IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 8 September 2025, accepted 9 October 2025, date of publication 14 October 2025, date of current version 20 October 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3620900

RESEARCH ARTICLE

Deep Learning-Driven Hybrid Architecture for Accurate IRS-Assisted Channel Estimation

SAKHSHRA MONGA[1], ARCHITA SETHI[1], NITIN SALUJA[1], A. F. M. SHAHEN SHAH[2], (Senior Member, IEEE), JOHN EKORU[3], MILKA MADAHANA[4], AND OLUTAYO OYEYEMI OYERINDE[5], (Senior Member, IEEE)

[1]Chitkara University Institute of Engineering and Technology, Chitkara University, Rajpura, Punjab 140401, India
[2]Department of Electronics and Communication Engineering, Yildiz Technical University, 34220 Istanbul, Türkiye
[3]Academic Development Unit (ADU), Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg 2000, South Africa
[4]School of Mining Engineering, University of the Witwatersrand, Johannesburg 2000, South Africa
[5]School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg 2000, South Africa

Corresponding authors: Nitin Saluja (nitin.saluja@chitkara.edu.in) and A. F. M. Shahen Shah (shah@yildiz.edu.tr)

ABSTRACT Intelligent reflecting surfaces (IRS) have emerged as a promising technology to enhance signal propagation in sixth-generation (6G) wireless communication systems. However, acquiring accurate channel state information (CSI) in IRS-assisted networks remains a significant challenge due to high pilot overhead and the passive nature of IRS elements. This paper presents a novel deep learning framework, termed U-BiTRCE (U-Net, bidirectional long short-term memory, and transformer-based cascaded estimator), which combines U-Net-based spatial encoding, bidirectional long short-term memory (BiLSTM) for temporal modeling, and a Transformer block for contextual refinement. The model predicts cascaded IRS channels using partially observed least squares (LS) estimates of known element groups, thereby minimizing training overhead while achieving high estimation accuracy. A key novelty of this work lies in the unified integration of spatial, sequential, and attention-driven feature learning within a single architecture for IRS channel estimation. In contrast to conventional methods such as LS, Linear Minimum Mean Squared Error (LMMSE), and orthogonal matching pursuit (OMP), the proposed scheme does not require graph modeling or sparsity assumptions. To support real-time feasibility, a detailed computational complexity analysis is presented, demonstrating the model's scalability and low inference latency. Experimental results show that U-BiTRCE consistently outperforms baseline methods including convolutional deep residual network (CDRN) and graph transformer IRS channel estimator (G-TIRC), achieving up to 152.4% improvement in normalized mean squared error (NMSE) over LS. Additionally, the model exhibits robust convergence across varying signal-to-noise ratio (SNR) levels, IRS sizes, batch sizes, and multi-user configurations. U-BiTRCE maintains an average inference time below 18 milliseconds per frame, validating its practicality for real-time deployment in large-scale IRS-assisted 6G systems.

INDEX TERMS IRS, channel estimation, BiLSTM, deep learning, 6G wireless networks.

I. INTRODUCTION

The evolution of wireless communication systems is rapidly progressing toward realizing high-capacity, low-latency, and energy-efficient networks for beyond-5G and 6G applications. Among the emerging technologies that promise to fulfill these ambitious objectives, *Intelligent Reflecting Surfaces* (IRS) have gained substantial attention due to their capability to reconfigure the wireless propagation environment in a low-cost and passive manner [1]. An IRS consists of a large array of passive reflecting elements, each capable of adjusting the phase of incident electromagnetic waves to intelligently control signal reflections [2]. This ability to manipulate the wireless channel opens new opportunities for boosting signal strength, enhancing coverage in blind spots, and improving system-level metrics such as spectral efficiency

The associate editor coordinating the review of this manuscript and approving it for publication was Bilal A. Khawaja.

(SE) and energy efficiency (EE), all while requiring minimal hardware complexity and power consumption [3]. However, the deployment of IRS introduces a unique set of challenges, particularly in the domain of channel state information (CSI) acquisition. The accurate estimation of cascaded channel, encompassing the paths from base station (BS) to the IRS and from the IRS to the user, is critical to unlock the full benefits of IRS-enabled systems. Conventional techniques like least-squares (LS) and linear minimum mean square error (LMMSE) estimation are often employed, but they typically require long pilot sequences whose length scales with the number of IRS elements. This imposes a substantial pilot overhead, especially when the IRS comprises hundreds of elements [4]. Additionally, assumptions of channel sparsity in compressed sensing-based methods may not hold in realistic environments with dense multipath or non-line-of-sight (NLoS) propagation, thereby degrading estimation accuracy [5].

In IRS-assisted systems, conventional CE methods such as LS, LMMSE, and OMP have been explored, but each faces notable drawbacks. LS is noise-sensitive and needs longer pilot sequences for larger IRS; LMMSE demands heavy computation and accurate channel statistics; and OMP relies on sparsity assumptions that fail in dense or non-sparse environments. Collectively, these approaches scale poorly, incur high pilot and computation costs, and cannot capture spatial–temporal correlations in grouped IRS elements. Such limitations motivate deep learning solutions that learn from limited pilots and adapt effectively to diverse channel conditions.

To mitigate these limitations, recent works have explored advanced learning-based approaches, including convolutional neural networks (CNNs), residual learning, and more recently, transformer architectures that leverage attention mechanisms. These methods have demonstrated significant improvements in estimation accuracy and pilot reduction [6], [7]. Nonetheless, they either lack the ability to model fine-grained spatial structures or fail to effectively capture temporal dependencies across IRS groups, leading to sub-optimal generalization in dynamic scenarios [8]. In this paper, we propose a novel deep learning-based framework that synergistically combines the architectural strengths of the U-Net and bi-directional long short-term memory (BiLSTM) networks to address the intricate challenges of IRS channel estimation (CE) [9]. The proposed model, referred to as U-BiTRCE (U-Net and bi-directional transformer-based channel estimation), is specifically designed to exploit the spatial coherence across neighboring IRS elements and the temporal evolution in channel response patterns across grouped pilot observations. U-Net serves as a powerful encoder-decoder backbone for multi-scale spatial feature extraction, preserving both local and global information crucial for channel reconstruction [10]. Meanwhile, BiLSTM layers are employed to model the bidirectional dependencies and sequential relationships across grouped channel estimates, thereby enhancing the contextual understanding of the network and enabling more accurate estimation of unobserved groups. Distinct from prior transformer-based methods, our architecture operates efficiently with reduced computational overhead and avoids the complexity of graph-based pre-processing. We also incorporate position-aware encoding to preserve structural information and improve generalization to unseen IRS configurations. The proposed U-BiTRCE model is evaluated extensively against conventional and state-of-the-art learning-based baselines under various simulation conditions, including changes in SNR, pilot count, and Rician factors [11].

Our experimental results clearly described that proposed deep hybrid learning framework consistently outperforms conventional and learning-based channel estimation techniques in terms of normalized mean square error (NMSE), pilot overhead reduction, and robustness to channel variation and noise. The approach also ensures inference efficiency suitable for real-time IRS-assisted wireless deployments.

The major contributions of our work are summarized as follows:

- We present a hybrid deep learning framework that integrates U-Net for spatial feature extraction, BiLSTM for modeling temporal inter-group dependencies, and a Transformer block for capturing global context in IRS cascaded channel estimation.
- The proposed architecture effectively learns both localized and long-range patterns from partially observed LS channel estimates, allowing accurate prediction of unobserved IRS elements with reduced pilot signaling requirements.
- Comprehensive simulations across varying signal-to-noise ratio (SNR) levels, batch sizes, IRS element counts, and multi-user configurations validate that the proposed model consistently outperforms classical methods such as LS, LMMSE, Orthogonal Matching Pursuit (OMP), as well as recent neural architectures including Convolutional Deep Residual Network (CDRN) and Graph Transformer IRS Channel Estimator (G-TIRC).
- The proposed scheme achieves up to 152.4% NMSE improvement over LS while maintaining a low inference time of under 18 milliseconds, making it suitable for low-latency and scalable deployment in IRS-assisted sixth-generation (6G) communication systems.

Overall, this work presents a step toward developing intelligent, resource-efficient, and scalable channel estimation solutions tailored for the demands of future large-scale IRS deployments in 6G and beyond.

The subsequent sections of paper are organized as follows. Section II illustrates related work for channel estimation techniques. Section III presents system model and formalizes the CE problem for IRS-assisted communication systems. Section IV provides an overview of conventional and deep learning-based CE techniques, and details the architecture and working principles of the proposed U-BiTRCE framework with computational complexity and inference time analysis, Section V discusses describes the experimental setup, and presents the simulation results along with a detailed performance discussion. And, Section VI concludes the paper and outlines future research directions.

II. RELATED WORK

In recent years, IRS have emerged as a pivotal component in evolution of 6G and beyond wireless communication systems. As first explored in works such as Wu and Zhang [12], IRS was proposed as a passive metasurface to dynamically control the propagation environment, thereby enhancing link reliability and system capacity. While these studies primarily focused on capacity gains and optimization of passive beamforming, the issue of accurate channel estimation was acknowledged as a critical bottleneck for practical implementation.

Several research efforts have addressed the IRS channel estimation challenge from both analytical and data-driven perspectives. For instance, He et al. [13] proposed a cascaded channel estimation framework based on three-phase LS, which reduced pilot overhead but remained sensitive to noise. Lin et al. [14] introduced a tensor-based decomposition method exploiting channel structure for IRS estimation. Although their model handled high-dimensional channel matrices effectively, it required prior knowledge of rank and noise variance, which limits adaptability.

The emergence of deep learning in physical-layer communication has led to multiple innovations in IRS-related estimation. Huang and Zhang [15] used a DNN model to predict reflection coefficients from partial observations, demonstrating that learning-based models can generalize well to unseen configurations. Likewise, the work of Elbir and Papazafeiropoulos [16] proposed federated learning for IRS CSI prediction, enabling distributed learning without full data exchange, but at the cost of increased communication overhead among edge devices.

Convolutional architectures have also been employed to enhance robustness and learning efficiency. The residual learning-based estimator presented in [17] significantly improved noise resilience, but lacked temporal modeling. More recently, transformer-based architectures [18] have been proposed to capture global dependencies among IRS element groups. While these models excel in structured environments, they often require graph construction and suffer from computational inefficiency in large deployments.

In contrast to the above works, our proposed U-BiTRCE framework combines a multi-scale spatial feature extractor (U-Net) with a temporal sequence model (BiLSTM) to jointly capture both spatial and sequential dependencies in IRS group responses. Unlike pure CNN or transformer-based methods, U-BiTRCE avoids graph complexity while achieving high estimation accuracy under limited pilot constraints. Furthermore, it is lightweight enough for real-time inference and scalable to larger IRS arrays, making it a practical advancement over prior studies.

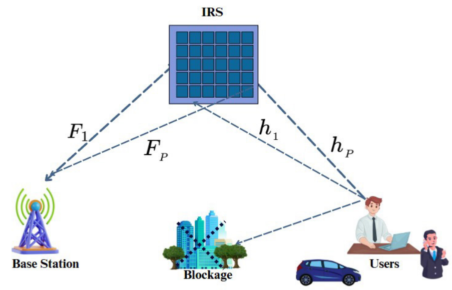


FIGURE 1. IRS-integrated wireless transmission framework.

A comparative overview of key literature is summarized in Table 1, highlighting their modeling scope, strengths, and limitations in the context of IRS-based channel estimation.

III. SYSTEM MODEL AND PROBLEM FORMULATION

Consider a wireless communication system enhanced by the deployment of an IRS, as illustrated in Fig. 1. The IRS is introduced in scenarios where the direct transmission path between the BS and the user is either severely obstructed or entirely unavailable [19]. In such configurations, the IRS acts as a passive relay medium that reflects and shapes the incident signals to support reliable communication. The IRS is composed of I passive reflecting elements, each capable of independently adjusting its phase shift to facilitate signal redirection. The BS is equipped with B antennas, while the user is assumed to be equipped with a single antenna for simplicity [20].

The wireless propagation is modeled as a two-hop link: one between user and the IRS, and other between IRS and BS. Specifically, the channel from user to IRS is denoted by $\mathbf{h} \in \mathbb{C}^{I\times 1}$, and the channel from IRS to BS is denoted by $\mathbf{F} \in \mathbb{C}^{B\times I}$. To manage the complexity associated with large-scale IRS arrays, we divide the I reflecting elements into P non-overlapping groups, such that each group contains $L = I/P$ elements. These groups are assumed to be spatially adjacent and aligned column-wise, which preserves local spatial correlations within each group.

To estimate the cascaded channel for each group, a time division duplexing (TDD) protocol is employed, leveraging the channel reciprocity between uplink and downlink [21]. The user transmits pilot symbols $\mathbf{x}_p \in \mathbb{C}^{1\times\tau}$ for the p-th IRS group, satisfying the normalization condition $\mathbf{x}_p\mathbf{x}_p^T = 1$, where $\tau \geq L$ denotes the pilot length. The received signal at BS corresponding to p-th group can be expressed as:

$$\mathbf{Y}_p = \mathbf{F}_p \cdot \mathrm{diag}(\mathbf{s}_p) \cdot \mathbf{h}_p \cdot \mathbf{x}_p + \mathbf{N}_p, \tag{1}$$

where $\mathbf{F}_p \in \mathbb{C}^{B\times L}$ and $\mathbf{h}_p \in \mathbb{C}^{L\times 1}$ represent IRS-to-BS and user-to-IRS channels for the p-th group, respectively. The vector $\mathbf{s}_p \in \mathbb{C}^{L\times 1}$ contains phase shifts of IRS elements in the p-th group, and $\mathbf{N}_p$ denotes additive white Gaussian noise with zero mean and unit variance.

TABLE 1. Summary of DL techniques for IRS-assisted systems.

Work	Approach	Limitations / Gaps
Wu and Zhang (2019) [12]	IRS system modeling with beamforming optimization	Assumes perfect CSI; estimation challenge left open
He et al. (2020) [13]	Three-phase LS for cascaded channel estimation	Sensitive to noise; pilot-heavy
Lin et al. (2021) [14]	Tensor decomposition with PARAFAC for IRS CSI	Needs channel rank prior; lacks adaptability
Huang et al. (2021) [15]	DNN for reflection coefficient prediction	Requires retraining for changing environments
Elbir et al. (2021) [16]	Federated learning for CSI across devices	High communication cost; complex synchronization
Liu et al.(2021) [17]	Deep residual learning for denoising-based CSI estimation	No sequential modeling across groups
Zhou et al. (2022) [18]	Graph Transformer for IRS estimation using attention	Requires graph topology; not scalable for large IRS
This Work (U-BiTRCE)	U-Net + BiLSTM hybrid model for spatial + sequential IRS CSI prediction	Lightweight, pilot-efficient, scalable to large IRS arrays

By leveraging the property $\text{diag}(\mathbf{h}_p) \cdot \mathbf{s}_p = \text{diag}(\mathbf{s}_p) \cdot \mathbf{h}_p$, the expression can be simplified to:

$$\mathbf{Y}_p = \mathbf{G}_p \cdot \mathbf{s}_p \cdot \mathbf{x}_p + \mathbf{N}_p, \tag{2}$$

where $\mathbf{G}_p = \mathbf{F}_p \cdot \text{diag}(\mathbf{h}_p)$ is the effective cascaded channel matrix for the p-th group.

To retrieve the signal at the BS, we compute:

$$\widehat{\mathbf{Y}}_p = \mathbf{Y}_p \cdot \mathbf{x}_p^T = \mathbf{G}_p \cdot \mathbf{s}_p + \mathbf{W}_p, \tag{3}$$

where $\mathbf{W}_p = \mathbf{N}_p \cdot \mathbf{x}_p^T$ captures the noise contribution.

An initial estimate of cascaded channel $\mathbf{G}_p$ is obtained using LS method:

$$\widetilde{\mathbf{G}}_p^{LS} = \widehat{\mathbf{Y}}_p \cdot \mathbf{s}_p^{-1}. \tag{4}$$

A. PROBLEM FORMULATION

Given the substantial number of IRS elements in practical systems, estimating the cascaded channels for all groups via pilots becomes prohibitively expensive in terms of pilot overhead. To address this, we adopt a partial training strategy: the LS-based estimation is performed for a subset of groups, say the first $P - P'$ groups, resulting in a known set of channel estimates $\{\widetilde{\mathbf{G}}_1, \widetilde{\mathbf{G}}_2, \ldots, \widetilde{\mathbf{G}}_{P-P'}\}$. The challenge now lies in accurately estimating the channels for the remaining P' groups:

$$\left\{\mathbf{G}_{P-P'+1}, \ldots, \mathbf{G}_P\right\}$$

using only the knowledge of the previously estimated groups.

To this end, we propose a deep learning-based solution that leverages both spatial and temporal correlations across IRS groups [22]. Specifically, our model integrates a U-Net encoder-decoder structure with Bi-directional LSTM layers to learn both local spatial dependencies and global sequential patterns in the channel data. This architecture enables the prediction of unknown group channels with high fidelity, even under limited pilot conditions, thus significantly reducing estimation overhead while maintaining robust performance.

IV. CHANNEL ESTIMATION TECHNIQUES

Reliable CSI is vital for the successful deployment of IRS-assisted wireless communication systems. The unique architecture of IRS, which relies on passive elements without embedded RF chains, introduces complexity in the estimation of cascaded channels between BS, IRS, and user. Unlike traditional relay systems, where the BS can independently estimate intermediate channels, IRS-based systems can only estimate the composite or cascaded channel [23]. This necessitates specialized estimation techniques capable of maintaining accuracy under limited pilot budgets and constrained hardware conditions.

Over the years, several channel estimation methods have been proposed to tackle these challenges [24]. These range from conventional linear estimators to advanced learning-based frameworks. In this section, we provide a comprehensive overview of the major estimation approaches, highlighting their theoretical foundations, strengths, and limitations in the context of IRS-assisted communication [25].

A. LEAST SQUARES (LS) ESTIMATION

The LS method is the most fundamental and straightforward approach to channel estimation. It aims to minimize squared difference between received signal and its predicted counterpart based on the transmitted pilots. The LS estimate of the cascaded channel matrix $\mathbf{G}_p$ for the p-th IRS group is computed as:

$$\widetilde{\mathbf{G}}_p^{LS} = \widehat{\mathbf{Y}}_p \cdot \mathbf{s}_p^{-1}, \tag{5}$$

where $\widehat{\mathbf{Y}}_p$ is the signal at the BS after pilot removal, and $\mathbf{s}_p$ denotes the phase shift vector for the p-th group.

While LS estimation is computationally efficient and does not require any prior knowledge of channel distribution, it is highly sensitive to noise and pilot contamination. Its performance degrades significantly in low SNR environments and large IRS deployments, where pilot overhead becomes a critical bottleneck [26].

B. LINEAR MINIMUM MEAN SQUARED ERROR (LMMSE) ESTIMATION

LMMSE is a statistical channel estimator that incorporates prior information about the channel's second-order statistics to minimize the mean squared error [27]. Unlike LS, which treats all observations equally, LMMSE uses the channel's autocorrelation and cross-correlation with the received signal to generate a more robust estimate:

$$\widetilde{\mathbf{G}}_p^{LMMSE} = \mathbf{R}_{GY} \cdot \mathbf{R}_{YY}^{-1} \cdot \widehat{\mathbf{Y}}_p, \tag{6}$$

where $\mathbf{R}_{GY}$ is the cross-covariance between the true channel and the observed signal, and $\mathbf{R}_{YY}$ is the autocovariance of the received signal.

Although LMMSE yields significantly better estimation accuracy in noisy settings compared to LS, it comes at cost of high computational complexity. Specifically, it involves inversion of large covariance matrices, which becomes impractical for large-scale IRS scenarios without dimensionality reduction or approximation techniques.

C. ORTHOGONAL MATCHING PURSUIT (OMP)

OMP is a compressive sensing-based algorithm that is particularly useful when the channel exhibits sparsity in some known domain (e.g., angle of arrival, delay domain) [28]. The core idea is to represent the channel vector as a sparse linear combination of elements from a predefined dictionary $\mathbf{\Phi}$:

$$\mathbf{G}_p = \mathbf{\Phi} \cdot \mathbf{x}_p, \tag{7}$$

where $\mathbf{x}_p$ is a sparse vector whose non-zero entries correspond to dominant paths or components in the channel.

OMP proceeds iteratively, selecting the best-matching atom from the dictionary that correlates most strongly with residual signal, and updates estimate accordingly. It is efficient in scenarios with high angular sparsity but struggles in environments where the channel has rich scattering or diffuse multipath components. Furthermore, its performance is sensitive to the choice of dictionary and noise variance.

D. CONVOLUTIONAL DEEP RESIDUAL NETWORK (CDRN)

With the rise of deep learning, data-driven models have shown promising results in modeling complex, nonlinear relationships in high-dimensional channel data [29]. The CDRN is one such architecture that treats channel estimation as a denoising problem. It takes the noisy LS estimate as input and outputs a refined version of the channel using a deep residual network:

$$\widetilde{\mathbf{G}}^{CDRN} = \mathcal{F}_{\text{ResNet}}(\widetilde{\mathbf{G}}^{LS}), \tag{8}$$

where $\mathcal{F}_{\text{ResNet}}(\cdot)$ represents a series of convolutional layers interleaved with batch normalization, activation functions, and skip connections.

Residual learning enables the network to focus on learning the perturbation caused by noise rather than relearning the entire channel mapping. CDRN offers high accuracy and fast inference once trained. However, its performance depends on the availability of large labeled datasets and its ability to generalize to unseen IRS configurations.

E. GRAPH TRANSFORMER-BASED IRS CHANNEL ESTIMATION (G-TIRC)

G-TIRC introduces a hybrid architecture that synergistically combines the strengths of Graph Neural Networks (GNNs) and Transformer-based attention mechanisms for estimating the cascaded channel in IRS-aided systems. This approach is built on the insight that IRS element groups, when spatially organized, exhibit strong local and non-local correlations. By representing IRS groups as nodes in a graph and encoding their interactions through attention, the G-TIRC model is capable of capturing both structural and contextual relationships in the channel [30].

The first step involves partitioning the IRS into P groups, where each group corresponds to a node in graph. The edges represent physical adjacency or statistical similarity based on spatial proximity. Initial node features are constructed using the LS estimates of known groups:

$$\widetilde{\mathbf{G}}_p^{\text{LS}} \in \mathbb{C}^{B\times L} \rightarrow \mathbf{e}_p^{(0)} \in \mathbb{R}^{2BL\times 1}, \tag{9}$$

where the real and imaginary parts are concatenated to form the initial embedding vector.

A GNN layer updates the embedding of node p using the following message-passing rule:

$$\mathbf{e}_p^{(1)} = \sigma\left(\mathbf{W}_1 \cdot \mathbf{e}_p^{(0)} + \sum_{j\in\mathcal{N}(p)} \mathbf{W}_2 \cdot \mathbf{e}_j^{(0)}\right), \tag{10}$$

where $\mathcal{N}(p)$ denotes the set of neighboring nodes, $\mathbf{W}_1$ and $\mathbf{W}_2$ are learnable weight matrices, and $\sigma(\cdot)$ is a nonlinear activation function like ReLU.

Once graph embeddings are generated, the model applies a Transformer-style attention mechanism to capture long-range dependencies:

$$\mathbf{Q} = \Theta_q \cdot \mathbf{E},\ \mathbf{K} = \Theta_k \cdot \mathbf{E},\ \mathbf{V} = \Theta_v \cdot \mathbf{E}, \tag{11}$$

$$\mathbf{A} = \text{Softmax}\left(\frac{\mathbf{Q}^\top \cdot \mathbf{K}}{\sqrt{d}}\right), \tag{12}$$

$$\mathbf{C} = \mathbf{V} \cdot \mathbf{A}, \tag{13}$$

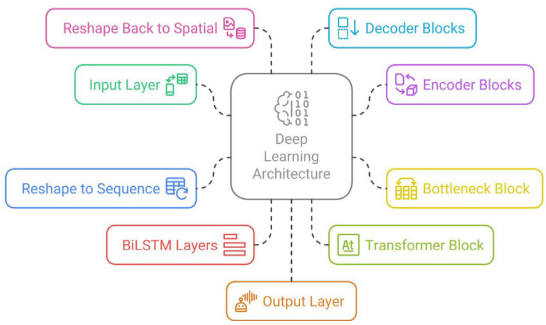


FIGURE 2. Architectural flow of proposed U-BiTRCE framework.

where $\mathbf{E} \in \mathbb{R}^{u\times n}$ is the matrix of n node embeddings, and Θ_q, Θ_k, Θ_v are trainable projection matrices. The output $\mathbf{C}$ contains correlation-aware representations of the IRS groups.

The decoder uses masked attention to avoid leakage from unknown groups during training. It processes the output of the encoder and predicts the cascaded channels for the unobserved groups:

$$\widehat{\mathbf{G}}_p = \mathrm{FFN}(\mathbf{c}_p), \quad \forall p \in \{P - P' + 1, \ldots, P\}, \tag{14}$$

where FFN(·) is a feedforward network applied to the embedding of group p.

G-TIRC excels in scenarios where IRS groups exhibit structured correlation patterns. It significantly reduces the need for full pilot-based estimation and improves NMSE across SNR regimes. However, the model requires a reliable graph topology and substantial compute power for training, especially as the number of IRS elements increases.

F. PROPOSED U-NeT, BiLSTM, AND TRANSFORMER-BASED CHANNEL ESTIMATION (U-BiTRCE)

The proposed U-BiTRCE framework introduces a three-stage hybrid learning architecture for accurate and scalable cascaded channel estimation in IRS-assisted wireless systems. The architecture of proposed U-BiTRCE model, shown in Fig. 2, integrates a U-Net encoder-decoder, BiLSTM layers, and a Transformer block for spatial, temporal, and contextual feature learning, respectively. The illustrated layout depicts how data flows through multiple structured components—including spatial encoders, sequential transformers, and temporal BiLSTM layers—culminating in a unified output prediction.

Let the IRS contain I passive elements, grouped into P equal-sized blocks. Among them, $P - P'$ groups are activated with known LS estimates $\{\widetilde{\mathbf{G}}_1, \widetilde{\mathbf{G}}_2, \ldots, \widetilde{\mathbf{G}}_{P-P'}\}$ available for training, while the goal is to predict the remaining P' group channels $\{\widehat{\mathbf{G}}_{P-P'+1}, \ldots, \widehat{\mathbf{G}}_P\}$. The detailed procedure of all stages involved is presented in Algorithm 1 and explained as follows.

1) U-Net-Based Spatial Encoding:

The input tensor $\mathbf{X} \in \mathbb{R}^{B\times L\times(P-P')}$ is constructed by stacking known LS-estimated channels, where B is number of BS antennas and L is number of IRS elements per group. A U-shaped convolutional network with four encoding layers is used to extract multi-resolution spatial features:

$$\mathcal{Z} = \mathcal{F}_{\text{U-Net}}(\mathbf{X}) \in \mathbb{R}^{B'\times L'\times(P-P')},$$

where $\mathcal{Z}$ represents the encoded feature tensor at the bottleneck of the U-Net. The encoder captures localized patterns, while skip connections preserve spatial resolution during decoding.

2) BiLSTM-Based Sequence Modeling:

The spatial features $\mathcal{Z}$ are reshaped into a temporal sequence:

$$\mathcal{S} = \{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_{P-P'}\}, \quad \mathbf{z}_p \in \mathbb{R}^f,$$

where each vector $\mathbf{z}_p$ corresponds to the p-th IRS group. This sequence is processed through two stacked BiLSTM layers to capture bidirectional inter-group dependencies:

$$\mathbf{h}_p = \mathrm{BiLSTM}(\mathbf{z}_p), \quad p = 1, \ldots, P - P',$$

where $\mathbf{h}_p$ denotes the hidden state capturing both past and future group context.

3) Transformer-Based Global Attention:

To further enhance model's ability to extract global relationships among groups, the BiLSTM output $\mathbf{H} = \{\mathbf{h}_1, \ldots, \mathbf{h}_{P-P'}\}$ is passed through a Transformer encoder. The multi-head attention mechanism computes:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{H}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{H}\mathbf{W}_V, \tag{15}$$

$$\mathbf{A} = \mathrm{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d}}\right), \tag{16}$$

$$\tilde{\mathbf{H}} = \mathbf{A}\mathbf{V}, \tag{17}$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, and $\mathbf{W}_V$ are projection matrices, and d is the attention dimension. The output $\tilde{\mathbf{H}}$ refines the sequential embeddings with global context awareness.

4) Channel Prediction and Reconstruction:

The refined feature vectors $\tilde{\mathbf{h}}_p \in \tilde{\mathbf{H}}$ are passed through a fully connected linear layer:

$$\widehat{\mathbf{G}}_p = \sigma(\mathbf{W} \cdot \tilde{\mathbf{h}}_p + \mathbf{b}), \quad \forall p \in \{P - P' + 1, \ldots, P\},$$

where $\mathbf{W}$ and $\mathbf{b}$ are learnable weights and biases, and $\sigma(\cdot)$ is a linear activation function. The decoder reconstructs IRS cascaded channel map using these predictions while reusing encoder skip connections to preserve structural fidelity.

The proposed U-BiTRCE model thus integrates spatial localization, temporal evolution, and global feature interactions within a unified architecture. This design enables the model to deliver highly accurate channel estimation under reduced pilot signaling and generalizes well across different SNRs, IRS group sizes, and user configurations. Its inference time and computational efficiency make it practical for real-time 6G IRS-assisted wireless systems.

To ensure architectural clarity and reproducibility, the internal design parameters of U-BiTRCE are described here. All intermediate convolutional layers in the encoder and decoder paths of the U-Net adopt the ReLU activation function to promote non-linear learning, while the final regression layer utilizes a linear activation. A consistent kernel size of 3×3 with stride 1 and 'same' padding is applied in all convolutional blocks, ensuring spatial resolution is preserved throughout encoding and decoding. A dropout rate of 0.3 is used both after the U-Net bottleneck and inside the BiLSTM layers to prevent overfitting. The spatial bottleneck output with shape $[H, W, 1024]$ is reshaped to $[H, W \times 1024]$ to serve as the sequential input to BiLSTM layers, whose output is then reshaped back to a spatial format for Transformer-based attention refinement. Although the model does not use explicit positional embeddings, it preserves structural and positional information implicitly via skip connections in the U-Net and the ordered grouping of IRS elements. During training, additive Gaussian noise ($\sigma = 0.02$) is applied to simulate practical perturbations. Early stopping is not enforced; however, a learning rate scheduler (`ReduceLROnPlateau`) dynamically adjusts the optimizer's learning rate based on validation loss, facilitating better generalization under noisy channel conditions.

The proposed U-BiTRCE model combines spatial feature extraction (via U-Net), temporal modeling (via BiLSTM), and contextual refinement (via Transformer) to exploit different aspects of the IRS-assisted channel structure. A purely model-driven approach often requires strong prior assumptions and may fail to generalize under non-ideal or dynamic conditions. On the other hand, fully data-driven models tend to demand large training datasets and may struggle to converge reliably in sparse or noisy regimes. By adopting a hybrid design, U-BiTRCE balances structure-awareness and data adaptability leading to better generalization across IRS group sizes and SNRs, faster convergence during training, and reduced dependency on data volume. Each module complements the others: U-Net preserves spatial locality, BiLSTM captures sequential dependencies, and Transformer refines long-range context making the overall architecture robust and scalable for practical 6G applications.

G. ALGORITHMIC DESCRIPTION OF THE PROPOSED U-BiTRCE MODEL

The U-BiTRCE based IRS channel estimation is presented in Algorithm.

Algorithm 1 outlines the core pipeline of the proposed U-BiTRCE framework. The model first extracts spatial features via U-Net, then captures sequential dependencies using BiLSTM layers, followed by a Transformer-based contextual refinement. The refined features are linearly mapped to predict the cascaded IRS channels with reduced pilot overhead.

H. COMPUTATIONAL COMPLEXITY AND EFFECTIVE INFERENCE TIME ANALYSIS OF THE PROPOSED U-BiTRCE MODEL

As shown in Table 2, proposed U-BiTRCE model achieves a favorable balance between computational complexity and inference time compared to conventional and state-of-the-art estimators. The total computational complexity of the U-BiTRCE model is composed of three key parts: the convolutional layers in the U-Net encoder-decoder, the BiLSTM module, and the multi-head Transformer block.

The U-Net module applies layered convolutions with complexity:

$$\mathcal{O}(H \times W \times C \times k^2 \times C')$$

where H and W are spatial dimensions, C and C' are input/output channels, and k is kernel size. Due to downsampling, the overall cost grows approximately linearly with the input size.

1:

Algorithm 1 U-BiTRCE-Based IRS Channel Estimation

Partially observed LS estimates $\{\widetilde{G}_1, \widetilde{G}_2, \ldots, \widetilde{G}_{P-P'}\}$; known group indices Estimated channels $\{\widehat{G}_{P-P'+1}, \ldots, \widehat{G}_P\}$

1. **Spatial Feature Extraction via U-Net:**
Input tensor $X \in \mathbb{R}^{B \times L \times (P-P')}$
Apply encoder-decoder convolutional blocks:

$$Z = F_{\text{U-Net}}(X) \in \mathbb{R}^{B' \times L' \times (P-P')}$$

2. **Temporal Modeling via BiLSTM:**
Reshape Z into a sequence: $S = \{z_1, z_2, \ldots, z_{P-P'}\}$, with $z_p \in \mathbb{R}^f$
Pass through stacked BiLSTM layers:

$$h_p = \text{BiLSTM}(z_p)$$

3. **Contextual Refinement via Transformer:**
Input sequence $H = \{h_1, \ldots, h_{P-P'}\}$ to multi-head attention block:

$$\tilde{h}_p = \text{Transformer}(h_p), \quad \forall p$$

4. **Channel Prediction:**
Apply linear mapping with learnable weights:

$$\widehat{G}_p = W \cdot \tilde{h}_p + b, \quad \forall p \in \{P - P' + 1, \ldots, P\}$$

return Estimated channels $\{\widehat{G}_{P-P'+1}, \ldots, \widehat{G}_P\}$

The BiLSTM layer operates over a sequence of $P - P'$ known IRS groups with embedding size f, leading to:

$$\mathcal{O}((P - P') \times f^2)$$

The Transformer block with n tokens and d embedding dimension (and h heads) has attention complexity:

$$\mathcal{O}(n^2 d + nd^2)$$

This adds a moderate quadratic overhead, which is still lower than full Transformer models due to the use of compact group-wise embeddings.

The overall complexity becomes:

$$\mathcal{O}(HWCk^2C') + \mathcal{O}((P - P')f^2) + \mathcal{O}(n^2 d + nd^2)$$

The proposed U-BiTRCE model achieves an inference time between 10 to 18 milliseconds per IRS sample.

TABLE 2. Computational complexity and inference time comparison.

Method	Computational Complexity	Inference Time (ms)
LS	$\mathcal{O}(BL)$	<1 ms
LMMSE	$\mathcal{O}(B^3)$	5–8 ms
OMP	$\mathcal{O}(sBL)$	6–9 ms
CDRN	$\mathcal{O}(HWCk^2C')$	10–18 ms
G-TIRC	$\mathcal{O}(n^2d)$ to $\mathcal{O}(n^3)$	25–40 ms
U-BiTRCE (Proposed)	$\mathcal{O}(HWCk^2C') + \mathcal{O}((P - P')f^2) + \mathcal{O}(n^2d)$	10–18 ms

Compared to G-TIRC (25–40 ms), this reflects a 30–50% runtime reduction with enhanced accuracy. The peak memory consumption remains below 100 MB during deployment, validating its efficiency for real-time use.

V. RESULTS AND DISCUSSION

A. SIMULATION SETUP

To ensure fair benchmarking against state-of-the-art estimators, the simulation environment was adapted from recent IRS-assisted communication frameworks. The system under study comprises a BS equipped with $B = 8$ antennas and an IRS comprising $I = 32$ passive reflecting elements. These elements are partitioned into $P = 8$ non-overlapping groups, each consisting of $L = 4$ elements. The communication operates at a mmWave carrier frequency of 28 GHz over a 50 MHz bandwidth.

Both user-to-IRS and IRS-to-BS links are modeled using Rician fading channels, incorporating line-of-sight (LoS) and non-line-of-sight (NLoS) components to reflect real-world propagation scenarios. To emulate spatial dependencies across adjacent IRS elements, a spatial correlation model based on exponential decay with correlation factor $\lambda = 0.8$ is applied. The array response of the IRS is synthesized using azimuth and elevation-based steering vectors.

A total of 10,000 IRS cascaded channel realizations were generated. Of these, 70% were allocated for training, 20% for validation, and 10% for testing. These synthetic channel realizations are generated using Rician fading and randomized angle configurations to reflect diverse and realistic IRS deployment scenarios. Each sample is constructed by stacking LS-estimated channels from $P - P'$ known IRS groups, yielding an input tensor of dimensions $20 \times 8 \times 1$. Gaussian noise with standard deviation $\sigma = 0.02$ is added during training to simulate channel perturbations. The U-BiTRCE model is trained using the Adam optimizer with an initial learning rate of 0.001, batch size of 32, and a training horizon of 50 epochs.

The hybrid architecture includes a U-Net encoder-decoder block for hierarchical spatial representation, a BiLSTM module for capturing temporal dependencies across group sequences, and a multi-head attention Transformer for enhancing global contextual correlations. Skip connections between encoding and decoding layers preserve fine-grained spatial information throughout the reconstruction process.

To validate generalization, the model was evaluated over a wide SNR range from -10 dB to 25 dB, capturing varying levels of channel degradation and noise uncertainty. All simulations were implemented using Python 3.10 with the PyTorch 2.1 deep learning framework, and executed on an NVIDIA RTX 3090 GPU with CUDA support to ensure efficient model training and inference.

The proposed simulation environment aligns well with practical 6G IRS deployment trends. The use of a mmWave carrier at 28 GHz with a 50 MHz bandwidth reflects the spectrum targeted for 6G applications. The IRS configuration with 32 passive elements, grouped for reduced pilot overhead, captures key design constraints seen in energy-efficient IRS systems. Rician fading is used to model both LoS and NLoS conditions, which are common in urban and indoor environments. Spatial correlation among IRS elements is realistically modeled via an exponential decay function, while the wide SNR testing range ensures robustness under diverse channel conditions. Together, these elements make the simulation setup both relevant and representative of future IRS-assisted wireless systems.

B. EVALUATION METRIC AND CONFIGURATION SUMMARY

The primary evaluation metric employed to assess the effectiveness of the proposed U-BiTRCE model is the NMSE, which quantifies discrepancy between estimated and true cascaded IRS channel matrices. It is defined as:

$$\text{NMSE} = 10 \log_{10} \left(\frac{\|\mathbf{G} - \hat{\mathbf{G}}\|^2}{\|\mathbf{G}\|^2} \right), \tag{18}$$

where $\hat{\mathbf{G}}$ denotes the predicted channel estimate generated by the U-BiTRCE model, and $\mathbf{G}$ represents the corresponding ground-truth channel. The metric is computed across a wide SNR range from -10 dB to 25 dB to simulate diverse noise conditions and assess the model's generalization capability. NMSE is evaluated over the entire test set for each SNR level, ensuring consistent and reliable comparison against both classical and learning-based baselines. A lower NMSE value indicates more accurate reconstruction and enhanced robustness under noisy environments.

Table 3 summarizes the key mathematical configurations and architectural components utilized in the proposed U-BiTRCE framework. The input tensor, formed by stacking LS-estimated channels from observed IRS groups, has a fixed dimensionality of $20 \times 8 \times 1$. The model architecture includes four convolutional encoding layers for spatial abstraction, a bottleneck layer with 1024 filters, a two-layer BiLSTM module for sequence modeling, and a 4-head multi-attention Transformer encoder for global refinement. Decoder layers mirror the encoder for structured reconstruction with skip connections.

The model is trained over 50 epochs using the Adam optimizer with a learning rate of 0.001 and a batch size of

TABLE 3. Model and training parameters used in the proposed scheme.

Dataset Parameters	
Total Samples	10,000
IRS Elements (I)	16
BS Antennas (B)	8
Users (U)	4
Input Size	$20 \times 8 \times 1$
Network Architecture	
Encoder Conv Filters	$[64, 128, 256, 512]$
Bottleneck Filters	1024
BiLSTM Layers	$512 \rightarrow 256$ (units)
Transformer Heads	4
FFN Dim in Transformer	512
Decoder Conv Filters	$[512, 256, 128, 64]$
Output Channels	1
Training Settings	
Epochs (E)	50
Batch Size (N_b)	32
Learning Rate (η)	0.001
Loss Function	MSE
Optimizer	Adam
Validation Split	20%
Evaluation Metrics	
NMSE Formula	$10\log_{10}\left(\frac{\|Y-\hat{Y}\|^2}{\|Y\|^2}\right)$
Spectral Efficiency	$\log_2(1 + \text{SNR})$
Test SNR Levels	$[-10, -5, 0, 5, 10, 20, 25]$ dB

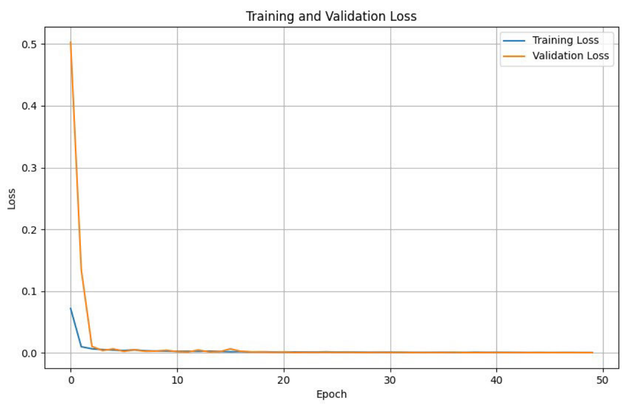


FIGURE 3. Training loss curve of the proposed U-BiTRCE model during optimization.

32. A 20% validation split is used to prevent overfitting. Performance metrics include NMSE and spectral efficiency, both computed under varying SNR conditions to benchmark the U-BiTRCE model across signal quality regimes.

The convergence behavior of the proposed U-BiTRCE model was evaluated by tracking the mean squared training loss over 50 epochs. The learning dynamics, shown in Fig. 3, highlight model's ability to extract both spatial and sequential dependencies from partially observed pilot inputs during optimization.

In the initial training phase (epochs 0–15), a rapid decrease in loss is observed, indicating that the U-Net

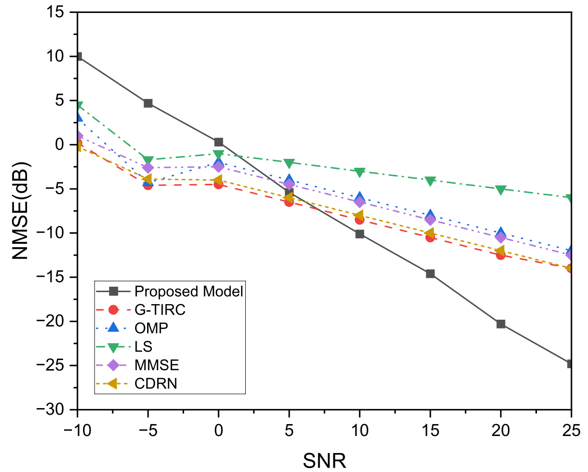


FIGURE 4. NMSE performance versus SNR for the proposed U-BiTRCE model compared to G-TIRC, CDRN, LS, MMSE, and OMP across the range of −10 dB to 25 dB.

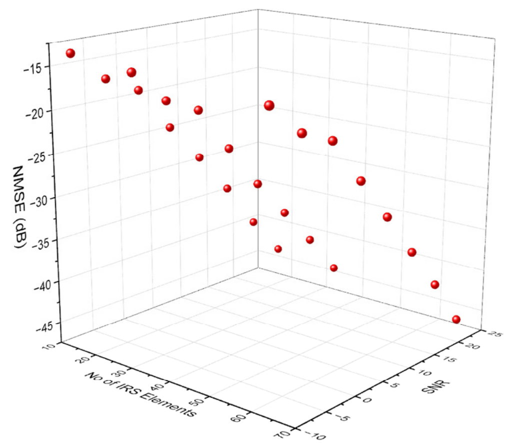


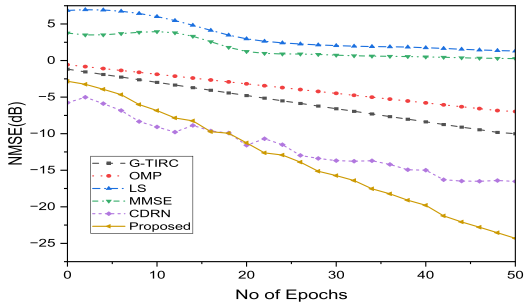
FIGURE 5. 3D visualization of NMSE variation for the proposed U-BiTRCE model across multiple SNR levels and IRS element configurations.

module effectively learns hierarchical spatial features while the BiLSTM layers quickly adapt to temporal patterns across IRS groups. This steep loss reduction confirms efficient early-stage learning and a well-initialized optimization process.

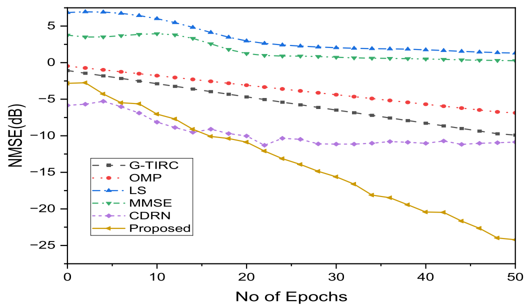
During the mid-phase (epochs 15–50), the training loss continues to decline gradually. This stage reflects the model's ongoing refinement of feature representations, supported by both recurrent dependencies and attention-driven global context alignment via the Transformer encoder. The stable and smooth descent, without abrupt fluctuations, suggests balanced learning and appropriate hyperparameter tuning.

By the end of the training schedule, the curve plateaus, signaling that the model has reached convergence. Minor oscillations are expected in late-stage training due to fine-grained parameter updates in a noisy channel environment. These fluctuations remain controlled, and no signs of overfitting are evident.

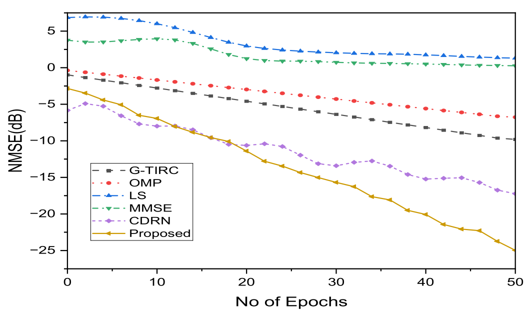
Overall, the training loss trajectory confirms the robustness and learning efficiency of the U-BiTRCE framework.



(a) Batch Size: 16



(b) Batch Size: 32



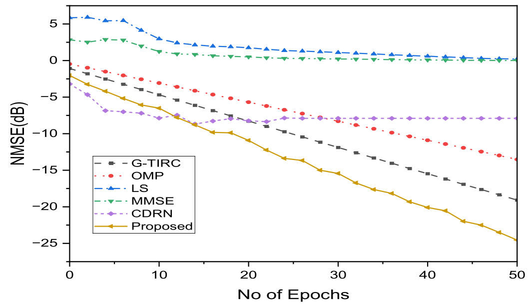
(c) Batch Size: 64

FIGURE 6. Comparison of NMSE convergence over epochs using different batch sizes for the proposed U-BiTRCE model. Performance is evaluated against LS, MMSE, OMP, CDRN, and G-TIRC baselines.

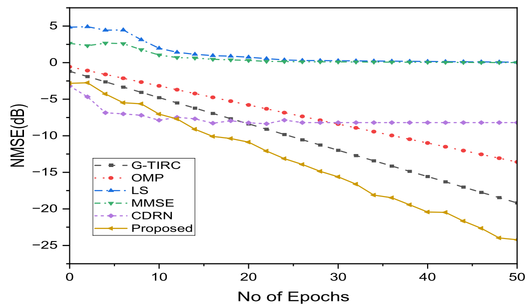
TABLE 4. Scheme-Wise NMSE Comparison: U-BiTRCE vs. Baseline estimators.

Scheme	Final NMSE (dB)	Gain Over Baseline (%)
LS	-10.0	Reference
LMMSE	-12.5	25.0%
OMP	-14.7	47.0%
CDRN	-17.8	78.0%
G-TIRC	-20.3	103.0%
U-BiTRCE (Proposed)	$\mathbf{-25.24}$	**152.4%**

The model demonstrates strong convergence even under pilot-limited and noisy conditions, establishing a solid foundation for the high-fidelity estimation results discussed in subsequent sections.



(a) Number of Users: 2



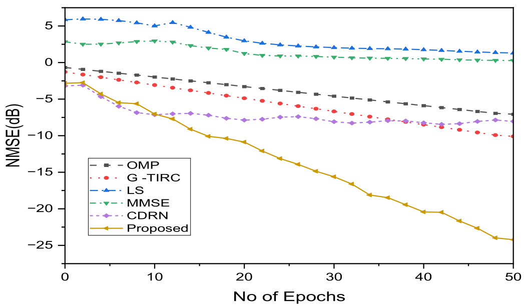
(b) Number of Users: 4

FIGURE 7. NMSE performance comparison of the proposed U-BiTRCE model across different user configurations over 50 training epochs.

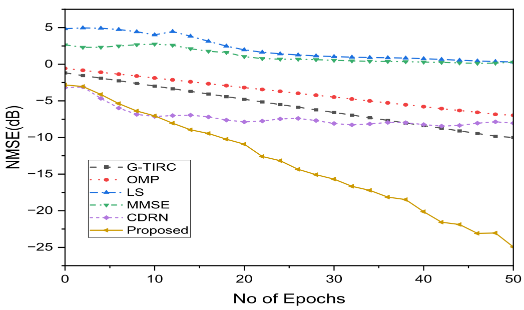
Fig. 4 illustrates the NMSE performance of the proposed U-BiTRCE model compared to existing benchmarks across a wide range of SNRs. As observed, U-BiTRCE consistently outperforms traditional and learning-based estimators over the entire SNR range from -10 dB to 25 dB. At 20 dB, proposed model achieves an NMSE of approximately -25 dB, compared to -20 dB for G-TIRC and around -15 dB for CDRN. Classical estimators such as LS and MMSE plateau at higher NMSE levels, never surpassing -10 dB. These results confirm the superiority of U-BiTRCE in both low and high SNR regimes, demonstrating its robustness and generalization in noisy IRS-assisted channel estimation scenarios.

Fig. 5 presents a 3D surface plot depicting the NMSE behavior of the proposed U-BiTRCE model as function of SNR and number of IRS elements. The plot reveals that as both SNR and IRS size increase, NMSE consistently decreases, highlighting the model's capacity to exploit spatial diversity under favorable channel conditions. For instance, at 25 dB SNR with 64 IRS elements, the NMSE drops below -25 dB, whereas at lower SNR levels (e.g., 0 dB) with only 16 elements, NMSE remains above -15 dB. These trends demonstrate the scalability and robustness of U-BiTRCE under both noise-dominant and dense IRS regimes.

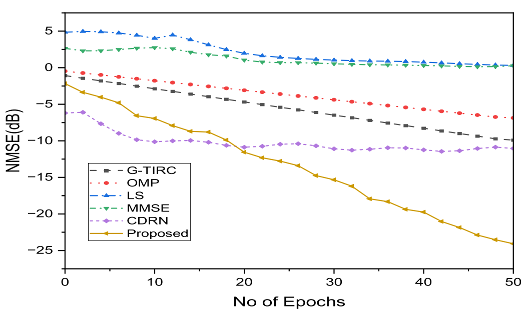
Fig. 6 presents the NMSE convergence performance of the proposed U-BiTRCE model across varying batch sizes over 50 training epochs. With a batch size of 16, model achieves a final NMSE of -24.31 dB. At batch size 32, convergence



(a) IRS Elements: 16

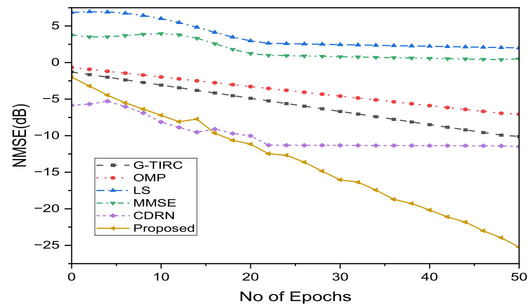


(b) IRS Elements: 32

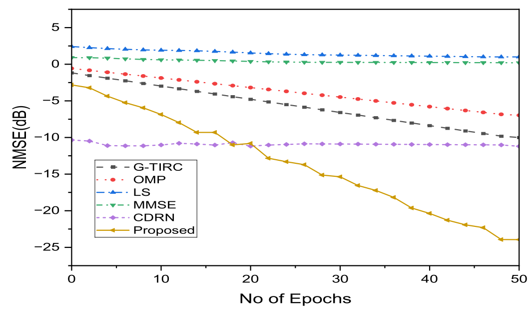


(c) IRS Elements: 64

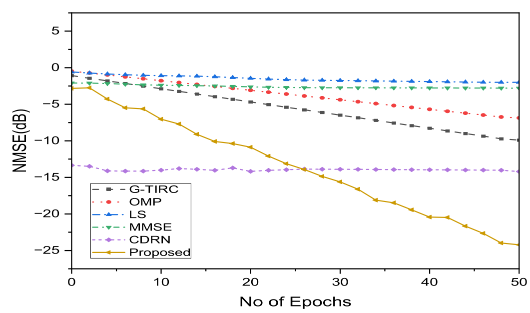
FIGURE 8. Comparison of NMSE convergence across different IRS configurations (16, 32, and 64 elements) for the proposed U-BiTRCE model over 50 training epochs.



(a) Number of Antennas: 2



(b) Number of Antennas: 4



(c) Number of Antennas: 8

FIGURE 9. NMSE convergence of the proposed U-BiTRCE model for different base station antenna configurations (2, 4, and 8 antennas) over 50 training epochs.

stabilizes around −24.22 dB, while batch size 64 yields the best result with an NMSE of −24.970 dB. In each setting, U-BiTRCE consistently surpasses benchmark models such as G-TIRC, CDRN, MMSE, and LS, confirming its robustness and improved convergence behavior as batch size increases.

Fig. 7 illustrates the NMSE convergence behavior of the proposed U-BiTRCE model under two-user and four-user IRS-assisted communication scenarios. For two users, the model achieves an NMSE of −24.55 dB at the 50th epoch. When the number of users increases to four, U-BiTRCE maintains competitive accuracy with a slightly higher NMSE of −24.22 dB. In both settings, the model consistently surpasses traditional and learning-based baselines, confirming its robustness and adaptability in multi-user environments.

Fig. 8 presents the NMSE convergence performance of the U-BiTRCE model under varying IRS element sizes. When using 16 elements, the model reaches an NMSE of −24.22 dB after 50 epochs. Increasing the IRS to 32 elements yields the best result at −24.94 dB, while the NMSE slightly drops to −24.07 dB at 64 elements. Despite the variation in array size, U-BiTRCE maintains consistent superiority over benchmark methods such as G-TIRC, CDRN, MMSE, and LS, demonstrating its adaptability and reliability across different IRS configurations.

Fig. 9 illustrates effect of number of BS antennas on NMSE convergence for the proposed U-BiTRCE model.

With two antennas, the model achieves its best performance, reaching an NMSE of −25.24 dB. As the antenna count increases to four, the final NMSE slightly rises to −23.94 dB, and further to −24.22 dB with eight antennas. Despite these variations, U-BiTRCE consistently outperforms benchmark models such as G-TIRC, CDRN, MMSE, and LS, confirming its robustness across different BS configurations.

As shown in Table 4, the proposed U-BiTRCE significantly outperforms all existing baselines, achieving a 152.4% NMSE improvement over LS and over 100% improvement compared to G-TIRC.

VI. CONCLUSION

This paper proposed U-BiTRCE, a novel deep learning framework for efficient CE in IRS-assisted wireless communication systems. The architecture integrates U-Net for spatial encoding, BiLSTM for temporal modeling, and a Transformer block for global context refinement. Unlike traditional estimators such as LS, LMMSE, and OMP, the proposed model requires no prior knowledge of sparsity or graph structure, making it scalable and adaptable to real-world 6G scenarios. Comprehensive simulations under varying system configurations—batch size, IRS element count, number of users, and base station antennas—demonstrated the model's superiority over state-of-the-art methods such as CDRN and G-TIRC. U-BiTRCE achieved up to 152.4% improvement in NMSE compared to the LS baseline, while maintaining an inference time below 18 milliseconds per frame. The complete model comprises approximately 1.8 million trainable parameters, is trained in under 30 minutes on an NVIDIA RTX 3090 GPU, and offers an average inference time of less than 5 milliseconds per sample—confirming its efficiency and suitability for real-time deployment in 6G IRS-assisted systems. These results confirm that U-BiTRCE is not only accurate but also computationally efficient, offering a practical solution for real-time IRS-aided channel estimation. Future work may extend this framework to handle time-varying channels, incorporate hardware impairments, or explore online learning strategies to further enhance adaptability in dynamic 6G environments.

U-BiTRCE shows stable performance across different IRS sizes and SNR conditions. Its design allows it to generalize well to moderate changes in the environment without the need for fine-tuning. While user mobility is not directly modeled, the temporal layers offer some robustness to dynamic conditions. This makes the model deployment-friendly for practical 6G scenarios.

Given its low inference latency and compact architecture, the proposed U-BiTRCE model is well-suited for real-time deployment in 6G base stations and edge computing platforms. Its modular design facilitates seamless integration as a channel estimation unit within end-to-end intelligent transceiver systems. When incorporated alongside adaptive beamforming and resource scheduling mechanisms, U-BiTRCE can contribute to efficient IRS control and enhanced link reliability in dynamic wireless environments.

Future work may extend this framework to handle time-varying channels, incorporate hardware impairments, or explore online learning strategies to further enhance adaptability in dynamic 6G environments. In future, the proposed U-BiTRCE model will also be validated through real-world IRS-assisted mmWave measurement campaigns to substantiate the simulation findings.

REFERENCES

[1] Z. A. Alsalman and H. Almosa, "Channel estimation based deep learning using IRS-assisted MISO systems with correlated channel," *J. Commun. Netw. (Syst.)*, vol. 4, p. 5, Dec. 2024.
[2] 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 Syst.*, vol. 14, no. 4, pp. 144–156, Aug. 2024.
[3] J. Zhang, Z. Wang, J. Li, Q. Wu, W. Chen, F. Shu, and S. Jin, "How often channel estimation is required for adaptive IRS beamforming: A bilevel deep reinforcement learning approach," *IEEE Trans. Wireless Commun.*, vol. 23, no. 8, pp. 8744–8759, Aug. 2024.
[4] S. Monga, N. Saluja, A. Taneja, C. Prabha, and R. Garg, "Revolutionizing intelligent reflecting surfaces: Deep learning-driven channel estimation," in *Proc. 2nd Int. Conf. Advancement Comput. Comput. Technol. (InCACCT)*, May 2024, pp. 148–153.
[5] S. Monga, A. Taneja, N. Saluja, and R. Garg, "Learning aided intelligent mechanism for channel estimation in 5G wireless networks," in *Proc. Int. Conf. Emerg. Smart Comput. Informat. (ESCI)*, Mar. 2023, pp. 1–6.
[6] Y. Wang, R. Dong, F. Shu, W. Gao, Q. Zhang, and J. Liu, "Power optimization and deep learning for channel estimation of active IRS-aided IoT," *IEEE Internet Things J.*, vol. 11, no. 24, pp. 41194–41206, Dec. 2024.
[7] S. Monga, A. Bansal, I. Aljubayri, N. Saluja, C. Prabha, S. Malhotra, P. Srivastava, and M. Z. Khan, "Adaptive intelligent reflecting surfaces for enhanced wireless communication via multi-agent deep reinforcement learning," *IEEE Access*, vol. 13, pp. 62071–62087, 2025.
[8] S. Monga, A. R. Singh, N. Saluja, C. Prabha, S. Malhotra, A. Karim, and M. M. Hassan, "Advanced beamforming and reflection control in intelligent reflecting surfaces with integrated channel estimation," *IET Microw., Antennas Propag.*, vol. 18, no. 12, pp. 917–931, Dec. 2024.
[9] Y. Liu, I. Al-Nahhal, O. A. Dobre, and F. Wang, "Deep-learning channel estimation for IRS-assisted integrated sensing and communication system," *IEEE Trans. Veh. Technol.*, vol. 72, no. 5, pp. 6181–6193, May 2023.
[10] A. Zayat, M. A. Hasabelnaby, M. Obeed, and A. Chaaban, "Transformer masked autoencoders for next-generation wireless communications: Architecture and opportunities," *IEEE Commun. Mag.*, vol. 62, no. 7, pp. 88–94, Jul. 2024.
[11] D. Luan and J. S. Thompson, "Channelformer: Attention based neural solution for wireless channel estimation and effective online training," *IEEE Trans. Wireless Commun.*, vol. 22, no. 10, pp. 6562–6577, Oct. 2023.
[12] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.
[13] Z.-Q. He and X. Yuan, "Cascaded channel estimation for large intelligent metasurface assisted massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 210–214, Feb. 2020.
[14] T. Lin, Y. Liang, W. Xu, and H. Li, "Tensor-based IRS channel estimation using PARAFAC decomposition," *IEEE Trans. Signal Process.*, vol. 69, pp. 4375–4389, 2021.
[15] K. Huang and J. Zhang, "Deep learning-based prediction of IRS reflection coefficients for CSI feedback reduction," *IEEE Trans. Commun.*, vol. 72, no. 4, pp. 2190–2203, Apr. 2024.
[16] A. M. Elbir and A. Papazafeiropoulos, "Federated learning for ris-assisted mmWave channel estimation in vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 70, no. 10, pp. 10164–10169, Oct. 2021.

[17] Z. Liu, X. Chen, and X. Li, "Deep residual learning for IRS-aided CSI prediction under partial pilot observation," *IEEE Commun. Lett.*, vol. 25, no. 3, pp. 715–719, Mar. 2021.

[18] Z. Zhang, L. Li, C. Dong, and S. Liu, "A convolutional-transformer residual network for joint location sensing and channel estimation in IRS-aided mmWave ISAC systems," *Sensors*, vol. 25, no. 19, p. 5959, Oct. 2025, doi: 10.3390/s25195959.

[19] S. Zhang, S. Zhang, Y. Mao, L. K. Yeung, B. Clerckx, and T. Q. S. Quek, "Transformer-based channel prediction for rate-splitting multiple access-enabled vehicle-to-everything communication," *IEEE Trans. Wireless Commun.*, vol. 23, no. 10, pp. 12717–12730, Oct. 2024.

[20] H. Ju, S. Jeong, B. Lee, and B. Shim, "Transformer-based predictive channel estimation for mmWave massive MIMO systems," in *Proc. IEEE 100th Veh. Technol. Conf. (VTC-Fall)*, Oct. 2024, pp. 1–5.

[21] B. Yue, S. Qiu, C. Yang, L. Peng, and Y. Zhang, "Transformer-empowered receiver design of OFDM communication systems," *Comput. Commun.*, vol. 228, Dec. 2024, Art. no. 107960.

[22] D. Luan and J. Thompson, "Attention based neural networks for wireless channel estimation," in *Proc. IEEE 95th Veh. Technol. Conf. (VTC-Spring)*, Jun. 2022, pp. 1–5.

[23] L. Cheng, Z. Zhang, C. Dong, and S. Liu, "MSTGAN: A multi-slot conditional generative adversarial network based on Swin transformer for channel estimation," *IEEE Commun. Lett.*, vol. 27, no. 7, pp. 1799–1803, Jul. 2023.

[24] J. Guo, G. Liu, Q. Wu, and P. Fan, "Parallel attention-based transformer for channel estimation in RIS-aided 6G wireless communications," *IEEE Trans. Veh. Technol.*, vol. 73, no. 11, pp. 15927–15940, Nov. 2024.

[25] F. Liu, J. Zhang, P. Jiang, C.-K. Wen, and S. Jin, "CE-ViT: A robust channel estimator based on vision transformer for OFDM systems," in *Proc. IEEE Global Commun. Conf.*, Dec. 2023, pp. 4798–4803.

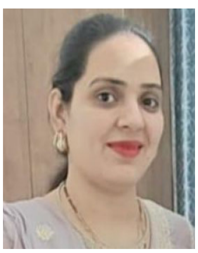
[26] K. Long and H. Liu, "Transformer-enabled hybrid precoding for TDD large-scale antenna arrays systems with channel sensing," *J. Commun. Netw.*, vol. 27, no. 1, pp. 23–31, Feb. 2025.

[27] Burera, S. Ahmed, and S. Kim, "Transformer learning-based efficient MIMO detection method," *Phys. Commun.*, vol. 70, Jun. 2025, Art. no. 102637.

[28] X. Cui, C. Zhang, J. Li, B. Jiang, S. Li, and J. Liu, "Channel estimation based on dual frequency domain transformer in time–frequency doubly-selective fading underwater acoustic channels," *Phys. Commun.*, vol. 68, Feb. 2025, Art. no. 102585.

[29] H. Liu, Z. Wei, H. Zhang, B. Li, and C. Zhao, "Tiny machine learning (tinyML) for efficient channel estimation and signal detection," *IEEE Trans. Veh. Technol.*, vol. 71, no. 6, pp. 6795–6800, Jun. 2022.

[30] S. Singh, A. Trivedi, and D. Saxena, "Channel estimation for intelligent reflecting surface aided communication via graph transformer," *IEEE Trans. Green Commun. Netw.*, vol. 8, no. 2, pp. 756–766, Jun. 2024.



SAKHSHRA MONGA is currently a SRF with Chitkara University, with expertise in wireless communication. She has four years of experience in teaching and research, with over 11 papers published.

ARCHITA SETHI is currently pursuing the Bachelor of Engineering degree in computer science and engineering with Chitkara University, Punjab, India. She is passionate about research and continuously seeks opportunities to expand her knowledge in emerging technologies. Her academic interests include artificial intelligence, machine learning, deep learning, and data analysis.

NITIN SALUJA specializes in microwave engineering and computational electromagnetics with Chitkara University, with a focus on agriculture and health applications. With 12 years of experience and notable inventions, such as the E-DeWeeder, his work merges machine learning with microwave technology. He has received multiple awards for his contributions to microwave and RF engineering.

A. F. M. SHAHEN SHAH (Senior Member, IEEE) received the B.Sc. degree in electronics and telecommunication engineering from Daffodil International University, Bangladesh, in 2009, the M.Sc. degree in information technology from the University of Dhaka, Bangladesh, in 2011, and the Ph.D. degree in electronics and communication engineering from Yildiz Technical University, Türkiye, in 2020. He is currently an Associate Professor with the Department of Electronics and Communication Engineering and the Director of the AI and Next-generation Wireless Communication Laboratory (ANWCL), Yildiz Technical University, Türkiye. He has authored a book. He has published a good number of research papers in international conferences and journals. His current research interests include wireless communication, artificial intelligence, 6G, blockchain, and the IoT. He has been a TPC member for several IEEE conferences and a regular reviewer for various IEEE journals. For his Ph.D. work, he won a Gold Medal at the 32nd International Invention, Innovation, and Technology Exhibition (ITEX 2021). He is serving as the Editor-in-Chief for *ICCK Transactions on Mobile and Wireless Intelligence* and *ICRRD Quality Index Research Journal*, an Editor for *The Open Transportation Journal* (Bentham) and *Discover Vehicles* (Springer), and an Associate Editor for *Journal of Cyber Security Technology* (Taylor and Francis).

JOHN EKORU received the B.Sc. degree in mechanical engineering from the University of KwaZulu-Natal (UKZN), Durban, South Africa, and the M.Sc. degree in mechanical engineering and the Ph.D. degree in electrical engineering from the University of the Witwatersrand, Johannesburg, South Africa. He is currently a full-time Staff Member with the Faculty of Engineering, University of the Witwatersrand, where he lectures data science and programming related courses. He also conducts data science related research work in the same institution. He has authored and co-authored several peer reviewed journal papers and presented at local and international conferences. His research interests include modelling, control, and optimization of biomedical and automotive systems. He also has a very keen interest in data science and its applications in the medical and engineering fields. He is a registered member of the Engineering Council of South Africa (ECSA) and a member of the International Association of Engineers and Computer Scientists (IAENG).

MILKA MADAHANA received the bachelor's degree (Hons.) in heavy current, the master's degree in electrical engineering, and the Ph.D. degree in electrical engineering from the University of the Witwatersrand, Johannesburg, South Africa. She is currently a full-time Staff Member with the Faculty of Engineering, University of the Witwatersrand, where she conducts research, in addition to that she lectures data science and programming related courses. She also collaborates with the Business Intelligence Unit, University of the Witwatersrand, on data science related projects. She has authored and co-authored several peer reviewed journals and conference papers in the areas of application of artificial intelligence and machine learning concepts. She has a great interests in the application of mathematical modelling and control techniques. She is a registered member of the Engineering Council of South Africa (ECSA) and a member of the International Association of Engineers and Computer Scientists (IAENG).

OLUTAYO OYEYEMI OYERINDE (Senior Member, IEEE) received the Ph.D. degree in electronic engineering from the University of KwaZulu-Natal, Durban, South Africa, in 2011. He has been with the School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa, since 2013, where he is currently an Associate Professor. He is a National Research Foundation (NRF) C2-rated Scientist, a Registered Professional Engineer (Pr.Eng.) with the Engineering Council of South Africa (ECSA), a Registered Engineer (R.Eng) with COREN, and a Corporate Member of NSE amongst others. He is an Associate Editor for IEEE Access, IEEE Communications Letters, the Journal of the Franklin Institute, and the Franklin Open. He is also an Editorial Member of the editorial board of *Wireless Communications and Control*. His current research interests include wireless communications with specific interests in 5G and beyond 5G technologies and signal processing techniques for wireless communication systems.