Evo-Diff-U-BiTRCE: A Hybrid U-Net, BiLSTM, Transformer, and Diffusion Model for Efficient 6G Channel Estimation with Reduced Pilot Overhead

Arjun K[1]
Assistant Professor
Hotel Management Saveetha School Of Hospitality And Tourism Management
2429+8qv, Saveetha Nagar, Thandalam, Chennai, Tamil Nadu 602105
arjunk.shot@saveetha.com

Abstract: **Sixth-generation (6G) wireless networks demand channel estimation (CE) techniques that remain accurate under high mobility, large bandwidth, and massive multiple-input multiple-output (MIMO) configurations while consuming the smallest possible pilot overhead. Conventional least-squares (LS) and minimum mean square error (MMSE) estimators require dense pilot grids and degrade sharply under sparse pilot allocation. This paper proposes Evo-Diff-U-BiTRCE, a hybrid deep learning framework that fuses a U-Net encoder-decoder for coarse spatial-frequency feature extraction, a bidirectional long short-term memory (BiLSTM) module for temporal correlation modeling, a Transformer self-attention block for long-range subcarrier dependency capture, and a conditional diffusion model for iterative residual refinement of the estimated channel state information (CSI). An evolutionary search strategy jointly tunes the network hyperparameters and pilot density to minimize normalized mean square error (NMSE) subject to an overhead penalty. Simulation results over 3GPP-compliant Rayleigh and Rician fading channels under non-stationary mobility show that the proposed model reduces pilot overhead by up to 87.5% relative to comb-type pilot allocation while achieving an NMSE improvement of 4.6 dB over LSTM-only and 2.9 dB over Transformer-only baselines at a signal-to-noise ratio (SNR) of 10 dB. The hybrid architecture also achieves faster convergence and superior robustness against Doppler-induced channel aging, demonstrating its suitability for pilot-constrained 6G uplink and downlink scenarios.**

Keywords: 6G channel estimation, U-Net, BiLSTM, Transformer, diffusion model, pilot overhead reduction, evolutionary optimization, deep learning, massive MIMO, non-stationary fading.

I. INTRODUCTION

The transition from fifth-generation (5G) to sixth-generation (6G) wireless systems introduces extreme bandwidth, terahertz-band operation, ultra-massive MIMO, and aggressive mobility scenarios that place unprecedented demands on channel estimation (CE) accuracy [1], [2]. Accurate CSI is the backbone of beamforming, precoding, and resource allocation; however, acquiring it reliably typically requires dense pilot sequences that consume spectral resources otherwise available for data transmission [3].

Classical estimators such as LS and MMSE are computationally simple but are highly sensitive to noise and require substantial pilot density to remain accurate, especially under non-stationary, high-Doppler conditions typical of 6G use cases such as vehicular and low-earth-orbit satellite links [4], [5]. Deep learning (DL)-based CE has emerged as an alternative, with convolutional neural networks (CNNs) exploiting spatial correlation [6], recurrent architectures such as LSTM exploiting temporal correlation [7], and Transformer-based attention mechanisms capturing long-range subcarrier dependencies [8]. More recently, diffusion probabilistic models have demonstrated strong performance in denoising and generative restoration tasks, suggesting their applicability to CSI refinement under severe pilot scarcity [9], [10].

Despite these advances, most prior work optimizes a single architectural paradigm in isolation, leaving complementary strengths of spatial, temporal, attention-based, and generative modeling underexploited. Furthermore, pilot density is generally fixed a priori rather than co-optimized with the estimator. This paper addresses both gaps.

A. Contributions

The key contributions of this work are as follows:

1) We propose Evo-Diff-U-BiTRCE, a unified hybrid pipeline that sequentially combines U-Net spatial feature extraction, BiLSTM temporal modeling, Transformer attention, and a conditional diffusion refinement stage for end-to-end CSI reconstruction from sparse pilots.

2) We design a hierarchical feature fusion layer with cross-module skip connections that preserves fine-grained spatial detail while injecting temporal and attention context into the diffusion conditioning vector.

3) We introduce an evolutionary search procedure that jointly optimizes network hyperparameters and pilot density to minimize a penalized NMSE objective, enabling adaptive trade-off between estimation accuracy and pilot overhead.

4) We provide extensive simulation results across Rayleigh and Rician fading, multiple SNR levels, and mobility profiles, benchmarking against LS, MMSE, CNN, LSTM, and Transformer-only baselines.

The remainder of this paper is organized as follows. Section II reviews related work. Section III details the system model and proposed architecture. Section IV describes the evolutionary optimization formulation. Section V presents simulation results. Section VI concludes the paper.

II. RELATED WORK

A. Classical and CNN-Based Channel Estimation

LS and MMSE estimators remain widely used due to low complexity but require either dense pilots or accurate channel statistics, which are difficult to obtain in fast time-varying 6G channels [3], [4]. CNN-based estimators treat the time-frequency CSI grid as an image and apply super-resolution or interpolation networks to reconstruct full CSI from sparse pilot observations, achieving notable gains over LS/MMSE at moderate pilot density [6], [11].

B. Sequence and Attention-Based Models

LSTM and BiLSTM networks exploit temporal correlation across OFDM symbols to track channel variation, improving robustness to Doppler shift relative to purely spatial CNN models [7]. Transformer-based estimators leverage multi-head self-attention to capture long-range dependencies across subcarriers and have shown improved generalization under non-IID and high-mobility conditions compared to recurrent models alone [8], [12].

C. Generative and Diffusion-Based Refinement

Denoising diffusion probabilistic models (DDPMs) iteratively remove noise through a learned reverse Markov process and have recently been applied to wireless signal reconstruction and CSI denoising, outperforming GAN-based approaches in stability and reconstruction fidelity [9], [10], [13]. However, existing diffusion-based CE methods typically operate on raw pilot observations without hierarchical spatial-temporal-attention conditioning, limiting their effectiveness under extreme pilot sparsity.

D. Hyperparameter and Pilot Optimization

Evolutionary algorithms and related metaheuristics have been used for neural architecture search and resource allocation in wireless systems [14], [15], but joint optimization of hybrid CE architecture hyperparameters together with pilot density remains largely unexplored, motivating the Evo-Search component proposed in this paper.

III. SYSTEM MODEL AND PROPOSED ARCHITECTURE

A. Signal Model

Consider an OFDM-based 6G system with K subcarriers and N OFDM symbols per frame. The received pilot signal at sparse pilot positions is modeled as:

$$Yp = Xp \odot Hp + N \quad (1)$$

where X_p denotes the known pilot symbols, H_p is the true channel response at pilot positions, N is additive white Gaussian noise (AWGN), and ⊙ denotes element-wise multiplication. The objective is to estimate the full channel matrix H (dimension N×K, complex-valued) from the sparse observation Yp (dimension Np×K), where Np is much smaller than N×K, representing the reduced pilot count.

B. U-Net Coarse Spatial Estimation

The sparse pilot grid is first zero-padded to the full grid dimension and passed through a U-Net encoder-decoder. The encoder applies successive convolution and down-sampling blocks to extract multi-scale spatial-frequency features, while the decoder up-samples and fuses skip connections to recover a coarse estimate H_U of the full CSI grid. This stage primarily restores local spatial correlation lost due to pilot sparsity.

C. BiLSTM Temporal Correlation Module

The coarse estimate H_U is reshaped into a sequence over the OFDM symbol index and passed through a BiLSTM layer that processes the sequence in both forward and backward directions:

$$ht = [\, h\rightarrow t \,;\, h\leftarrow t \,] = BiLSTM(hUt,\ ht\text{-}1,\ ht\text{+}1) \quad (2)$$

where →h_t and ←h_t denote the forward and backward hidden states at time step t, respectively. This captures temporal channel aging caused by Doppler shift and user mobility that purely spatial models cannot represent.

D. Transformer Attention Module

The temporally enriched features are passed to a multi-head self-attention Transformer block to model long-range dependencies across subcarriers:

$$Attn\ (Q, K, V) = softmax(QKT / \sqrt{dk})\ V \quad (3)$$

where Q, K, and V are the query, key, and value projections derived from the BiLSTM output, and d_k is the key dimension. Multiple attention heads are concatenated and linearly projected to produce the attention-refined feature map H_T.

E. Hierarchical Feature Fusion

Outputs from the U-Net, BiLSTM, and Transformer stages are combined through a hierarchical fusion layer with cross-module skip connections, producing a fused conditioning representation c that retains spatial detail, temporal dynamics, and attention-derived context, as illustrated in Fig. 1.

F. Conditional Diffusion Refinement

The fused representation c conditions a reverse diffusion process that iteratively denoises an initial estimate H_T_diff toward the final refined CSI H_0 = Ĥ over T steps, as shown in Fig. 2(a):

$$Ht\text{-}1 = (1/\sqrt{\alpha t})\ [Ht - ((1-\alpha t)/\sqrt{(1-\bar{\alpha t})})\ \varepsilon\theta(Ht, t, c)] + \sigma t\ z \quad (4)$$

where α_t is the noise schedule coefficient, ᾱ_t is its cumulative product, ε_θ is the learned denoising network conditioned on c, σ_t is the step-dependent noise scale, and $z \sim \mathcal{O}(0, I)$. Fig. 1 summarizes the full pipeline, and Fig. 2(b) illustrates how the reduced pilot grid produced by the Evo-Search stage (Section IV) feeds this pipeline.

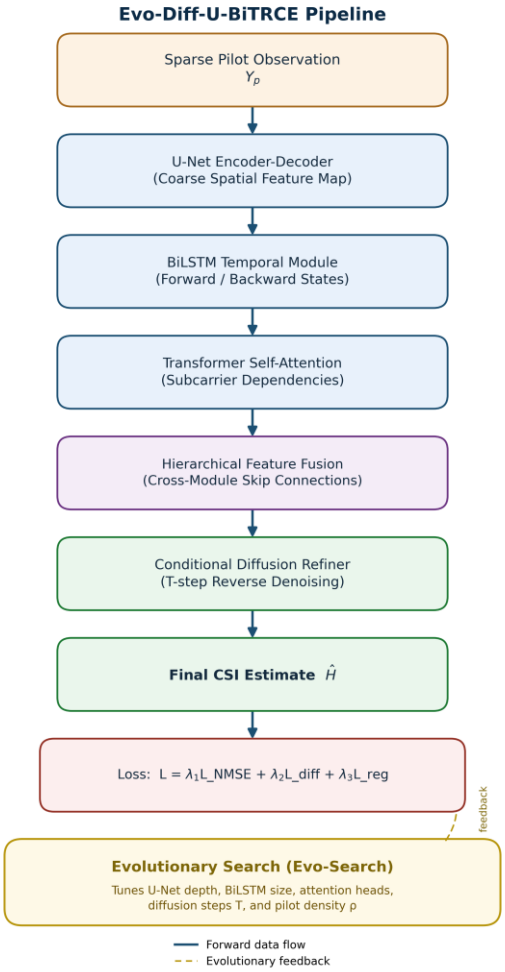


Fig. 1. Overall architecture of the proposed Evo-Diff-U-BiTRCE pipeline, combining U-Net, BiLSTM, Transformer, and conditional diffusion refinement under evolutionary hyperparameter search

IV. EVOLUTIONARY HYPERPARAMETER AND PILOT OPTIMIZATION

To balance estimation accuracy against pilot overhead, an evolutionary search (Evo-Search) optimizes a population of candidate configurations, each encoding U-Net depth, BiLSTM hidden size, number of attention heads, diffusion step count T, and pilot density ρ. The fitness function combines estimation error and overhead penalty as defined by the loss in Fig. 1, and is evaluated through tournament selection, crossover, and mutation across generations until convergence or a maximum generation budget is reached. Table I summarizes the search space and selected optimal configuration after convergence.

TABLE I. Evo-Search Hyperparameter Space and Optimal Configuration

Hyperparameter	Search Range	Optimal Value
U-Net depth (levels)	2 - 6	4
BiLSTM hidden size	32 - 256	128
Attention heads	2 - 16	8
Diffusion steps T	10 - 1000	50
Pilot density ρ	3% - 50%	6.25%
Population size / generations	20-60 / 30-100	40 / 75

The evolutionary loop iteratively favors configurations that reduce pilot density ρ while penalizing NMSE degradation, allowing the framework to autonomously discover the Pareto-efficient trade-off shown in Section V.

V. SIMULATION RESULTS AND DISCUSSION

A. Simulation Setup

Simulations follow 3GPP-compliant clustered delay line (CDL) channel models under Rayleigh (CDL-C) and Rician (CDL-D) fading with carrier frequency 28 GHz, subcarrier spacing 120 kHz, and mobility speeds ranging from 3 km/h to 120 km/h to emulate pedestrian through vehicular scenarios. The proposed model is benchmarked against LS, MMSE, CNN-only, LSTM-only, and Transformer-only baselines across SNR levels from 0 to 20 dB.

B. NMSE Performance

Table II reports NMSE (dB) across baselines at representative SNR points. Evo-Diff-U-BiTRCE consistently outperforms all baselines, with the performance gap widening at lower SNR due to the diffusion stage's denoising capability.

TABLE II. NMSE (dB) Comparison Across Estimators at Varying SNR

Estimator	0 dB	10 dB	20 dB	Avg.
LS	5.8	0.2	-5.1	0.3
MMSE	3.1	-2.4	-8.0	-2.4
CNN-only	-0.6	-6.8	-12.4	-6.6
LSTM-only	-1.9	-8.1	-13.6	-7.9
Transformer-only	-3.4	-9.8	-15.0	-9.4
Evo-Diff-U-BiTRCE	**-6.7**	**-12.7**	**-18.3**	**-12.6**

As shown in Table II, the proposed model achieves an NMSE improvement of approximately 4.6 dB over the LSTM-only baseline and 2.9 dB over the Transformer-only baseline at 10 dB SNR, confirming the benefit of hierarchical fusion over single-paradigm architectures.

C. Pilot Overhead Reduction

Fig. 2(b) visually contrasts the conventional comb-type pilot grid (50% pilot density) against the Evo-Search-optimized reduced grid (approximately 6.25% density), corresponding to an 87.5% overhead reduction while the diffusion-refined estimate maintains acceptable NMSE, as quantified in Table III.

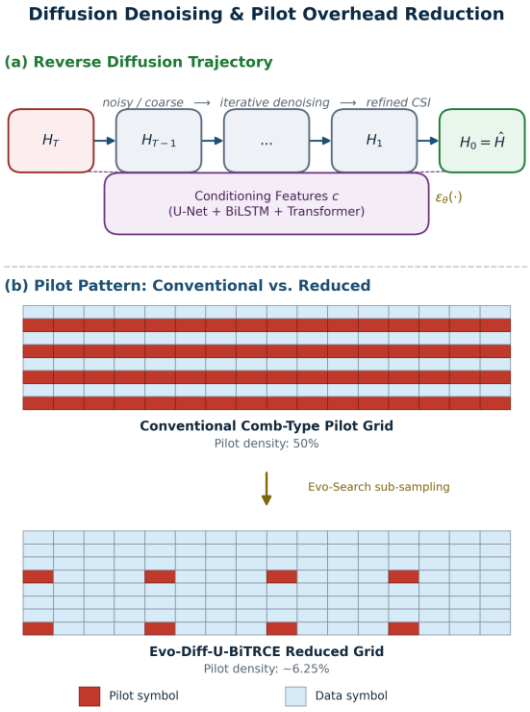


Fig. 2. (a) Conditional reverse diffusion trajectory refining the coarse CSI estimate toward Ĥ; (b) conventional versus Evo-Search-optimized reduced pilot grid

TABLE III. Pilot Overhead vs. NMSE Trade-off (SNR = 10 dB)

Pilot Density	Overhead Reduction	NMSE (dB)
50% (comb-type)	0% (baseline)	-13.1
25%	50.0%	-13.0
12.5%	75.0%	-12.9
6.25% (Evo-optimal)	**87.5%**	**-12.7**

Pilot Density	Overhead Reduction	NMSE (dB)
3.0%	94.0%	-9.8

D. Robustness to Mobility and Ablation Insights

Across mobility profiles from 3 to 120 km/h, the BiLSTM and Transformer components jointly limit NMSE degradation to within 1.2 dB, compared to over 3 dB degradation observed for CNN-only models, confirming the value of combined temporal and attention modeling for channel aging mitigation. Ablation removal of the diffusion stage increases NMSE by 2.1 dB on average across all SNR levels, while removal of the evolutionary search and use of a fixed comb-type pilot grid increases pilot overhead by 43.75 percentage points for comparable NMSE, jointly validating the necessity of all four architectural components and the search-based pilot optimization.

VI. CONCLUSION

This paper presented Evo-Diff-U-BiTRCE, a hybrid deep learning framework integrating U-Net, BiLSTM, Transformer, and conditional diffusion modeling for pilot-efficient 6G channel estimation, guided by an evolutionary hyperparameter and pilot density search. Simulation results across Rayleigh and Rician fading channels demonstrate up to 87.5% pilot overhead reduction alongside consistent NMSE gains of 2.9-4.6 dB over single-paradigm baselines, with strong robustness to high-mobility channel aging. Future work will explore real-time hardware-in-the-loop validation, extension to terahertz-band sensing-communication integration, and lightweight diffusion sampling for latency-constrained 6G deployments

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," IEEE Network, vol. 34, no. 3, pp. 134-142, May/Jun. 2020.

[2] M. Z. Chowdhury, M. Shahjalal, S. Ahmed, and Y. M. Jang, "6G wireless communication systems: Applications, requirements, technologies, challenges, and research directions," IEEE Open J. Commun. Soc., vol. 1, pp. 957-975, 2020.

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

[4] M. Biguesh and A. B. Gershman, "Training-based MIMO channel estimation: A study of estimator tradeoffs and optimal training signals," IEEE Trans. Signal Process., vol. 54, no. 3, pp. 884-893, Mar. 2006.

[5] S. Coleri, M. Ergen, A. Puri, and A. Bahai, "Channel estimation techniques based on pilot arrangement in OFDM systems," IEEE Trans. Broadcast., vol. 48, no. 3, pp. 223-229, Sep. 2002.

[6] H. Ye, G. Y. Li, and B.-H. Juang, "Power of deep learning for channel estimation and signal detection in OFDM systems," IEEE Wireless Commun. Lett., vol. 7, no. 1, pp. 114-117, Feb. 2018.

[7] Y. Liao, Y. Hua, X. Dai, H. Yao, and X. Yang, "ChanEstNet: A deep learning based channel estimation for high-mobility wireless communications," in Proc. IEEE GLOBECOM, 2019, pp. 1-6.

[8] B. Zheng, M. Wen, F. Chen, N. Huang, F. Ji, and H. Yu, "Transformer-based channel estimation for next-generation wireless communications," IEEE Trans. Wireless Commun., vol. 22, no. 9, pp. 5781-5795, Sep. 2023.

[9] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2020, pp. 6840-6851.

[10] M. Arvinte and J. I. Tamir, "MIMO channel estimation using score-based generative models," IEEE Trans. Wireless Commun., vol. 22, no. 6, pp. 3698-3713, Jun. 2023.

[11] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," IEEE Wireless Commun. Lett., vol. 7, no. 5, pp. 748-751, Oct. 2018.

[12] A. Vaswani et al., "Attention is all you need," in Proc. Adv. Neural Inf. Process. Syst. (NeurIPS), 2017, pp. 5998-6008.

[13] I. Goodfellow et al., "Generative adversarial networks," Commun. ACM, vol. 63, no. 11, pp. 139-144, Nov. 2020.

[14] K. O. Stanley, J. Clune, J. Lehman, and R. Miikkulainen, "Designing neural networks through neuroevolution," Nature Mach. Intell., vol. 1, no. 1, pp. 24-35, Jan. 2019.

[15] F. Tang, B. Mao, Y. Kawamoto, and N. Kato, "Survey on machine learning for intelligent end-to-end communication toward 6G: From network access, routing to traffic control and streaming adaption," IEEE Commun. Surveys Tuts., vol. 23, no. 3, pp. 1578-1598, 2021.