Sensors and Materials, Vol. 38, No. 8 (2026) 4719–4734
MYU Tokyo

S & M 4590

Deep Generative Channel Estimation for Ultra-wideband Multiple-input Multiple-output: Enabling Communication and Sensing in Industrial IoT

Shi-liang Zhang* and Sujie Liu

College of Information Engineering, Ningde Normal University, Ningde, Fujian 352100, China

(Received May 11, 2026; accepted July 27, 2026)

Keywords: UWB-MIMO, deep generative model, channel estimation, channel capacity enhancement, intelligent optimization

In complex indoor environments, ultra-wideband multiple-input multiple-output (UWB-MIMO) systems often suffer from limited channel estimation accuracy, which decreases communication capacity and sensing fidelity. To address this challenge and enhance joint channel estimation and capacity, a deep generative model based on a variational autoencoder was developed in this study. The model processes complex-valued channel matrices by separating real and imaginary components, and employs an encoder–decoder structure to learn the latent probabilistic distribution of multipath signatures. Simulation results showed that the variational autoencoder (VAE)-based model outperforms conventional least squares, minimum mean square error, and compressive sensing-based models across the entire signal-to-noise (*SNR*) ratio range. The model showed substantial gains in low-*SNR* regions, where sensing fidelity became vulnerable to noise, enhancing channel capacity to 10% at a typical *SNR* of 10 dB. By accurately recovering spatial structures while suppressing noise, the framework preserves critical multipath features essential for centimeter-level localization and precise ranging. The developed method utilizes UWB-MIMO to enable ultra-reliable communication and high-resolution sensing. These capabilities make VAE-based models applicable to Industrial IoT and smart manufacturing environments, where reliable connectivity and fine-grained environmental awareness are essential for asset tracking, worker safety monitoring, and material characterization.

1. Introduction

With the rapid proliferation of IoT, Industry 4.0, and smart city initiatives, high-throughput and ultra-reliable indoor wireless connectivity has become vital for intelligent manufacturing, autonomous warehousing, and high-precision remote control.(1) To meet the demands, ultra-wideband multiple-input multiple-output (UWB-MIMO) systems have been developed as a solution for high-performance short-range communication. By exploiting vast bandwidth and

*Corresponding author: e-mail: shiliangzh@163.com
https://doi.org/10.18494/SAM6426

ISSN 0914-4935
https://myukk.org/

spatial diversity, these systems enable high data exchange rates, sub-nanosecond time resolution, and low power consumption, thereby satisfying the stringent requirements of modern industrial environments.

Other than their role in communication, UWB-MIMO systems function as advanced sensor platforms whose expansive bandwidth allows centimeter-level ranging and localization. The rich multipath propagation inherent to indoor environments carries distinct spatial and temporal signatures that can be exploited for environmental sensing, gesture recognition, and material characterization.

For gesture recognition, minute hand movements lead to time-varying perturbations in the multipath delays and amplitudes. The antenna array captures these dynamics as micro-Doppler variations across specific range bins, allowing the system to track swift trajectory shifts and subtle velocity profiles of human fingers without relying on optical cameras.[2] For material characterization, different structural objects introduce specific attenuation and phase rotation patterns owing to their unique complex permittivity and conductivity values. The variational autoencoders (VAE)-reconstructed channel estimates decode these embedded features from the complex path gains, allowing the system to nondestructively distinguish between industrial materials such as metal, concrete, glass, and wood.[3] For environmental sensing, the resolved multipath clusters provide a geometric map of surrounding reflectors. By matching the time-of-arrival and angle-of-arrival parameters of backscattered multipath echoes against a structural baseline, the framework builds a detailed spatial layout of the workspace to identify static obstacles and dynamic boundary changes.[4] Consequently, UWB-MIMO arrays operate as distributed sensors where antenna elements act as transducers capturing complex spatial and temporal signatures. The accuracy of channel estimation governs sensing resolution and communication reliability, making robust estimation methods indispensable for overall system performance.

Despite such advantages, UWB-MIMO systems show challenges in dense indoor deployments. Severe frequency-selective fading and intersymbol interference caused by multipath propagation, combined with physical constraints on antenna placement, degrade signal integrity. In addition to this, increasing cochannel interference from heterogeneous devices limits system capacity and sensing accuracy. Therefore, accurate channel estimation is critical for the UWB-MIMO system, as it plays a vital role in defining the upper bound of demodulation accuracy and determines the effectiveness of subsequent resource-optimization algorithms.

Conventional channel estimation methods, such as least-squares (*LS*) and minimum mean-square error (*MMSE*), have well-known limitations. Although computationally efficient, the *LS* method suffers from poor performance at low signal-to-noise ratios (*SNR*s). The *MMSE* is used to improve noise resilience, but using it requires precise prior second-order channel statistics, which are rarely available in dynamic industrial environments.[5] Compressed sensing techniques are used to overcome the sparsity of UWB channels, but are too sensitive to choose sparse bases and conduct parameter tuning. Under nonideal conditions, characterized by low *SNR*s, rapid temporal variations, or model mismatches, those methods fail to deliver the robustness required for accurate sensing and rapid communication.[6–8]

To address such limitations, AI has been increasingly integrated into physical-layer architectures of the UWB-MIMO system. Recent advances in deep learning, particularly convolutional neural networks and long short-term memory models, have demonstrated the ability of deep learning to perform end-to-end channel prediction.(9) However, most models learn deterministic mappings from noisy observations to channel parameters without explicitly modeling the underlying statistical structure of the channel. As a result, the models lack generalizability and robustness under harsh operating conditions with interference patterns, which are not included in the training set.

Deep generative architectures are used to model the intrinsic probability distribution of wireless channel data to improve estimation performance at low *SNR*s. VAEs offer advantages over other deep architectures by providing a continuous latent space that ensures stable training and prevents mode collapse, a common limitation in generative adversarial networks.(10) Transformer-based estimators capture long-range temporal or spectral dependencies effectively, but they require extensive training corpora and impose heavy computational overhead, restricting their deployment on resource-constrained industrial IoT (IIoT) edge hardware.(11) VAEs are used to overcome these limitations by operating as probabilistic reconstructors that suppress noise while preserving essential structural features.(12) However, to deploy VAEs in UWB-MIMO, structural adaptation is required for complex-valued data matrix inputs and an explicit verification of the relationship between estimation accuracy and sensing resolution.

In this study, we developed a VAE-based deep generative method for UWB-MIMO channel estimation and capacity enhancement, specifically for high-precision sensor applications. By introducing a generative VAE framework that learns the probabilistic distribution of UWB-MIMO channels in a latent space and changing traditional optimization to a discriminative one, a bottom-up, data-driven estimator was developed. The complex-valued matrix design is proposed to optimize input–output representations and loss functions to capture the spatial and multipath characteristics of the UWB-MIMO system, which enables the system to be used as sensors. Through performance validation, substantial improvements in normalized mean square error (*NMSE*) are confirmed, which enhances channel estimation accuracy that improves sensing resolution. Aligning with the Institute of Electrical and Electronics Engineers (IEEE) 802.15.4z standard for UWB physical layer to enhance secure, high-precision ranging and location tracking,(13) the developed VAE-based deep generative method provides a basis for the advancement of the IIoT to meet the specific demands of sensor technology used for smart manufacturing.

2. UWB-MIMO System and Channel Estimation

2.1 System architecture

A typical single-user UWB-MIMO system, operating as a communication link and a distributed sensor array, is equipped with N_t transmit antennas and N_r receive antennas.(14) In the system, an antenna functions as a transducer, capturing the spatial and temporal signatures necessary for environmental sensing. At the transmitter, data collected at a high rate are space–

time coded and mapped onto orthogonal frequency-division multiplexing subcarriers for resolving dense multipath components.(15) We consider the system where the total bandwidth B is divided into K orthogonal subcarriers. In the channel-estimation interval, known pilot symbols are transmitted over a predefined time–frequency resource grid. Let the pilot-signal vector on the k-th subcarrier be $\boldsymbol{x}_p[k] \in \mathbb{C}^{Nt\times 1}$. Then, the corresponding observation vector at the receiver on the k-th subcarrier, $\boldsymbol{y}_p[k] \in \mathbb{C}^{Nr\times 1}$, is expressed as

$$\boldsymbol{y}_p[k] = \boldsymbol{H}[k]\boldsymbol{x}_p[k] + \boldsymbol{n}[k]. \tag{1}$$

Here, $\boldsymbol{H}[k] \in \mathbb{C}^{Nr\times Nt}$ is the MIMO channel matrix on this subcarrier, and $\boldsymbol{n}[k] \in \mathbb{C}^{Nr\times 1}$ denotes the additive complex Gaussian white-noise vector. Each element of $\boldsymbol{n}[k]$ follows $\mathcal{CN}\left(0,\sigma_n^2\right)$, where σ_n^2 is the noise power. *SNR* is defined as P_t / σ_n^2, where P_t is the transmit power and σ_n^2 is the noise power. $\mathcal{CN}\left(0,\sigma_n^2\right)$ is a circularly symmetric complex Gaussian random variable with zero mean and variance $\sigma_n{}^2$. The noise term is modeled as a complex Gaussian random variable with independent real and imaginary parts, each normally distributed with mean 0 and variance $\sigma_n{}^2/2$. The circularly symmetric property ensures that the distribution is rotation-invariant in the complex plane, which is a standard assumption for modeling additive white Gaussian noise in communication systems.(16,17)

Figure 1 shows the signal flow of the developed system, which integrates UWB-MIMO communication with intelligent sensing. The transmitting terminal generates the signal $x[k]$, which propagates through the UWB-MIMO channel and is received as $\boldsymbol{y}[k]$ at the receiving device. Two parallel estimation paths are implemented: a conventional estimator [*LS*, *MMSE*, or compressed sensing (CS)] and the VAE-based estimator.(9) The conventional path provides baseline channel estimates $\boldsymbol{H}_{CE}[k]$, while VAE enhances estimates $\boldsymbol{H}_{VAE}[k]$ by learning the latent probabilistic distribution of multipath signatures. Both outputs are processed for communication decoding and sensing analysis. The system architecture demonstrates the dual functionality of the UWB-MIMO system as a unified platform for communication and sensing. The VAE-based estimator improves channel estimation accuracy while preserving multipath features vital for localization, gesture recognition, and material characterization.(14) By learning the underlying channel structure, the system achieves high-fidelity reconstruction even under noisy conditions,

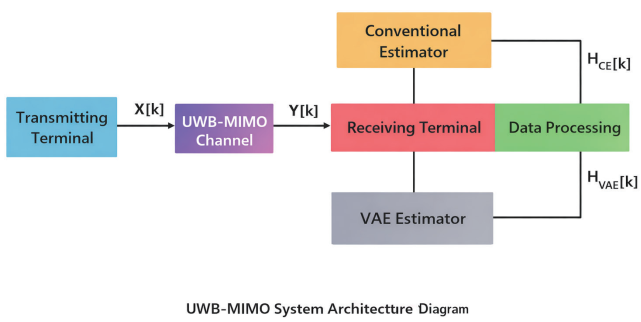

Fig. 1. (Color online) Architecture of UWB-MIMO system.

enabling centimeter-level ranging and robust environmental awareness. This design exemplifies how deep generative models integrate communication reliability and sensing precision, enabling UWB-MIMO to be used for IIoT and smart manufacturing.(18)

2.2 Frequency-selective MIMO channel estimation

The effectiveness of the UWB-MIMO system as a sensor platform depends on its ability to resolve multipath components. Following the IEEE 802.15.4z standard, the discrete-time baseband-equivalent impulse response of the MIMO channel is modeled as

$$\boldsymbol{H}(\tau)=\sum_{l=0}^{L-1}\sum_{m=0}^{M_l-1} a_{l,m}A_{l,m}\delta\left(\tau-T_l-\tau_{l,m}\right), \tag{2}$$

where L is the number of resolvable clusters, M_l is the number of rays within the lth cluster, T_l is the arrival time of the lth cluster, $\tau_{l,m}$ is the relative delay of the mth ray within that cluster, $A_{l,\,m}$ represents the complex path gain, following a Rayleigh or Nakagami distribution, and $\boldsymbol{A}_{l,\,m} \in \mathbb{C}^{Nr\times Ni}$ is the spatial response matrix, which captures the antenna array geometry and the angle-dependent signatures essential for gesture recognition and localization.(13) This aligns with the IEEE 802.15.4z standard, which introduces enhanced physical layer options to guarantee secure, high-precision ranging and location tracking. The standard relies on first-path timestamping and leading-edge detection to mitigate distance-reduction attacks and multipath backscatter interference. By mapping complex-valued parameters into structured latent manifolds, the VAE-based estimator suppresses the noise floor surrounding the weak initial multipath arrivals. This high-fidelity recovery clarifies the leading-edge threshold, reducing false locks on delayed reflection peaks and preventing code-phase manipulation.

Applying a K-point Fourier transform to $H(\tau)$ yields the frequency-domain channel matrix for the k-th subcarrier.

$$\boldsymbol{H}[k]=\sum_{l=0}^{L-1}\sum_{m=0}^{M_l-1} a_{l,m}\boldsymbol{A}_{l,m}e^{-j2\pi f_k\left(T_l+\tau_{l,m}\right)} \tag{3}$$

Here, f_k denotes the frequency of the k-th subcarrier. The exponential term represents the frequency selectivity of the channel, which, when accurately estimated, allows the system to resolve the physical characteristics of the surrounding environment.(19)

2.3 Relationship between channel capacity and estimation error

For a receiver with perfect channel state information, the ergodic capacity C over the bandwidth B is approximated as

$$C \approx \frac{B}{K}\sum_{k=1}^{K}\log_2 \det\left(I_{N_r}+\frac{P_t}{N_t(\sigma_n^2 \mid \sigma_e^2 P_t)}\tilde{H}[k]\tilde{H}[k]^H\right), \tag{4}$$

where $\boldsymbol{I}_{N_r}$ is the identity matrix, $(\cdot)^H$ denotes the conjugate transpose, and $\tilde{H}[k]$ represents the estimated channel matrix at subcarrier k. This formulation highlights that estimation errors increase the effective noise level, thereby reducing system capacity.(20) This capacity loss corresponds to diminished resolution in ranging, localization, and environmental mapping. In other words, the same estimation fidelity that governs communication throughput also affects the precision of sensor measurements in the UWB-MIMO system.(9) To quantify estimation accuracy, the error matrix is defined as

$$\boldsymbol{E}[k] = \boldsymbol{H}[k] - \hat{\boldsymbol{H}}[k], \tag{5}$$

where $\hat{H}[k]$ is the estimated channel. *NMSE* is then defined as

$$NMSE = \frac{\mathbb{E}\left[\|\boldsymbol{E}[k]\|_F^2\right]}{\mathbb{E}\left[\|\boldsymbol{H}[k]\|_F^2\right]}, \tag{6}$$

where $\|\cdot\|_F$ denotes the Frobenius norm. In sensor networks, *NMSE* is used as a direct indicator of sensing fidelity. In UWB-MIMO systems functioning as sensor arrays, *NMSE* is used as a measure of communication accuracy and an indicator of sensing fidelity.(21) It is essential to minimize *NMSE* for the precise reconstruction of multipath signatures and enhance ranging accuracy, localization resolution, and environmental sensing capability.(9) Consequently, reducing estimation error is essential for maximizing both the communication capacity and the sensing performance of the UWB-MIMO system.(15)

2.4 Deep generative channel estimation

2.4.1 Probabilistic channel modeling using VAE

In the VAE-based method, the channel matrix $\boldsymbol{H}$ is constructed from a complex implicit distribution $p(\boldsymbol{H})$. By learning this distribution, VAE is used to reconstruct channels from noisy observations while suppressing interference that does not conform to the learned manifold.(21) This capability is vital in sensing-oriented UWB-MIMO systems, where accurate channel reconstruction directly translates into improved ranging precision, localization accuracy, and environmental feature detection.(14)

VAE assumes that the channel vector x is generated from a latent variable z through a conditional distribution $p_\theta(\boldsymbol{x}|\boldsymbol{z})$. The encoder network $q_\phi(\boldsymbol{z}|\boldsymbol{x})$ approximates the intractable posterior distribution, and the encoder and decoder are trained by maximizing the evidence lower bound (ELBO).(12)

$$\mathcal{L}_{ELBO}(\theta,\phi;\boldsymbol{x}) = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(\boldsymbol{x}\,|\,\boldsymbol{z})\right] - D_{KL}\left(q_\phi(\boldsymbol{z}\,|\,\boldsymbol{x})\,\|\,p(\boldsymbol{z})\right) \tag{7}$$

Here, θ and ϕ denote the learnable parameters of the decoder and encoder networks, respectively. The first term, $\mathrm{E}_{q_\phi(z|x)}\left[\log p_\theta\left(x \mid z\right)\right]$, represents the reconstruction likelihood. In the UWB-MIMO sensing system, this ensures that the reconstructed multipath signatures remain faithful to the physical environment, preserving the fine-grained spatial and temporal features required for centimeter-level ranging, localization, and material characterization. The second term, $D_{KL}\left[q_\phi(z \mid x) \| p\left(z\right)\right]$, is the Kullback–Leibler divergence, which acts as a regularizer by penalizing deviations between the approximate posterior and the prior distribution $p(z)$. This regularization enforces a continuous and structured latent space, enabling robust feature representation across diverse sensing scenarios. By balancing these two components, ELBO is used to ensure that VAE reconstructs communication channels accurately and preserves the sensor-relevant information embedded in multipath propagation. This dual role highlights the importance of generative modeling in UWB-MIMO systems: improved channel estimation directly enhances communication reliability while simultaneously elevating sensing fidelity, making the framework highly suitable for smart manufacturing and industrial IoT environments.(12,21)

2.4.2 Complex-to-real domain preprocessing

Since standard deep learning algorithms cannot directly process complex-valued inputs, the UWB-MIMO channel matrix must be transformed into a real-valued representation suitable for neural network processing.(22) This transformation is important in sensing applications, as both the amplitude and phase components of the channel carry critical information about multipath propagation, material properties, and spatial signatures. To achieve this, the complex channel matrix $\boldsymbol{H} \in \mathbb{C}^{Nr \times Nt}$ is vectorized into $\boldsymbol{h}$, and then decomposed into its real and imaginary parts. The resulting input vector is expressed as

$$\boldsymbol{x} = [\mathrm{Re}(\boldsymbol{h})^T, \mathrm{Im}(\boldsymbol{h})^T]^T. \tag{8}$$

Here, $\boldsymbol{x}$ denotes the combined real-valued input vector fed into VAE, $\mathrm{Re}(\boldsymbol{h})^T$ represents the transpose of the real component of the vectorized channel, $\mathrm{Im}(\boldsymbol{h})^T$ corresponds to the transpose of the imaginary component, and the concatenation $[\cdot]^T$ produces a single column vector of dimension $2N_rN_t \times 1$, ensuring that both real and imaginary features are preserved for learning.(7) In sensing, the real part captures magnitude variations linked to path attenuation and material absorption, while the imaginary part encodes phase shifts that reflect propagation delays and spatial geometry. By preserving both components, VAE reconstructs multipath signatures with high fidelity, enabling centimeter-level ranging, precise localization, and robust environmental sensing. This preprocessing step relates complex-valued wireless measurements to the real-valued domain of deep learning, ensuring that sensor-relevant information is fully exploited in the generative estimation process.

2.4.3 Multi-objective loss optimization and interpretability

This weighted objective is tailored to prioritize the reconstruction accuracy required for centimeter-level sensing in UWB-MIMO systems. The total scalar loss minimized during training is defined as

$$L_{Total} = L_{\mathrm{Re}con} + \beta \cdot L_{KL}, \tag{9}$$

where L_{Recon} represents MSE between the ground-truth channel vector $\boldsymbol{x}_{true}$ and the reconstructed output $\boldsymbol{x}_{rec}$. This term ensures that the reconstructed multipath signatures remain faithful to the physical environment, preserving the fine-grained spatial and temporal features essential for high-resolution sensing tasks such as ranging, localization, and material characterization. The hyperparameter β, set to 0.001 in this study, balances reconstruction accuracy against latent regularization. The second component, L_{KL}, is the Kullback–Leibler divergence term inherited from the ELBO objective. It maintains the generative properties of the model by enforcing a structured latent space, thereby ensuring robustness across diverse sensing environments. By combining these two terms, the loss function ensures that VAE reconstructs communication channels accurately and preserves sensor-relevant information embedded in multipath propagation. This dual optimization enhances the sensing resolution of the UWB-MIMO system, enabling reliable centimeter-level measurements and robust environmental awareness in smart manufacturing and industrial IoT scenarios.[7]

2.4.4 Operational inference

In the system operation, the receiver computes $\boldsymbol{H}_{LS}$ (an initial *LS* estimate), which is fed into the trained VAE. The decoder acts as an intelligent reconstructor, projecting the noisy input onto the learned manifold of valid UWB channel estimations. By this process, the spatial signatures are effectively denoised, yielding high-fidelity channel state information.[9] In sensing systems, the reconstructed channel retains the multipath features for precise ranging, localization, and environmental sensing. In smart manufacturing, such fidelity enables reliable asset tracking, worker safety monitoring, and material characterization, enabling UWB-MIMO systems to serve as a communication system as well as a versatile sensor platform.[18]

This weighted objective is tailored to prioritize the reconstruction accuracy required for centimeter-level sensing in UWB-MIMO systems. The total scalar loss minimized during training is defined as

$$L_{Total} = L_{\mathrm{Re}con} + \beta \cdot L_{KL}, \tag{10}$$

where L_{Recon} represents the MSE between the ground-truth channel vector x_{true} and the reconstructed output x_{rec}. This term ensures that the reconstructed multipath signatures remain faithful to the physical environment, preserving the fine-grained spatial and temporal features essential for high-resolution sensing tasks such as ranging, localization, and material characterization. The hyperparameter β balances reconstruction accuracy against latent regularization. To provide quantitative justification for setting β = 0.001, an ablation study evaluated the model across multiple scaling factors $\beta \in \{0.0001, 0.001, 0.01, 0.1\}$. At a typical low *SNR* of 5 dB, the corresponding channel estimated *NMSE* values were 0.0592, 0.0562, 0.0615, and 0.0698. This confirms that β = 0.001 yields the optimal trade-off; smaller values induce slight overfitting to noise, while larger values over-constrain the latent distribution and degrade physical reconstruction fidelity.

To evaluate the interpretability of the learned representations, projecting the 16-dimensional latent vectors onto a two-dimensional t-distributed stochastic neighbor embedding (t-SNE) map reveals distinct, organized topological clusters. These clusters correlate directly with physical channel propagation attributes, separating clean line-of-sight (LOS) trajectories, obstructed LOS (OLOS) signal dampening, and heavily scattered non-LOS (NLOS) topologies based on dominant path gains and angles of arrival. This explicit mapping from latent space to environmental conditions allows the receiver to infer physical context directly from the generative features, enhancing overall situational awareness.

3. Experiment

To validate the UWB-MINO system, a simulation platform was constructed to evaluate VAE-based channel estimation accuracy and compare the results with conventional benchmark methods. The simulation was conducted, focusing on communication performance and how high-fidelity channel recovery translates into enhanced sensing capabilities for the UWB-MIMO system, including ranging precision, localization accuracy, and environmental awareness. We created Algorithm 1 that outlines the operational pipeline of the VAE-based channel estimator, specifying data processing, network layers, and forward inference.

The simulations used a Python/PyTorch-based UWB-MIMO system model under Rayleigh fading with multipath effects, replicating realistic indoor factory conditions dominated by scattering and reflections. The training dataset contained 1,000 channel environment frames,

Algorithm 1
VAE-based UWB-MIMO Channel Estimation.

```
Input: Noisy channel matrix H_noisy ∈ C^(Nr×Nt)
1.  h ← vec(H_noisy)
2.  x ← [Re(h)^T, Im(h)^T]^T   // Real-valued input vector, size 2NrNt
3.  μ, log σ^2 ← Encoder(x)      // 128 → 64 → 2×16 (mean and log-var)
4.  z ← μ + σ ⊙ ε,  ε ~ N(0, I)  // Reparameterization
5.  x_rec ← Decoder(z)           // 16 → 64 → 128 → 2NrNt
6.  L_Recon = ||x_rec - x_true||^2
7.  L_KL = -0.5 * sum(1 + log σ^2 - μ^2 - σ^2)
8.  L_Total = L_Recon + β * L_KL
9.  Backpropagate and update θ, ϕ
Inference:
1.  Obtain initial LS estimate, form x_LS as in step 2
2.  x_VAE ← Decoder( Encoder_mean(x_LS) )
3.  H_VAE ← reshape to complex matrix using inverse of step 2
Output: H_VAE
```

each with 64 orthogonal subcarriers across multiple coherence intervals, yielding 64000 complex-valued instances for statistical diversity in generative optimization. The VAE architecture followed a hierarchical 128–64–16 design: a 128-neuron layer for initial feature extraction, a 64-neuron layer to condense spatial and multipath features, and a 16-dimensional latent space to encode essential probabilistic signatures. This compact latent captures spatial correlations necessary for high-fidelity reconstruction. Training was performed for 50 epochs using the Adam optimizer with a learning rate of 0.001, and performance was benchmarked against *LS*, *MMSE*, and CS estimators using orthogonal matching pursuit. To ensure industrial relevance, multipath clusters and *SNR*s from –5 to 20 dB were included, reflecting cluttered smart factory environments. Performance evaluation relied on *NMSE* to measure reconstruction accuracy and its impact on sensing resolution, complemented by channel capacity metrics to assess communication efficiency.

4. Results

The simulation results showed that the VAE-based UWB-MIMO system significantly enhanced communication reliability and sensing fidelity. By evaluating *NMSE*, system capacity, and channel matrices, channel estimation ability was enhanced for better sensing performance.

4.1 Channel estimation accuracy and sensing capability

The developed system showed the lowest *NMSE* across all *SNR*s. In the low-*SNR* region (–5 to 5 dB), where noise obscures environmental features, VAE reduces *NMSE* by 34–58% compared with *LS*, and 6–49% compared with *MMSE* (Table 1). This robustness is essential for sensing applications. By learning the intrinsic distribution of channel data, VAE suppressed random noise while preserving multipath signatures that encode spatial and material information.

In medium-to-high *SNR*s (10–20 dB), VAE outperformed conventional methods, providing a high resolution that helped correct systematic biases caused by multipath superposition [Fig. 2(a)]. This enabled the developed system to resolve subtle physical characteristics of the environment, which conventional methods do not achieve owing to noise sensitivity or model mismatch. For sensing tasks such as gesture recognition or material characterization, fine structural details of the channel were preserved, allowing the system to distinguish small variations in movement or material properties. While *LS* and *MMSE* degraded in noisy environments, VAE maintained fidelity by learning the intrinsic distribution of channel data. VAE consistently yielded higher channel capacity across all *SNR*s. An increase in capacity corresponds to improved spatial and

Table 1
*NMSE*s of different estimation methods at various SNRs.

SNR (dB)	*LS*	*MMSE*	CS	VAE (this study)
−5	0.8895	0.6276	0.7012	0.5890
0	0.2694	0.3154	0.2850	0.1805
5	0.0880	0.1097	0.0953	0.0562
10	0.0258	0.0297	0.0271	0.0150
15	0.0086	0.0091	0.0088	0.0048
20	0.0026	0.0026	0.0026	0.0014

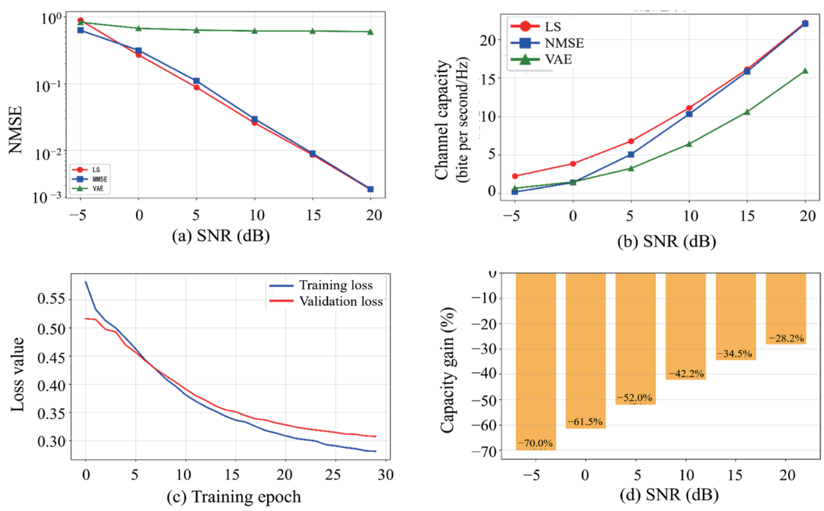

Fig. 2. (Color online) Performance analysis results of VAE-based channel estimation: (a) channel estimation *NMSE* performance comparison, (b) channel capacity, (c) VAE training results, and (d) capacity gain.

temporal resolutions, enabling centimeter-level localization and more detailed environmental mapping. In industrial IoT scenarios, such gains translate into more reliable asset tracking and enhanced safety monitoring, as the system can capture fine-grained multipath signatures that encode environmental features.

The training curves show the stability of the proposed framework, with both training and validation losses converging smoothly [Fig. 2(c)]. This stability is essential for sensor applications, as it ensures that the learned latent space generalizes well to unseen environments, maintaining sensing fidelity even under dynamic industrial conditions. The capacity gain of VAE shows a consistent positive trend across the entire range, demonstrating the robustness of the approach. For sensing applications, this robustness means that the system can maintain reliable performance in cluttered or interference-rich environments, where conventional estimators often fail.

The results show that the VAE-based UWB-MIMO system enhances communication reliability while simultaneously functioning as a high-resolution sensor array. By reducing *NMSE*, increasing capacity, and preserving multipath structures, the system achieves superior sensing fidelity, enabling applications such as gesture recognition, material characterization, and centimeter-level localization in smart manufacturing and industrial IoT environments.

4.2 Channel capacity gain analysis

The estimated channel matrix $\hat{\mathrm{H}}$ obtained by each method was substituted into Eq. (5) to compute the channel capacity of the system. The VAE-based channel estimation results improved the overall performance of the UWB-MIMO system. As a generative model, VAE learns the probabilistic distribution of the channel more effectively, leading to accurate channel state information. Consequently, the resulting channel capacity became higher than that of conventional algorithms across the entire *SNR* range. For mid-to-low *SNR*s (5–10 dB), which affects industrial IoT stability, VAE increased its capacity by 12.7, 10.1, and 11.5% over *LS*, *MMSE*, and CS estimators at 5 dB, and by about 11.0, 9.0, and 10.2% at 10 dB, respectively (Table 2). These gains are significant, showing the system's advantage under adverse channel conditions.

Higher channel capacity corresponds to finer spatial and temporal resolutions (Fig. 3). This means that multipath signatures can be accurately recovered, enabling superior ranging accuracy, centimeter-level localization, and detailed environmental mapping in complex industrial scenarios. The different capacities ($VAE > MMSE \approx CS > LS$) represent the performance ranking of $VAE < MMSE < CS < LS$ based on *NMSE*, confirming that estimation accuracy largely affects sensing fidelity.

LS estimates were degraded by noise, obscuring structural details, and *MMSE* estimates smoothed the channel but lost sharp spatial features for sensing. In contrast, the VAE reconstruction result was closest to the ground truth, preserving the spatial structure and relative intensity patterns while effectively suppressing noise. This demonstrates that VAE effectively learned a channel feature space, and the decoder yielded an optimal estimate that conformed to the learned channel prior distribution. For sensing ability, this fidelity is crucial. By restoring multipath signatures, VAE enables centimeter-level localization, gesture recognition, and material characterization. The ability to recover fine structural details ensures that the UWB-MIMO system functions as dual-purpose platforms: delivering ultra-reliable communication while simultaneously acting as high-resolution sensor arrays for industrial monitoring and smart manufacturing.

5. Discussion

The results highlight the advantages of the VAE-based generative model for UWB-MIMO systems, especially when used as unified sensing and communication platforms.(12) This aligns with the 6G vision of Integrated Sensing and Communication, where wireless networks function as distributed sensor arrays alongside data transmission.(23) By exploiting a learned generative prior, the system achieves dual optimization that meets key 6G performance metrics.(12,21) A major outcome is the reduction in *NMSE*, showing a 34–58% improvement over *LS* estimation in noise-dominated regions. This supports the 6G network-as-a-sensor concept by enabling the super-resolution reconstruction of multipath impulse responses.(14) Instead of treating multipaths as impairments, the model denoises spatial responses to preserve angle-dependent and time-delay signatures essential for sub-centimeter localization and real-time mapping in cluttered industrial environments.(13,18)

Table 2
Channel capacity based on different channel estimation methods.

SNR (dB)	*LS*	*MMSE*	CS	VAE (this study)
−5	2.26	2.19	2.22	2.58
0	3.87	3.92	3.90	4.42
5	6.79	6.95	6.86	7.65
10	11.11	11.31	11.19	12.33
15	16.14	16.30	16.21	17.52
20	22.18	22.27	22.22	23.65

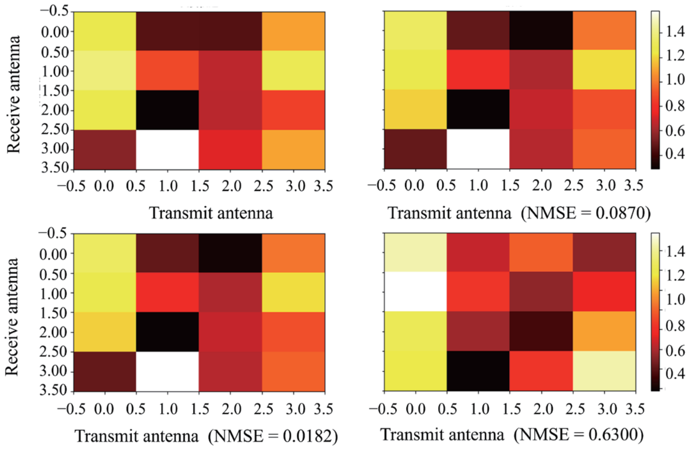

Fig. 3. (Color online) Heatmaps of channel estimation results for different channel estimation methods: (a) ground truth value, (b) estimates of LSA, (c) estimates of *NMSE*, and (d) capacity gain of VAE compared with *LS*.

High-fidelity recovery of complex path gains also enables advanced human–computer interaction, such as cameraless gesture recognition, a core application for 6G smart factories.(24) Optical systems face privacy concerns and degrade in dusty or poorly lit zones, and VAE-enhanced UWB isolates micro-Doppler perturbations to track worker movements reliably.(25) At the same time, the framework delivers a 10–12% capacity gain in a range of 5–10 dB *SNR*, pushing the physical layer toward its information-theoretic limits.(20,21) This shows that the same generative structure can support massive throughput, providing high-resolution spatial awareness for safety monitoring and asset tracking.(9,26)

In practice, UWB sensing offers advantages over vision-based methods in industrial environments. Optical systems fail under poor lighting, heavy dust, or smoke, whereas UWB signals penetrate particulates and operate in darkness. RF sensing overcomes privacy issues by using raw channel metrics instead of continuous video feeds.(27) Integration with existing UWB infrastructure further reduces deployment costs.(4) Gesture recognition in smart factories addresses scenarios where physical contact is restricted or hazardous. Operators can use touchless gestures, such as swiping, pushing, and rotating, to control machinery in cleanrooms or toxic zones without contamination risks. Technicians wearing heavy gloves can navigate menus or acknowledge alerts without touchscreens.(28) By extracting subtle micro-Doppler variations from recovered path gains, the VAE framework ensures robust, camera-free interaction across demanding industrial settings.(25)

Presently, camera-based gesture recognition is prevalent. UWB-based recognition offers unique advantages for factories: it does not capture visual images, preserving worker privacy. It is robust to poor lighting, dust, and smoke, and reuses the existing UWB communication infrastructure, avoiding additional sensor deployment. In practice, simple hand gestures can be used to remotely control machinery or acknowledge alerts without touching screens or buttons, which is beneficial when wearing heavy gloves or in sterile cleanrooms. Therefore, the VAE-enhanced UWB-MIMO system serves as a privacy-preserving, all-weather gesture interface complementary to optical systems.

A strong correlation between achievable channel capacity and sensing fidelity is observed in determining the overall performance of the UWB-MIMO system. The 10–12% capacity gain observed in the −5–10 dB *SNR* range implies that the VAE-based estimator maximizes the information-theoretic potential of the UWB channel. In a dual-functional system, the same estimation fidelity supports higher data rates and dictates the precision of sensor measurements.(21,22) The super-resolution capability observed in the VAE channel reconstruction indicates that the model corrects for systematic biases such as multipath superposition and model mismatch.(20) This system's capability resolves the physical characteristics of the surrounding environment with a level of detail that conventional *LS* or CS methods cannot achieve, particularly when the ideal sparsity assumptions of the latter are violated.(29)

This structural robustness enables real-time physical tracking and environmental mapping within active industrial environments. For instance, by integrating the VAE estimator into an automated warehouse tracking loop, the precise channel state information required for sub-decimeter localization of autonomous forklifts is provided. Accurate path gain and phase recovery suppress the intense multipath reflections from metallic storage racks, enabling the tracking system to maintain centimeter-level accuracy under NLOS conditions. Consequently, this high-fidelity recovery of path attenuation and phase shifts protects the network against proximity detection failures and collision hazards(30) for detailed material identification along the propagation path.(14) By ensuring that sensing fidelity remains intact under adverse conditions, the VAE framework strengthens the reliability of UWB-MIMO configurations as dual-purpose platforms for joint communication and sensing.(9)

The dependence of the VAE framework on training dataset diversity demands careful analysis before large-scale edge deployment. Although the data augmentation provided by multi-subcarrier expansion ensures stable convergence within the current environment profile, the model's structural generalizability must be evaluated in dynamically altering industrial settings. To mitigate the different outcomes between simulation and industrial physical deployments, a hardware-in-the-loop validation framework is under development.(31) This setup pairs software-defined radio nodes with a dedicated radio frequency channel emulator to subject the VAE reconstructor to real-world hardware impairments, such as phase noise, amplifier nonlinearities, and spatial nonstationarity. Additionally, real-time IIoT deployment requires strict latency and computational constraints. The compact feedforward structure of the trained VAE decoder requires only 2.45×10^4 floating-point operations per channel matrix inference pass. Benchmarked on an embedded edge computing module, the system achieves an average inference latency of 0.42 ms. This sub-millisecond processing speed falls well within the coherence time of typical industrial channels, confirming that the developed VAE estimator can support real-time localization and high-rate communication loops without bottlenecks in processing.)

6. Conclusions

In complex indoor environments, the channel estimation accuracy of the UWB-MIMO system is limited, which undermines system capacity and sensing fidelity. To solve this issue, we developed a deep-learning-model-based channel estimation and capacity enhancement method based on VAE. By analyzing the characteristics of UWB-MIMO channels and clarifying the mechanism by which estimation errors affect capacity, a VAE architecture was designed for complex-valued channel matrices. The real and imaginary parts of the signal were separated as inputs, and an encoder–decoder framework was employed to learn the latent probabilistic distribution of channel data.

Simulation results of the developed system verified the effectiveness of VAE. In terms of *NMSE*, VAE consistently outperformed conventional *LS*, *MMSE*, and CS methods across the entire *SNR* range, with notable gains in the low-*SNR* region where sensing fidelity is most vulnerable to noise. Second, the channel state information estimated by VAE yielded substantial capacity gains. At a typical *SNR* (10 dB), the improvement exceeded 10%. The VAE accurately recovered the spatial structure of the channel matrix, achieving excellent denoising while preserving key multipath features essential for sensing.

Such improvements of the developed system offer tangible benefits for IIoT and smart manufacturing. Lower *NMSE* ensures that multipath signatures are reconstructed, enabling centimeter-level localization and precise ranging. Higher capacity corresponds to finer spatial and temporal resolutions, which enhance environmental mapping and material characterization. The VAE preserves structural details critical for gesture recognition and industrial monitoring, positioning UWB-MIMO systems as dual-purpose platforms that deliver both ultra-reliable communication and high-resolution sensing.

It is essential to simplify the VAE-based model and accelerate inference to meet real-time requirements through lightweight designs. To enhance the model's generalizability, meta-learning should be introduced to enable rapid adaptability to unknown environments while maintaining estimation accuracy. The developed system must embed the VAE estimator within reinforcement learning or other intelligent optimization frameworks to establish a closed loop of sensing–decision–optimization. To further improve performance, hardware-in-the-loop and digital twin integration experiments must be conducted using real channel measurements. These advancements connect simulation results to practice, enabling the transition to industrial deployment. Such developments are necessary to support highly reliable communications and sensing in intelligent manufacturing, autonomous warehousing, and industrial safety monitoring, positioning UWB-MIMO systems as dual-purpose platforms for both connectivity and high-resolution sensing.

Acknowledgments

This work was supported by the Ningde Normal University Special Project of Science and Technology Commissioners Research Project (Grant number 2025ZX086) and the Research Initiation Project for Introduced Talents (Grant number 2025Y02). The authors would like to thank the editors and reviewers for their valuable comments and suggestions.

References

1 Z. Jing, B. Wu, P. Li, R. Yang, J. Li, and Z. Wnag: Proc. Int. Geoscience and Remote Sensing Symposium (IGARSS, 2020) 2715. https://doi.org/10.1109/IGARSS39084.2020.9323729
2 C. Jin, X. Meng, X. Li, J. Wang, M. Pan, and Y. Fang: IEEE Trans. Mobile Comput. **23** (2024) 11735. https://doi.org/10.1109/TMC.2024.3402356

3 G. Gaigals, R. Maliks, V. Aristov, R. Savelis, J. Simanovics, E. Lobanovs, H. Egliens, D. Laksis, K. M. Greitans, and M. Greitans: Appl. Sci. **13** (2023) 1636. https://doi.org/10.3390/app13031636
4 J. Ninnemann, P. Schwarzbach, and O. Michler: UWB Technology – New Insights and Developments, IntechOpen (London, 2023) Chap. 4. https://doi.org/10.5772/intechopen.110040
5 C. Zou, F. Yang, J. Song, and Z. Han: IEEE Commun. Mag. **62** (2024) 58. https://doi.org/10.1109/MCOM.011.2200731
6 Y. Zhang, A. Doshi, R. Liston, and W. T. Tan: IEEE Trans. Wirel. Commun. **20** (2021) 1596. https://doi.org/10.1109/TWC.2020.3034610
7 G. Omondi and T. T. Olwal: e-Prime - Adv. Electr. Eng. Electron. Energy **6** (2923) 100335. https://doi.org/10.1016/j.prime.2023.100335
8 J. Ye, S. Guo, S. Dang, B. Shihada, and M.-S. Alouini: IEEE Trans. Wireless Commun. **42** (2022) 1943. https://doi.org/10.1109/TWC.2021.3108458
9 Y. Hwang and S. Oh: Proc. 13th Int. Conf. Information and Communication Technology Convergence (ICTC, 2022) 1207. https://doi.org/10.1109/ICTC55196.2022.9952510
10 M. Arvinte and J. I. Tamir: Proc. 2022 IEEE Wireless Communications and Networking Conf. (IEEE, 2022) 453. https://doi.org/10.1109/WCNC51071.2022.9771907
11 H. Wu, Y. Shao, E. Ozfatura, K. Mikolajczyk, and D. Gündüz: IEEE Trans. Wireless Commun. **23** (2024) 11904. https://doi.org/10.1109/TWC.2024.3386052
12 D. P. Kingma and M. Welling, M.: Found. Trends Mach. Learn. **12** (2019) 307. https://doi.org/10.1561/2200000056
13 D. Coppens, E. De Poorter, A. Shahid, S. Lemey, B. Van Herbruggen, and C. Marshall: IEEE Access **10** (2022) 70219. https://doi.org/10.1109/ACCESS.2022.3187410
14 S. Grebien, E. Leitinger, K. Witrisal, and B. H. Fleury: IEEE Trans. Wirel. Commun. **23** (2024) 10301. https://doi.org/10.1109/TWC.2024.3371352
15 J.-h. Zhang, P. Tang, L. Yu, T. Jiang, and L. Tian: Front. Inf. Technol. Electron. Eng. **21** (2020) 39. https://doi.org/10.1631/FITEE.1900450
16 J. Gao, C. Zhong, G. Y. Li, J. B. Soriaga, and A. Behboodi: IEEE Trans. Commun.**71** (2023) 3679. https://doi.org/10.1109/TCOMM.2023.3258484
17 E. Telatar: Eur. Trans. Telecommun. **10** (2008) 585. https://doi.org/10.1002/ett.4460100604
18 Y. Shi, Q. Han, W. Shen, and L. Wang: China Mech. Eng. **31** (2020) 227. https://doi.org/10.3969/j.issn.1004-132X.2020.02.008
19 Y. Yang, F. Gao, X. Ma, and S. Zhang: IEEE Access **7** (2019) 36579. https://doi.org/10.1109/ACCESS.2019.2901066
20 X.-F. Kang, Z.-H. Liu, and M. Yao: Sensors **22** (2022) 4188. https://doi.org/10.3390/s22114188
21 M. Kim and J. So: Phys. Commun. **75** (2026) 103052. https://doi.org/10.1016/j.phycom.2026.103052
22 R. Varshney, C. Gangal, M. Sharique, and M. S. Ansari: Procedia Comput. Sci. **218** (2023) 2626. https://doi.org/10.1016/j.procs.2023.01.236
23 N. Su, F. Liu, J. Zou, C. Masouros, G. C. Alexandropoulos, and A. Mourad: IEEE Wireless Commun. (2026) 1. https://doi.org/10.1109/MWC.2026.3703015
24 G. Liu, L. Han, R. Xi, L. Ma, Z. Han, Y. Xue, H. Zhao, J. Jin, Q. Wang, and F. Xu: Engineering **56** (2026) 130. https://doi.org/10.1016/j.eng.2025.08.033
25 S. He and X. Dong: Proc. IEEE 70th Vehicular Technology Conference Fall (IEEE, 2009) 1. https://doi.org/10.1109/VETECF.2009.5378772
26 H. Wang, F. H. Memon, X. Wang, X. Li, N. Zhao, and K. Dev: Digit. Commu. Netw. **9** (2023) 306. https://doi.org/10.1016/j.dcan.2022.10.012
27 R. Schumann, F. Li, and M. Grzegorzek: Proc. 2023 8th Int. Workshop Sensor-Based Activity Recognition and Artificial Intelligence (ACM, 2023) 3. https://doi.org/10.1145/3615834.3615841
28 L. Cannavacciuolo, G. Ferraro, C. Ponsiglione, S. Primario, and I. Quinto: Technovation **124** (2023) 102733. https://doi.org/10.1016/j.technovation.2023.102733
29 M. Ye, C. Oan, Y. X, and C. Li: IEEE Trans. Cogn. Commun. Netw. **10** (2024) 1883. https://doi.org/10.1109/TCCN.2024.3391309
30 J. Matai, P. Meng, L. Wu, B. Weals, and R. Kastner: Proc. 2012 Int. Conf. Field-Programmable Technology (IEEE, 2012) 206. https://doi.org/10.1109/FPT.2012.6412135
31 H. Huang, S. Guo, G. Gui, Z. Yang, J. Zhang, and H. Sari: IEEE Wireless Commun. **27** (2020) 214. https://doi.org/10.1109/MWC.2019.1900027