEARNING IN WIRELESS COMMUNICATIONS

The article in [9] presents a comprehensive review of the integration of deep learning (DL) in wireless communications, a field poised to undergo significant transformations with the advent of emerging applications like virtual reality and the Internet of Things. The paper identifies two primary methodologies for applying DL in wireless communications. The first is DL-based architecture design, which represents a departure from the classical model-based block design that has dominated wireless communications for decades. This approach leverages the capabilities of DL to design more flexible and efficient communication systems. The second methodology is DL-based algorithm design, illustrated through various examples of techniques conceived for 5G and beyond. These include novel neural network architectures and the use of communication expert knowledge to enhance system performance. The paper discusses the principles, key features, and performance gains of these methodologies, providing a clear understanding of how DL can revolutionize wireless communication systems. Additionally, the authors highlight open problems and future research opportunities, emphasizing the synergistic relationship between DL and wireless communications. This review is expected to inspire novel ideas and significant contributions towards the development of intelligent wireless communication systems.

A fine-grained localization based on massive MIMO-OFDM system presented in [10] focuses on the high angular resolution that massive MIMO systems can provide, which is pivotal for fine-grained indoor localization necessary for indoor location-based services (ILBS). Instead of directly using channel state information (CSI) for localization, which is common in many systems, the paper emphasizes the extraction of multipath components (MPCs) from the CSI. These MPCs are obtained through the space-alternating generalized expectation-maximization (SAGE) algorithm, which is a more refined approach to understanding the multipath environment of indoor spaces. The authors propose a generalized fingerprinting system that employs various single-metric and hybrid-metric schemes. Their paper thoroughly evaluates the impact of different antenna topologies, the size of the training set, the number of antennas, and the effective signal-to-noise ratio (SNR) on the localization accuracy.

3) RECONFIGURABLE INTELLIGENT SURFACES IN MIMO SYSTEMS

The paper in [11] provides a crucial overview of the application of reconfigurable intelligent surfaces in future MIMO systems. This paper delves into two innovative MIMO system designs that incorporate RIS to boost performance and spectral efficiency. The first design integrates RIS with the Vertical Bell Labs Layered Space-Time (VBLAST) scheme [12], enhancing the nulling and canceling-based suboptimal detection procedure. This integration not only improves performance but also significantly increases spectral efficiency by transmitting additional bits through the phase adjustment of RIS elements. The second design revisits Alamouti's scheme [13], utilizing RIS elements to redesign it with a single radio frequency signal generator at the transmitter side, thereby enhancing its bit error rate (BER) performance. The paper presents Monte Carlo simulations to demonstrate the effectiveness of these system designs, showing that they outperform reference schemes in terms of BER performance and spectral efficiency. This research highlights the potential of RIS in revolutionizing MIMO communication systems, paving the way for more efficient and high-performing wireless communication technologies.

The authors in [14] propose a method that leverages a reconfigurable intelligent surface (RIS) to enable joint localization and synchronization using only downlink transmissions from a single base station (BS). They introduce a direct maximum likelihood (ML) estimator for determining the position and clock offset, which requires a good initialization typically achieved through exhaustive grid search. To streamline this process, the paper also presents a decoupled, relaxed estimator that does not necessitate prior knowledge of the clock offset. This estimator serves as a preliminary step to provide a good starting point for the ML optimization. The paper's results demonstrate that the proposed method can achieve the Cramér-Rao lower bound for localization accuracy, even without optimized beamforming and RIS control matrix, and with moderate system parameters. This suggests that the use of RIS in mmWave MISO systems could significantly enhance localization and synchronization capabilities, which are critical for the deployment of efficient and accurate 5G networks.

4) KNOWLEDGE DISTILLATION IN DEEP LEARNING

Alkhulaifi et al. [15] examine the concept of knowledge distillation in the context of deep learning, particularly addressing the challenge of deploying large deep learning models on resource-limited devices like mobile phones and embedded systems. The authors present a detailed overview

of various knowledge distillation techniques, which involve training a smaller, more efficient model (the student) using information derived from a larger, more complex model (the teacher). This process allows the compact student model to achieve performance close to that of the larger model while being more suitable for deployment in resource-constrained environments.

In their paper, they introduce a new metric, termed the distillation metric, to evaluate and compare different knowledge distillation techniques. This metric considers both the size of the models and their accuracy scores, providing a comprehensive measure of the effectiveness of various distillation approaches. The authors' survey leads to several insightful conclusions about the current state of knowledge distillation in deep learning, including the identification of current challenges and potential directions for future research. Their work highlights the significance of knowledge distillation as a critical tool in optimizing deep learning models for practical applications, especially in scenarios where computational resources are limited.

The paper in [16] presents an innovative approach to fingerprint-based indoor positioning systems (F-IPS). The paper suggests using convolutional neural networks (CNN) as they are better at learning the overall topology of fingerprinting images and capturing highly abstract features. However, CNNs are computationally expensive and require significant storage, making them impractical for resource-limited devices. To address this, the authors proposed incorporating knowledge distillation (KD) into CNN-based IPS (CNN-IPS). The KD-CNN-IPS shows a notable improvement in localization performance, with a significant percentage of positioning errors within 2 meters, and operates faster than the teacher model. It also outperforms a CNN-IPS of the same architecture, demonstrating the effectiveness of knowledge distillation in enhancing the performance of resource-efficient CNN models for indoor positioning.

B. CONTRIBUTIONS

This paper introduces an innovative deep learning-based approach for UE positioning in massive MIMO systems, enhanced by Flying Reconfigurable Intelligent Surfaces. The study emphasizes the transformative impact of Massive MIMO technology, which substantially increases the capacity and efficiency of wireless networks through extensive antenna arrays. We present a knowledge distillation model where a comprehensive teacher model is trained on massive MIMO data to predict UE positions, which in turn guides a more compact student model. This results in a system that not only retains high accuracy but also benefits from reduced complexity and resource requirements. Our findings reveal that the distilled student model achieves a significant reduction in computational load while maintaining precise positioning capabilities, showcasing the potential for practical deployment in future wireless communication systems.

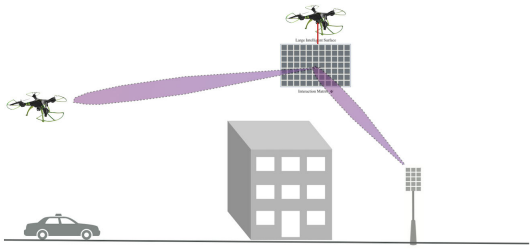

FIGURE 1. Flying reconfigurable intelligent surface (FRIS) [17].

The remainder of this paper is organized as follows. Section II provides an overview of Flying Reconfigurable Intelligent Surfaces (FRIS) and their potential applications in wireless communications. Section III introduces the concept of knowledge distillation and its applications in deep learning. Section IV presents the proposed system, including the data generation and preprocessing steps, the knowledge distillation model, and the evaluation metrics. Section V presents the results and discussion, followed by the conclusion in Section VI.

II. FLYING RECONFIGURABLE INTELLIGENT SURFACES

Reconfigurable Intelligent Surfaces (RISs) represent an emerging technology in wireless communications. They are essentially smart surfaces with the ability to shape electromagnetic waves, aiming to improve the communication quality, coverage, and energy efficiency. RISs are composed of a large number of passive elements that can be adjusted (reconfigured) to change the phase shifts of the incident signals. By doing so, they can steer the signals in desired directions or focus them on specific points, making them particularly useful for enhancing wireless communications in complex environments.

By integrating RISs with aerial platforms, such as drones or unmanned aerial vehicles (UAVs), we can further extend their potential applications, such as delivering high-quality wireless signals to remote or densely populated areas and enhancing network coverage in dead zones [17]. In this context, we refer to the RISs mounted on aerial platforms as Flying Reconfigurable Intelligent Surfaces (FRISs) shown in Figure 1. FRISs can dynamically adapt their configurations based on the changing wireless environment and user requirements. This adaptability allows them to optimize the signal propagation conditions in real-time, leading to improved communication performance and enhanced UE positioning accuracy in massive MIMO system.

To illustrate the potential of FRISs, we consider a drone scenario shown in Figure 2 using DeepMIMO [18] where a UE is located in a dead zone, i.e., a region with no direct line-of-sight (LOS) path to the BS. In this case, the UE can only receive signals from the BS through reflections from the surrounding objects. The received signals are usually weak and distorted, leading to poor communication quality and inaccurate UE positioning. To address this issue, we can

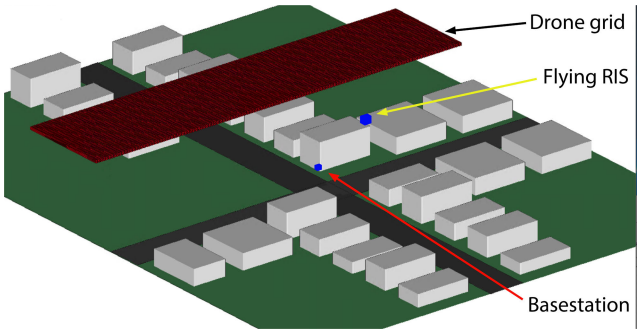

FIGURE 2. Bird-eye view of the drone scenario [17].

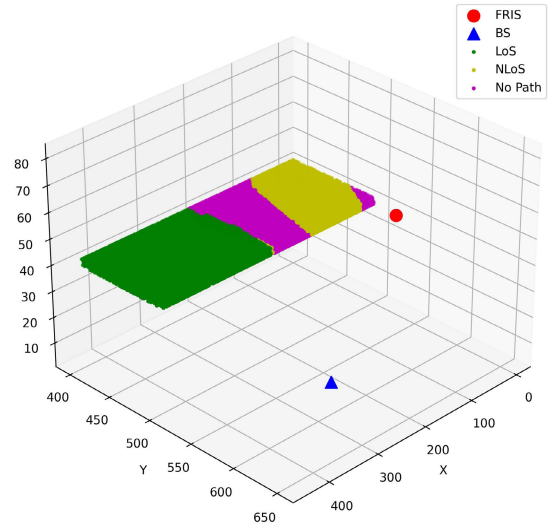

FIGURE 3. Locations of FRIS, BS, and UEs based on LoS from BS.

deploy a FRIS near the UE to reflect the signals from the BS towards the UE. The FRIS can be configured to steer the signals in the desired direction and focus them on the UE. By doing so, the FRIS can improve the signal quality and enhance the UE positioning accuracy.

Figure 3 shows the locations of the FRIS, BS, and UEs in a drone scenario. The FRIS is mounted on a drone hovering at a height of h_{FRIS} above the ground. The BS is located at a height of h_{BS} above the ground where $h_{FRIS} > h_{BS}$. The UE is located at a height of h_{UE}. The FRIS is equipped with N passive elements, each of which can be configured to adjust the phase shift of the incident signals. The FRIS is connected to the BS through a wireless link with a distance of d_{BS} and a wireless link with a distance of d_{UE}. From Figure 3, we can see that the UE is located in a dead zone. In this case, the UE can only receive signals from the BS through reflections from the surrounding objects. The received signals are usually weak and distorted, leading to poor communication quality and inaccurate UE positioning. To address this issue, we can deploy a FRIS near the UE to reflect the signals from the BS towards the UE. The FRIS can be configured to steer the signals in the desired direction and focus them on the UE,

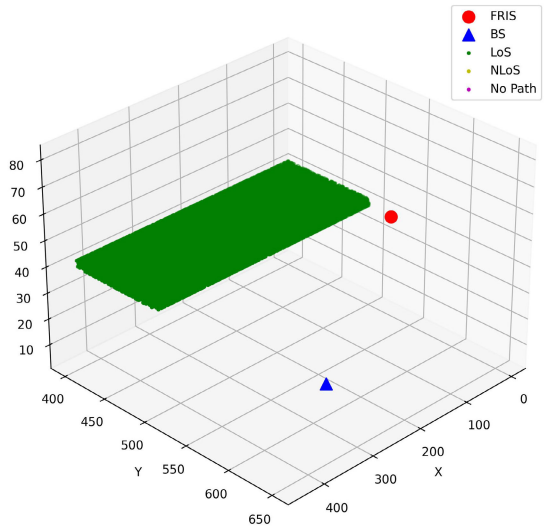

FIGURE 4. Locations of FRIS, BS, and UEs based on LoS from FRIS.

as shown in Figure 4. By doing so, the FRIS can improve the signal quality and enhance the UE positioning accuracy.

III. KNOWLEDGE DISTILLATION BASED DEEP LEARNING MODEL

The concept of Knowledge Distillation (KD) has been proposed as a way to improve the training and deployment of deep neural networks [19]. It involves transferring the knowledge from a complex and computationally expensive model, often referred to as the ''teacher'' model, to a simplified and more efficient model, known as the ''student'' or ''distilled'' model. By using knowledge distillation, the student model can learn from the teacher model's expertise and achieve similar performance with reduced computational complexity.

The use of knowledge distillation in the training and deployment of deep neural networks has shown promising results in improving the efficiency and performance of communication systems, including massive MIMO systems. By distilling the knowledge from complex and computationally expensive models, such as those used in massive MIMO systems, to simplified and more efficient models, massive MIMO systems can be optimized for better efficiency and performance [20]. Additionally, by applying knowledge distillation techniques, lightweight networks can achieve higher accuracy and perform complex tasks such as image classification, object detection, and face recognition more effectively.

Knowledge distillation for UE positioning in massive MIMO systems, especially when using FRIS, is a sophisticated topic that combines various cutting-edge technologies in wireless communications and deep learning. Given the complexity of massive MIMO systems and the dynamic nature of FRIS, predicting the 3D position of UE can be challenging. A large and complex deep learning model (the

teacher) can be trained to predict the UE position based on various features like DoA, Phase, ToA, and received signal strength. However, deploying such a large model in real-time systems can be computationally expensive. This is where knowledge distillation comes in. By distilling the knowledge from the teacher model to a smaller and more efficient model (the student), we can achieve similar performance with reduced computational complexity. The student model can be deployed in real-time systems, such as those used in massive MIMO systems, to predict the UE position based on the same features used by the teacher model. This can significantly improve the efficiency and performance of massive MIMO systems, especially when using FRIS.

A. TEACHER MODEL

In the teacher model, the deep neural network is trained on massive MIMO data to predict the UE position. Given its size, it captures intricate patterns and relationships in the data.

Given an input feature vector X (e.g., signal features like DoA, Phase, ToA, etc.), the forward propagation through the neural network can be represented by a series of transformations:

$$H^{(0)} = X \tag{1}$$

For each layer l from 1 to L (where L is the total number of layers):

$$H^{l} = \sigma(W^{(1)}H^{(l-1)} + b^{(l)}) \tag{2}$$

where $H^{(l)}$ is the output of layer l, $W^{(l)}$ and $b^{(l)}$ are the weight matrix and bias vector for layer l, respectively, and σ is an activation function such as Rectified Linear Unit (ReLU) or sigmoid. To train the teacher model, a suitable loss function is chosen based on the task. For a regression task like predicting the 3D position, the Mean Squared Error (MSE) is commonly used:

$$\mathcal{L}_{\text{teacher}} = \frac{1}{N}\sum_{i=1}^{N}\left((x_{\text{t}_i} - x_{T_i})^2 + (y_{\text{t}_i} - y_{T_i})^2 + (z_{\text{t}_i} - z_{T_i})^2\right) \tag{3}$$

where $(x_{T_i}, y_{T_i}, z_{T_i})$ are the predictions from the teacher model and $(x_{\text{t}_i}, y_{\text{t}_i}, z_{\text{t}_i})$ are the true labels.

To optimize the model's parameters, gradients are computed using backpropagation and then updated using optimization algorithms like Stochastic Gradient Descent (AGD), Adam, or Root Mean Square Propagation (RMSprop). The teacher model, once trained, provides predictions on the dataset. These predictions, often softened using a temperature parameter, are then used in conjunction with the actual labels to train the student model in the knowledge distillation process.

The temperature parameter, often denoted as T, is a hyperparameter used to adjust and "soften" the output probabilities of the teacher model. By softening the outputs, we can capture the model's uncertainty and transfer not just the hard predictions, but also the relative confidences of the model across different classes.

In a typical softmax function, which is often used in the output layer of classification models, the probability. In a typical softmax function, which is often used in the output layer of classification models, the probability p_i of a class i is given by:

$$p_i = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \tag{4}$$

where z_i is the logit (raw output) of the model for class i.

In the context of knowledge distillation, the softmax function is modified by introducing the temperature parameter:

$$p_i = \frac{\exp\left(\frac{z_i}{T}\right)}{\sum_j \exp\left(\frac{z_j}{T}\right)} \tag{5}$$

In higher temperature $T > 1$, the probabilities become "softer", meaning they move closer to being uniform. This can reveal more information about the model's uncertainties and confidences across classes. It's particularly useful in knowledge distillation, as it allows the student model to learn more from the teacher's outputs, including the nuances and relative confidences. Whereas in lower temperature $T < 1$, the probabilities become "sharper", pushing the maximum probability closer to 1 and the rest closer to 0 resulting in more deterministic outputs. The softened probabilities from the teacher model (obtained using a higher temperature) are used to train the student model. This allows the student model to learn from the teacher's relative confidences across classes, not just the hard predictions. The same temperature value is typically used to soften the student's outputs during training. The loss is then computed between the softened outputs of the teacher and student, helping the student model to generalize better.

B. STUDENT MODEL

The student model is a typically smaller and more lightweight neural network compared to the teacher model. It is designed to be efficient and deployable, especially in resource-constrained environments. The primary goal of the student model is to mimic the behavior of the teacher model while retaining most of its accuracy. Given an input feature vector X, the forward propagation through the student neural network can be represented by a series of transformations.

$$H_S^{(0)} = X \tag{6}$$

For each layer l from 1 to L_S (where L_S is the total number of layers in the student model):

$$H_S^{(l)} = \sigma(W_S^{(l)}H_S^{(l-1)} + b_S^{(l)}) \tag{7}$$

where $H_S^{(l)}$ is the output of layer l in the student model. $W_S^{(l)}$ and $b_S^{(l)}$ are the weight matrix and bias vector for layer l in the student model, respectively and σ is an activation function. The loss function often consists of two parts, distillation loss, $\mathcal{L}_{\text{distill}}$ which measures the difference between the student's

predictions and the softened outputs from the teacher model, and ground truth loss, $\mathcal{L}_{\text{true}}$ which measures the difference between the student's predictions and the actual ground truth labels where:

$$\mathcal{L}_{\text{distill}} = \frac{1}{N}\sum_{i=1}^{N}\left((x_{T_i} - x_{S_i})^2 + (y_{T_i} - y_{S_i})^2 + (z_{T_i} - z_{S_i})^2\right) \tag{8}$$

where $(x_{T_i}, y_{T_i}, z_{T_i})$ are the softened predictions from the teacher model for the i^{th} sample.

$$\mathcal{L}_{\text{true}} = \frac{1}{N}\sum_{i=1}^{N}\left((x_{\text{t}_i} - x_{S_i})^2 + (y_{\text{t}_i} - y_{S_i})^2 + (z_{\text{t}_i} - z_{S_i})^2\right) \tag{9}$$

The overall loss used to train the student model is a weighted combination of the distillation loss and the ground truth loss:

$$\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_{\text{distill}} + (1-\alpha)\mathcal{L}_{\text{true}} \tag{10}$$

where α is a hyperparameter that balances the two loss components.

IV. PROPOSED SYSTEM

This section delineates the proposed system for UE positioning in Massive MIMO systems, enhanced with FRIS. The system comprises several integral components: data generation, data preprocessing, and the implementation of a KD based deep learning model. These components work in tandem to predict the 3D positions of UEs with high accuracy and efficiency.

A. DATA GENERATION

DeepMIMO [18] is a data generation framework that can be used to generate synthetic channel data for massive MIMO systems. It can be used to generate data for various scenarios, including drone scenarios. In this study, we use DeepMIMO to generate synthetic channel data for a drone scenario. The scenario consists of a single BS at a height of 6m, a single FRIS at a height of 80m, and 4 drone grids at a 4 different heights of 40m, 40.8m, 41.6m and 42.4m. Each drone grid consists of 67,456 drones where each drone acts as a single UE with an isotropic antenna. The total number of EUs in all four drone grids is 269,824. The BS is equipped with a uniform linear array (ULA) of 32 antennas. The FRIS is equipped with a ULA of 64 antennas. The BS and FRIS are connected through a wireless link with a distance of 200m. The UE is located in a dead zone. In this case, the UE can only receive signals from the BS through reflections from the surrounding objects. The received signals are usually weak and distorted, leading to poor communication quality and inaccurate UE positioning. To address this issue, we can deploy a FRIS near the UE to reflect the signals from the BS towards the UE. The FRIS can be configured to steer the signals in the desired direction and focus them on the UE, as shown in Figure 4. By doing so, the FRIS can improve the signal quality and enhance the UE positioning accuracy.

TABLE 1. Parameters for the drone scenario.

Parameter	BS	FRIS
Active BS	1 ($id = 1$)	1 ($id = 2$)
User row first	1	1
User row last	496	496
Antenna shape	(32,1,1)	(64,1,1)
Number of paths	1	1
OFDM subcarriers	512	512

B. DATA PREPROCESSING

In the data preprocessing stage, the system extracts relevant information from the dataset generated by DeepMIMO. Specifically, it focuses on user paths and locations, which are essential for understanding the communication channel characteristics in a MIMO system. The data is extracted into arrays and dictionaries, providing a structured format for further processing. The 'user.paths' dictionary contains information about the angle of arrival, phase, time of arrival, and power of the signal, which are crucial for characterizing the wireless channel. On the other hand, 'user.location' provides the 3D coordinates of the user UE. These pieces of information are vital for training models that predict UE positions based on channel characteristics.

Once the data is extracted, the system transforms it into a more manageable and analytical format using Pandas DataFrames. Each user's data is represented as a row, with columns for the X, Y, and Z coordinates, as well as the channel characteristics like angle of arrival, phase, time of arrival, and power. This tabular format enables easier manipulation, visualization, and application of machine learning models. Additionally, it facilitates the integration of standard data preprocessing and machine learning libraries, enhancing the workflow's efficiency. The system also ensures that all relevant channel characteristics and user location information are included, providing a comprehensive dataset for model training and evaluation.

Finally, the data splits into training and testing sets, ensuring that the model can be trained on one subset of the data and evaluated on another, unseen subset. This split is crucial for assessing the model's generalization performance. Furthermore, the system applies standardization to the features and targets, transforming them to have a mean of zero and a standard deviation of one. This normalization step is crucial for models like neural networks, which are sensitive to the scale of the input features. By ensuring that all features are on the same scale, the training process becomes more stable and faster, and it helps in preventing the dominance of features with larger scales over those with smaller scales.

C. KD BASED MODEL FOR UE POSITIONING

The Distiller class in Algorithm 1, as presented in the paper, is an innovative approach designed to transfer knowledge

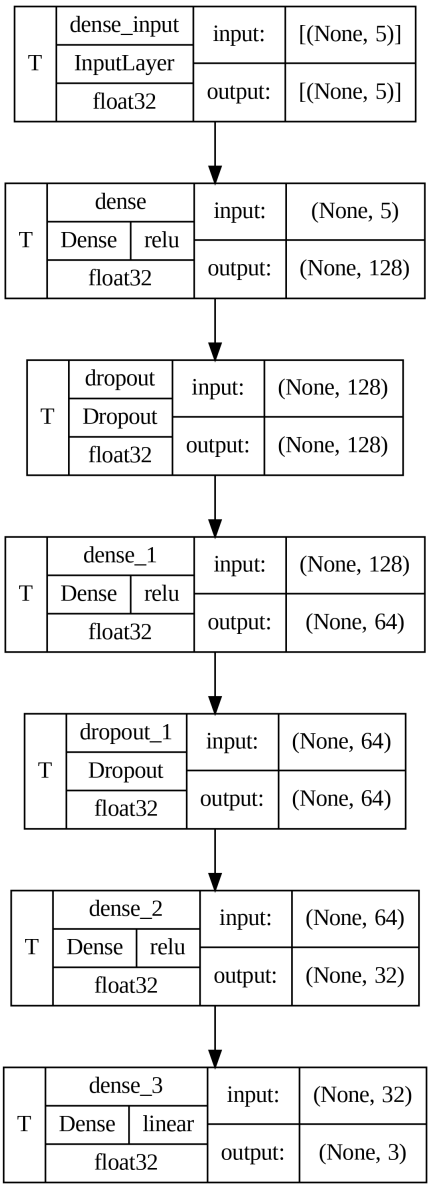

FIGURE 5. The architecture of teacher model.

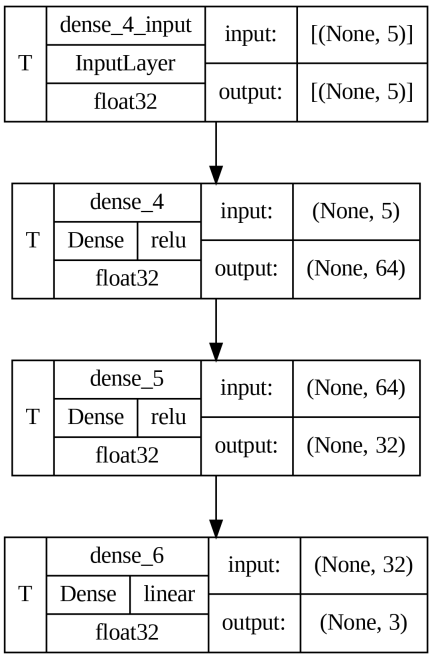

FIGURE 6. The architecture of student model.

from a complex, well-trained teacher model M_{teacher} to a simpler, more efficient student model M_{student}. This process is particularly valuable in scenarios where deploying large models is not feasible due to resource constraints.

1) INITIALIZATION AND COMPILATION

The algorithm begins with the initialization phase, where an instance of the distiller class is created. This class inherits from the Keras model and requires two primary inputs: the teacher model and the student model. The teacher model is a pre-trained, larger neural network with high predictive accuracy, while the student model is typically smaller and less complex. The aim is for the student model to learn from the teacher model's outputs, thus achieving comparable performance with a reduced computational footprint. Following initialization, the compilation stage configures the distiller for training. Key parameters in this phase include the optimizer, metrics for performance evaluation, and loss functions for both the student and the distillation process. The student loss function calculates the loss between the actual targets and the student's predictions, while the distillation loss function measures the discrepancy between the softened predictions of the teacher and the student models. The concept of 'softened predictions' involves adjusting the outputs of the models using a temperature parameter, allowing the student to learn not only the hard predictions but also the relative confidence levels across different outputs from the teacher model.

2) TRAINING STEP

In the training phase, the algorithm processes data in batches. For each batch, it first computes the predictions from the teacher model. These predictions serve as a guide or a form of advanced insight for the student model. The student model then makes its own predictions based on the same input. The core of the knowledge distillation process occurs here, where two types of losses are computed: the student loss and the distillation loss. The student loss, as mentioned earlier, is the difference between the student model's predictions and the actual labels. The distillation loss, however, is more nuanced; it involves comparing the student model's predictions against the teacher model's softened outputs. This loss is crucial as it encapsulates the essence of knowledge distillation — transferring the intricate patterns and insights the teacher model has learned to the student model. The total loss, used to update the student model, is a weighted sum of these two losses,

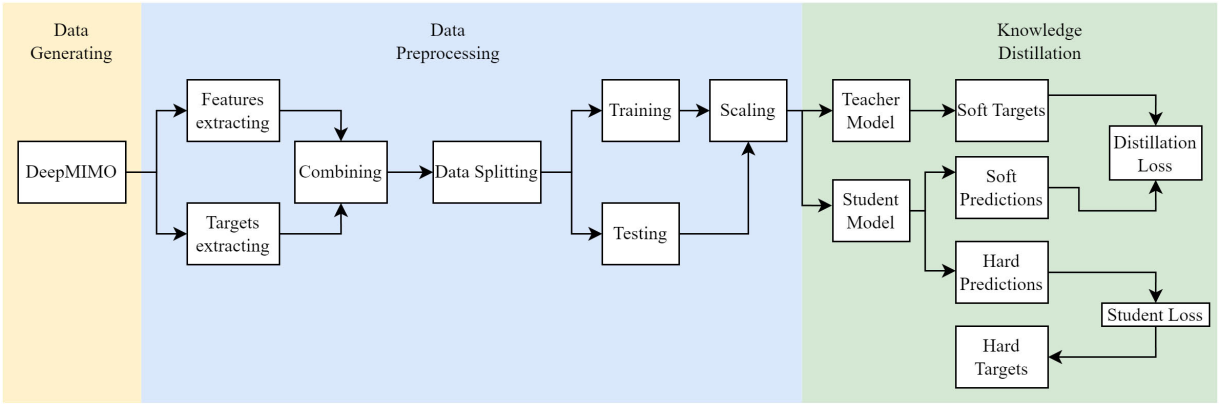

FIGURE 7. Flowchart of the proposed system.

governed by the alpha parameter. This parameter balances the importance given to the student loss versus the distillation loss. The training step is iterative and involves updating the student model's parameters to minimize the total loss, thereby gradually improving the student model's performance to approximate that of the teacher model.

3) TESTING STEP

The final phase is the testing step, which defines how the distiller model evaluates its performance. Here, the student model, now trained with knowledge distilled from the teacher, is used to make predictions on a separate test dataset. This phase is crucial as it assesses how well the student model has learned and its ability to generalize to new, unseen data. The testing step involves computing the forward pass of the student model to obtain predictions, which are then compared against the actual labels to update the metrics specified during the compilation phase. This step is devoid of any training or updates to the model's parameters and is solely focused on evaluation.

V. RESULTS AND DISCUSSION

Table 2 presents a comparative analysis of simulation parameters and their corresponding outcomes across two distinct sets of experiments, designed to evaluate the performance of a teacher model and its distilled counterpart, as well as a base model. In both Set 1 and Set 2, the teacher model undergoes a substantial training regimen of 1000 epochs to establish a robust predictive framework, intended to serve as a benchmark or a high-fidelity model from which knowledge is distilled.

In contrast, the distilled and base models are trained for significantly fewer epochs, only 100, focusing on efficiency and testing of how well these models can approximate the teacher model's performance with less computational effort. The batch size for Set 1 is double that of Set 2, at 512 compared to 256. This larger batch size could potentially afford more stable gradient estimates during training but at the cost of higher memory requirements and longer processing times per epoch quantifying the average magnitude of errors between the models' predictions and the actual values. The Mean Absolute Error (MAE) is employed as a performance metric, defined as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

where n is the number of observations, y_i is the actual value of an observation, $\hat{y}_i$ is the predicted value and $|y_i - \hat{y}_i|$ is the absolute error for each observation. The distilled

Algorithm 1 Distiller Class for Knowledge Distillation

Require: Student model M_{student}, Teacher model M_{teacher}
Require: Student loss function $\mathcal{L}_{\text{student}}$, Distillation loss function $\mathcal{L}_{\text{distill}}$
Require: Learning rate α, Temperature T
Initialization: Initialize the Distiller with M_{student} and M_{teacher}
Compilation: Configure Distiller with optimizer, metrics, $\mathcal{L}_{\text{student}}$, $\mathcal{L}_{\text{distill}}$, α, T
Training Step:
for each batch (x, y) in training data **do**
  Compute teacher predictions: $Y_{\text{teacher}} = M_{\text{teacher}}(x)$
  Compute student predictions: $Y_{\text{student}} = M_{\text{student}}(x)$
  Compute student loss: $\mathcal{L}_S = \mathcal{L}_{\text{student}}(y, Y_{\text{student}})$
  Compute distillation loss: $\mathcal{L}_D = \mathcal{L}_{\text{distill}}(\text{Softmax}(Y_{\text{teacher}}/T), \text{Softmax}(Y_{\text{student}}/T))$
  Total loss: $\mathcal{L}_{\text{total}} = \alpha\mathcal{L}_S + (1-\alpha)\mathcal{L}_D$
  Update M_{student} to minimize $\mathcal{L}_{\text{total}}$
  Update metrics
end for
Testing Step:
for each batch (x, y) in test data **do**
  Compute predictions: $Y_{\text{pred}} = M_{\text{student}}(x)$
  Update metrics
end for
Ensure: Trained M_{student} and performance metrics

TABLE 2. Simulation parameters and results.

Parameter	Set 1	Set 2
Teacher model epochs	1000	1000
Distilled model epochs	100	100
Base model epochs	100	100
Batch size	512	256
Distilled model MAE	4.14	4.03
Base model MAE	6.2	4.99
Teacher model running time	647.89 sec	638.75 sec
Distilled and base models running time	13.6 sec	35.63 sec

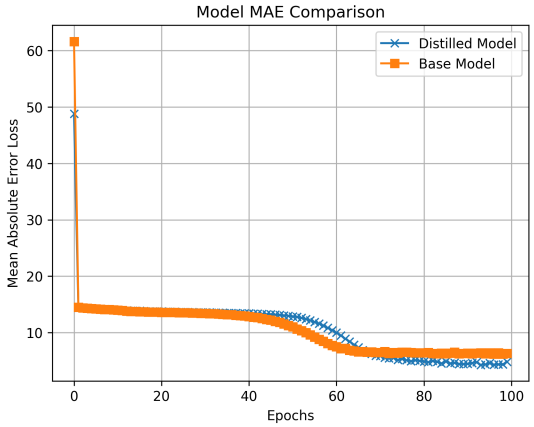

FIGURE 8. Models MAE comparison for set 1.

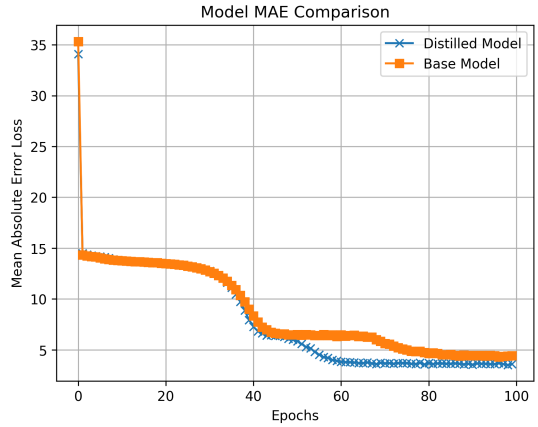

FIGURE 9. Models MAE comparison for set 2.

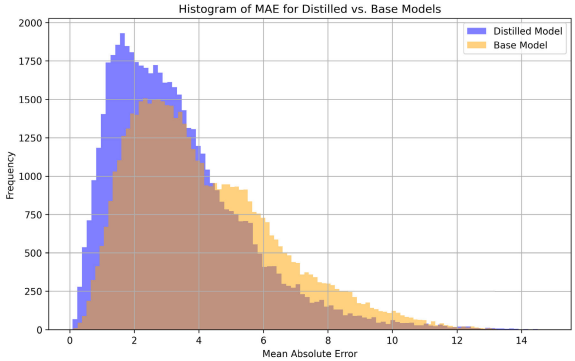

FIGURE 10. Histogram of MAE for distilled vs. base models.

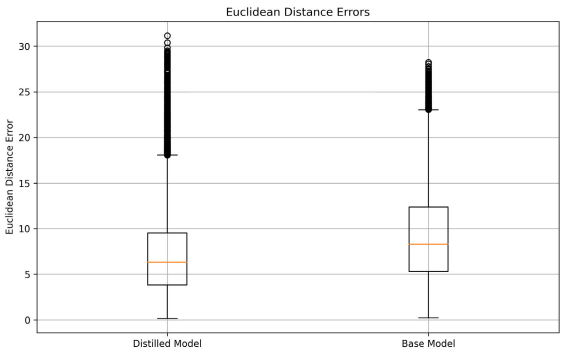

FIGURE 11. Euclidean distance errors.

model exhibits a lower MAE than the base model in both sets, with Set 2 showing a marginal improvement over Set 1 (4.03 versus 4.14). This suggests that the process of knowledge distillation may have been more effective in Set 2, possibly due to the adjusted batch size. The base model's MAE is noticeably higher, especially in Set 1 (6.2), indicating a less accurate performance compared to the distilled model.

Figures 8, and 9 illustrate the MAE comparison between the distilled and base models for Set 1 and Set 2, respectively. The distilled model's MAE in both sets starts to perform better than the base model after approximately 60 epochs, with the gap widening as the number of epochs increases. This suggests that the distilled model's performance improves as it is trained for more epochs, while the base model's performance remains relatively stagnant.

Running times provide insight into the computational efficiency of the models. For the teacher model, the running times are relatively close across the two sets (647.89 seconds for Set 1 and 638.75 seconds for Set 2), which might be expected given the identical number of epochs. However, there is a marked difference in the combined running times of the distilled and base models between the two sets (13.6 seconds for Set 1 and 35.63 seconds for Set 2). This substantial increase in running time for Set 2 is attributed to the halved batch size, necessitating more iterations per epoch, and points to a trade-off between model performance and computational efficiency.

The histogram shown in Figure 10 illustrates the distribution of MAE values for the distilled and base models in Set 1. The distilled model's MAE is noticeably lower than the base model's MAE, with the majority of the distilled model's MAE values falling between 0 and 10. The base model's MAE values are more spread out, with a significant number of values falling between 8 and 12. This suggests that the distilled model is more accurate than the base model, with a lower average magnitude of errors between the models' predictions and the actual values. The peaks of the two models

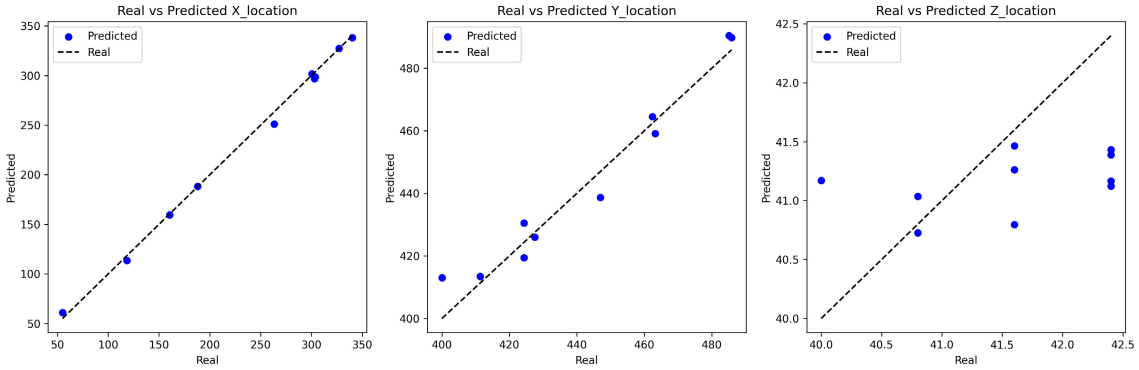

FIGURE 12. Real vs predicted targets for 10 random UEs.

are also noticeably different, with the distilled model's peak occurring at approximately 1.7 and the base model's peak occurring at approximately 2.2. This further indicates that the distilled model is more accurate than the base model.

The Euclidean distance error (EDE) is another metric used to evaluate the performance of the models. It is defined as the Euclidean distance between the predicted and actual positions of the UE. The EDE is calculated as:

$$\text{EDE} = \sqrt{(x_{\text{t}_i} - x_{T_i})^2 + (y_{\text{t}_i} - y_{T_i})^2 + (z_{\text{t}_i} - z_{T_i})^2}$$

where $(x_{T_i}, y_{T_i}, z_{T_i})$ are the model predictions and $(x_{\text{t}_i}, y_{\text{t}_i}, z_{\text{t}_i})$ are the true targets. Figure 11 shows the EDE for distilled and base models in Set 2. The median value for the distilled model is around 6 and 7.5 for the base model. Although the distilled model has more outliers, the upper and lower whiskers are shorter than the base model, indicating that the distilled model has a more consistent performance. The box plot also shows that the base model has a higher median value and a wider spread of values, indicating that the distilled model is more accurate than the base model.

Figure 12 shows the real vs predicted targets for 10 random UEs in Set 2. The predicted targets are close to the real targets, indicating that the model is able to predict the UE positions accurately. The model's performance is consistent across the 10 UEs, with the predicted targets falling close to the real targets for all 10 UEs. This suggests that the model is able to generalize well and predict the UE positions accurately for different UEs.

Table 3 reveals an intricate evaluation of KD efficacy, indicated by the varying MAE across different experimental sets. The MAE reflects the average magnitude of errors in the student model's estimates of the UE positions within massive MIMO systems. Adjusting the KD parameters—alpha and temperature—yielded distinct impacts on the student model's performance. Notably, a low alpha of 0.1 paired with a high temperature value of 15 produced a MAE of 4.02, closely aligning with the base model's accuracy. This configuration suggests an optimal synergy between adherence to the teacher model's soft targets and the ground truth, enabling the student model to learn effectively without overfitting to the teacher's outputs. The nuanced relationship between these parameters and the resulting MAE underscores the potential of KD in refining deep learning models for complex spatial estimations in wireless networks.

TABLE 3. Simulation parameters for set 3 to Set 9.

	Alpha	Temperature	Student MAE
Set 3	0.1	3	3.89
Set 4	0.1	5	3.95
Set 5	0.1	10	3.67
Set 6	0.1	15	4.02
Set 7	0.5	10	4.05
Set 8	1	10	3.2
Set 9	2	10	3.92

The application of KD has shown to be a promising technique for enhancing the computational efficiency of deep learning models tasked with positioning in massive MIMO systems. The student models, which were subjected to the distillation process, displayed a notable reduction in MAE while requiring less computational resources compared to more complex base models. For instance, a student model with an alpha setting of 0.1 and a temperature of 10, despite experiencing a slight increase in MAE to 3.67, presents itself as an efficient alternative. This trade-off between computational demand and positioning accuracy is at the heart of KD, aiming to craft models that are both practical for deployment and capable of maintaining a high degree of precision. The flexibility in adjusting the KD parameters is critical, as it allows for the tailoring of student models to fit the specific performance and resource allocation requirements of emerging wireless communication systems.

The trends exhibited in Table 3 offer valuable insights into the application of KD within the domain of massive MIMO systems. As these systems evolve to accommodate the burgeoning requirements of modern telecommunications, the need for accurate yet computationally efficient models becomes increasingly important. KD addresses this need by enabling student models to approximate the accuracy of their more complex teacher counterparts. This study

lays the groundwork for further optimization, suggesting that future research could explore more refined adjustments to the KD parameters. This could lead to discoveries of even more effective parameter configurations, further enhancing model performance. Additionally, the application of these optimized models to real-world scenarios, such as dynamic UE positioning, could significantly benefit the telecommunications industry, particularly in the streamlined deployment of 5G and the forthcoming 6G networks.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have presented a novel approach to UE positioning in massive MIMO systems using knowledge distillation. The proposed system leverages the high spatial resolution of massive MIMO channels to accurately predict the 3D position of UE. The system consists of a teacher model and a distilled (student) model. The teacher model is a large and complex deep neural network trained on massive MIMO data to predict the UE position. The distilled model is a smaller and more lightweight neural network designed to mimic the behavior of the teacher model while retaining most of its accuracy. The distiller is used to transfer the knowledge from the teacher model to the distilled model. It takes the teacher model and the student model as inputs and outputs a distilled model that is more efficient and accurate than the base model. The proposed system was evaluated on a drone scenario generated using DeepMIMO. The results show that the distilled model is more accurate than the base model, with a lower average magnitude of errors between the models' predictions and the actual values. The distilled model also has a more consistent performance across different UEs, indicating that it is able to generalize well and predict the UE positions accurately for different UEs. The proposed system can be used to improve the efficiency and performance of UE positioning in massive MIMO systems, especially when using FRIS.

A significant strength of this model is its remarkable accuracy in UE positioning, a critical factor for optimizing network performance and enhancing user experience in 5G and beyond. This heightened accuracy stems from the model's ability to intricately analyze complex signal patterns using advanced deep learning techniques. It effectively harnesses the vast spatial information provided by massive MIMO systems, ensuring precise localization even in environments with challenging signal propagation characteristics, such as urban landscapes with multiple obstacles and reflective surfaces. Another notable advantage is the model's scalability. Designed to accommodate the expanding scale and complexity of modern wireless networks, it can be seamlessly integrated into larger systems without substantial modifications. This scalability is crucial for adapting to the ever-growing number of devices and the increasing demand for high-speed, reliable wireless connectivity.

Furthermore, the model exemplifies a significant leap in computational efficiency, particularly in its distilled version. While the comprehensive teacher model encapsulates a wealth of knowledge from extensive training on massive datasets, the distilled student model retains this depth of understanding in a more compact and computationally efficient form. This efficiency is pivotal for practical deployment, especially in scenarios where computational resources are limited or costly. It facilitates the model's integration into existing infrastructure without necessitating extensive hardware upgrades, making it a cost-effective solution for network operators. Moreover, the incorporation of FRIS in the model represents an innovative approach to enhancing signal quality and coverage. By intelligently manipulating electromagnetic waves, FRIS can optimize signal propagation and mitigate common issues such as interference and signal degradation. This integration not only bolsters the model's positioning capabilities but also opens new avenues for improving overall network performance. In summary, the combination of high accuracy, scalability, computational efficiency, and the novel use of FRIS sets this model apart as a groundbreaking solution in the field of wireless communications.

Despite the promising results demonstrated by our knowledge distillation-based deep learning model for UE positioning in massive MIMO systems, there are inherent limitations that need to be acknowledged. Firstly, the computational complexity of the model, particularly the teacher component, is significant. While knowledge distillation effectively reduces this complexity in the student model, the initial training and setup of the teacher model require substantial computational resources, which might not be feasible in all practical scenarios. Additionally, the performance of the model heavily relies on the quality and volume of the training data. In scenarios where training data is limited or not sufficiently representative of real-world conditions, the model's accuracy and reliability may be compromised. Another limitation is the model's generalizability. While it shows high efficiency in controlled scenarios, its adaptability to diverse and dynamically changing real-world environments, such as urban areas with complex signal propagation patterns, has not been fully tested.

To address these limitations, our future work will focus on several key areas. Firstly, we plan to explore more efficient training algorithms and network architectures that could reduce the computational burden of the teacher model without compromising its performance. This could involve the use of more advanced optimization techniques or lightweight neural network structures. Secondly, efforts will be made to enhance the model's robustness to diverse data sets, including those with limited or noisy data. Techniques such as data augmentation, transfer learning, and semi-supervised learning methods could be employed to improve the model's performance in less-than-ideal data conditions. Furthermore, to improve the model's generalizability, we aim to test and refine it in a variety of real-world environments. This would involve deploying the model in different settings, collecting real-world performance data, and iteratively refining the model based on this feedback.

Finally, bridging the gap between theoretical models and practical applications will be a crucial part of our future work. This will involve close collaboration with industry partners to test the model in real-world telecom scenarios. By doing so, we can gain valuable insights into the model's practical limitations and advantages, which can guide further refinements. Additionally, considering the rapid evolution of wireless communication technologies, we intend to continuously update and adapt our model to align with emerging standards and practices, such as 6G networks. This ongoing adaptation will ensure that our research remains relevant and useful in the ever-evolving landscape of wireless communications. Through these efforts, we aim not only to enhance the technical capabilities of our model but also to ensure its practical applicability and value in real-world scenarios.

REFERENCES

[1] M. H. Siddiqui, K. Khurshid, I. Rashid, A. A. Khan, and K. Ahmed, "Optimal massive MIMO detection for 5G communication systems via hybrid n-bit heuristic assisted-VBLAST," *IEEE Access*, vol. 7, pp. 173646–173656, 2019.

[2] M. E. Diago-Mosquera, A. Aragón-Zavala, and G. Castañón, "Bringing it indoors: A review of narrowband radio propagation modeling for enclosed spaces," *IEEE Access*, vol. 8, pp. 103875–103899, 2020.

[3] Z. Wang, C. Li, and Y. Yin, "A meta-surface antenna array decoupling (MAAD) design to improve the isolation performance in a MIMO system," *IEEE Access*, vol. 8, pp. 61797–61805, 2020.

[4] N. K. Kundu and M. R. McKay, "Channel estimation for reconfigurable intelligent surface aided MISO communications: From LMMSE to deep learning solutions," *IEEE Open J. Commun. Soc.*, vol. 2, pp. 471–487, 2021.

[5] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[6] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[7] D. Dardari, P. Closas, and P. M. Djurić, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Veh. Technol.*, vol. 64, no. 4, pp. 1263–1278, Apr. 2015.

[8] F. Wen, H. Wymeersch, B. Peng, W. P. Tay, H. C. So, and D. Yang, "A survey on 5G massive MIMO localization," *Digit. Signal Process.*, vol. 94, pp. 21–28, Nov. 2019.

[9] L. Dai, R. Jiao, F. Adachi, H. V. Poor, and L. Hanzo, "Deep learning for wireless communications: An emerging interdisciplinary paradigm," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 133–139, Aug. 2020.

[10] C. Li, S. De Bast, E. Tanghe, S. Pollin, and W. Joseph, "Toward fine-grained indoor localization based on massive MIMO-OFDM system: Experiment and analysis," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5318–5328, Mar. 2022.

[11] N. S. Perovic, L.-N. Tran, M. Di Renzo, and M. F. Flanagan, "Achievable rate optimization for MIMO systems with reconfigurable intelligent surfaces," *IEEE Trans. Wireless Commun.*, vol. 20, no. 6, pp. 3865–3882, Jun. 2021.

[12] P. W. Wolniansky, G. J. Foschini, G. D. Golden, and R. A. Valenzuela, "V-BLAST: An architecture for realizing very high data rates over the rich-scattering wireless channel," in *Proc. URSI Int. Symp. Signals, Syst., Electron.*, Oct. 1998, pp. 295–300.

[13] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[14] A. Fascista, A. Coluccia, H. Wymeersch, and G. Seco-Granados, "RIS-aided joint localization and synchronization with a single-antenna mmWave receiver," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Jun. 2021, pp. 4455–4459.

[15] A. Alkhulaifi, F. Alsahli, and I. Ahmad, "Knowledge distillation in deep learning and its applications," *PeerJ Comput. Sci.*, vol. 7, p. e474, Apr. 2021.

[16] A. B. Mazlan, Y. H. Ng, and C. K. Tan, "A fast indoor positioning using a knowledge-distilled convolutional neural network (KD-CNN)," *IEEE Access*, vol. 10, pp. 65326–65338, 2022.

[17] N. Abuzainab, M. Alrabeiah, A. Alkhateeb, and Y. E. Sagduyu, "Deep learning for THz drones with flying intelligent surfaces: Beam and handoff prediction," in *Proc. IEEE Int. Conf. Commun. Workshops*, Jun. 2021, pp. 1–6.

[18] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.

[19] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," 2015, *arXiv:1503.02531*.

[20] K. Kong, W.-J. Song, and M. Min, "Knowledge distillation-aided end-to-end learning for linear precoding in multiuser MIMO downlink systems with finite-rate feedback," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 11095–11100, Oct. 2021.

ABDULLAH AL-AHMADI received the B.S. degree in electrical engineering from Saba University, Yemen, in 2005, and the M.S. and Ph.D. degrees in electrical engineering from Universiti Teknologi Malaysia, in 2008 and 2012, respectively. Since 2013, he has been an Assistant Professor with the Department of Electrical Engineering, College of Engineering, Majmaah University, Saudi Arabia. His research interests include wireless communications, massive MIMO, Bayesian networks, and machine learning.