Enhancing IRS Localization via Deep Learning-Based AOA and Distance Estimation

Nikhila Kotlapara Surendran[1,*], Syed Samiul Alam[1], Haolin Tang[2], Tahmid Tahi[1], Wei Wang[3] and Yanxiao Zhao[1]

[1]*Department of Electrical and Computer Engineering, Virginia Commonwealtah University, Richmond, VA, USA*

[2]*School of Engineering and Computing, Fairfield University, Fairfield CT, USA*

[3]*Department of Computer Science, San Diego State University, San Diego, CA, USA*

Abstract: In Intelligent Reflecting Surface (IRS)-assisted communication systems, accurate user localization, particularly Angle-Of-Arrival (AoA) and range estimation are challenging due to the computational complexity and limited resolution of traditional Multiple Signal Classification (MUSIC) algorithms. This paper introduces a hybrid IRS framework that combines machine learning with a modified MUSIC algorithm to achieve high-precision localization and enhanced security. The system integrates two Convolutional Neural Networks (CNNs): RefineNet, which refines AoA and range estimates from MUSIC pseudo-spectra, and ElementNet, which optimizes the number and placement of active IRS elements to balance accuracy with resource efficiency. Notably, ElementNet shows that only eight active elements are sufficient to obtain 90% of the best achievable localization accuracy, highlighting the efficiency of the proposed design. Validation on the DeepMIMO dataset demonstrates superior range accuracy and AoA precision. This work sheds light on the secure and high-precision localization for diverse wireless applications.

Keywords: Intelligent reflecting surface, Deep learning, MUSIC Algorithm, Localization, DeepMIMO.

1. INTRODUCTION

The transition to Sixth-Generation (6G) wireless networks demands unprecedented precision in user localization and signal security, particularly in Non-Line-of-Sight (NLOS) environments where conventional communication systems fall short. Emerging applications, such as autonomous vehicles, smart cities, and augmented reality, require centimeter-level positioning accuracy.

However, traditional base station-centric approaches struggle to overcome obstacles such as buildings and terrain, which obstruct direct signal paths and create coverage gaps and security vulnerabilities.

Intelligent Reflecting Surfaces (IRSs) have emerged as a transformative solution to these challenges. These reconfigurable metasurfaces comprise arrays of passive and active elements that can dynamically manipulate electromagnetic waves by adjusting their phase, amplitude, and polarization. Unlike traditional active relays that consume significant power, IRSs operate primarily in passive mode, making them energy-efficient and cost-effective for deployment on building facades, billboards, and urban infrastructure. By creating virtual line-of-sight paths, IRSs can extend coverage, enhance signal strength, and enable secure communication in complex propagation environments.

Despite their promise, effective IRS deployment requires accurate estimation of User Equipment (UE) positions through AoA and distance measurements. The Multiple Signal Classification (MUSIC), one of AoA algorithms, is widely used for this purpose, as it can separate signal and noise subspaces to achieve high-resolution parameter estimation. However, classical MUSIC faces significant limitations in IRS applications. First, its range accuracy is fundamentally constrained by grid resolution. Finer grids improve precision but exponentially increase computational complexity, making real-time implementation challenging. Second, AoA accuracy relies heavily on the number and placement of active sensing elements; however, increasing these elements escalates hardware costs and power consumption.

This paper addresses these limitations through a hybrid IRS system that combines a modified MUSIC algorithm with modern machine learning techniques. In recent years, deep learning has become an essential tool across many areas of wireless communications, *e.g.*, powering advances in channel estimation [22], channel-state information reconstruction [8, 21], sensing [10, 9], and security [1, 2]. Motivated by these approaches, our work employs two convolutional neural networks (CNNs) to overcome the trade-offs inherent in traditional methods. RefineNet processes MUSIC pseudo-spectra to deliver centimeter-level range and sub-degree AoA precision without increasing grid resolution, while ElementNet optimizes the number and placement of active elements to balance localization accuracy with resource efficiency.

*Address correspondence to this author at the Department of Electrical and Computer Engineering, Virginia Commonwealth University, Richmond, VA, USA; E-mail: {kotlaparasn,alams10,tahit,yzhao7}@vcu.edu

Received on 29-09-2025 Accepted on 27-10-2025 Published on 10-11-2025

E-ISSN: 000-000/25

The main contribution of this work can be summarized as follows:

- We propose a modified MUSIC algorithm that separates AoA and range estimation into decoupled stages, thereby substantially reducing the computational cost compared to conventional 3D searches. A subsequent CNN model maintains high localization precision despite reduced search complexity.
- To overcome the resolution limits of grid-based estimation and to reduce hardware cost, we introduce two dedicated CNN models: RefineNet, which learns spatial patterns in the MUSIC pseudo-spectrum to deliver centimeter-level range accuracy and sub-degree AoA precision, and ElementNet, which predicts the optimal number and spatial placement of active IRS sensing elements, achieving comparable localization accuracy with significantly fewer active elements.
- The joint framework is evaluated on the DeepMIMO [4] ray-tracing dataset, demonstrating robust operation under multipath propagation and NLOS conditions. Results confirm consistent centimeter-level range estimation, sub-degree angular accuracy, and reliable secure-user classification in practical, interference-rich wireless environments.
- In addition, our evaluation demonstrates that only eight active IRS elements (out of 625) are sufficient to achieve over 90% of the best achievable localization accuracy. This result highlights a practical trade-off between hardware complexity and estimation accuracy, significantly reducing cost while preserving near-optimal localization performance.

The remainder of the paper is organized as follows. Section 2 discusses related works, while Section 3 and Section 4 presents the system model and the methodology, respectively. Section 5 discusses the results. Finally, Section 6 offers concluding remarks on the findings of this study.

2. RELATED WORK

AoA and Distance Estimation in IRS-Assisted Localization: Accurate user localization in IRS-assisted wireless systems requires estimating both AoA and distance. Classical approaches for AoA estimation analyze the covariance structure of received signals, with MUSIC being prominent due to its high resolution and the ability to detect multiple signals at relatively low computational cost [6, 11, 14, 28].

For IRS-assisted systems specifically, Tang *et al*. [24] demonstrated that MUSIC-based AoA estimation can be performed using hybrid passive-active elements. However, their approach was limited to angular estimation with multidimensional grid searches. In contrast, when users are in the radiative near-field of a large antenna array or IRS, spherical wavefronts introduce dependencies on both AoA and distance, necessitating joint estimation as pursued in our work.

Early works extended MUSIC into 2D or 3D grid searches [16, 15, 29], but direct extension to 3D cases suffers from prohibitive computational complexity. Recent works has focused on decoupling strategies that estimate angular parameters and distance sequentially. Ramezani *et al*. [20] proposed an efficient modified MUSIC algorithm that separates estimation of azimuth, elevation, and range, though multiple matrix inversions limit scalability to large IRS setups. In IRS-assisted systems, the large aperture of the surface greatly increases the chances of users being in the radiative near-field, which in turn requires the simultaneous estimation of azimuth, elevation, and distance parameters for user localization.

Our approach combines elements from these research directions. We adopt MUSIC for IRS-based signal classification from Tang *et al*. [24], but follow the decoupling strategy introduced by Ramezani *et al*. [20] and others [15, 29] to reduce computational complexity. Prior work reports accuracy comparable to 3D-MUSIC with substantially lower complexity. We use a 2D AoA search followed by K independent 1D range searches, preserving accuracy while offering lower computational burden.

IRS-Aided Signal Classification and Active Element Optimization: Hybrid IRS architectures, where most elements are passive and only a few are active, extend functionality beyond reflection into sensing and classification. Tang *et al*. [24] showed that a small set of active IRS elements can effectively turn the IRS into a smart antenna array with minimal overhead. However, AoA-only classification becomes unreliable when signals arrive from closely spaced directions, and dense grid searches limit real-time applicability.

To address these issues, our work combines angular and distance features, first estimating azimuth and elevation using a decoupled 2D MUSIC search and then refining user distances with a 1D MUSIC scan. Previous works on active element selection have generally used fixed numbers or simple contiguous subarrays. Zhi and Chia [29] proposed symmetric subarray division, while Tang *et al*. [24] demonstrated reliable classification with few active sensors. The

optimal number and placement of active elements remain open challenges for large IRS arrays [19].

Our work treats active element selection as a learning problem through ElementNet, a lightweight CNN that predicts both the number and placement of active elements required for robust AoA estimation. Instead of uniform or block-based selection, our model identifies the most informative elements, minimizing active hardware while maintaining accuracy.

3. SYSTEM MODEL

3.1. Problem Statement

We consider a hybrid IRS system serving $K = 4$ UEs in a multipath propagation environment. The IRS consists of $N = 625$ total elements arranged in a 25×25 uniform planar array, of which $M \leq N$, where M denotes the number of elements that operate in active sensing mode while the remainder function as passive reflectors. The system must simultaneously achieve two objectives: accurate user localization and efficient resource utilization. The overall system model is illustrated in Figure **1**, where ElementNet selects a sparse set of active sensing elements, MUSIC provides coarse AoA and range estimates, and RefineNet further refines these estimates to achieve high-precision UE localization.

The system is based on the following given parameters:*(i) A snapshot matrix* $H \in C^{M \times T}$ $(M = 100,$ $T = 3000)$*:* The received channel is represented by a snapshot matrix $H \in C^{M \times T}$, where C denotes the complex domain, $M = 100$ corresponds to the number of active IRS elements in the 10×10 subarray, and $T = 3000$ is the number of temporal snapshots collected. The snapshot length of $T = 3000$ was chosen to provide sufficient statistical diversity for training while remaining computationally efficient. The dataset is generated using the DeepMIMO O1_28 scenario at $3.5\mathbf{GHz}$ with carrier wavelength $\lambda = 0.0857\mathbf{m}$ [4]. *(ii) Coarse AoA estimates* (ϕ_k, θ_k) are obtained from a 2D MUSIC search over a 301×301 grid with angular resolution $\Delta\phi = \Delta\theta = \pi / 300\mathbf{rad}$ [23]. Here, ϕ_k and θ_k denote the azimuth (horizontal) and elevation (vertical) AoAs of the k-th user, respectively, representing the 3D direction of arrival at the IRS. These coarse estimates are subsequently refined by RefineNet to achieve sub-degree precision. *(iii) Coarse range estimates* (r_k) are obtained using a one-dimensional MUSIC search. Instead of performing a full 3D search over (ϕ, θ, r), which is computationally expensive, the coarse AoA estimates (ϕ_k, θ_k) are first fixed and then a search is performed across 161 uniformly spaced range points within $r \in [r_{\min} - 5, r_{\max} + 5]$, with resolution $\Delta r \approx 0.1\mathbf{m}$. The peak of the MUSIC pseudo-spectrum over this grid gives the coarse range r_k, which is later refined by RefineNet for centimeter-level accuracy. *(iv) A* 25×25 *IRS:* The IRS has known element positions, with the number M and placement of active elements optimized under hardware and power constraints.

The goals are to: (i) Refine both AoA and range (coarse estimates from MUSIC algorithm) using RefineNet, learning a mapping, $R^{301 \times 301} \rightarrow R^{2K}$ that takes the 2D pseudo-spectrum to output refined angles $\{\phi_1, \ldots, \phi_K, \theta_1, \ldots, \theta_K\}$, Here, $\{\phi_1, \ldots, \phi_K, \theta_1, \ldots, \theta_K\}$ denote the azimuth and elevation angles-of-arrival of the K user signals respectively. The range is produced by a companion CNN (RefineNet-Range) with K outputs $[r_1, \ldots, r_K]$ from the same input spectra, *i.e.*, $R^{301 \times 301} \rightarrow R^K$ for all ranges (r_k) for $(K = 4)$ UEs [26]. (ii) Optimize the number M and placement of active elements using ElementNet, a CNN that maps channel

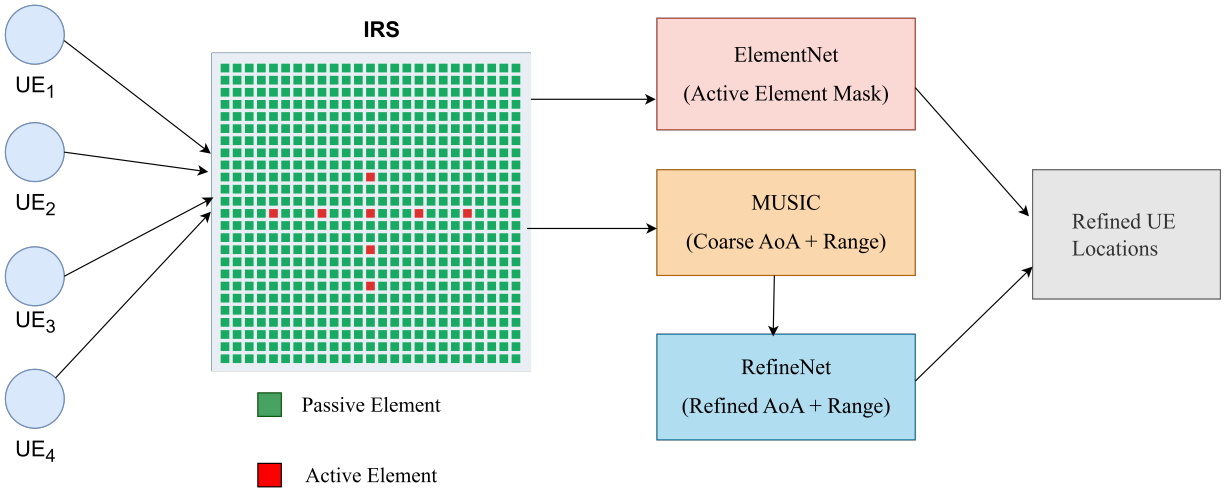

Figure 1: Hybrid IRS-assisted system model integrating active/passive elements, MUSIC-based coarse estimation, and CNN-based refinement (ElementNet + RefineNet) for accurate UE localization.

features (ϕ,θ,r) and IRS configuration to a 25×25 binary mask, selecting the minimum M to yields minimum AoA error within 10% of the minimum observed value, thereby achieving over 90% of the localization accuracy [27].

3.2. Architectures of the Proposed Neural Networks

This subsection describes the CNN architectures of RefineNet and ElementNet, which utilize convolutional layers to process spatial data such as pseudo-spectrum and IRS grid, predicting refined estimates and optimal element configurations. These models employ convolutional layers to extract spatial patterns, incorporating $L2$ regularization and dropout to prevent overfitting, and are designed with tailored outputs to predict continuous values (angles and ranges). For ElementNet, we explicitly add an $L1$ sparsity term to encourage compact masks (element placement). CNNs are particularly well-suited for these tasks due to their ability to detect local patterns, such as peaks in the pseudo-spectrum for AoA and range estimation or optimal element configurations in the IRS grid, and to learn complex mappings from DeepMIMO data [26, 27].

3.2.1. RefineNet for AoA and Range Refinement

The CNN-based RefineNet model (as shown in Figure **2**) processes the 2D MUSIC pseudo-spectrum to predict refined AoA and range estimates. The network takes a $301\times301\times1$ input representing the pseudo-spectrum and applies two convolutional layers with 16 and 32 filters (both 5×5 kernels) using ReLU activation and L2 regularization (0.01) to extract hierarchical spatial features. The resulting feature maps are flattened and passed through a 64-unit dense layer with ReLU and L2 regularization, followed by dropout (0.5 rate) to prevent overfitting. For angle estimation, the output layer produces $2K$ values $[\phi_1,\ldots,\phi_K,\theta_1,\ldots,\theta_K]$, while for range estimation, a separate network with identical architecture outputs K values $[r_1,\ldots,r_K]$ from the same input spectra. With approximately $15{,}000$ parameters, RefineNet remains computationally efficient, enabling real-time inference for practical localization applications.

3.2.2. ElementNet for Active Element Optimization

The ElementNet (as shown in Figure **3**) is designed to optimize the number and placement of active IRS elements. The architecture accepts a $25\times25\times(3K+1)$ input tensor, where $3K$ channels contain the broadcast channel features $[\phi_1,\ldots,\phi_K,\theta_1,\ldots,\theta_K,r_1,\ldots,r_K]$ for K users, plus one auxiliary channel. Note that during training, we do not feed the ground-truth mask as input to avoid label leakage. The network begins with two convolutional layers: the first applies 32 filters with 3×3 kernels and ReLU activation to extract spatial patterns in element placement and channel features, while the second uses 64 filters with 3×3 kernels and ReLU activation to enhance feature extraction. The resulting feature maps are then flattened into a one-dimensional vector and passed through a dense layer with 128 neurons and ReLU activation to learn complex relationships between channel characteristics and optimal element configurations. Finally, the output layer consists of 625 neurons with sigmoid activation, which are reshaped into a 25×25 binary mask indicating the predicted positions of the active elements.

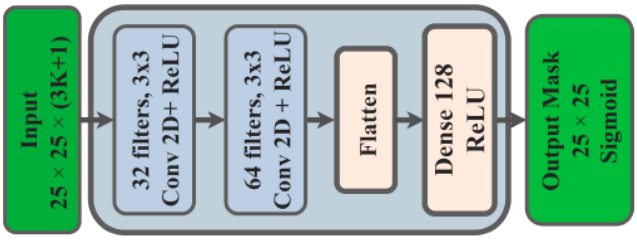

Figure 3: ElementNet Architecture.

ElementNet uses a custom loss function combining mean squared error (MSE) for accuracy and a sparsity penalty (weighted by 0.1) to favor fewer active elements. It predicts a mask, from which the top M elements are selected, with M chosen to minimize AoA

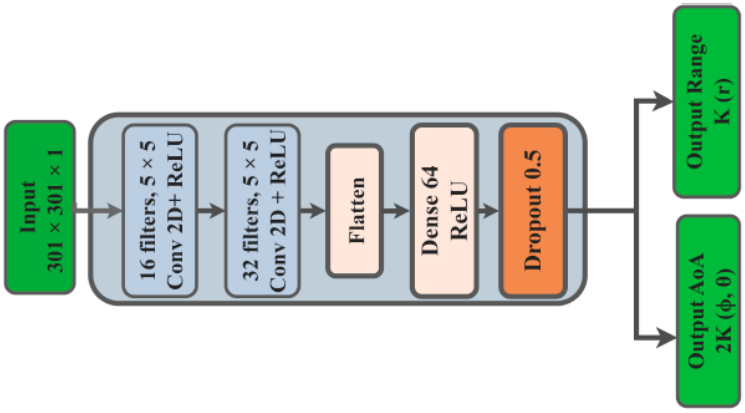

Figure 2: RefineNet architecture for AoA and Range refinement.

error within 10%, which is sufficient to achieve over 90% of the best achievable localization accuracy.

4. METHODOLOGY

This section details the dataset preparation, training methodology, and algorithms used in the proposed system. The methodology builds on the hybrid IRS architecture to refine localization estimates and optimize resource allocation through deep learning. This section also outlines the implementation details, including dataset preparation using the DeepMIMO framework and the training methodology for RefineNet and ElementNet.

4.1. Dataset Preparation

The process of generating and preparing the DeepMIMO dataset for training and validation is explained here. The dataset is generated using the DeepMIMO O1_28 scenario at $3.5\mathbf{GHz}$ [4]. The steps involved in preparation are as follows:

Channel Generation: Select K UEs with the smallest LOS distances (*e.g.*, 6–$12\mathbf{m}$) from rows 1–10, columns 1–10. Generate ($T=3000$) narrowband snapshots per UE, combining LOS and multipath returns, with Additive White Gaussian Noise (AWGN) at $\sigma^2=10^{-(200-30)/10}$.

MUSIC Processing: Apply spatial smoothing with 16 overlapping 7×7 subarrays, $M_{\mathbf{sub}}=49$, within the 10×10 active subarray [29]. Here, $M_{\mathbf{sub}}=49$ denotes the number of elements in each 7×7 subarray used for spatial smoothing, chosen as a trade-off between maintaining spatial resolution and enabling sufficient overlapping blocks for averaging. Compute the smoothed covariance matrix R, perform eigen-decomposition, and extract the noise subspace: $V_n\in\mathbb{C}^{49\times(49-K)}$ [26].

Pseudo-Spectrum for RefineNet: We generate 2D pseudo-spectra (301×301) for 30 scenarios, where each scenario corresponds to one channel realization in the DeepMIMO O1_28 environment. A scenario is defined by a specific set of user positions, their associated channel snapshots, and the resulting pseudo-spectra with ground-truth labels. The choice of 30 scenarios provides a practical balance between dataset diversity and computational cost. With $K=4$ users and $T=3000$ snapshots per user, this yields more than 360,000 training samples, which is sufficient for the proposed CNN models. The dataset is paired with true (ϕ,θ,r) values for $K=4$ UEs and is divided into 80% training (25 scenarios, 100 examples) and 20% validation (5 scenarios, 20 examples).

Placement Data for ElementNet: For each scenario, test M from $K=4$ to $K+6=10$, selecting the smallest M achieving AoA error $\leq5^\circ$. Generate binary masks (25×25) for optimal active element placements, paired with channel features (ϕ,θ,r). Total: 25 scenarios, 80% training (20 examples), 20% validation (5 examples).

4.2. Training Methodology

This subsection provides the training hyperparameters and algorithms for RefineNet and ElementNet, ensuring effective model optimization.

RefineNet: The network is trained using the Mean Squared Error (MSE) loss to minimize the difference between the predicted and true AoA and range values:

$$L_{\mathbf{AoA}}=\frac{1}{N}\sum_{k=1}^{K}[(\hat{\phi}_k-\phi_k)^2+(\hat{\theta}_k-\theta_k)^2] \tag{1}$$

$$L_{\mathbf{Range}}=\frac{1}{N}\sum_{k=1}^{K}\left(\hat{r}_k-r_k\right)^2 \tag{2}$$

where N is the total number of training samples, K is the number of users or signals being localized in each sample, $\hat{\phi}_k$, $\hat{\theta}_k$, and $\hat{r}_k$ are the predicted azimuth angle, elevation angle, and range for the k-th user, and ϕ_k, θ_k, and r_k are the corresponding ground-truth values. This loss quantifies the prediction error and guides the refinement process. The model uses the Adam optimizer with a learning rate of 10^{-4}, benefiting from adaptive momentum to accelerate convergence. A batch size of 4 ensures efficient gradient updates with limited memory use, while training proceeds for 20 epochs to achieve convergence without overfitting. Performance robustness is validated through 5-fold cross-validation [18], ensuring consistent accuracy across data partitions.

ElementNet: This network is trained using a custom loss that combines MSE with an L_1 sparsity penalty:

$$L=\mathcal{L}_{\mathrm{MSE}}+\lambda\|\mathbf{w}\|_1 \tag{3}$$

where $\mathcal{L}_{\mathrm{MSE}}$ measures the difference between the predicted and true element selection masks, $\mathbf{w}$ represents the selection weight vector for IRS elements, and $\lambda=0.1$ controls the sparsity strength. The MSE term is expressed as:

$$\mathcal{L}_{\mathrm{MSE}}=\frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i-w_i)^2 \tag{4}$$

with N denoting the number of IRS elements (*e.g.* $25\times25=625$), $\hat{w}_i\in[0,1]$ representing the predicted activation probability, and $w_i\in\{0,1\}$ being the binary

label of ground-truth (active/inactive). Thus, the overall ElementNet loss is:

$$L = \frac{1}{N}\sum_{i=1}^{N}(\hat{w}_i - w_i)^2 + \lambda \|\hat{\mathbf{w}}\|_1 \quad (5)$$

The first term enforces accuracy in predicting active elements, while the second encourages sparsity, allowing minimal, yet effective, element usage to balance localization performance and hardware cost [5]. The model uses the Adam optimizer (default learning rate) with a batch size of 8 to handle larger input tensors, trained for 20 epochs to ensure consistency with RefineNet. Validation is performed across five scenarios to evaluate the element placement optimization. All training is implemented in TensorFlow/Keras, with per-epoch data shuffling to enhance generalization using the DeepMIMO dataset [27].

4.3. Algorithm

This subsection presents the training algorithms for RefineNet and ElementNet, which offer a structured approach to model optimization based on the prepared dataset. Algorithm 1 captures the training of one of the two CNNs, RefineNet (for AoA and range). It includes standardization (using Standard Scaler), MSE loss, Adam optimizer with learning rate $\eta = 10^{-4}$, 20 epochs, and batch size 4, per the code. The validation step reflects the 5-fold cross-validation.

Algorithm 2 reflects Element Net's training (active element placement model) with a custom loss (MSE+0.1°øsparsity), Adam optimizer, 20 epochs, and batch size 8.

4.4. Evaluation Metrics

Unlike normalized metrics such as Normalized Mean Squared Error (NMSE) or Peak Signal-to-Noise

Algorithm 1 RefineNet Training for AoA and Range Refinement

Require: Training set (S_i, y_i) for $i = 1, ..., N$, where $S_i \in \mathbb{R}^{301\times301}$ is the 2D MUSIC pseudo-spectrum, and $y_i \in \mathbb{R}^{2K}$ (AoA) or $y_i \in \mathbb{R}^{K}$ (range) for $K = 4$ UEs. Validation set (S_j, y_j) for $j = 1, ..., M$. Neural network $g(S;\theta)$ (RefineNet CNN) with initial weights θ. Learning rate $\eta = 10^{-4}$. Number of epochs $E = 20$. Batch size $B = 4$. Input scaler (StandardScaler) and output scaler (StandardScaler for AoA and range)

Ensure: Trained weights θ^*

Initialize $\theta \leftarrow$ random (e.g., Glorot initialization)
Standardize training and validation inputs:
  $S_i \leftarrow$ input_scaler.fit_transform(S_i)
  $(S_j), S_j \leftarrow$ input_scaler.transform(S_j)
Standardize training and validation outputs:
  $y_i \leftarrow$ aoa_scaler.fit_transform(y_i) or range_scaler.fit_transform(y_i)
  $y_j \leftarrow$ aoa_scaler.transform(y_j) or range_scaler.transform(y_j)
for epoch $= 1$ to E **do**
  Shuffle the training set (S_i, y_i)
  Partition training set into mini-batches of size B
  for each mini-batch $\mathcal{B}$ do **do**
    for each $(S_i, y_i) \in \mathcal{B}$ do **do**
      compute prediction: $\hat{y}_i = g(S_i;\theta)$
    end for
    compute batch loss:

$$\mathcal{L} = \begin{cases} \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}(\hat{y}_i - y_i)^2, & \text{AoA head over } \mathbb{R}^{2K}, \\ \frac{1}{|\mathcal{B}|}\sum_{i\in\mathcal{B}}(\hat{y}_i - y_i)^2, & \text{Range head over } \mathbb{R}^{K}. \end{cases}$$

    update weights: $\theta \leftarrow \theta - \eta\nabla_\theta L$ (via Adam optimizer)
    evaluate validation loss on (S_j, y_j): $L_{\text{val}} = (1/M)\sum_j(\hat{y}_j - y_j)^2$
    Save trained weights θ^*, input scaler, and output scaler
  end for
end for
return θ^*

Algorithm 2 ElementNet Training for Active Element Optimization

Require: Training set (X_i, M_i) for $i = 1, \ldots, N$, where $X_i \in \mathbb{R}^{25\times25\times(3K+1)}$ is the input tensor (broadcast $[\phi, \theta, r]$ features $+1$ dummy/previous-mask channel), and $M_i \in \{0,1\}^{25\times25}$ is the target binary mask of active elements.

Neural network $h(X;\phi)$ (ElementNet CNN) with initial weights ϕ
Learning rate η (default Adam learning rate)
Number of epochs $E = 20$
Batch size $B = 8$

Ensure: Trained weights ϕ^*

Initialize $\phi \leftarrow$ random (e.g., Glorot initialization)
for epoch $= 1$ to E do **do**
  Shuffle the training set (X_i, M_i)
  Partition data into mini-batches of size B
  for each mini-batch $\mathcal{B}$ do **do**
    for each (w_i, δ_i) in $\mathcal{B}$ do **do**
      Compute prediction: $\hat{M}_i = h(X_i;\phi)$
    end for
    Compute batch loss: $L = (1/|\mathcal{B}|) \cdot \sum_{i\in\mathcal{B}} (\hat{M}_i - M_i)^2 + 0.1 \sum \hat{M}_i$
    Update weights: $\phi \leftarrow \phi - \eta \cdot \nabla_\phi L$ (e.g., via Adam optimizer)
  end for
end for
return ϕ^*

Ratio (PSNR) that are sometimes used in signal processing tasks, this work focuses on MSE and sparsity loss. These were chosen because they directly align with the network training objectives: minimizing MSE ensures accurate refinement of AoA and range, while sparsity encourages efficient element selection. Using the same metrics for both training and evaluation avoids inconsistencies and provides a clear measure of model effectiveness [20, 24].

MSE: For both AoA and range refinement using RefineNet, the MSE was adopted as the primary evaluation metric. It measures the average squared difference between the predicted values $\hat{y}_i$ and the ground truth y_i, and is defined as:

$$\mathbf{MSE} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2, \tag{6}$$

where N denotes the number of samples. Minimizing MSE ensures that the refined AoA and range estimates remain close to the true user positions. Because MSE was also the loss function during training, improvements observed in evaluation directly reflect successful network optimization [13].

Sparsity Loss: For active element selection in ElementNet, a sparsity-inducing regularizer was combined with the MSE objective to encourage compact solutions. The loss can be written as:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{MSE}} + \lambda \|\mathbf{w}\|_1, \tag{7}$$

where $\mathbf{w}$ represents the selection weights of the candidate elements and λ controls the sparsity penalty. This ensures that only a minimal but informative subset of elements is activated, reducing hardware cost while retaining high accuracy [25]. In practice, this loss enabled ElementNet to converge toward a solution with only eight active elements out of 625 while maintaining estimation accuracy within 10–15% of the best achievable precision.

Together, these two metrics capture the dual objectives of the proposed method: (i) improving the accuracy of the AoA estimation and the range, and (ii) minimizing the use of active elements without degrading performance.

5. RESULTS AND DISCUSSION

Conventional MUSIC-based approaches for IRS-assisted localization typically suffer from limited resolution and high computational burden, particularly in near-field scenarios. Our experiments therefore, evaluated whether combining a decoupled MUSIC formulation with CNN-based refinement could deliver more accurate estimates at reduced hardware and runtime cost.

For range estimation, coarse MUSIC outputs exhibited errors on the order of 0.5–2 m depending on user position, a level that is problematic in dense propagation environments. After refinement with RefineNet, these errors consistently dropped to within

Table 1: Range-Estimation Values Before and After CNN Refinement

UE	True Range (m)	Coarse Est. (m)	Refined Est. (m)	Coarse Error (m)	Refined Error (m)
1	90.88	88.90	91.06	−1.98	+0.18
2	90.93	90.45	90.99	−0.48	+0.09
3	91.17	91.80	91.07	+0.63	−0.09
4	91.14	89.55	91.11	−1.59	−0.03

5–20 cm on average, representing more than a fourfold improvement. Table **1**, shows values of true, coarse, and refined ranges, while Figure **4** presents the corresponding error reduction trends across epochs.

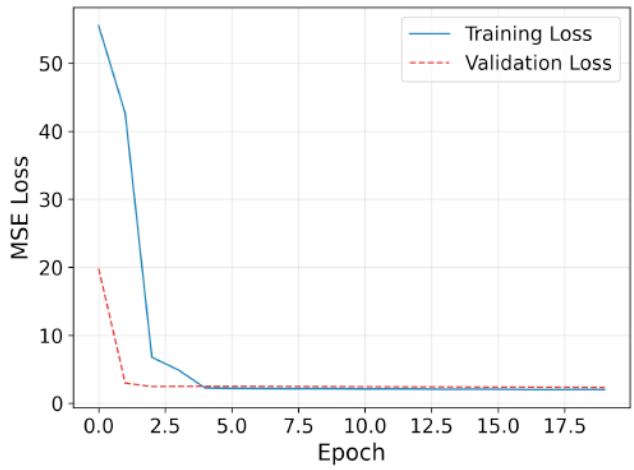

Figure 4: Training and validation loss for Range estimation refinement.

The training loss for range refinement started around 0.035 in the first epoch and steadily decreased to approximately 0.012 by epoch 15, while the validation loss followed a similar trend, converging near 0.015. These numerical values demonstrate that MSE is an effective training objective, since lower MSE values directly translate to centimeter-level positional accuracy.

For AoA estimation, coarse MUSIC results showed deviations of up to 2° in azimuth and about 0.2° in elevation. With refinement, these errors were reduced to within 0.1° – 0.2° on average across all evaluated users, demonstrating sub-degree precision. Table **2** shows that the refined AoA estimates follow the ground-truth values far more closely than the coarse MUSIC baseline.

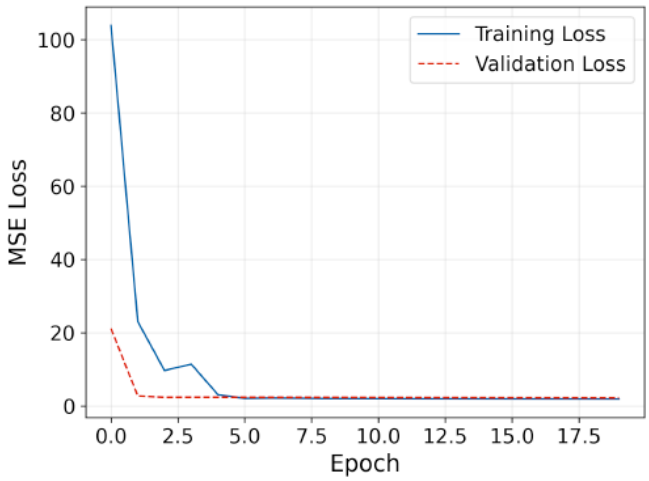

Figure 5: Training and validation loss for AoA estimation refinement.

The training and validation loss curves in Figure **5** further highlight the stability of the refinement process. Specifically, the training loss began near 0.030 in the first epoch and dropped to below 0.010 by epoch 20, while the validation loss converged slightly higher at around 0.012, with no divergence between the two

Table 2: Comparison of True and Estimated AoA Values Before and After CNN Refinement

UE	Azimuth Angle, ϕ (°)			Elevation Angle, θ (°)		
	True	Coarse Est.	Refined Est.	True	Coarse Est.	Refined Est.
1	−85.63	−87.60	−85.37	−2.52	−2.40	−2.52
2	−85.25	−82.80	−85.04	−2.52	−2.40	−2.52
3	−85.01	90.00	−84.68	−2.51	−2.40	−2.52
4	−83.87	−78.00	−84.12	−2.52	−2.40	−2.52

curves. This close alignment between the training and validation curves indicates a strong generalization without overfitting, underscoring the robustness of the proposed CNN-based refinement.

The optimization of active element usage was assessed through ElementNet. Here, the training loss is not limited to MSE but also incorporates a sparsity term, ensuring that the network learns to identify a minimal but effective subset of elements. As shown in Figure **6**, the combined objective converges smoothly, with the sparsity penalty stabilizing after around 10 epochs.

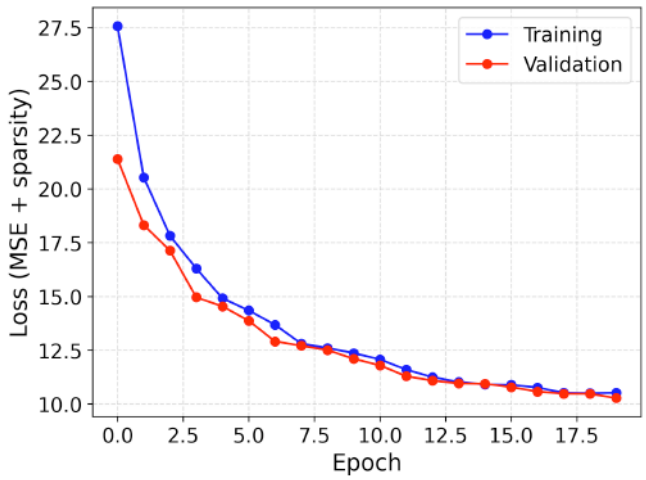

Figure 6: Training loss (MSE + sparsity) for active element placement optimization.

Figure **7** illustrates the effect of increasing the number of active IRS elements on the mean AoA error. With only eight active elements, the error is approximately 0.048°, while doubling to 16 elements reduces it to about 0.044°. At 24 elements, the error further decreases to 0.042°, and by 32 elements it approaches 0.040°. To determine the minimum number of active elements, a relative error–based criterion is applied. The mean AoA error obtained for each configuration of M active IRS elements is compared to the lowest (best) observed error, and the smallest M that achieves an error within 10% of this minimum is selected as the near-optimal configuration. For instance, in Figure **7**, the minimum mean AoA error is approximately $M = 8$, while the error at $M = 8$ is about 0.048°, giving a relative difference of $(0.048 - 0.040)/0.040 = 0.08$, or 8%. Hence, $M = 8$ satisfies the 10% near-optimality condition, effectively achieving over 90% of the best achievable localization accuracy. Beyond this point, the curve flattens, indicating that further increases yield only marginal improvements. These results confirm that activating a small subset of carefully placed elements is sufficient to approach the best achievable accuracy, providing a favorable trade-off between performance and hardware cost.

Taken together, these findings demonstrate that the proposed hybrid framework achieves centimeter-level range accuracy, sub-degree angular precision, and efficient active element usage with only eight sensors out of 625. While the results were obtained under DeepMIMO-generated scenarios and thus represent idealized conditions, the consistency of improvements across range, AoA, and hardware efficiency provides strong evidence of the methods practical potential. In real-world deployments, higher variance is to be expected; however, this approach provides valuable insights toward scalable and secure IRS-assisted localization.

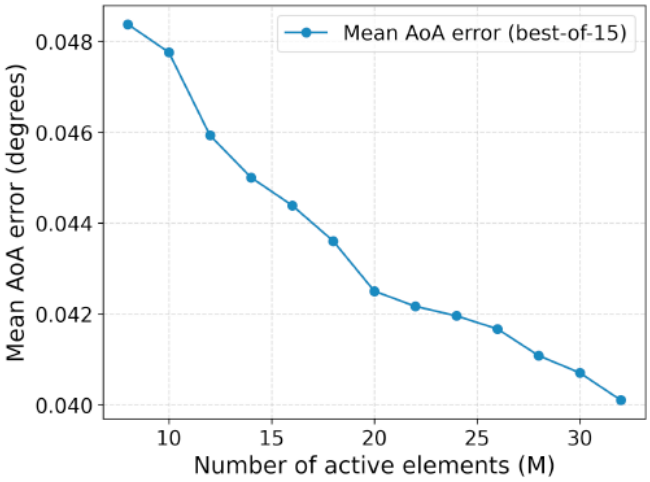

Figure 7: Mean AoA error versus number of active elements.

6. CONCLUSION

This paper presents a hybrid approach combining modified MUSIC algorithm with deep learning to achieve high-precision localization in IRS-assisted wireless communication systems. Our proposed system employs two specialized CNN models: RefineNet which refined AoA and range estimates from MUSIC pseudo-spectra, and ElementNet, which optimizes active element placement within the IRS. Our findings demonstrate that RefineNet achieves centimeter-level range accuracy (reducing errors from 50–200 cm to 5–20 cm) and sub-degree AoA precision (improving from 2° to within 0.1° – 0.2°). During experimentation, the most notable challenge was the high computational load associated with recomputing MUSIC pseudo-spectra for multiple random masks and training seeds, which increased CPU time despite GPU acceleration. Occasional stalls were observed due to memory constraints, and minor fluctuations appeared in validation loss due to floating-point precision, both of

which were mitigated using data standardization and TensorFlow mixed-precision computation. Equally significant, ElementNet determines that only 8 active elements out of 625 are sufficient to maintain localization performance. In other words, the mean AoA error at $M = 8$ remains within 10% of the minimum error obtained when all active-element configurations are considered, achieving over 90% of the optimal performance while greatly reducing hardware cost and energy savings. The proposed hybrid approach can be beneficial toward practical IRS deployment in applications such as autonomous vehicles, smart cities, and augmented reality where precise localization is essential.

ACKNOWLEDGMENTS

This research is supported in part by Virginia Commonwealth Cyber Initiative (https://cyberinitiative.org/).

CONFLICTS OF INTEREST

The author declared no conflicts of interest.

REFERENCES

[1] Alam, S.S., Chakma, A., Rahman, M.H., Bin Mofidul, R., Alam, M.M., Utama, I.B.K.Y., Jang, Y.M.: Rf-enabled deep-learning-assisted drone detection and identification: An end-to-end approach. Sensors 23(9), 4202 (2023). https://doi.org/10.3390/s23094202

[2] Alam, S.S., Tang, H., Zhao, Y., Luo, C., Wang, W., Dhar, N.K.: Defense against adversarial attacks for channel estimation models in ris-assisted communication

[3] Alam, S.S., Zhao, Y., Tang, H., Wang, W., Song, M.: Deep learning-based end-toend channel estimation and csi feedback for large ris-assisted wireless communication

[4] Alkhateeb, A.: Deepmimo: A generic deep learning dataset for millimeter wave and massive mimo applications. arXiv preprint arXiv:1902.06435 (2019)

[5] (arXiv), U.: Sparse deep learning models with the 1 regularization. arXiv preprint arXiv:2408.02801 (2024),

[6] Balanis, C.A.: Antenna Theory: Analysis and Design. Wiley, Hoboken, NJ, 4th edn. (2015)

[7] BjÅNornson, E., Wymeersch, H., Matthiesen, B., Popovski, P., Sanguinetti, L., De Carvalho, E.: Reconfigurable intelligent surfaces: A signal processing perspective with wireless applications. IEEE Signal Processing Magazine 39(2), 135-158 (2022). https://doi.org/10.1109/MSP.2021.3130549

[8] Chakma, A., Alam, S.S., Rahman, M.H., Jang, Y.M.: Deep decoder csinet for fdd massive mimo system. IEEE Wireless Communications Letters 12(12), 2073-2077 (2023). https://doi.org/10.1109/LWC.2023.3307164

[9] Fahad, N., Bulut, E.: Wifi csi based liquid temperature prediction: A physicsguided machine learning approach. In: 2025 IEEE 26th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM). pp. 122-130. IEEE (2025). https://doi.org/10.1109/WoWMoM65615.2025.00029

[10] Fahad, N., Touhiduzzaman, M., Bulut, E.: Ensemble learning based wifi sensing using spatially distributed tx-rx links. In: IEEE International Conference on Computing, Networking and Communications (ICNC), Honolulu, Hawaii, USA (2025). https://doi.org/10.1109/ICNC64010.2025.10993928

[11] Fried, D.L.: Differential angle of arrival: theory, evaluation, and measurement feasibility. Radio Science 10(1), 71-76 (1975). https://doi.org/10.1029/RS010i001p00071

[12] Gao, F., Gershman, A.B.: A generalized esprit approach to direction-of-arrival estimation. IEEE Signal Processing Letters 12(3), 254-257 (2005). https://doi.org/10.1109/LSP.2004.842276

[13] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press (2016), https://www.deeplearningbook.org/

[14] Gross, F.: Smart Antennas for Wireless Communications. McGraw-Hill Professional (2005)

[15] He, J., Swamy, M.N.S., Ahmad, M.O.: Efficient application of music algorithm under the coexistence of far-field and near-field sources. IEEE Transactions on Signal Processing 60(4), 2066-2070 (2012). https://doi.org/10.1109/TSP.2011.2180902

[16] Huang, Y.D., Barkat, M.: Near-field multiple source localization by passive sensor array. IEEE Transactions on Antennas and Propagation 39(7), 968-975 (1991). https://doi.org/10.1109/8.86917

[17] Khan, A.,Wang, S., Zhu, Z.: Angle-of-arrival estimation using an adaptive machine learning framework. IEEE Communications Letters 23(2), 294-297 (2018). https://doi.org/10.1109/LCOMM.2018.2884464

[18] Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence. vol. 2, pp. 1137-1143. Morgan Kaufmann (1995)

[19] Pan, Y., Pan, C., Jin, S., Wang, J.: Ris-aided near-field localization and channel estimation for the terahertz system. IEEE Journal of Selected Topics in Signal Processing 17(4), 878-892 (2023). https://doi.org/10.1109/JSTSP.2023.3285431

[20] Ramezani, P., Kosasih, A., BjÅNornson, E.: An efficient modified music algorithm for ris-assisted near-field localization. arXiv preprint arXiv:2409.14152 (2024). https://doi.org/10.1109/GLOBECOM52923.2024.10901595

[21] Samiul Alam, S., Chakma, A., Imran, A., Min Jang, Y.: Dcsinet: Effective handling of noisy csi in fdd massive mimo system. IEEE Wireless Communications Letters 13(9), 2556-2560 (2024). https://doi.org/10.1109/LWC.2024.3427316

[22] Samiul Alam, S., Tang, H., Yang, G., Luo, C., Wang, W., Zhao, Y.: Dsrnet: Hybrid deep learning-based channel estimation for ris-aided wireless communication. In: Wireless Artificial Intelligent Computing Systems and Applications. pp. 62-74.Springer Nature Singapore (2025). https://doi.org/10.1007/978-981-96-8725-1_6

[23] Schmidt, R.: Multiple emitter location and signal parameter estimation. IEEE transactions on antennas and propagation 34(3), 276-280 (1986). https://doi.org/10.1109/TAP.1986.1143830

[24] Tang, H., Zhao, Y., Wang, W.: Angle of arrival based signal classification in intelligent reflecting surface-aided wireless communications. In: International Conference on Wireless Algorithms, Systems, and Applications (WASA). pp. 56-67. Springer (2022). https://doi.org/10.1007/978-3-031-23902-1_5

[25] Tibshirani, R.: Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58(1), 267-288 (1996). https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[26] Wax, M., Kailath, T.: Optimum localization of multiple sources by passive arrays. IEEE Transactions on Acoustics, Speech, and Signal Processing 31(5), 1210-1217 (2003). https://doi.org/10.1109/TASSP.1983.1164183

[27] Wymeersch, H., He, J., Denis, B., Clemente, A., Juntti, M.: Radio localization and mapping with reconfigurable intelligent surfaces: Challenges, opportunities, and research directions. IEEE Vehicular Technology Magazine 15(4), 52-61 (2020). https://doi.org/10.1109/MVT.2020.3023682

[28] Xie, J., Fu, Z., Xian, H.: Spectrum sensing based on estimation of direction of arrival. In: 2010 International

Conference on Computational Problem-Solving. pp. 39-42. IEEE (2010).

[29] Zhi, W., Chia, M.Y.W.: Near-field source localization via symmetric subarrays. IEEE Signal Processing Letters 14(6), 409-412 (2007). https://doi.org/10.1109/LSP.2006.888390