IET Communications

WILEY

The Institution of Engineering and Technology

ORIGINAL RESEARCH OPEN ACCESS

Intelligent Reflecting Surface-Aided Wireless Networks: Deep Learning-Based Channel Estimation Using ResNet+UNet

Sakhshra Monga[1] | Aditya Pathania[1] | Nitin Saluja[1] | Gunjan Gupta[2] | Ashutosh Sharma[3,4]

[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India | [2]Department of Electrical, Electronics and Computer Engineering (DEECE), Cape Peninsula University of Technology, Cape Town, South Africa | [3]Business School, Henan University of Science and Technology, Luoyang, Henan, China | [4]School of Computer Science, University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India

Correspondence: Sakhshra Monga (sakhshra@chitkara.edu.in) | Gunjan Gupta (guptag@cput.ac.za)

Received: 18 March 2025 | **Revised:** 14 July 2025 | **Accepted:** 3 August 2025

Funding: The authors received no specific funding for this work.

ABSTRACT

Accurate channel estimation is essential for optimising intelligent reflecting surface-assisted multi-user communication systems, particularly in dynamic indoor environments. Conventional techniques such as least squares (LS), linear minimum mean square error (LMMSE), and orthogonal matching pursuit (OMP) suffer from noise sensitivity and fail to effectively capture spatial dependencies in high-dimensional intelligent reflecting surface (IRS)-assisted channels. To overcome these limitations, this work proposes a deep learning-driven ResNet+UNet framework that refines initial LS estimates using residual learning and multi-scale feature reconstruction. While UNet enhances channel estimation through hierarchical processing, efficiently decreasing noise and enhancing estimate accuracy, ResNet gathers spatial features. Simulation results show that the proposed method significantly outperforms existing methods across various performance metrics. In NMSE versus signal-to-noise ratio assessments, the proposed approach surpasses convolutional deep residual network (CDRN) by 59%, OMP by 81%, LMMSE by 114%, and LS by 115%. When IRS elements are modified, it overcomes CDRN by 60%, OMP by 78%, LS by 107%, and LMMSE by 110%. Along with this, recommended structure performs more effectively than CDRN by 39%, OMP by 44%, LS by 122%, and LMMSE by 129% across various antenna configurations. The proposed approach is particularly beneficial for augmented reality (AR) applications, where real-time, high-precision channel estimation ensures seamless data streaming and ultra-low latency, enhancing immersive experiences in AR-based communication and interactive environments. These results illustrate the proposed method's scalability and resilience, making it a suitable choice for next-generation IRS-assisted wireless communication networks.

1 | Introduction

Intelligent reflecting surfaces (IRS) have emerged as a transformative technology capable of shaping wireless communication channels between users and the base station (BS) to optimise system performance. The introduction of advanced electromagnetic materials has enabled the development of reconfigurable and passive metasurfaces, making IRS deployment practical and sustainable [1]. The phase transition of incoming signals can be modified by separately adjusting each of the many passive reflecting elements that make up an IRS. By carefully adjusting these phase shifts, the reflection pattern can be optimised to

Aditya Pathania and Nitin Saluja contributed equally to this work.

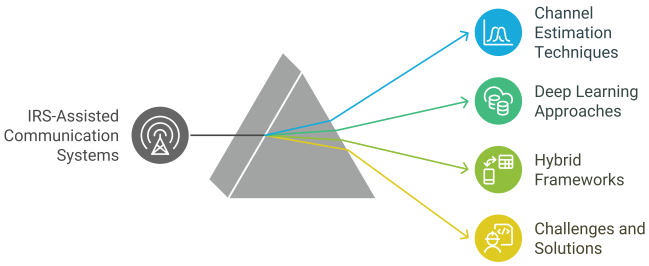


FIGURE 1 | Problems, estimates, and optimisation approaches in IRS networks.

create favourable wireless channels [2]. This improves transmission quality while maintaining low energy consumption. By utilising these characteristics, IRS can be included into wireless communication systems to boost spectrum and energy efficiency through the implementation of passive beamforming techniques. Signal-to-noise ratio (SNR) at the target receiver can be significantly improved by correctly adjusting the phase shifts [3].

Given these advantages, IRS-assisted communication systems have garnered substantial interest from both academia and industry. Research in this domain has focused on maximising network capacity and received SNR, enhancing energy and spectral efficiency, and securing physical-layer communications. However, the effectiveness of IRS depends critically on the accuracy of channel state information (CSI), which is essential for efficient beamforming [4]. Most existing studies assume perfect CSI knowledge, which is rarely feasible in real-world scenarios. Unlike conventional wireless systems, IRS is inherently passive and lacks the capability to transmit or receive training sequences or process signals directly. Consequently, individual channel state information for IRS-user and IRS-BS links is not separately available; instead, only a combined user-to-IRS-to-BS cascaded channel can be estimated [5].

This introduces two major challenges. First, the cascaded IRS-assisted channel deviates from the conventional Rayleigh fading model. An optimal minimum mean square error (MMSE) estimator requires solving complex multidimensional integrals, making it computationally infeasible for practical deployment. While linear MMSE (LMMSE) and least squares (LS) estimators are commonly used, their performance significantly lags behind the optimal MMSE estimator, leading to suboptimal channel estimation accuracy as in Figure 1. Secondly, cascaded channel exhibits a high-dimensional structure since IRS usually has plenty of reflecting elements. For large-scale IRS implementations, estimating such a channel using conventional methods such as LS or LMMSE becomes computationally expensive and impractical. Beyond wireless communication, the proposed IRS-based channel estimation framework holds significant potential for augmented reality (AR) applications. By ensuring low-latency, high-accuracy channel estimation, it enables real-time AR rendering and seamless data transmission, crucial for immersive AR experiences [6, 7]. Such challenges indicate the necessity for novel channel estimation techniques developed specifically for IRS-assisted wireless communication systems in order to promote their effective and useful implementation in next-generation networks [8].

Several optimisation techniques have been proposed to address the low channel estimation performance of IRS-assisted systems. For single-user scenarios, a certain approach utilises the use of a binary reflection-controlled least squares (LS) estimate method [9]. This method allows the base station (BS) to estimate the cascaded channel without interference by activating only one IRS element per time slot while keeping others inactive. However, received signal-to-noise ratio (SNR) at the BS is low because of the fact that only one element is active at a time, resulting in poor estimate accuracy. In addition, this method creates unnecessary latency for big IRS arrays [10].

In order to tackle these issues, an alternative method employs minimal variance unbiased estimator based on the discrete Fourier transform (DFT) to activate all reflecting elements in a particular period of time, boosting accuracy. Added to that, two methods based on parallel factors (PARAFAC) improve channel estimation using least squares bilinear estimate with Khatri–Rao factorisation [11]

Efforts have also been directed toward reducing the training overhead in IRS-assisted systems. A subsurface-based channel estimation method was introduced, wherein the IRS is partitioned into multiple independent subsurfaces, each sharing a common phase shift [12]. This significantly lowers the training complexity. Furthermore, a cascaded channel estimation framework based on sparse matrix factorisation and matrix completion has been proposed. For MIMO systems, channel estimation has been formulated as a sparse matrix recovery problem, employing compressed sensing to enhance estimation efficiency. A three-phase estimation framework was designed for multi-user systems, reducing estimation duration by leveraging redundancies among different users' reflecting channels. Moreover, an innovative IRS architecture featuring a single active radio frequency (RF) chain has been introduced, further minimising training overhead [13]. Although with these advances, the issue of obtaining high-accuracy channel estimate continues to be dealt with, calling for more practical and efficient methods.

In IRS-assisted communications, deep learning (DL) has only recently become an effective replacement for conventional model-driven approaches [14]. Passive beamforming, which is phase shift optimisation powered by DRL, and DRL-based security tactics for IRS-MUC systems are a few instances of DL applications in this field. However, these methods usually rely on flawless CSI, which is frequently unachievable. As denoising is an essential component of channel estimation, deep residual learning (DReL) has drawn attention because of its ability to enhance network performance and speed up training [15].

Motivated by these advancements, this research employs DReL to enhance channel estimation accuracy in IRS-assisted multi-user communication (IRS-MUC) systems. Our approach takes into consideration a typical IRS configuration without such improvements, in contrast to previous methods that need additional active sensors on the IRS. The proposed approach refines LS-based channel estimates using a neural network-based denoiser, unlike MMSE-based techniques that use them as initial coarse values [16]. Our technique is model-free and sufficiently adaptable to be used in many scenarios since it does not depend on previous statistical data.

TABLE 1 | Comparison of deep learning-based channel estimation approaches.

Study	Architecture	Residual connections	Multiscale fusion	IRS phase optimisation
Liu et al. [1]	CNN	No	Limited	No
Taha et al. [14]	CNN+UNet	No	Yes	No
Kundu and McKay [17]	Neural network	No	Limited	No
Guo et al. [18]	CNN	No	No	No
Zhou et al. [20]	ResNet+UNet	Yes	Yes	No
Ren et al. [21]	Model-based + CMMSE	No	No	Yes
Xu et al. [22]	LS + Codebook Opt.	No	No	Yes
Proposed	ResNet+UNet	Yes	Yes	Yes

1.1 | Related Work and Our Contributions

Several studies have explored deep learning for IRS-assisted channel estimation, mainly using convolutional neural network (CNN) or CNN+UNet architectures. Liu et al. [1] proposed a CNN-based denoising network, but without residual connections, limiting its ability to capture spatial dependencies in large IRS setups. Taha et al. [14] used a CNN+UNet model, but it also lacked residual learning, which affected performance under low-SNR or rapidly changing IRS environments. Kundu and McKay [17], and Guo et al. [18], also studied CNN-based models but did not incorporate multiscale fusion or phase optimisation mechanisms.

Compared to these works, our ResNet+UNet framework introduces three key improvements. First, residual connections enable the model to learn the difference between the LS estimate and the true channel, improving noise resilience [16, 19]. Second, multiscale feature fusion through UNet skip connections helps preserve spatial detail during reconstruction [14, 17]. Third, we incorporate dynamic IRS phase shift prediction, which most prior methods do not address [2]. This makes our approach more suitable for real-time deployment in large-scale IRS-MUC systems.

To clarify the distinctions, Table 1 summarises the key differences between the proposed framework and previous deep learning-based approaches.

We propose a hybrid ResNet+UNet-based channel estimation framework which employs deep feature extraction and hierarchical reconstruction to optimise performance with the aim to further enhance channel estimation accuracy and resiliency. While UNet's encoder–decoder layout increases multiscale feature representations for efficient channel estimation, ResNet successfully captures spatial correlations in channel matrices. By learning effective error correction functions, our model solves the drawbacks of traditional approaches and significantly improves accuracy even in dynamic IRS situations. The suggested strategy is a scalable solution for next-generation IRS-assisted communication systems, as it improves channel estimates and makes real-time phase shift tuning easier.

- In this research, a deep learning-driven technique for accurately extracting channel coefficients from noisy pilot-based observations utilising ResNet+UNet is described. Our method substantially enhances estimation accuracy by employing residual learning to refine coarse LS-based estimates in accordance with the MMSE principle.
- To implement the indicated structure, we set up a CNN-based structure that combines ResNet and UNet for feature extraction and noise reduction. Our model involves a CNN-based denoising block with an element-wise subtraction method, effectively capturing spatial channel features while reducing additive noise. This combination encourages resilience and flexibility under diverse channel conditions.
- While neural network-based estimators are usually challenging to study, we build ResNet+UNet approach as a mathematical function and receive an explicit analytical equation for estimator. This theoretical investigation demonstrates that the proposed estimator achieves performance on par with the LMMSE technique while far outperforming LS and orthogonal matching pursuit (OMP)-based approaches.
- To verify the effectiveness of our suggested method, we run extensive simulations, evaluating its performance in terms of NMSE across different SNR levels and channel dimensions. In addition, the ability of the model to enhance estimation accuracy can be seen by visual representations of learnt feature maps, which also show the noise suppression procedure.
- The deep learning-based convolutional deep residual network (CDRN) technique, LS, LMMSE, OMP, along with other traditional channel estimation techniques are thoroughly contrasted with our suggested ResNet+UNet-based estimator. Results show that, especially in high-noise situations with constrained pilot resources, our method delivers higher channel estimate accuracy.

The remainder of the paper is structured as follows: Section 2 presents system model for IRS-MUC, detailing the underlying mathematical framework and the integration of IRS phase shift optimisation. Section 3 discusses existing channel estimation techniques, including LS, LMMSE, OMP and CDRN, serving as benchmarks for performance evaluation and introduces the proposed ResNet+UNet-based approach, emphasising its deep learning-driven enhancement of LS estimates. Section 4 provides an extensive simulation results, comparing the proposed approach against conventional methods to assess its accuracy, robustness and computational efficiency. Finally, Section 5 concludes this paper with key findings and future research directions.

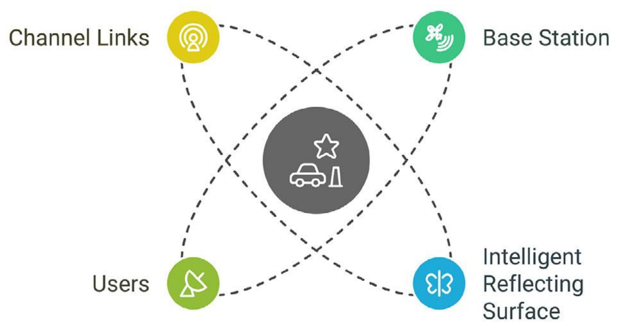


FIGURE 2 | Structure of IRS-aided communication with base station and users.

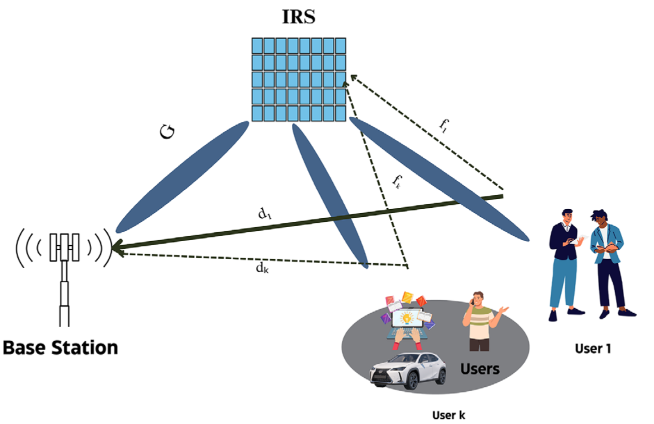


FIGURE 3 | IRS-assisted wireless communication system with base station and users.

2 | System Model and Problem Formulation

2.1 | System Model

We consider IRS-MUC system operating under a time-division duplex (TDD) protocol. The system consists of a BS equipped with an M-element antenna array, an IRS composed of N passive reflecting elements, and K single-antenna users as depicted in Figure 2. IRS enhances communication by dynamically adjusting phase shifts of reflected signals [23]. In TDD systems, channel reciprocity allows the estimation of downlink CSI using uplink channel measurements, thereby improving beamforming at both BS and IRS.

Received signals at the BS are influenced by two key links: direct channel between the users and BS and IRS-assisted reflected channel, which consists of two cascaded segments, namely, the user-to-IRS and the IRS-to-BS links as shown in Figure 3. The direct channel between user k and BS is denoted as $\mathbf{d}_k \in \mathbb{C}^{M\times 1}$, while channel between user k and the IRS is depicted by $\mathbf{f}_k \in \mathbb{C}^{N\times 1}$. The IRS-to-BS channel is given by $\mathbf{G} \in \mathbb{C}^{M\times N}$.

To control signal reflections, the IRS applies a phase-shift matrix defined as

$$\mathbf{R} = \mathrm{diag}(\mathbf{r}) \in \mathbb{C}^{N\times N}, \tag{1}$$

where $\mathbf{r} = [\beta e^{j\theta_1}, \beta e^{j\theta_2}, \dots, \beta e^{j\theta_N}]^T$, with $\beta \in [0, 1]$ representing reflection amplitude and θ_n being phase shift applied to nth IRS element [17].

The effective cascaded IRS-assisted channel is given by

$$\mathbf{B}_k = \mathbf{G} \cdot \mathrm{diag}(\mathbf{f}_k) \in \mathbb{C}^{M\times N}. \tag{2}$$

Thus, the overall channel matrix for user k, incorporating both direct and reflected paths, is expressed as

$$\mathbf{H}_k = [\mathbf{d}_k, \mathbf{B}_k] \in \mathbb{C}^{M\times(N+1)}. \tag{3}$$

During uplink channel estimation phase, pilot signals from users are received at BS [18]. Received signal at BS for cth IRS phase-shift configuration is given by,

$$\mathbf{S}_c = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{p}_c \mathbf{u}_k^H + \mathbf{V}_c, \tag{4}$$

where $\mathbf{p}_c = [1, \mathbf{r}_c^T]^T \in \mathbb{C}^{(N+1)\times 1}$ represents the IRS phase-shift vector, $\mathbf{u}_k$ is the pilot sequence transmitted by user k, and $\mathbf{V}_c$ represents an additive Gaussian noise matrix.

By aggregating multiple pilot signal observations across different IRS phase-shift configurations, the channel estimation problem is formulated as

$$\mathbf{X}_k = \mathbf{H}_k \mathbf{P} + \mathbf{Z}_k, \tag{5}$$

where $\mathbf{P}$ is the matrix containing the phase-shift vectors across multiple sub-frames, and $\mathbf{Z}_k$ represents the noise component. This system model mathematically characterises IRS-assisted communication, setting the foundation for efficient channel estimation and signal optimisation in IRS-MUC systems [24].

2.2 | Problem Formulation

Accurate channel estimation in IRS-MUC systems is challenging due to their passive nature of IRS elements, preventing direct CSI acquisition and necessitating advanced techniques to estimate high-dimensional cascaded channel efficiently.

Considered IRS-MUC system comprises a BS equipped with an M-element antenna array, an IRS with N passive reflecting elements, and K single-antenna users [19]. Direct channel between user k and BS is represented as $\mathbf{d}_k \in \mathbb{C}^{M\times 1}$, while user-to-IRS and IRS-to-BS channels are denoted by $\mathbf{f}_k \in \mathbb{C}^{N\times 1}$ and $\mathbf{G} \in \mathbb{C}^{M\times N}$, respectively. The IRS applies phase shifts to the incoming signals using a diagonal phase-shift matrix $\mathbf{R} = \mathrm{diag}(\mathbf{r}) \in \mathbb{C}^{N\times N}$, where $\mathbf{r} = [e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}]^T$ represents the phase shifts introduced by IRS elements. Overall cascaded channel between user k and BS via IRS can be expressed as

$$\mathbf{B}_k = \mathbf{G} \cdot \mathrm{diag}(\mathbf{f}_k) \in \mathbb{C}^{M\times N}. \tag{6}$$

Thus, the complete channel matrix for user k incorporating both direct and reflected links is given by

$$\mathbf{H}_k = [\mathbf{d}_k, \mathbf{B}_k] \in \mathbb{C}^{M\times(N+1)}. \tag{7}$$

During the uplink training phase, pilot signals transmitted by the users arrive at the BS through both the direct and IRS-assisted reflected links. The received pilot signal at the BS for the cth IRS phase-shift configuration can be expressed as,

$$\mathbf{S}_c = \sum_{k=1}^{K} \mathbf{H}_k \mathbf{p}_c \mathbf{u}_k^H + \mathbf{V}_c, \tag{8}$$

where $\mathbf{p}_c = [1, \mathbf{r}_c^T]^T \in \mathbb{C}^{(N+1)\times 1}$ represents the IRS phase-shift configuration for the cth training instance, $\mathbf{u}_k$ is the pilot sequence assigned to user k, and $\mathbf{V}_c$ is an additive noise matrix representing background interference.

By stacking pilot signals over multiple IRS configurations, the received signal matrix for user k is rewritten as,

$$\mathbf{X}_k = \mathbf{H}_k \mathbf{P} + \mathbf{Z}_k, \tag{9}$$

where $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_C] \in \mathbb{C}^{(N+1)\times C}$ represents the IRS phase-shift matrix across C training instances, and $\mathbf{Z}_k$ accounts for noise in the received signal.

Channel estimation in IRS-assisted systems aims to reconstruct $\mathbf{H}_k$ from noisy pilot observations $\mathbf{X}_k$ while addressing challenges such as non-Rayleigh fading, passive IRS elements, and high-dimensionality constraints. An efficient framework must accurately capture channel characteristics, mitigate noise, and optimise phase shifts for scalable multi-user IRS deployments.

3 | Channel Estimation Techniques

3.1 | Existing Channel Estimation Techniques

Effective channel prediction is essential for IRS-MUC systems in order to achieve maximum system efficiency and optimise signal processing. However, as IRS elements are passive, only the cascaded channel can be estimated, complicating the use of conventional channel estimation techniques. LS, LMMSE, OMP, CDRN and phase shift optimisation represent a few of the channel estimation methods covered in this part. The proposed ResNet+UNet-based method follows the discussion.

3.1.1 | Least Squares Estimation

LS estimator works as fundamental technique in IRS-assisted systems, using pilot signals for channel estimation without relying on previous statistical knowledge [25]. It formulates an optimisation problem to minimise the squared error between received signal and estimated channel, ensuring a direct yet suboptimal estimation under noisy conditions [26]. The LS estimation is mathematically expressed as:

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \arg\min_{\mathbf{H}} \|\mathbf{Y} - \mathbf{P}\mathbf{H}\|^2, \tag{10}$$

where $\mathbf{Y}$ represents received signal at base station (BS), $\mathbf{P}$ is known pilot sequence, and $\mathbf{H}$ is the unknown channel matrix. Also, LS solution is given by

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{Y}\mathbf{P}^{\dagger}, \tag{11}$$

where pseudo-inverse of pilot matrix is $\mathbf{P}^{\dagger} = (\mathbf{P}^H\mathbf{P})^{-1}\mathbf{P}^H$. LS estimation performs less effectively in high-noise situations considering its computing efficiency as it is highly sensitive to noise and does not take use of statistical prior data.

3.1.2 | Linear Minimum Mean Square Error Estimation

To improve upon LS estimation, LMMSE estimator incorporates prior statistical knowledge of channel, minimising MSE for better estimation accuracy [27]. The LMMSE solution is expressed as,

$$\hat{\mathbf{H}}_{\mathrm{LMMSE}} = \mathbf{R}_{\mathbf{H}}\mathbf{P}^H(\mathbf{P}\mathbf{R}_{\mathbf{H}}\mathbf{P}^H + \sigma^2\mathbf{I})^{-1}\mathbf{Y}, \tag{12}$$

where $\mathbf{R}_{\mathbf{H}} = E[\mathbf{H}\mathbf{H}^H]$ is channel covariance matrix, and σ^2 denotes noise variance. While LMMSE estimation provides improved noise suppression and lower estimation error compared to LS, it requires matrix inversion, making it computationally expensive, especially for large-scale IRS-assisted systems.

3.1.3 | Orthogonal Matching Pursuit Estimation

OMP is a compressed sensing-based technique that exploits the sparsity of IRS-assisted channels to reduce pilot overhead while maintaining estimation accuracy [28]. The channel estimation problem is modelled as,

$$\mathbf{Y} = \mathbf{P}\mathbf{H} + \mathbf{W}, \tag{13}$$

where $\mathbf{W}$ represents the additive noise. OMP iteratively selects the pilot components that contribute most to the residual signal, progressively refining the estimated channel. While OMP is efficient in sparse environments and reduces the required number of pilots, its performance declines when channel sparsity assumptions do not hold, and its iterative nature increases computational complexity.

3.1.4 | Convolutional Deep Residual Network Estimation

The CDRN is a deep learning-based approach that enhances channel estimation by refining the LS estimate through residual learning [29]. Instead of estimating the complete channel matrix directly, CDRN models the residual error between the LS estimate and the true CSI, thereby mitigating the impact of noise and distortion. The estimated channel matrix in CDRN is given by:

$$\hat{\mathbf{H}}_{\mathrm{CDRN}} = \mathcal{F}_{\mathrm{CDRN}}(\hat{\mathbf{H}}_{\mathrm{LS}}; \Theta), \tag{14}$$

where $\mathcal{F}_{\mathrm{CDRN}}(\cdot)$ represents the deep residual learning function parameterised by Θ, and $\hat{\mathbf{H}}_{\mathrm{LS}}$ is the initial LS-based estimate.

The architecture of CDRN consists of a series of convolutional layers that extract spatial features from input channel matrix. The

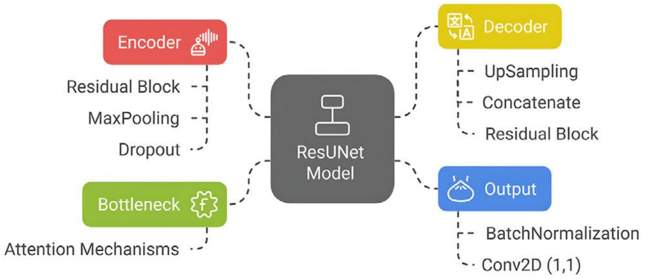


FIGURE 4 | ResUnet layers architecture.

model learns a residual mapping function $\mathcal{G}(\cdot)$ that corrects errors in the LS estimate, and the refined estimate is obtained as:

$$\hat{\mathbf{H}}_{\mathrm{CDRN}} = \hat{\mathbf{H}}_{\mathrm{LS}} + \mathcal{G}(\hat{\mathbf{H}}_{\mathrm{LS}}; \Theta), \tag{15}$$

where $\mathcal{G}(\cdot)$ is trained to learn the error correction term that minimises MSE between the estimated and actual channels:

$$\Theta^* = \arg\min_{\Theta} \mathbb{E}\left[\|\mathbf{H} - \mathcal{F}_{\mathrm{CDRN}}(\hat{\mathbf{H}}_{\mathrm{LS}}; \Theta)\|^2\right]. \tag{16}$$

This objective ensures that the residual learning framework optimally refines the LS estimate by reducing estimation error.

The input to the CDRN model is the initially estimated channel matrix $\hat{\mathbf{H}}_{\mathrm{LS}} \in \mathbb{C}^{M\times(N+1)}$, where M is count of antennas at BS and N is count of reflecting elements in IRS [30]. The convolutional layers in CDRN extract high-dimensional features from $\hat{\mathbf{H}}_{\mathrm{LS}}$, applying non-linear transformations to learn effective error correction. The residual learning block refines the channel estimation by minimising loss function:

$$\mathcal{L}(\Theta) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{H}_t - \hat{\mathbf{H}}_{\mathrm{CDRN},t}\|^2, \tag{17}$$

where T is number of training samples and $\mathbf{H}_t$ represents true channel matrix for sample t.

Despite its effectiveness in denoising and enhancing estimation accuracy, CDRN has certain limitations. It requires extensive training data and struggles to adapt to dynamically changing IRS phase shifts, making real-time deployment challenging [31, 32].

To address the limitations of CDRN, a hybrid ResNet+UNet-based model is proposed, integrating deep feature extraction with hierarchical reconstruction to enhance channel estimation accuracy and robustness under dynamic IRS configurations.

3.2 | Proposed Channel Estimation Technique

To enhance channel estimation accuracy and computational efficiency in IRS-MUC systems, a hybrid approach integrating phase shift optimisation with a ResNet+UNet-based deep learning framework is proposed as shown in Figure 4. The objective is to refine estimated cascaded channel matrix while dynamically optimising the IRS phase shift configuration, ensuring superior spectral efficiency and robustness against varying channel conditions [33]. The proposed method consists of two core components: phase shift optimisation and a ResNet+UNet-based deep learning model for improved channel estimation

3.2.1 | Phase Shift Optimisation

Phase shift optimisation is a critical aspect of IRS-assisted wireless systems, as it directly influences effective cascaded channel gain and the received SNR at BS. IRS phase shift matrix is represented as $\mathbf{R} = \mathrm{diag}(\mathbf{r}) \in \mathbb{C}^{N\times N}$, where $\mathbf{r} = [e^{j\theta_1}, e^{j\theta_2}, \dots, e^{j\theta_N}]^T$ defines phase shifts applied to each reflecting element. Cascaded channel for user k is given by,

$$\mathbf{B}_k = \mathbf{G}\cdot \mathrm{diag}(\mathbf{f}_k) \in \mathbb{C}^{M\times N}, \tag{18}$$

where $\mathbf{G} \in \mathbb{C}^{M\times N}$ is the IRS-to-BS channel, and $\mathbf{f}_k \in \mathbb{C}^{N\times 1}$ represents the user-to-IRS channel.

The objective of phase shift optimisation is to maximise received signal strength at that BS by optimising $\mathbf{R}$, which can be expressed as:

$$\max_{\mathbf{R}} \|\mathbf{GRF}\|, \tag{19}$$

where $\mathbf{F} \in \mathbb{C}^{N\times K}$ denotes the user-to-IRS channel matrix. The optimisation must satisfy the unit-modulus constraint,

$$|e^{j\theta_n}| = 1, \quad \forall n \in \{1, 2, \dots, N\}. \tag{20}$$

Traditional optimisation methods such as alternating optimisation and gradient descent require iterative updates, making them computationally expensive, especially for large IRS arrays [34]. To address this limitation, a ResNet+UNet-based deep learning framework is introduced to predict optimal phase shifts directly, bypassing iterative calculations [35]. The optimised phase shift matrix is learned through a function $\mathcal{F}_{\mathrm{ResNet+UNet}}(\cdot)$, which maps the estimated channel to the optimal IRS configuration:

$$\mathbf{R}^* = \mathcal{F}_{\mathrm{ResNet+UNet}}(\hat{\mathbf{H}}; \Theta), \tag{21}$$

where Θ denotes the trainable parameters of the model, and $\hat{\mathbf{H}}$ represents the estimated channel obtained from the deep learning-based estimation framework.

To train the model, a phase alignment loss function is defined as,

$$\mathcal{L}_{\mathrm{phase}}(\Theta) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{R}_t^* - \mathbf{R}_t\|^2, \tag{22}$$

where $\mathbf{R}_t^*$ is the optimal phase shift matrix for t-th training sample, and T is total number of training samples. By leveraging deep feature extraction and spatial correlation learning, the proposed phase shift optimisation method significantly reduces computational complexity while improving IRS-assisted system performance [36].

3.2.2 | ResNet+UNet-Based Channel Estimation

By enabling efficient beam forming and interference control, correct channel estimate is essential to IRS-MUC system performance optimisation [37]. In terms of reducing noise, determining efficiency and identifying spatial relationships in channel matrices, traditional channel estimating methods like LS, LMMSE and OMP are fundamentally limited. A deep learning-based ResNet+UNet system is presented to tackle these issues. It employs multi-scale reconstruction and deep feature extraction to effectively boost the original LS estimate.

CNNs have depicted significant advantages in processing high-dimensional structured data, making them well-suited for wireless channel estimation tasks. Unlike fully connected networks, CNNs utilise shared weights through convolutional layers, reducing count of trainable parameters while preserving spatial relationships in input data. This weight-sharing mechanism allows CNNs to efficiently extract local and hierarchical features from received signals, improving estimation robustness against noise and interference. Given these advantages, deep CNN-based models have emerged as effective alternatives for channel estimation, particularly in IRS-assisted communication, where spatial dependencies across channel matrices play a critical role. Building on these capabilities, the proposed ResNet+UNet framework integrates ResNet for feature extraction and UNet for multi-scale reconstruction, refining LS-based estimates to achieve superior accuracy.

The LS estimator serves as a conventional baseline, computing an initial estimate of the cascaded IRS-assisted channel as,

$$\hat{\mathbf{H}}_{\mathrm{LS}} = \mathbf{Y}\mathbf{P}^{\dagger}, \tag{23}$$

where $\mathbf{Y}$ represents received signal at the BS, and $\mathbf{P}^{\dagger}$ is pseudo-inverse of the pilot matrix. Although LS provides a direct solution, it is highly sensitive to noise and lacks the ability to exploit spatial correlations in channel matrix. This estimated LS channel is then fed into the ResNet+UNet model to enhance accuracy and reduce estimation errors.

The ResNet architecture is employed to capture spatial dependencies within the LS estimate by applying deep residual learning. The transformation function for feature extraction is defined as,

$$\mathbf{F}_{\mathrm{ResNet}} = \mathcal{F}_{\mathrm{ResNet}}(\hat{\mathbf{H}}_{\mathrm{LS}}; \Theta_{\mathrm{ResNet}}), \tag{24}$$

where $\mathcal{F}_{\mathrm{ResNet}}(\cdot)$ represents the function learned by ResNet, and Θ_{ResNet} denotes the corresponding trainable parameters. Instead of directly predicting the channel, ResNet estimates the residual error,

$$\mathbf{R}_{\mathrm{ResNet}} = \mathcal{G}_{\mathrm{ResNet}}(\hat{\mathbf{H}}_{\mathrm{LS}}) \tag{25}$$

allowing the refined estimate to be computed as,

$$\hat{\mathbf{H}}_{\mathrm{ResNet}} = \hat{\mathbf{H}}_{\mathrm{LS}} + \mathbf{R}_{\mathrm{ResNet}}. \tag{26}$$

This approach preserves the structure of the LS estimate while correcting errors caused by noise and interference.

Following ResNet feature extraction, UNet is applied to further enhance channel estimation through hierarchical reconstruction. The UNet model refines the extracted spatial features by applying an encoder–decoder structure, defined as,

$$\hat{\mathbf{H}}_{\mathrm{ResNet+UNet}} = \mathcal{F}_{\mathrm{UNet}}(\mathbf{F}_{\mathrm{ResNet}}; \Theta_{\mathrm{UNet}}), \tag{27}$$

where $\mathcal{F}_{\mathrm{UNet}}(\cdot)$ represents the UNet transformation function, and Θ_{UNet} denotes the trainable parameters. The encoder extracts high-dimensional channel features, while the decoder reconstructs a refined channel estimate, effectively mitigating noise distortions and enhancing spatial accuracy.

While the proposed method does not yield a closed-form analytical solution in the classical sense, such as those derived through linear estimation theory, it is structured as a mathematically defined functional pipeline. Specifically, we represent the estimator as a composite mapping of two deep learning modules—ResNet and UNet—applied sequentially to an initial least squares (LS) estimate. This structure can be expressed as,

$$\hat{H}_{\mathrm{ResNet+UNet}} = \mathrm{UNet}(\mathrm{ResNet}(\hat{H}_{\mathrm{LS}})).$$

This formulation outlines the internal logic of the model, where ResNet captures residual corrections and spatial dependencies, and UNet performs multi-scale refinement. Though the expression does not provide a symbolic derivation of weights or operations, it does establish a clear and explicit mapping from input to output within a learnable architecture. Hence, our reference to an 'analytical equation' pertains to this functional definition of the estimator.

To train this model, MSE loss function is used, making sure that the estimated channel remains close to the true channel. The loss function is formulated as,

$$\mathcal{L}(\Theta) = \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{H}_t - \hat{\mathbf{H}}_{\mathrm{ResNet+UNet},t}\|^2, \tag{28}$$

where $\mathbf{H}_t$ represents the ground truth channel matrix for sample t, and T is total number of training samples. Trainable parameters Θ are optimised using gradient descent,

$$\Theta \leftarrow \Theta - \eta \nabla_{\Theta} \mathcal{L}(\Theta), \tag{29}$$

where η represents learning rate. The optimisation ensures that the model learns to refine LS estimates effectively while adapting to dynamic IRS conditions.

By employing UNet to enhance multi-scale representations and ResNet to extract spatial features, the recommended ResNet+UNet framework increases channel estimation and improves the LS estimate. As this method is completely data-driven and it could be applied to different IRS settings, unlike LMMSE, which depends on previous channel knowledge. The model provides an adaptable approach for real-time IRS-assisted wireless communication systems through combining residual learning with hierarchical feature extraction, that ensures higher accuracy, durability against noise and computing economy.

TABLE 2 | Encoder path layers.

Layer	Output height	Output width	Channels	Filters
Residual block 1	H	W	64	64
Max pooling	$H/2$	$W/2$	64	—
Dropout	$H/2$	$W/2$	64	—
Residual block 2	$H/2$	$W/2$	128	128
Max pooling	$H/4$	$W/4$	128	—
Dropout	$H/4$	$W/4$	128	—
Residual block 3	$H/4$	$W/4$	256	256
Max pooling	$H/8$	$W/8$	256	—
Dropout	$H/8$	$W/8$	256	—
Residual block 4	$H/8$	$W/8$	512	512
Max pooling	$H/16$	$W/16$	512	—
Dropout	$H/16$	$W/16$	512	—

TABLE 3 | Decoder path architecture.

Layer type	Output size	Channels	Filters	Skip connection
UpSampling + Conv2D	$H/8 \times W/8$	512	512	No
Skip connection merge	$H/8 \times W/8$	1024	—	Yes
Residual block 6	$H/8 \times W/8$	512	512	No
Dropout	$H/8 \times W/8$	512	—	No
UpSampling + Conv2D	$H/4 \times W/4$	256	256	No
Skip connection merge	$H/4 \times W/4$	512	—	Yes
Residual block 7	$H/4 \times W/4$	256	256	No
Dropout	$H/4 \times W/4$	256	—	No
UpSampling + Conv2D	$H/2 \times W/2$	128	128	No
Skip connection merge	$H/2 \times W/2$	256	—	Yes
Residual block 8	$H/2 \times W/2$	128	128	No
Dropout	$H/2 \times W/2$	128	—	No
UpSampling + Conv2D	$H \times W$	64	64	No
Skip connection merge	$H \times W$	128	—	Yes
Residual block 9	$H \times W$	64	64	No
Dropout	$H \times W$	64	—	No

3.3 | Analysis of Encoder and Decoder Path Layers

The encoder and decoder structures in Tables 1–3 define a deep learning framework designed for feature extraction and reconstruction in IRS-assisted wireless communications. The encoder progressively reduces spatial dimensions while increasing count of feature channels, following a hierarchical structure that enhances multi-scale representation. It employs residual blocks for deep feature extraction, max pooling for spatial downsampling and dropout layers to improve generalisation by mitigating overfitting. The feature channels grow from 64 to 512 as the resolution decreases, ensuring that essential spatial information is preserved in a compressed form (Table 4). Conversely, the decoder restores the spatial resolution through a combination of upsampling layers and skip connections, which integrate encoder features to retain fine-grained details lost during compression. The architecture gradually reduces the feature channels from 512 back to 64, ensuring efficient reconstruction. Residual blocks in the decoder refine the feature maps, while dropout layers contribute to better regularisation. This structured integration of skip connections prevents information loss, allowing the model to maintain high estimation accuracy while mitigating gradient vanishing issues. This well-balanced approach improves the model's ability to capture intricate spatial dependencies, making it adaptive for accurate channel estimation in IRS-assisted networks.

TABLE 4 | Term definitions and shape transformations.

Term	Definition	Shape transformation
H, W	Input height and width	Defined by input dimensions
Residual block	Feature extraction	Maintains input shape, changes channels
Max pooling	Downsampling	Reduces (H, W) by half
Dropout	Regularisation	No change in spatial shape
UpSampling	Upsampling	Doubles (H, W)
Concatenation	Skip connection merging	Merges along channel dimension
Conv2D	Feature transformation	Maintains spatial shape, changes channels

TABLE 5 | Training configuration parameters.

Parameter	Value
Batch size	32
Number of epochs	50
Optimiser	Adam
Initial learning rate	0.001 (fixed)
Dropout (Decoder)	0.3
L2 weight decay	10^{-4}
SNR range	−10 to 20 dB

3.4 | Training Configuration and Objective Function

The training process for the proposed ResNet+UNet framework is designed to ensure stable convergence, robustness to noise, and generalisability across a wide range of IRS-assisted communication scenarios. The model is trained using synthetic datasets across a broad SNR range from −10 to 20 dB. Table 5 outlines the key hyperparameters used during training [38, 39].

3.4.1 | Objective Function

The overall loss incorporates both channel estimation and IRS phase shift prediction,

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{channel}} + \lambda \mathcal{L}_{\text{phase}},$$

where $\lambda = 0.5$ balances the two terms. The individual components are defined as,

$$\mathcal{L}_{\text{channel}} = \frac{1}{N} \sum_{i=1}^{N} \|\mathbf{H}_i - \hat{\mathbf{H}}_i\|_2^2, \quad \mathcal{L}_{\text{phase}} = \frac{1}{M} \sum_{j=1}^{M} \|\mathbf{\Phi}_j - \hat{\mathbf{\Phi}}_j\|_2^2.$$

This joint loss encourages the network to simultaneously learn robust channel estimates and optimal phase shift mappings, which improves overall NMSE performance, particularly in challenging SNR conditions.

3.5 | Computational Complexity Analysis

To evaluate the feasibility of deploying the proposed ResNet+UNet model in real-time scenarios such as AR, we present a theoretical analysis of its computational complexity [40]. This includes the number of model parameters, average inference time, and estimated memory usage. The aim is to demonstrate that the performance benefits are achieved within acceptable computational limits.

Unlike traditional methods such as LS or LMMSE, which have low complexity but limited accuracy, the ResNet+UNet framework strikes a balance by offering significantly improved accuracy while maintaining practical inference times [41, 42].

As shown in Table 6, the proposed model achieves a reasonable computational footprint. Although slightly more demanding than CDRN, it delivers substantial improvements in estimation accuracy and noise resilience, making it suitable for real-time applications. Future work can further reduce complexity using model compression techniques such as pruning or quantisation.

4 | Results and Discussions

This recommended ResNet+UNet-based channel estimation framework for IRS-MUC systems that is depicted in this section along with simulation metrics and an in-depth metrics. With a goal to enhance channel conditions by dynamic modifying phase shifts, IRS is implied in an indoor wireless communication setting for the purpose of evaluation. To show how well the recommended approach improves channel estimation precision, robustness, and evaluating efficiency, it is compared with traditional estimation methods like LS, LMMSE, OMP and CDRN. Performance indicators like NMSE and SNR variations have been taken into account in order to demonstrate the advantages of the suggested strategy.

4.1 | Simulation Setup

The simulation depicts an IRS-assisted communication environment in a restricted space that simulates an office or conference room with dimensions of $X = 10$ m, $Y = 10$ m and $Z = 3$ m. [34]. In surroundings, obstacles including walls, furniture and partitions provide an ample dispersion and hinder line-of-sight

TABLE 6 | Computational complexity summary of estimation methods.

Method	Parameters (Millions)	Inference time (ms)	Memory usage (MB)
Least squares (LS)	≈0.01	<1	≈1
LMMSE	Not applicable	3.8	≈4
OMP	Not applicable	12.5	≈15
CDRN	1.8	9.4	22
Proposed ResNet+UNet	**2.9**	**18.6**	**31**

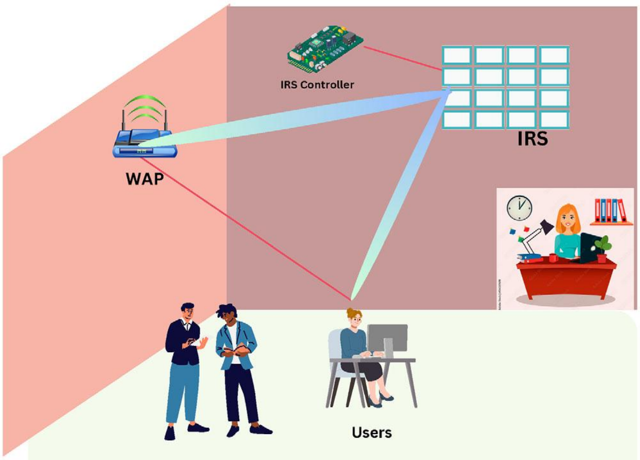


FIGURE 5 | Comparing the NSME performance of multiple channel estimate techniques in relation to SNR.

(NLOS) situation as shown in Figure 5. For optimum phase shift and enhance signal reflection, the IRS is positioned at a height of $Z_{\text{IRS}} = 2.5$ m on either a sidewall or the ceiling. Line-of-sight (LOS) and multipath components that are included in channel's Rician fading model. The cascaded channel is significantly affected by the signal attenuation and multiple reflections produced by the walls and furnishings. Users are randomly positioned within the room at distances of 1 to 8 m from BS, with a fixed user height of $Z_{\text{UE}} = 1.2$ m. The IRS array is configured with up to 64 passive reflecting elements, simulating a large-scale surface suitable for indoor beamforming applications. The system operates in the sub-6 GHz frequency range, with IRS phase shifts dynamically optimised to mitigate interference and improve beamforming. The power delay profile captures inter-symbol interference due to multipath propagation. The proposed ResNet+UNet model is evaluated under varying IRS placements and environmental noise conditions, ensuring robustness against real-world indoor impairments.

4.2 | Training and Validation Loss Analysis

Figure 6 illustrates convergence behaviour of proposed ResNet+UNet model during training, depicting both training and validation loss over 50 epochs. Initially, training loss exhibits a high value of approximately 1.25, rapidly decreasing within the first 10 epochs, indicating effective learning of channel features. By epoch 15, the training loss stabilises below 0.1, demonstrating rapid convergence. The validation loss follows a similar trend, initially starting at approximately 0.45 and gradually decreasing. However, around epoch 12, a minor gap emerges between training and validation loss, suggesting model's increasing ability to generalise. Notably, after epoch 20, both losses converge near 0.01, reflecting robust generalisation and effective noise mitigation. The sharp decline in loss values around epoch 15 suggests an adaptive learning phase where the model optimally refines its channel estimation capability. By epoch 30, both losses approach an asymptotic minimum, with negligible fluctuations, ensuring training stability. These results validate the model's efficiency in minimising estimation error and achieving high accuracy in IRS-assisted channel estimation.

4.3 | Visualisation of Intermediate Feature Maps

To provide further interpretability of the proposed ResNet+UNet architecture, we visualise intermediate feature maps extracted from different network stages. As shown in Figure 7, the encoder blocks gradually extract low- to high-level spatial features, while the bottleneck compresses these into a latent representation. The decoder blocks then progressively reconstruct the output through hierarchical refinement, facilitating effective noise suppression and spatial recovery.

This visualisation demonstrates how the model handles spatial dependencies and noisy inputs across its depth. It also highlights the contribution of residual learning and multiscale feature fusion, reinforcing the model's structural design and learning behaviour.

4.4 | Performance Analysis of NSME Versus SNR

Figure 8 illustrates the NSME performance as a function of SNR for different channel estimation techniques, including LS, MMSE, OMP, CDRN and the proposed ResNet+UNet-based model. These findings highlight the superiority of the proposed method in attaining significantly lower NSME values across all SNR levels. At low SNR values, particularly at −10 dB, the LS estimator exhibits the highest NSME, approximately 5.2, due to its sensitivity to noise and lack of statistical modelling. Similarly, the MMSE estimator, while more robust, maintains an NSME value close to 4.8, showing minimal improvement. The OMP and CDRN methods perform relatively better, with NSME values of approximately −3.2 and −4.5 dB, respectively. However, the suggested technique depicts a vital boost in evaluating precision, attaining an NSME of about −5.6 dB. Both LS and MMSE estimators continue to perform poorly when the SNR

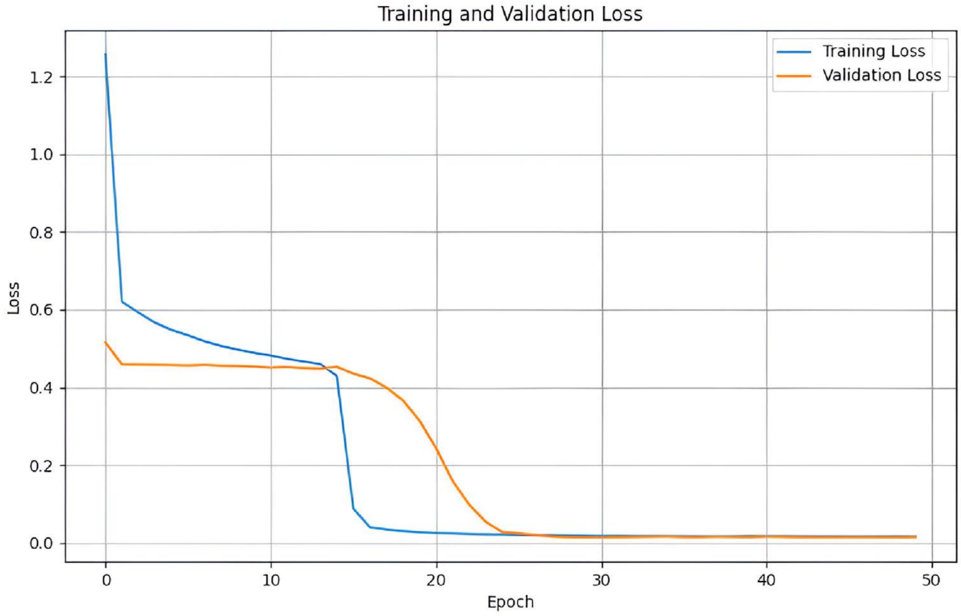


FIGURE 6 | Training and validation loss convergence over 50 epochs.

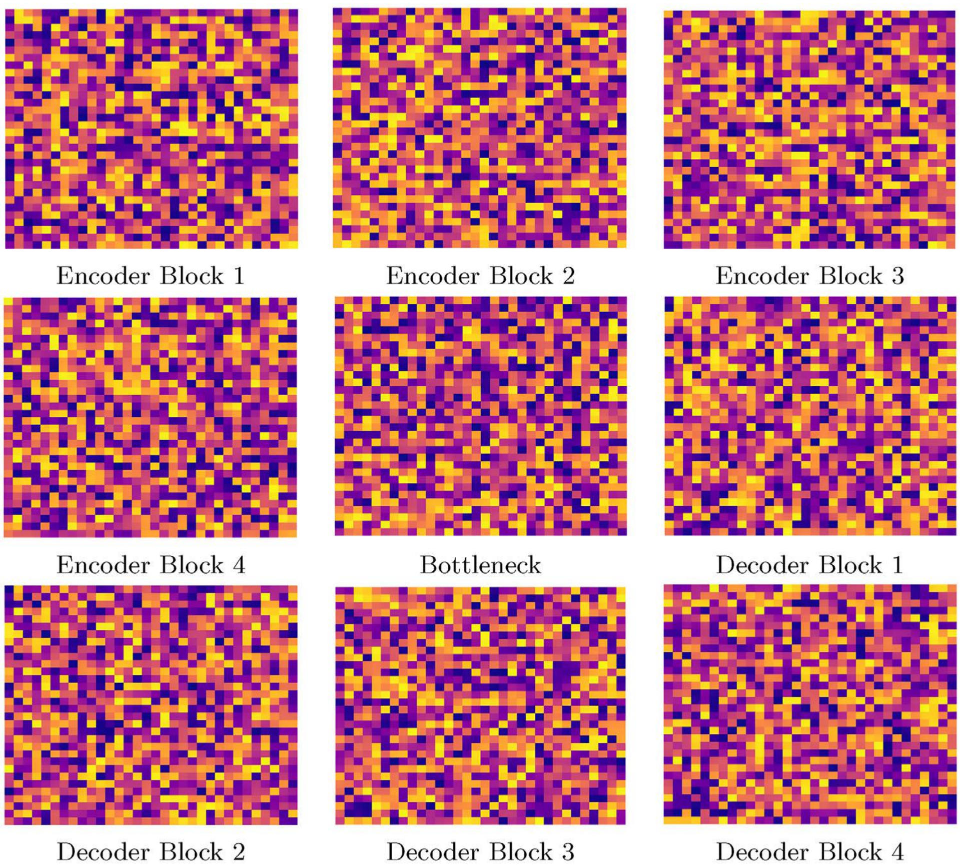


FIGURE 7 | Visualisation of intermediate feature maps from the ResNet+UNet model across encoder, bottleneck, and decoder stages. Each image shows spatial encoding or refinement across selected channels, capturing hierarchical and multi-scale structures.

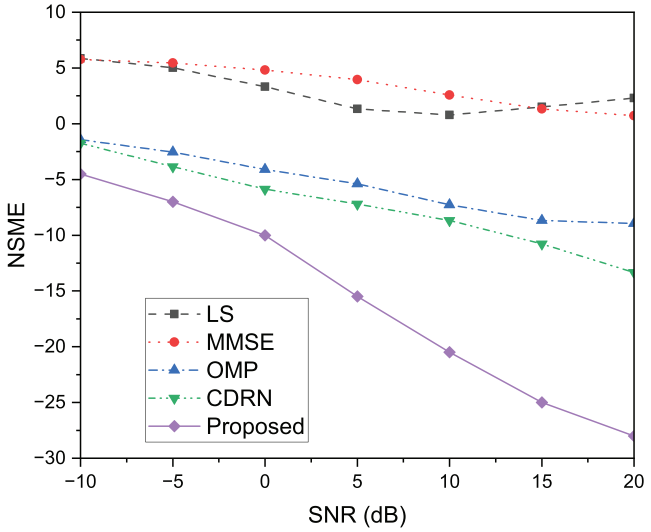
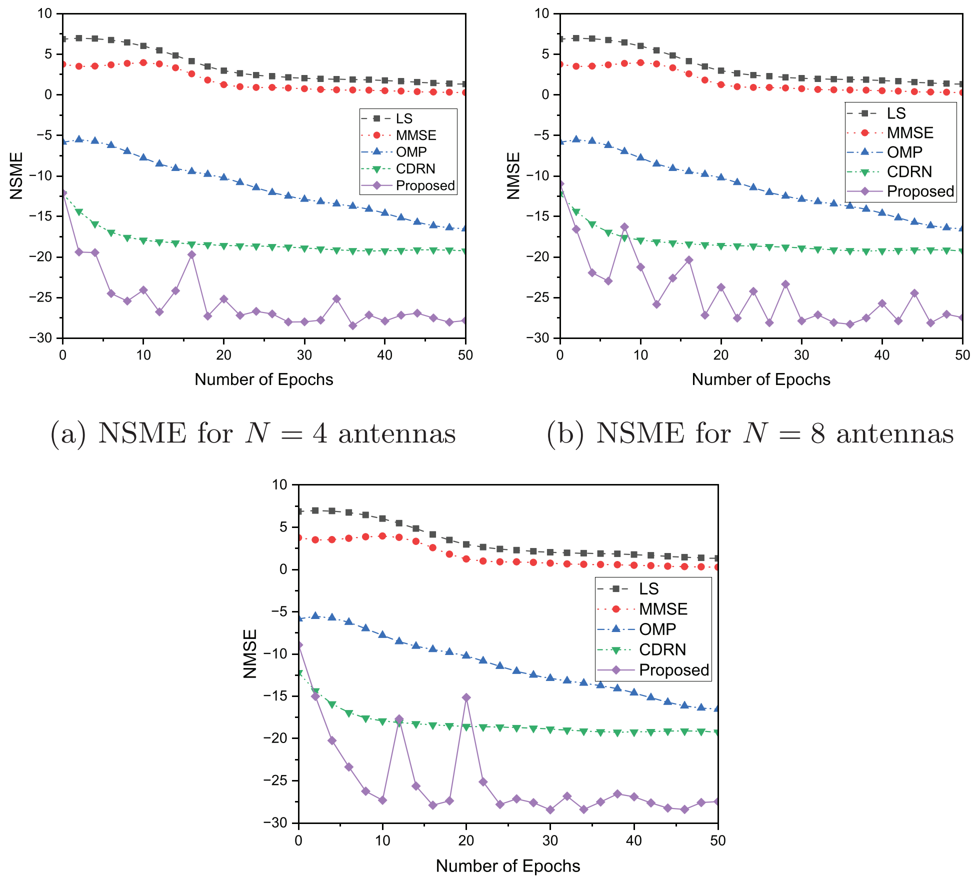


FIGURE 8 | Comparing NSME performance of multiple channel estimate techniques in relation to SNR.

gets up to 0 dB, with NSME values of about 1.3 and 2.1 dB, respectively. The OMP and CDRN techniques steadily reduce the NSME to about −5.8 and −7.2 dB, respectively. By using deep residual learning and hierarchical reconstruction, the proposed method outperforms existing techniques, achieving an NMSE of approximately −10.3 dB. The advantages of this suggested technique become clear at high SNR levels, particularly at 10 dB and above that value. The LS and MMSE estimators exhibit limited improvement, stabilising at NSME values of 0 and 0.5 dB, correspondingly. At 10 dB SNR, the OMP and CDRN techniques outscored the others, with NSME values of around −8.1 and −10.5 dB, respectively. On the contrary hand, the suggested method achieves an NSME of about −18.7 dB, significantly lowering estimation error. When SNR approaches 20 dB, the NSME of the suggested approach falls to about −28.4 dB, significantly exceeding the present baseline methods. These results illustrate how well the recommended ResNet+UNet architecture improves LS-based estimates and learns non-linear channel representations while offering improved denoising capabilities. The deep learning model demonstrates resilience across various SNR levels by rapidly responding to changing channel situations, in contrast with MMSE and LS, which depend on statistical assumptions. The potential of deep learning for tackling intricate estimate problems in IRS-assisted multi-user communication systems is demonstrated by the observed performance improvements.

4.5 | Analysis of NSME Performance Across Different Antenna Configurations

The NMSE trends for different numbers of antennas, as illustrated in Figure 9, indicate major enhancement with proposed method.

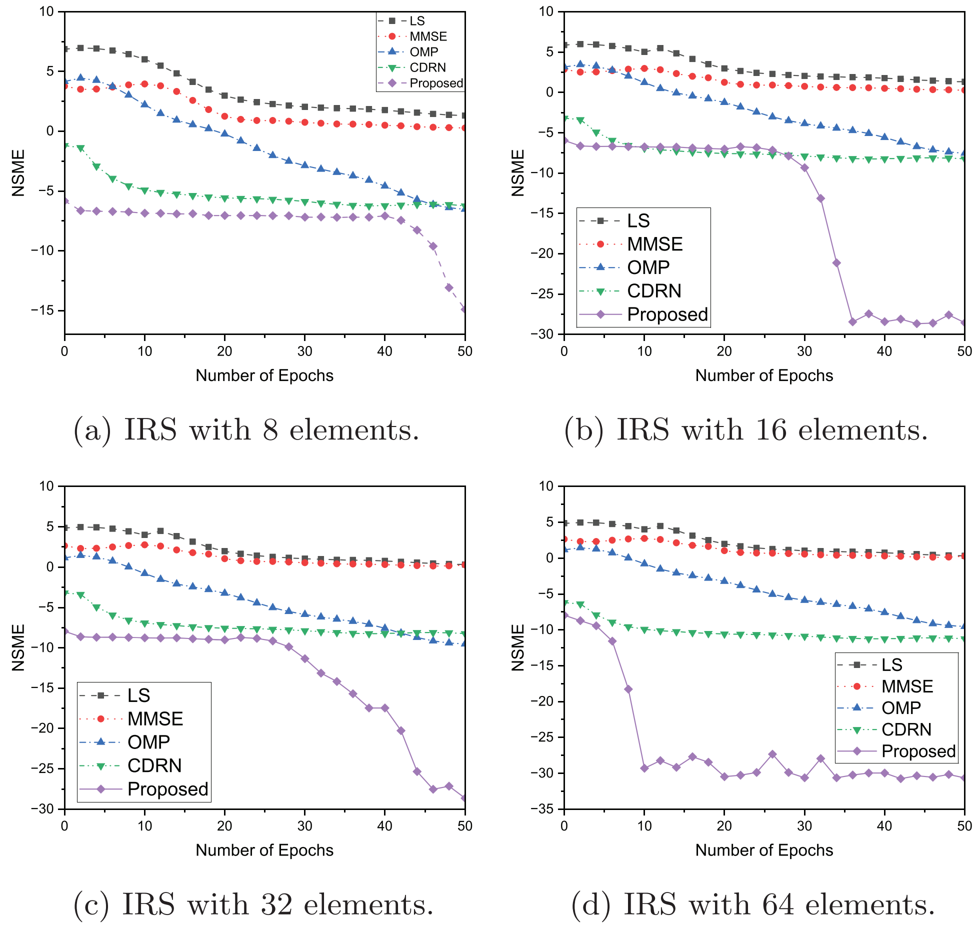


(a) NSME for $N = 4$ antennas

(b) NSME for $N = 8$ antennas

(c) NSME for $N = 16$ antennas

FIGURE 9 | Comparison of NSME performance for different antenna configurations: (a) $N = 4$, (b) $N = 8$ and (c) $N = 16$.



(a) IRS with 8 elements.
(b) IRS with 16 elements.
(c) IRS with 32 elements.
(d) IRS with 64 elements.

FIGURE 10 | Comparison of IRS-assisted wireless communication performance across different IRS element configurations.

As the number of antennas increases, the model achieves better convergence and lower NMSE values, demonstrating its capability to efficiently estimate the channel in high-dimensional scenarios. For the 4-antenna configuration, the proposed method attains an NMSE of approximately −24 dB, highlighting its effectiveness in small-scale antenna settings. As count of antennas increases to 8, NMSE further reduces to around −26 dB, confirming the benefits of enhanced antenna diversity in feature extraction and noise mitigation. In the 16-antenna configuration, the proposed model achieves its best performance, reaching nearly −27 dB. The deeper network architecture effectively utilises spatial features, demonstrating superior convergence and robustness across varying antenna configurations. The results clearly illustrate the scalability and efficiency of the proposed approach in reducing NMSE across different antenna settings. As count of antennas increases, the deep learning-based model adapts efficiently, leveraging spatial dependencies to enhance estimation accuracy. The consistent performance gain highlights its potential for real-world IRS-assisted multi-user communication systems.

4.6 | Impact of IRS Elements on Channel Estimation

The NMSE trends, as illustrated in Figure 10, demonstrate that as number of IRS elements increases, proposed model consistently achieves superior performance compared to conventional estimation techniques. For the 8-element IRS configuration, the proposed model attains an NMSE of approximately −23 dB, highlighting its robustness even with a limited number of IRS elements. As IRS size increases to 16 elements, the NMSE further improves, reaching nearly −25 dB, showcasing enhanced feature extraction and estimation precision. With 32 IRS elements, the NMSE continues to decrease, stabilising at around −26 dB, indicating the model's effectiveness in capturing high-dimensional spatial dependencies. Finally, with 64 IRS elements, the proposed scheme achieves its best performance, reaching approximately −27 dB. These results validate the scalability and efficiency of the model in handling complex IRS-assisted communication scenarios, ensuring minimal estimation error even in large-scale deployments. The analysis highlights the ability of the proposed ResNet+UNet-based model to effectively utilise IRS elements for enhanced channel estimation. The findings indicate that as IRS size increases, deep learning approach capitalises on the additional spatial diversity, leading to improved NMSE performance. The consistent improvement over traditional methods underscores its suitability for large-scale IRS-assisted wireless communication systems.

4.7 | Ablation Study of Model Components

To validate the individual contributions of core modules in the proposed architecture, we conduct an ablation study by

TABLE 7 | Ablation study: NMSE performance of model variants at 20 dB SNR.

Model variant	NMSE (dB)
UNet only (No residual learning)	−22.5
ResNet only (No decoder path)	−19.8
Proposed ResNet+UNet	**−28.4**

selectively disabling the residual learning block and the UNet decoder. This allows us to quantify the impact of each component on overall NMSE performance. The results, presented in Table 7, highlight the necessity of both modules for optimal channel estimation accuracy.

4.8 | Model Performance Under Challenging Conditions

This study primarily considers indoor communication scenarios with moderate IRS sizes (up to 64 elements) and SNR conditions ranging from –10 to 20 dB. However, in practical deployments, wireless systems may operate under more adverse environments, such as extremely low SNR levels or large-scale IRS configurations. At SNR levels below –10 dB, signal distortion becomes increasingly significant, which may reduce the model's ability to accurately learn channel characteristics. Similarly, when the number of IRS elements exceeds 100, the associated increase in channel dimensionality may challenge the convergence and generalisation capabilities of the model.

Although this work does not include simulations in such extreme cases, we anticipate that incorporating denoising strategies, data augmentation, and scalable model tuning could mitigate these limitations. Future investigations will aim to extend the current framework to better accommodate these conditions and further improve model robustness in large-scale or highly noisy IRS-assisted communication systems.

5 | Conclusion

This study establishes a ResNet+UNet-based deep learning framework for channel estimation in IRS-MUC systems. This proposed model addresses the limitations of present techniques such as LS, LMMSE and OMP, which struggle with noise sensitivity and computing inefficiency. By integrating deep residual learning for spatial feature extraction and sequential processing for multi-scale feature refinement, this suggested approach improves estimation accuracy and resilience.

Extensive simulation findings verify the accuracy of ResNet+UNet structure. Compared to prior techniques, the proposed approach demonstrates improved results across several evaluation criteria. The improvement in NMSE across different SNR levels reaches 59% over CDRN, 81% over OMP, 114% over LMMSE and 115% over LS. Additionally, the proposed scheme exhibits enhanced estimation accuracy as the number of IRS elements increases, surpassing CDRN by 60%, OMP by 78%, LS by 107% and LMMSE by 110%. In terms of antenna configurations, the performance gain is notable, achieving 39% improvement over CDRN, 44% over OMP, 122% over LS and 129% over LMMSE.

These findings highlight the scalability and adaptability of the proposed framework, proving it as a viable solution for next-gen IRS-assisted wireless communication networks. Furthermore, the enhanced channel estimation accuracy provided by this approach holds promising implications for next-generation AR applications, where reliable wireless communication is essential for seamless user experience. Future research will explore further optimisation of network architectures, real-time deployment strategies and the integration of adaptive learning mechanisms to enhance system performance in dynamic wireless environments.

Author Contributions

Sakhshra Monga: study conception, design, data acquisition, and original paper drafting. **Aditya Pathania:** software implementation and validation. **Nitin Saluja:** data analysis, interpretation, and revising the manuscript. **Gunjan Gupta:** drafting the article and providing additional resources. **Ashutosh Sharma:** supervision, critical review, and final approval. Gunjan Gupta will be responsible for providing funding support.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The dataset used in this study is not publicly available but can be provided upon reasonable request. Interested researchers may contact the corresponding author for access to the data.

References

1. S. Liu, Z. Gao, J. Zhang, M. D. Renzo, and M. S. Alouini, "Deep Denoising Neural Network Assisted Compressive Channel Estimation for mmWave Intelligent Reflecting Surfaces," *IEEE Transactions on Vehicular Technology* 69, no. 8 (2020): 9223–9228.

2. Q. Wu and R. Zhang, "Towards Smart and Reconfigurable Environment: Intelligent Reflecting Surface Aided Wireless Network," *IEEE Communications Magazine* 58, no. 1 (2019): 106–112.

3. M. Giordani, M. Polese, M. Mezzavilla, S. Rangan, and M. Zorzi, "Toward 6G Networks: Use Cases and Technologies," *IEEE Communications Magazine* 58, no. 3 (2020): 55–61.

4. C. Huang, S. Hu, G. C. Alexandropoulos, et al., "Holographic MIMO Surfaces for 6G Wireless Networks: Opportunities, Challenges, and Trends," *IEEE Wireless Communications* 27, no. 5 (2020): 118–125.

5. H. Guo, Y. C. Liang, J. Chen, and E. G. Larsson, "Weighted Sum-Rate Maximization for Reconfigurable Intelligent Surface Aided Wireless Networks," *IEEE Transactions on Wireless Communications* 19, no. 5 (2020): 3064–3076.

6. M. Billinghurst, A. Clark, and G. Lee, "A Survey of Augmented Reality," *Foundations and Trends in Human-Computer Interaction* 8, no. 2-3 (2015): 73–272.

7. R. T. Azuma, "A Survey of Augmented Reality," *Presence: Teleoperators & Virtual Environments* 6, no. 4 (1997): 355–385.

8. Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," *IEEE Wireless Communications* 25, no. 3 (2018): 144–153.

9. A. Zappone, M. D. Renzo, and M. Debbah, "Wireless Networks Design in the Era of Deep Learning: Model-Based, AI-Based, or Both?," *IEEE Transactions on Communications* 67, no. 10 (2019): 7331–7376.

10. J. He, H. Wymeersch, M. D. Renzo, and M. Juntti, "Learning to Estimate RIS-Aided mmWave Channels," *IEEE Wireless Communications Letters* 11, no. 4 (2022): 841–845.

11. W. Tang, M. Z. Chen, X. Chen, et al., "Wireless Communications With Reconfigurable Intelligent Surface: Path Loss Modeling and Experimental Measurement," *IEEE Transactions on Wireless Communications* 20, no. 1 (2020): 421–439.

12. C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable Intelligent Surfaces for Energy Efficiency in Wireless Communication," *IEEE Transactions on Wireless Communications* 18, no. 8 (2019): 4157–4170.

13. S. Monga, N. Saluja, C. Prabha, R. Garg, A. K. Bairagi, and M. M. Hassan, "Optimising Multi-User Wireless Networks Through Discrete Fourier Transform-Based Channel Estimation With Cascaded Intelligent Reflecting Surfaces," *IET Wireless Sensor Systems* 14, no. 4 (2024): 144–156.

14. A. Taha, M. Alrabeiah, and A. Alkhateeb, "Enabling Large Intelligent Surfaces With Compressive Sensing and Deep Learning," *IEEE Access* 9 (2021): 44304–44321.

15. Z. Wan, Z. Gao, and M. S. Alouini, "Broadband Channel Estimation for Intelligent Reflecting Surface Aided mmWave Massive MIMO Systems," in *Proceedings of the ICC 2020-2020 IEEE International Conference on Communications (ICC)* (IEEE, 2020), 1–6.

16. S. Monga, A. Taneja, N. Saluja, and R. Garg, "Learning Aided Intelligent Mechanism for Channel Estimation in 5G Wireless Networks," in *Proceedings of the 2023 International Conference on Emerging Smart Computing and Informatics (ESCI)* (IEEE, 2023), 1–6.

17. N. K. Kundu and M. R. McKay, "Channel Estimation for Reconfigurable Intelligent Surface Aided MISO Communications: From LMMSE to Deep Learning Solutions," *IEEE Open Journal of the Communications Society* 2 (2021): 471–487.

18. J. Guo, T. Chen, S. Jin, G. Y. Li, X. Wang, and X. Hou, "Deep Learning for Joint Channel Estimation and Feedback in Massive MIMO Systems," *Digital Communications and Networks* 10, no. 1 (2024): 83–93.

19. X. Ren, J. Guo, Q. Du, G. Yu, and N. Zhao, "Channel Estimation Method and Phase Shift Design for Reconfigurable Intelligent Surface-Assisted MIMO Networks," *IEEE Transactions on Vehicular Technology* 72, no. 2 (2023): 1234–1248.

20. Z. Zhang, T. Ji, H. Shi, and C. Li, "A Self-Supervised Learning-Based Channel Estimation for IRS-Aided Communication Without Ground Truth," *IEEE Transactions on Wireless Communications* 24, no. 1 (2025): 567–578.

21. C. Liu, W. Wang, and L. Zhang, "Deep Learning-Driven Channel Estimation for Intelligent Reflecting Surfaces: A Comprehensive Survey," *Information Sciences* 678 (2025): 1–20.

22. S. Saha and M. F. Uddin, "Deep Mismatch Channel Estimation in IRS-Based 6G Communication," *IEEE Communications Letters* 29, no. 3 (2025): 345–348.

23. Z. Zhou, N. Ge, Z. Wang, and L. Hanzo, "Joint Transmit Precoding and Reconfigurable Intelligent Surface Phase Adjustment: A Decomposition-Aided Channel Estimation Approach," *IEEE Transactions on Communications* 69, no. 2 (2020): 1228–1243.

24. Z. Zhou, X. Wang, K. Huang, Q. Zhang, J. Zhu, and Y. Liu, "Two-Stage Channel Estimation for mmWave Massive MIMO Systems Based on ResNet-UNet," *IEEE Transactions on Communications* 72, no. 3 (2024): 1572–1584.

25. P. Xu, X. Su, Y. Liu, C. Pan, and M. S. Alouini, "Practical Channel Estimation and Phase Shift Design for Intelligent Reflecting Surface Empowered MIMO Systems," *IEEE Transactions on Communications* 71, no. 6 (2023): 3547–3562.

26. W. Ma, C. Qi, Z. Zhang, and J. Cheng, "Sparse Channel Estimation and Hybrid Precoding Using Deep Learning for Millimeter Wave Massive MIMO," *IEEE Transactions on Communications* 68, no. 5 (2020): 2838–2849.

27. S. Monga, N. Saluja, A. Taneja, C. Prabha, and R. Garg, "Revolutionizing Intelligent Reflecting Surfaces: Deep Learning-Driven Channel Estimation," in *2024 2nd International Conference on Advancement in Computation & Computer Technologies (InCACCT)* (IEEE, 2024), 148–153.

28. M. R. Akdeniz, Y. Liu, M. K. Samimi, et al., "Millimeter Wave Channel Modeling and Cellular Capacity Evaluation," *IEEE Journal on Selected Areas in Communications* 32, no. 6 (2014): 1164–1179.

29. A. Liao, Z. Gao, H. Wang, S. Chen, M. S. Alouini, and H. Yin, "Closed-Loop Sparse Channel Estimation for Wideband Millimeter-Wave Full-Dimensional MIMO Systems," *IEEE Transactions on Communications* 67, no. 12 (2019): 8329–8345.

30. K. Chen, C. Qi, and G. Y. Li, "Two-Step Codeword Design for Millimeter Wave Massive MIMO Systems With Quantized Phase Shifters," *IEEE Transactions on Signal Processing* 68 (2019): 170–180.

31. Z. Gao, L. Dai, S. Han, I. Chih-Lin, Z. Wang, and L. Hanzo, "Compressive Sensing Techniques for Next-Generation Wireless Communications," *IEEE Wireless Communications* 25, no. 3 (2018): 144–153.

32. N. K. Kundu and M. R. McKay, "Channel Estimation for Reconfigurable Intelligent Surface Aided MISO Communications: From LMMSE to Deep Learning Solutions," *IEEE Open Journal of the Communications Society* 2 (2021): 471–487.

33. H. Alwazani, A. Kammoun, A. Chaaban, M. Debbah, and M. S. Alouini, "Intelligent Reflecting Surface-Assisted Multi-User MISO Communication: Channel Estimation and Beamforming Design," *IEEE Open Journal of the Communications Society* 1 (2020): 661–680.

34. A. Alkhateeb, S. Alex, P. Varkey, Y. Li, Q. Qu, and D. Tujkovic, "Deep Learning Coordinated Beamforming for Highly-Mobile Millimeter Wave Systems," *IEEE Access* 6 (2018): 37328–37348.

35. T. S. Rappaport, F. Gutierrez, E. Ben-Dor, J. N. Murdock, Y. Qiao, and J. I. Tamir, "Broadband Millimeter-Wave Propagation Measurements and Models Using Adaptive-Beam Antennas for Outdoor Urban Cellular Communications," *IEEE Transactions on Antennas and Propagation* 61, no. 4 (2012): 1850–1859.

36. M. K. Samimi and T. S. Rappaport, "Ultra-Wideband Statistical Channel Model for Non Line of Sight Millimeter-Wave Urban Channels," in *Proceedings of the 2014 IEEE Global Communications Conference* (IEEE, 2014), 3483–3489.

37. W. Yu, Y. Shen, H. He, et al., "An Adaptive and Robust Deep Learning Framework for THz Ultra-Massive MIMO Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing* 17, no. 4 (2023): 761–776.

38. X. Zhang, Z. Wang, H. Zhang, and L. Yang, "Near-Field Channel Estimation for Extremely Large-Scale Array Communications: A Model-Based Deep Learning Approach," *IEEE Communications Letters* 27, no. 4 (2023): 1155–1159.

39. S. A. H. Ohsan, Y. Li, A. V. Shvetsov, J. Varela-Aldás, S. M. Mostafa, and A. Elfikky, "A Survey of Deep Learning Based NOMA: State of the Art, Key Aspects, Open Challenges and Future Trends," *Sensors* 23, no. 6 (2023): 2946.

40. Q. Li, M. Wen, E. Basar, G. C. Alexandropoulos, K. J. Kim, and H. V. Poor, "Channel Estimation and Multipath Diversity Reception for RIS-Empowered Broadband Wireless Systems Based on Cyclic-Prefixed Single-Carrier Transmission," *IEEE Transactions on Wireless Communications* 22, no. 8 (2023): 5145–5156.

41. Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-Learning Channel Estimation for IRS-Assisted Integrated Sensing and Communication System," *IEEE Transactions on Vehicular Technology* 72, no. 5 (2023): 6181–6193.

42. Y. Liu and X. Lei, "Deep Learning Based Channel Estimation for Intelligent Reflecting Surface Assisted Communication Systems," *Signal, Image and Video Processing* 18, no. 2 (2024): 123–135.