Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

CINet: Communication-aware Inter-modality Network for Distributed Multi-Modal MRI Reconstruction

Shuo Yan
Beijing University of Posts and Telecommunications
Yu Bai
Beijing University of Posts and Telecommunications
Zheng Zhang
Beijing University of Posts and Telecommunications
Bo Zhang
zbo@bupt.edu.cn

Beijing University of Posts and Telecommunications
Wendong Wang
Beijing University of Posts and Telecommunications

Research Article

Keywords: Multi-modal MRI reconstruction, accelerated MRI, distributed reconstruction, cross-modal information, communication constraints, task-oriented communication

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11216972/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

CINet: Communication-aware Inter-modality Network for Distributed Multi-Modal MRI Reconstruction

Shuo Yan[1], Yu Bai[1], Zheng Zhang[2], Bo Zhang[1*], Wendong Wang[1]

[1*]State Key Laboratory of Networking and Switching Technology, School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing, 100876, China.

[2]School of Intelligent Engineering and Automation, Beijing University of Posts and Telecommunications, Beijing, 100876, China.

*Corresponding author(s). E-mail(s): zbo@bupt.edu.cn;

Abstract

AI-enabled medical imaging is increasingly moving toward distributed acquisition, computing, and inference, where information required by a downstream task may reside at different nodes. In multi-modal magnetic resonance imaging (MRI) reconstruction, this situation arises when a fully sampled auxiliary contrast is not locally available but can provide complementary information for reconstructing an undersampled target contrast. Existing methods generally assume direct access to the auxiliary modality and therefore do not consider the communication constraints involved in transferring such information between distributed nodes. To address this problem, we propose the Communication-aware Inter-modality Network (CINet) for distributed multi-modal MRI reconstruction, which jointly considers auxiliary representation transmission and cross-modal reconstruction under communication constraints. Specifically, a Task-Oriented Communication Encoder reweights multi-scale auxiliary features and generates compact representations under a prescribed communication budget. The resulting representations are transmitted through a noisy channel, and a Target-Conditioned Selective Fusion module selectively incorporates the received auxiliary information according to the current target representation. CINet thus coordinates

compact auxiliary information transmission with its downstream utilization for target reconstruction. Experiments on multi-modal prostate MRI demonstrate that CINet achieves competitive or superior reconstruction performance under different undersampling patterns and acceleration factors while preserving local structural details. Further evaluations under varying communication budgets and channel signal-to-noise ratios characterize the communication–reconstruction trade-off under different transmission conditions.

Keywords: Multi-modal MRI reconstruction, accelerated MRI, distributed reconstruction, cross-modal information, communication constraints, task-oriented communication

1 Introduction

The growing deployment of artificial intelligence (AI) across distributed devices, edge platforms, and cloud infrastructures is placing new demands on communication systems. In many distributed AI applications, intermediate representations rather than raw observations need to be exchanged between spatially separated nodes. Since these representations can be high-dimensional and heterogeneous in their relevance to the downstream task, transmitting them without considering communication constraints may incur substantial overhead and expose task-critical information to channel degradation. This has motivated increasing interest in communication-efficient and task-oriented transmission, where learned representations are adapted to communication constraints and downstream task requirements [1–5]. Recent studies further show that different semantic features and modalities contribute unequally to task performance, motivating importance-aware feature transmission and multimodal semantic communication [6–8]. These developments suggest that, for distributed AI systems, the central question is no longer simply how to transmit information reliably, but which information should be communicated and how the received information should be exploited by the downstream model.

Distributed medical imaging provides a representative setting in which this communication–inference coupling becomes particularly relevant. Magnetic resonance

imaging (MRI) routinely acquires multiple contrasts that provide complementary anatomical and tissue information [9, 10]. However, acquiring multiple fully sampled contrasts increases examination time, motivating accelerated multi-contrast reconstruction in which a fully sampled auxiliary contrast assists the recovery of an undersampled target contrast [11, 12]. When acquisition, storage, and reconstruction are distributed across different imaging or computing nodes, the auxiliary contrast may not be locally available at the reconstruction site. Existing multi-site MRI studies have considered distributed reconstruction and heterogeneous contrast availability [13–15], but the communication required to deliver auxiliary information to the reconstruction node has received much less attention. In such settings, reconstruction quality is jointly determined by the information retained for transmission, the channel condition, and the way the received auxiliary representation is incorporated into the local target reconstruction.

Meanwhile, substantial progress has been made in exploiting auxiliary information once both MRI contrasts are locally available. Early methods incorporated cross-modal priors through image- or feature-level fusion [16–19]. MTrans introduced Transformer-based interaction to capture long-range cross-modal dependencies [20], while DUN-SA and FEFA addressed inter-modal misalignment through spatial and deep feature alignment [21, 22]. More recent approaches further model complementary information across spatial and frequency domains. FSMNet selectively integrates auxiliary spatial and frequency features [12], MMR-Mamba employs target-guided interaction with spatial–frequency fusion [23], and A^{2}MC-MRI incorporates anatomy-aware priors into a deep-unrolling framework [24]. Despite their effectiveness, these methods primarily address how auxiliary information should be fused after it is already available at the reconstruction site. The cost and reliability of delivering that information are generally outside their formulation.

This separation between communication and reconstruction becomes problematic when the auxiliary modality is remotely located. Directly transmitting complete images or full-dimensional intermediate features can be communication-intensive, whereas aggressive compression may discard task-relevant information. Moreover, even after transmission, not all received auxiliary information is equally useful to the current target reconstruction. Existing semantic and task-oriented communication methods have demonstrated the value of compact and importance-aware representation transmission [25, 26], but they are mainly designed for general image delivery or semantic inference and do not explicitly consider cross-modal reconstruction. Multi-modal MRI reconstruction therefore poses a coupled problem with two complementary requirements: the auxiliary representation should be compactly encoded according to the available communication resources before transmission, and the received information should be selectively utilized according to the current target representation after transmission. This motivates jointly considering these two aspects for communication-aware distributed multi-modal MRI reconstruction.

To address this problem, we propose the Communication-aware Inter-modality Network (CINet), a communication-aware reconstruction framework that jointly considers auxiliary representation transmission and cross-modal reconstruction for distributed multi-modal MRI. CINet couples the communication process with the downstream reconstruction task instead of assuming unrestricted access to the auxiliary modality. Specifically, a Task-Oriented Communication Encoder (TCE) is introduced at the auxiliary node to reweight multi-scale auxiliary features and generate compact representations under a prescribed communication budget. These representations are transmitted through a noisy channel and recovered at the reconstruction node by a Feature Decoder (FD). A Target-Conditioned Selective Fusion (TSF) module then incorporates the received auxiliary information according to the current target representation. In this way, TCE controls the representation delivered

through the communication link, while TSF determines how the received information contributes to reconstruction, enabling joint consideration of communication constraints and cross-modal information utilization. Experiments on multi-modal prostate MRI show that CINet achieves competitive or superior reconstruction performance under different undersampling patterns and acceleration factors. Further evaluations under varying communication budgets and channel SNRs characterize the communication–reconstruction trade-off under different transmission conditions.

The main contributions of this work are summarized as follows:

- We formulate a communication-constrained multi-modal MRI reconstruction setting in which the fully sampled auxiliary modality resides at a separate auxiliary node rather than being directly accessible at the reconstruction node. This formulation explicitly considers the communication budget and channel condition when exploiting cross-modal information for target reconstruction.
- We propose CINet with complementary pre- and post-transmission mechanisms for auxiliary information transfer. TCE reweights auxiliary features and generates compact representations under a prescribed communication budget, while TSF selectively incorporates the received auxiliary information according to the current target representation. Together, they integrate compact auxiliary representation transmission with target-conditioned cross-modal reconstruction within a unified framework.
- We conduct extensive experiments on multi-modal prostate MRI under different undersampling patterns and acceleration factors. CINet consistently outperforms the compared reconstruction methods, while evaluations under varying communication budgets and channel SNRs further demonstrate its effectiveness and robustness under communication constraints.

2 Methods

2.1 Problem Formulation and System Model

Given an undersampled target-modality image x_T^u and a fully sampled auxiliary-modality image x_A, multi-modal MR reconstruction aims to recover the fully sampled target image x_T by exploiting information from the auxiliary contrast. In this work, T2-weighted (T2w) MRI is considered as the target modality and T1-weighted (T1w) MRI as the auxiliary modality. Existing multi-modal reconstruction methods commonly assume that both contrasts are locally available at the reconstruction site. However, in practical clinical imaging systems, image acquisition, storage, and reconstruction resources can be distributed across scanners, departments, clinical institutions, or edge/cloud computing nodes. Distributed and cloud-based MRI reconstruction has been explored to move computationally demanding reconstruction beyond individual scanners [27], while multi-institutional MRI scenarios may involve modality-specific data distributed across different clinical sites [15]. We therefore consider a distributed multi-modal reconstruction setting in which the auxiliary contrast is available at an auxiliary node, whereas the undersampled target acquisition is locally available at a reconstruction node. The auxiliary information must consequently be communicated between the two nodes before it can assist target reconstruction, making the available communication resources and channel condition part of the reconstruction problem. The resulting system is illustrated in Fig. 1.

At the auxiliary node, x_A is encoded into three-scale feature representations $\{F_A^l\}_{l=1}^3$, where

$$F_A^l \in \mathbb{R}^{C_l \times H_l \times W_l}, \tag{1}$$

and C_l, H_l, and W_l denote the number of channels, height, and width at scale l, respectively. Direct transmission of the complete intermediate representations can

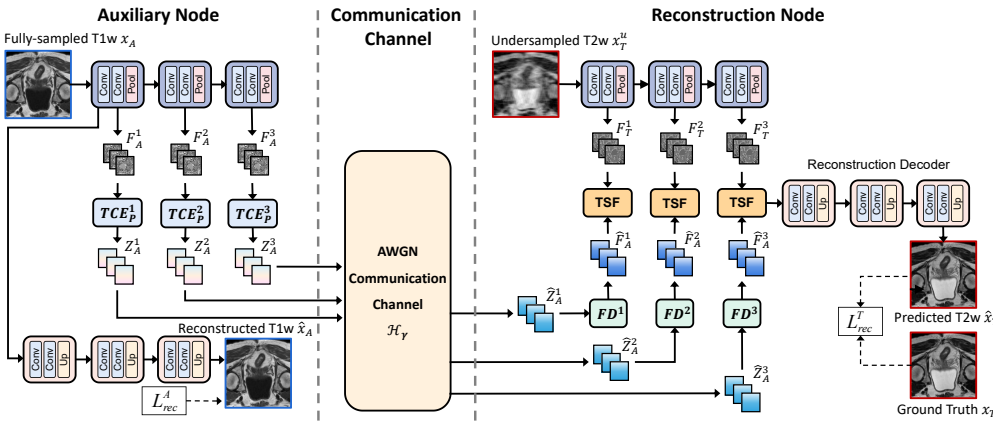


Fig. 1 Overview of the proposed CINet. The fully sampled auxiliary modality is encoded at the auxiliary node and transmitted as compact multi-scale representations through a noisy communication channel. At the reconstruction node, the received auxiliary features are recovered and selectively integrated with the undersampled target features for target reconstruction. TCE denotes the Task-Oriented Communication Encoder, FD the Feature Decoder, and TSF the Target-Conditioned Selective Fusion module.

incur substantial communication overhead due to their high dimensionality. To control the amount of information transmitted between the two nodes, we introduce a normalized communication-symbol budget $\rho \in (0, 1]$, implemented by controlling the channel dimension of the transmitted representation. At each scale, F_A^l is encoded into a compact latent representation

$$Z_A^l \in \mathbb{R}^{C_l' \times H_l \times W_l}, \tag{2}$$

where

$$C_l' = \max\left(1, \lfloor \rho C_l \rfloor\right). \tag{3}$$

Thus, $\rho = 1$ preserves the original channel dimension, whereas decreasing ρ reduces the number of transmitted feature channels and imposes a tighter communication constraint.

The compact representations are transmitted to the reconstruction node through a noisy wireless link. Following a point-to-point communication model, the link is modeled as an additive white Gaussian noise (AWGN) channel $\mathcal{H}_\gamma$, where γ denotes

the channel signal-to-noise ratio (SNR). The received latent representation at scale l is expressed as

$$\hat{Z}_A^l = \mathcal{H}_\gamma \left(Z_A^l \right), \tag{4}$$

where $\mathcal{H}_\gamma(\cdot)$ encompasses power normalization and AWGN corruption. At the reconstruction node, $\hat{Z}_A^l$ is decoded to obtain the recovered auxiliary feature $\hat{F}_A^l$. In parallel, the locally available x_T^u is encoded into three-scale target features $\{F_T^l\}_{l=1}^3$. The recovered auxiliary and local target representations are subsequently combined for target reconstruction.

The overall reconstruction process is formulated as

$$\hat{x}_T = \mathcal{R}_\Theta \left(x_T^u, \mathcal{H}_\gamma \left(\mathcal{E}_\rho(x_A) \right) \right), \tag{5}$$

where $\mathcal{E}_\rho(\cdot)$ denotes auxiliary representation encoding under communication budget ρ, $\mathcal{H}_\gamma(\cdot)$ models transmission under channel condition γ, and $\mathcal{R}_\Theta(\cdot)$ denotes the target reconstruction process parameterized by Θ. Unlike conventional multi-modal reconstruction, auxiliary information is not assumed to be freely accessible at the reconstruction node. The resulting problem therefore involves two coupled questions: which auxiliary information should be communicated under constrained communication resources, and how the received information should be utilized for target reconstruction.

2.2 Overall Architecture of CINet

An overview of CINet is shown in Fig. 1. The auxiliary and target encoders extract three-scale representations $\{F_A^l\}_{l=1}^3$ and $\{F_T^l\}_{l=1}^3$, respectively. CINet then performs auxiliary representation encoding, communication, feature recovery, and cross-modal fusion within a unified reconstruction framework.

At the auxiliary node, each F_A^l is processed by a Task-Oriented Communication Encoder (TCE), which reweights the auxiliary feature and compresses its channel dimension according to the communication budget ρ, producing the latent representation Z_A^l. The resulting representations are power-normalized and transmitted through $\mathcal{H}_\gamma$, yielding $\hat{Z}_A^l$ at the reconstruction node. A Feature Decoder (FD) restores the channel dimension from C_l' to C_l to obtain the recovered auxiliary feature $\hat{F}_A^l$. During training, an auxiliary decoder additionally reconstructs $\hat{x}_A$ from the auxiliary encoder features and is supervised by x_A, providing reconstruction supervision for the auxiliary branch.

At the reconstruction node, each $\hat{F}_A^l$ and the corresponding target feature F_T^l are fed into a Target-Conditioned Selective Fusion (TSF) module. TSF generates a target-conditioned selection gate from F_T^l and $\hat{F}_A^l$, and introduces the selected auxiliary information into the target branch through residual fusion, producing F_C^l. The fused representations are subsequently processed by the reconstruction decoder to obtain $\hat{x}_T$. In this way, TCE regulates the auxiliary representation before transmission, while TSF regulates the contribution of the received auxiliary information after transmission. Detailed formulations are presented in the following subsections.

2.3 Task-Oriented Communication Encoder

Direct transmission of the complete auxiliary representations requires communicating all feature channels, leading to unnecessary overhead under constrained communication resources. To obtain a compact representation tailored to the reconstruction task, we introduce the Task-Oriented Communication Encoder (TCE), which combines channel-wise feature reweighting with budget-controlled channel compression and is jointly optimized with the reconstruction network, as illustrated in Fig. 2(a).

Given the auxiliary feature $F_A^l \in \mathbb{R}^{C_l \times H_l \times W_l}$ at scale l, TCE first aggregates its spatial information through global average pooling (GAP). The resulting channel

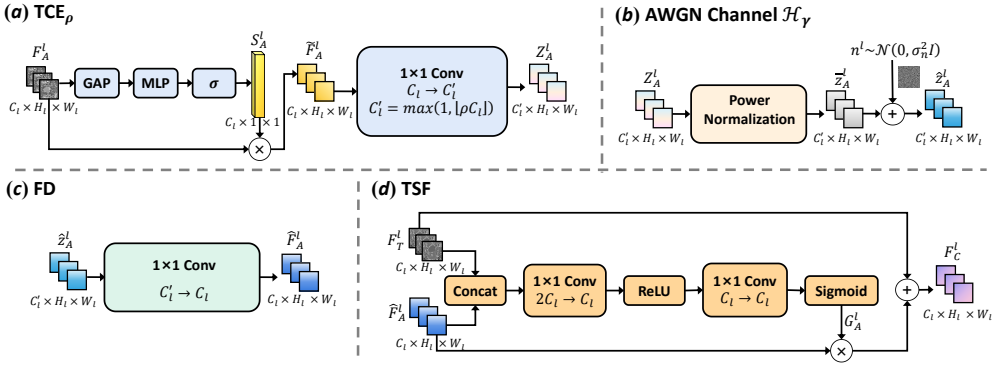


Fig. 2 Detailed architectures of the communication and fusion modules in CINet. (a) Task-Oriented Communication Encoder (TCE), (b) AWGN communication channel, (c) Feature Decoder (FD), and (d) Target-Conditioned Selective Fusion (TSF).

descriptor is passed through a multilayer perceptron (MLP) followed by a sigmoid function to obtain the channel-wise weights:

$$s_A^l = \sigma\left(\mathrm{MLP}\left(\mathrm{GAP}(F_A^l)\right)\right), \qquad s_A^l \in [0,1]^{C_l \times 1 \times 1}. \tag{6}$$

The auxiliary representation is reweighted accordingly:

$$\widetilde{F}_A^l = F_A^l \odot s_A^l, \tag{7}$$

where $\odot$ denotes element-wise multiplication with broadcasting over the spatial dimensions. The reweighted representation is subsequently projected by a 1×1 convolution into

$$Z_A^l = \phi_\rho^l\left(\widetilde{F}_A^l\right), \qquad Z_A^l \in \mathbb{R}^{C_l' \times H_l \times W_l}, \tag{8}$$

where $\phi_\rho^l(\cdot)$ denotes the channel projection from C_l to C_l'. According to Eq. (3), C_l' is determined by the communication budget ρ, thereby explicitly controlling the dimensionality of the representation transmitted at each scale.

2.4 Communication Channel and Feature Decoding

The compact representations $\{Z_A^l\}_{l=1}^3$ are transmitted to the reconstruction node through a noisy wireless channel. To satisfy an average transmit-power constraint, each Z_A^l is power-normalized before transmission, as illustrated in Fig. 2(b). We treat each element of Z_A^l as one real-valued channel symbol. Let

$$M_l = C_l' H_l W_l \tag{9}$$

denote the number of transmitted feature symbols at scale l. The normalized representation is given by

$$\bar{Z}_A^l = \sqrt{P} \frac{Z_A^l}{\sqrt{\frac{1}{M_l} \left\| Z_A^l \right\|_2^2 + \epsilon}}, \tag{10}$$

where P denotes the average transmit power and ϵ is a small constant for numerical stability. Accordingly, the normalized representation satisfies the average transmit-power constraint

$$\frac{1}{M_l} \left\| \bar{Z}_A^l \right\|_2^2 \leq P. \tag{11}$$

We model the wireless link as an additive white Gaussian noise (AWGN) channel. The received representation is expressed as

$$\hat{Z}_A^l = \bar{Z}_A^l + n^l, \qquad n^l \sim \mathcal{N}(0, \sigma_n^2 I), \tag{12}$$

where n^l denotes additive Gaussian noise with variance σ_n^2. The channel condition is characterized by the signal-to-noise ratio (SNR):

$$\gamma = \frac{P}{\sigma_n^2}, \qquad \gamma_{\mathrm{dB}} = 10 \log_{10} \gamma. \tag{13}$$

The communication budget ρ and channel SNR γ characterize two complementary aspects of the communication setting: the former controls the dimensionality of the transmitted representation, whereas the latter characterizes the channel noise level.

After transmission, $\hat{Z}_A^l$ remains in the compressed communication space with C_l' channels. To recover a feature representation compatible with the reconstruction branch, the Feature Decoder (FD) projects $\hat{Z}_A^l$ back to C_l channels through a 1×1 convolution, as illustrated in Fig. 2(c):

$$\hat{F}_A^l = D^l\left(\hat{Z}_A^l\right), \qquad \hat{F}_A^l \in \mathbb{R}^{C_l \times H_l \times W_l}, \tag{14}$$

where $D^l(\cdot)$ denotes the 1×1 channel projection from C_l' to C_l. The recovered auxiliary feature $\hat{F}_A^l$ is then combined with the corresponding local target feature F_T^l through the TSF module.

2.5 Target-Conditioned Selective Fusion

After communication and feature decoding, the recovered auxiliary representation $\hat{F}_A^l$ is available at the reconstruction node. To avoid uniformly incorporating the recovered auxiliary feature into the target branch, we introduce the Target-Conditioned Selective Fusion (TSF) module. At each scale l, TSF^l jointly uses the local target feature F_T^l and the recovered auxiliary feature $\hat{F}_A^l$ to generate a selection gate, as illustrated in Fig. 2(d).

Specifically, F_T^l and $\hat{F}_A^l$ are first concatenated along the channel dimension. The concatenated representation is projected from $2C_l$ to C_l channels by a 1×1 convolution followed by a ReLU activation. A second 1×1 convolution followed by a sigmoid function generates the selection gate

$$G_A^l = \sigma\left(W_2^l \delta\left(W_1^l\left[F_T^l, \hat{F}_A^l\right]\right)\right), \qquad G_A^l \in [0,1]^{C_l \times H_l \times W_l}, \tag{15}$$

where $[\cdot,\cdot]$ denotes channel-wise concatenation, $\delta(\cdot)$ denotes the ReLU activation, and $\sigma(\cdot)$ denotes the sigmoid function. Here, W_1^l and W_2^l denote the 1×1 convolutional mappings from $2C_l$ to C_l and from C_l to C_l, respectively.

The recovered auxiliary feature is modulated by the selection gate:

$$F_S^l = G_A^l \odot \hat{F}_A^l, \tag{16}$$

where $\odot$ denotes element-wise multiplication and F_S^l denotes the selected auxiliary feature. The selected feature is subsequently incorporated into the target representation through residual fusion:

$$F_C^l = F_T^l + F_S^l = F_T^l + G_A^l \odot \hat{F}_A^l. \tag{17}$$

The resulting F_C^l retains the target representation through the residual path while selectively incorporating the recovered auxiliary feature, and is subsequently passed to the reconstruction decoder.

2.6 Reconstruction Objective

CINet is trained end-to-end with reconstruction supervision on both the target and auxiliary branches. The target reconstruction loss measures the difference between the reconstructed T2w image $\hat{x}_T$ and its fully sampled ground truth x_T:

$$\mathcal{L}_{\text{rec}}^T = \mathbb{E}\left[\|x_T - \hat{x}_T\|_1\right]. \tag{18}$$

The auxiliary reconstruction branch is supervised in the same manner:

$$\mathcal{L}_{\text{rec}}^A = \mathbb{E}\left[\|x_A - \hat{x}_A\|_1\right], \tag{19}$$

where $\hat{x}_A$ denotes the reconstructed auxiliary image.

The overall reconstruction objective is formulated as

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{rec}}^{T} + \lambda_A \mathcal{L}_{\text{rec}}^{A}, \tag{20}$$

where λ_A balances the auxiliary reconstruction term relative to the target reconstruction objective. The entire framework is optimized end-to-end under this joint reconstruction objective.

3 Experiments

3.1 Experimental Setup

Dataset and preprocessing. We evaluate CINet on a multi-modal prostate MRI dataset comprising 7,300 registered T1w–T2w image pairs from 387 patients. The images were acquired using 1.5 T and 3 T scanners from multiple vendors, including Siemens, GE, and Philips. T1w and T2w images were registered using ITKElastix. Since the original field of view varied across acquisitions, the central regions of the slices were cropped to standardize the field of view. Image intensities were linearly scaled to $[-1, 1]$, and all slices were resized to 256×256. The fully sampled T2w images were used as ground truth, while the corresponding T1w images served as the auxiliary modality. Five-fold cross-validation was performed in this work, and the results were averaged over the five folds.

Accelerated acquisitions were simulated by retrospectively undersampling the fully sampled T2w images in k-space. Quantitative evaluations were conducted under Cartesian and radial undersampling at acceleration factors $R = 3.3$ and 5. Additional qualitative results under Cartesian undersampling at $R = 10$ were included to examine reconstruction quality under more severe undersampling. The same undersampling configurations were used for all compared methods under each evaluation setting.

Implementation details. CINet is implemented in PyTorch and trained on two NVIDIA RTX 3090 GPUs. We use the Adam optimizer to train the network for 500 epochs. The learning rate is initialized to 1×10^{-3}, kept fixed for the first 100 epochs, and linearly decayed to zero over the remaining 400 epochs. The auxiliary reconstruction weight λ_A is set to 0.2.

Communication settings. Unless otherwise specified, the communication budget is set to $\rho = 0.5$, reducing the transmitted channel dimension to approximately 50% of the original auxiliary feature dimension, and the channel SNR is fixed at 10 dB during both training and testing. To evaluate the effect of the communication budget, we consider $\rho \in \{0.25, 0.5, 0.75, 1.0\}$ at a fixed SNR of 10 dB, with a separate model trained for each budget. Channel robustness is evaluated using the model trained at 10 dB by varying the test SNR over $\{0, 5, 10, 15, 20\}$ dB while fixing $\rho = 0.5$.

Evaluation metrics. Reconstruction quality is evaluated using normalized mean squared error (nMSE), structural similarity index measure (SSIM), and peak signal-to-noise ratio (PSNR), consistent with the evaluation protocol used for this dataset. Lower nMSE and higher SSIM and PSNR indicate better reconstruction quality. All metrics are computed between the reconstructed T2w images and their corresponding fully sampled ground truth.

3.2 Comparison with State-of-the-Art Methods

Quantitative comparison. We compare CINet with representative reconstruction baselines and cross-modal MR image restoration methods adapted to the target reconstruction task. The comparison includes ReconCNN, Multi-pix2pix [28], Multi-pGAN [29], ReconGAN, Multi-ReconGAN, Hi-Net [30], DAGAN [31], SANet [32], and MD-DAN [33]. Recent multi-modal and multi-contrast MRI reconstruction methods, including MTrans [20], FEFA [22], FSMNet [12], MMR-Mamba [23], A^{2}MC-MRI [24], DAINet [34], and SGF-MRI [35], are further included for comparison.

Table 1 Quantitative results under Cartesian undersampling. Results are reported at acceleration factors $R = 3.3$ and 5. The best results are shown in bold.

Method	$R = 3.3$			$R = 5$		
	nMSE↓	SSIM↑	PSNR↑	nMSE↓	SSIM↑	PSNR↑
ReconCNN	1.073	0.884	29.802	1.617	0.833	27.922
Multi-pix2pix [28]	1.599	0.858	28.014	1.850	0.840	27.456
Multi-pGAN [29]	1.544	0.868	28.260	1.690	0.850	27.802
ReconGAN	1.090	0.884	29.772	3.267	0.745	24.882
Multi-ReconGAN	0.780	0.919	31.316	1.076	0.894	29.873
Hi-Net [30]	0.772	0.922	31.339	1.059	0.895	29.899
DAGAN [31]	1.144	0.916	30.748	1.937	0.876	28.675
SANet [32]	0.820	0.914	31.073	1.197	0.884	29.454
MD-DAN [33]	0.836	0.915	31.395	1.056	0.885	29.656
MTrans [20]	0.791	0.887	30.120	1.087	0.879	29.173
FEFA [22]	0.802	0.894	30.374	1.043	0.887	29.386
FSMNet [12]	0.753	0.913	30.589	0.988	0.896	29.807
MMR-Mamba [23]	0.760	0.906	31.062	1.012	0.893	29.594
A^2MC-MRI [24]	0.747	0.918	30.793	0.970	0.896	30.021
DAINet [34]	0.751	0.921	31.447	0.963	0.903	30.156
SGF-MRI [35]	0.738	0.926	31.836	0.951	0.906	30.219
CINet (Ours)	**0.732**	**0.930**	**31.921**	**0.937**	**0.907**	**30.238**

For consistent evaluation, all methods use the same data partitions, undersampling patterns, acceleration factors, and evaluation metrics described in Sec. 3.1. The compared methods are implemented or reproduced according to their original formulations and configured for T1w-assisted T2w reconstruction. Conventional multi-modal methods directly access the auxiliary modality during reconstruction, whereas CINet communicates auxiliary representations under the default setting of $\rho = 0.5$ and an SNR of 10 dB. Quantitative comparisons are conducted under Cartesian and radial undersampling at acceleration factors $R = 3.3$ and $R = 5$ using nMSE, SSIM, and PSNR.

Tables 1 and 2 report the quantitative results under Cartesian and radial undersampling, respectively. Under Cartesian undersampling, CINet achieves the best performance across all three metrics at both acceleration factors. At $R = 3.3$, CINet

Table 2 Quantitative results under Radial undersampling. Results are reported at acceleration factors $R = 3.3$ and 5. The best results are shown in bold.

Method	$R = 3.3$			$R = 5$		
	nMSE↓	SSIM↑	PSNR↑	nMSE↓	SSIM↑	PSNR↑
ReconCNN	1.069	0.868	29.752	1.838	0.790	27.281
Multi-pix2pix [28]	1.687	0.844	27.838	2.042	0.810	26.995
Multi-pGAN [29]	1.472	0.854	28.432	1.901	0.820	27.244
ReconGAN	1.472	0.832	28.424	1.953	0.785	27.066
Multi-ReconGAN	0.879	0.901	30.698	1.217	0.874	29.232
Hi-Net [30]	0.876	0.902	30.728	1.164	0.874	29.326
DAGAN [31]	1.152	0.907	30.777	2.157	0.849	27.905
SANet [32]	0.970	0.899	30.502	1.210	0.869	28.949
MD-DAN [33]	0.792	0.901	31.067	1.164	0.864	29.200
MTrans [20]	0.971	0.895	29.549	1.282	0.887	29.464
FEFA [22]	0.941	0.901	29.648	1.156	0.889	29.618
FSMNet [12]	0.871	0.913	30.541	1.098	0.895	29.796
MMR-Mamba [23]	0.903	0.907	30.286	1.066	0.892	29.752
A^2MC-MRI [24]	0.796	0.919	30.892	1.041	0.891	29.813
DAINet [34]	0.759	0.925	31.347	1.037	0.904	29.956
SGF-MRI [35]	0.728	**0.932**	**31.706**	1.031	0.907	29.974
CINet (Ours)	**0.723**	0.930	31.681	**1.025**	**0.909**	**29.993**

obtains an nMSE of 0.732, an SSIM of 0.930, and a PSNR of 31.921 dB. At the more challenging $R = 5$, CINet maintains the best overall performance with an nMSE of 0.937, an SSIM of 0.907, and a PSNR of 30.238 dB. Compared with SGF-MRI, the strongest competing method under this setting, CINet reduces nMSE from 0.951 to 0.937 and improves SSIM and PSNR from 0.906 and 30.219 dB to 0.907 and 30.238 dB, respectively.

Under radial undersampling, CINet remains competitive across both acceleration factors. At $R = 3.3$, CINet achieves the lowest nMSE of 0.723 and obtains an SSIM of 0.930 and a PSNR of 31.681 dB, ranking second on the latter two metrics behind SGF-MRI. At $R = 5$, CINet achieves the best performance across all three metrics, with an nMSE of 1.025, an SSIM of 0.909, and a PSNR of 29.993 dB. Compared with recent multi-modal reconstruction methods such as MTrans, FSMNet, MMR-Mamba, and

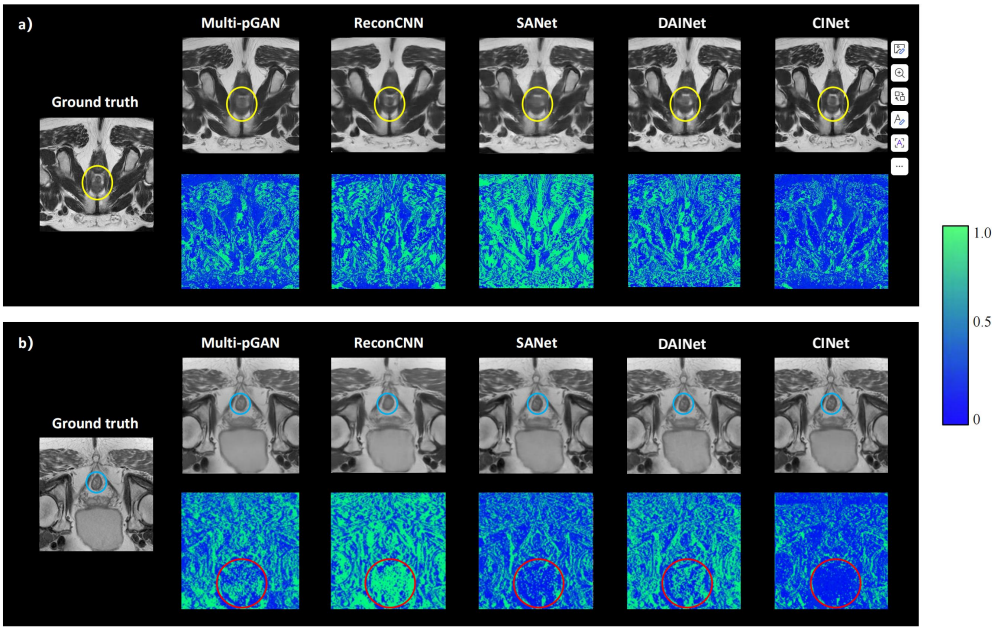


Fig. 3 Qualitative results under Cartesian undersampling at $R = 10$**.** Reconstructed T2w images and corresponding difference maps are shown for two representative examples.

A^{2}MC-MRI, CINet maintains favorable reconstruction performance as the acceleration factor increases. These results indicate that CINet can effectively exploit the communicated auxiliary representations despite the communication constraints imposed on the auxiliary branch.

Qualitative comparison. To complement the quantitative evaluation, Figs. 3–6 compare CINet with Multi-pGAN [29], ReconCNN, SANet [32], and DAINet [34]. All methods are evaluated using the same test images, undersampling settings, regions of interest, and display ranges. The reconstructed images are presented together with difference maps and enlarged regions to examine both reconstruction discrepancies and local structural preservation.

Figs. 3 and 5 show the reconstructed images and corresponding difference maps under Cartesian undersampling at $R = 10$ and radial undersampling at $R = 3.3$, respectively. ReconCNN exhibits pronounced residual responses in several highlighted

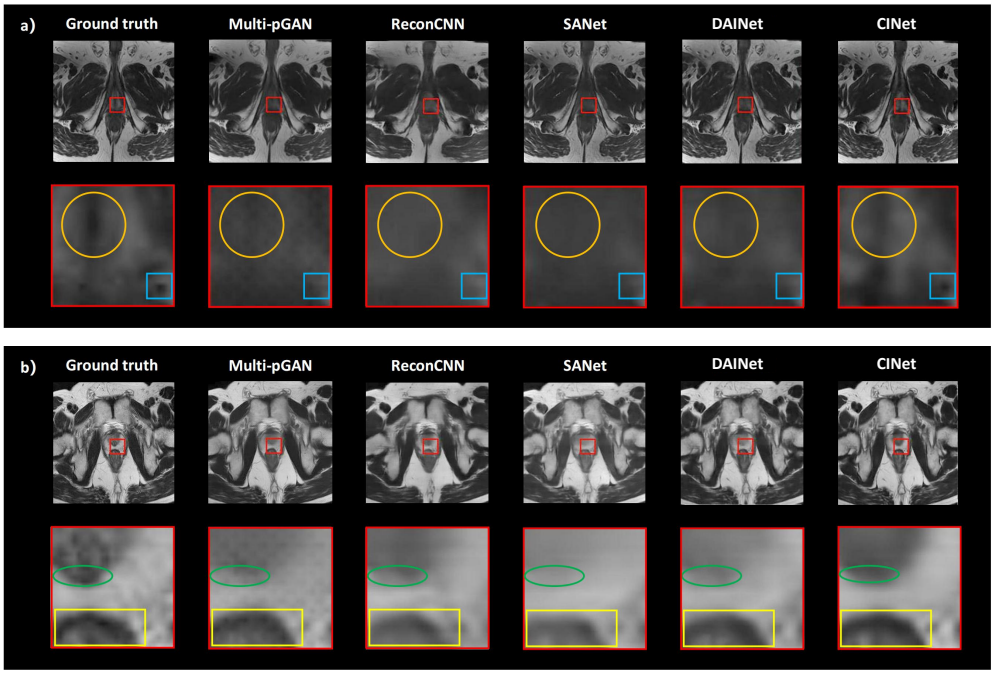


Fig. 4 Local details under Cartesian undersampling at $R = 10$**.** Reconstructed images and enlarged regions are shown for two representative examples.

regions, reflecting the difficulty of recovering local structures solely from the undersampled target modality. The multi-modal methods generally reduce these discrepancies by incorporating auxiliary information, although residual structures remain visible for Multi-pGAN, SANet, and DAINet. In comparison, CINet exhibits weaker residual responses in the highlighted regions and more closely reproduces the corresponding structures in the fully sampled reference.

The enlarged regions in Figs. 4 and 6 provide a complementary view of local reconstruction quality. ReconCNN tends to produce smoother local appearances, while Multi-pGAN retains visible blurring around several highlighted structures. SANet and DAINet recover more structural information but still exhibit discrepancies around subtle boundaries and intensity transitions. CINet more closely preserves these local structures and weak intensity variations. Together with the difference maps, these

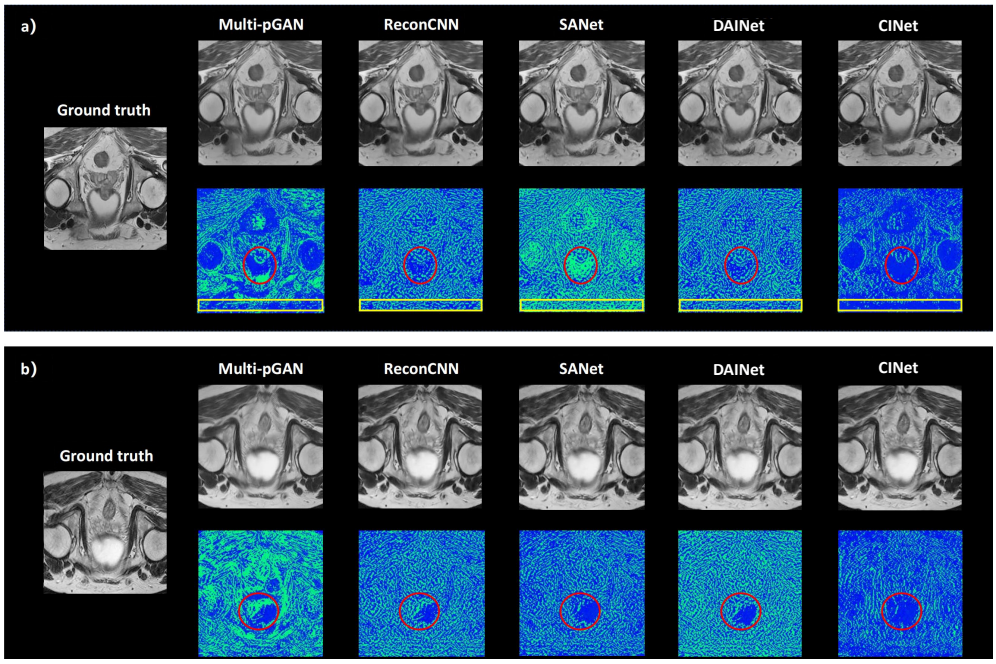


Fig. 5 Qualitative results under radial undersampling at $R = 3.3$**.** Reconstructed T2w images and corresponding difference maps are shown for two representative examples.

observations support the quantitative results and show that the communicated auxiliary representations remain effective for recovering local structural information under different undersampling patterns.

3.3 Performance under Communication Constraints

We further evaluate CINet under different communication budgets and channel conditions. All experiments are conducted under Cartesian undersampling at $R = 5$. For the communication-budget experiments, the channel SNR is fixed at 10 dB and the communication budget ρ is varied from 0.25 to 1.0. For the channel-SNR experiments, ρ is fixed at 0.5, and the model trained at an SNR of 10 dB is evaluated at different test SNRs. The results are reported in Table 3.

Effect of communication budget. The left part of Table 3 shows the reconstruction performance under different values of ρ. Increasing ρ from 0.25 to 0.50 substantially

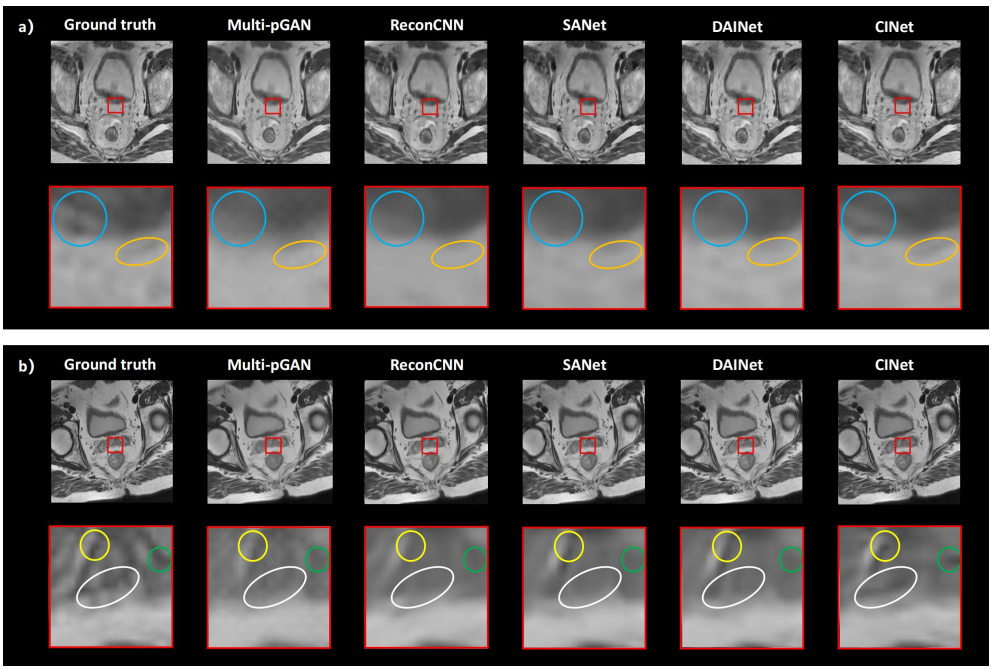


Fig. 6 Local details under radial undersampling at $R = 3.3$. Reconstructed images and enlarged regions are shown for two representative examples.

Table 3 Performance under different communication conditions. Communication-budget experiments use an SNR of 10 dB, while channel-SNR experiments use $\rho = 0.5$ with a model trained at 10 dB.

Communication Budget				**Channel SNR**			
ρ	nMSE↓	SSIM↑	PSNR↑	SNR (dB)	nMSE↓	SSIM↑	PSNR↑
0.25	1.204	0.889	29.157	0	1.146	0.895	28.914
0.50	1.025	0.909	29.993	5	1.098	0.903	29.562
0.75	1.020	0.912	30.002	10	1.025	0.909	29.993
1.00	1.019	0.913	30.008	15	1.021	0.912	30.021
				20	1.017	0.914	30.045

improves all three metrics, reducing nMSE from 1.204 to 1.025, while increasing SSIM from 0.889 to 0.909 and PSNR from 29.157 to 29.993 dB. Further increasing ρ produces only marginal improvements. In particular, increasing ρ from 0.50 to 1.00 improves PSNR by only 0.015 dB and SSIM by 0.004, while nMSE decreases from 1.025 to 1.019. These results indicate that severe compression at $\rho = 0.25$ removes useful auxiliary information, whereas $\rho = 0.50$ already retains most of the information required for

target reconstruction. We therefore adopt $\rho = 0.50$ as the default communication budget, providing a favorable balance between transmitted representation dimensionality and reconstruction performance.

Effect of channel SNR. The right part of Table 3 evaluates the sensitivity of CINet to channel conditions. The model is trained at 10 dB and evaluated without retraining at SNRs ranging from 0 to 20 dB. Reconstruction performance improves consistently as the test SNR increases. Specifically, PSNR increases from 28.914 dB at 0 dB to 29.993 dB at 10 dB, while SSIM increases from 0.895 to 0.909 and nMSE decreases from 1.146 to 1.025. The improvement becomes considerably smaller above the training SNR, with PSNR reaching 30.021 and 30.045 dB at 15 and 20 dB, respectively. This behavior is consistent with the reduced perturbation of the transmitted auxiliary representations at higher SNRs and suggests that reconstruction performance gradually approaches saturation as the channel condition improves.

Taken together, the two experiments characterize complementary aspects of communication-constrained reconstruction. The communication budget controls the dimensionality of the auxiliary representation available for transmission, whereas the channel SNR determines the level of noise affecting the transmitted representation. CINet maintains stable reconstruction performance once a moderate communication budget and channel quality are available, supporting the use of $\rho = 0.50$ and an SNR of 10 dB as the default communication setting in the remaining experiments.

3.4 Ablation Study

We conduct ablation experiments to evaluate the contributions of the main components of CINet, including TCE, TSF, and auxiliary reconstruction supervision $\mathcal{L}_{\text{rec}}^{A}$. All variants are evaluated under Cartesian undersampling at $R = 5$ with the default communication setting of $\rho = 0.5$ and an SNR of 10 dB. The results are reported in Table 4.

Table 4 Ablation results of the main CINet components. All variants are evaluated under Cartesian undersampling at $R = 5$, with $\rho = 0.5$ and an SNR of 10 dB.

Configuration	TCE	TSF	$\mathcal{L}_{\text{rec}}^{A}$	nMSE↓	SSIM↑	PSNR↑
Direct fusion	×	×	×	1.590	0.843	28.046
+ TCE	✓	×	×	1.237	0.885	29.072
+ TSF	×	✓	×	1.368	0.880	28.983
+ TCE + TSF	✓	✓	×	1.061	0.903	29.951
CINet	✓	✓	✓	**1.025**	**0.909**	**29.993**

For the direct-fusion configuration, the auxiliary transmission pipeline is retained, including dimensional projection, channel transmission, and feature decoding, while the channel-wise reweighting in TCE, the selective fusion in TSF, and auxiliary reconstruction supervision are removed. The recovered auxiliary features are directly incorporated into the target branch through residual fusion. This configuration preserves the same communication budget and channel condition as the complete model, allowing the proposed components to be evaluated under consistent communication settings.

As shown in Table 4, both TCE and TSF improve reconstruction performance when introduced individually. Compared with direct fusion, TCE reduces nMSE from 1.590 to 1.237 and improves SSIM and PSNR from 0.843 and 28.046 dB to 0.885 and 29.072 dB, respectively. Similarly, introducing TSF yields an nMSE of 1.368, an SSIM of 0.880, and a PSNR of 28.983 dB. These results indicate that both pre-transmission feature regulation and target-conditioned utilization of the received auxiliary information contribute to reconstruction performance.

Combining TCE and TSF further improves the results to an nMSE of 1.061, an SSIM of 0.903, and a PSNR of 29.951 dB, outperforming either component alone across all three metrics. This consistent improvement supports the complementary roles of processing the auxiliary representation before transmission and selectively utilizing it after reception. Finally, introducing auxiliary reconstruction supervision further improves the three metrics to 1.025, 0.909, and 29.993 dB, respectively. Although the additional gain is relatively modest, the consistent improvement across all metrics

indicates that auxiliary reconstruction supervision provides further benefit when TCE and TSF are jointly employed.

4 Discussion

The experimental results demonstrate that auxiliary-modality information can remain effective for multi-modal MRI reconstruction even when it is not directly available at the reconstruction node. Existing multi-modal reconstruction methods typically assume that the fully sampled auxiliary modality can be directly accessed and integrated with the undersampled target modality [20, 23]. CINet relaxes this assumption by transmitting compact auxiliary representations between distributed nodes before cross-modal fusion. Despite the communication constraint, CINet consistently outperforms the compared reconstruction methods under different undersampling patterns and acceleration factors. The qualitative results further show reduced reconstruction discrepancies and improved preservation of local structural details. These findings suggest that transmitting task-relevant intermediate representations can provide an effective alternative to directly transferring or locally accessing the complete auxiliary modality.

The communication experiments provide further insight into the deployment of CINet under constrained communication conditions. Varying ρ characterizes the relationship between the dimensionality of the transmitted representation and reconstruction quality, while the SNR experiments examine the sensitivity of reconstruction to channel degradation. Similar trade-offs between transmitted representation size and reconstruction fidelity have been investigated in learned image communication systems, where bandwidth and channel conditions directly affect reconstruction quality [4, 5]. Unlike channel-adaptive communication approaches that explicitly condition the model on channel states [4, 5], CINet focuses on task-oriented auxiliary representation transmission and target-conditioned utilization for MRI reconstruction.

This formulation is particularly relevant to distributed imaging scenarios in which acquisition, storage, and reconstruction resources are separated across scanners or computing nodes. Previous studies have demonstrated the feasibility of distributed and cloud-based MRI reconstruction by separating image acquisition from computationally demanding reconstruction [27]. CINet extends this perspective by considering how complementary information from another MR contrast can be communicated and utilized when it is not locally available at the reconstruction node.

Several limitations should be acknowledged. First, the current communication link is modeled as an AWGN channel. Although this controlled model allows the effects of channel noise to be studied independently, practical communication systems may additionally involve fading, interference, packet loss, and latency. Second, the current experiments assume spatially registered T1w and T2w images. Residual inter-modal misalignment may affect the reliability of auxiliary information and should be considered in more heterogeneous imaging environments. Third, the communication budget ρ controls the dimensionality of the transmitted feature representation rather than an explicit bit rate; quantization, entropy coding, and protocol overhead are therefore beyond the scope of the present study. Finally, the current evaluation is conducted on paired prostate MRI data. Further validation across additional anatomies, acquisition protocols, and multi-institutional datasets will be necessary to assess the generalizability of the framework. Future work will investigate more realistic channel models, explicit rate-constrained transmission, robustness to inter-modal misalignment, and broader clinical validation.

5 Conclusion

This paper presented CINet, a communication-aware framework for multi-modal MRI reconstruction in distributed imaging scenarios. Unlike conventional approaches that assume direct access to the fully sampled auxiliary modality, CINet communicates

compact auxiliary representations through a noisy channel and exploits the received information to assist target reconstruction. TCE selectively reweights and compresses multi-scale auxiliary features before transmission, while TSF adaptively incorporates the recovered auxiliary information according to the current target representation. Experiments on multi-modal prostate MRI show that CINet outperforms the compared reconstruction methods across different undersampling patterns and acceleration factors, while preserving fine structural details in the reconstructed images. Further experiments under varying communication budgets and channel SNRs demonstrate the effectiveness of CINet under communication constraints. Overall, these results establish the feasibility of communication-aware cross-modal information transfer for distributed multi-modal MRI reconstruction.

Abbreviations

MRI	Magnetic resonance imaging
T1w	T1-weighted
T2w	T2-weighted
CINet	Communication-aware Inter-modality Network
TCE	Task-Oriented Communication Encoder
FD	Feature Decoder
TSF	Target-Conditioned Selective Fusion
AWGN	Additive white Gaussian noise
SNR	Signal-to-noise ratio
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity index measure
nMSE	Normalized mean squared error
GAP	Global average pooling
MLP	Multilayer perceptron

Declarations

Acknowledgements

Not applicable.

Author contributions

SY contributed to conceptualization, methodology, software, formal analysis, investigation, validation, visualization, and writing of the original draft. YB contributed to validation and writing, review and editing. ZZ contributed to formal analysis, investigation, validation, and writing of the original draft. BZ contributed to methodology, supervision, funding acquisition, and writing, review and editing. WDW contributed to conceptualization, supervision, project administration, funding acquisition, and writing, review and editing. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Natural Science Foundation of China (No. 62676045) and the Beijing Natural Science Foundation (No. L182034).

Data availability

The data used in this study are not publicly available due to privacy and institutional restrictions. The de-identified data may be available from the corresponding author upon reasonable request and subject to approval by the relevant institution.

Ethics approval and consent to participate

The study used retrospective, fully de-identified clinical MRI data. According to the institutional policy of the participating institution, additional ethical approval and informed consent were not required for this secondary analysis of de-identified data.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

References

[1] E. Bourtsoulatze, D.B. Kurka, D. Gündüz, Deep joint source-channel coding for wireless image transmission. IEEE Transactions on Cognitive Communications and Networking **5**(3), 567–579 (2019). https://doi.org/10.1109/TCCN.2019.2919300

[2] H. Xie, Z. Qin, G.Y. Li, B.H. Juang, Deep learning enabled semantic communication systems. IEEE Transactions on Signal Processing **69**, 2663–2675 (2021). https://doi.org/10.1109/TSP.2021.3071210

[3] D.B. Kurka, D. Gündüz, Bandwidth-agile image transmission with deep joint source-channel coding. IEEE Transactions on Wireless Communications **20**(12), 8081–8095 (2021). https://doi.org/10.1109/TWC.2021.3090048

[4] C. Bian, Y. Shao, D. Gündüz, *DeepJSCC-l++: Robust and Bandwidth-Adaptive Wireless Image Transmission*, in *2023 IEEE Global Communications Conference (GLOBECOM)* (2023), pp. 3148–3154. https://doi.org/10.1109/GLOBECOM54140.2023.10436878

[5] H. Wu, Y. Shao, C. Bian, K. Mikolajczyk, D. Gündüz, Deep joint source-channel coding for adaptive image transmission over mimo channels. IEEE Transactions on Wireless Communications **23**(10), 15002–15017 (2024). https://doi.org/10.1109/TWC.2024.3422794

[6] G. Zhang, Q. Hu, Z. Qin, Y. Cai, G. Yu, X. Tao, A unified multi-task semantic communication system for multimodal data. IEEE Transactions on Communications **72**(7), 4101–4116 (2024). https://doi.org/10.1109/TCOMM.2024.3364990

[7] Y. Zhou, P. Dai, B. He, M. Ren, M. Luo, J. Chen, Joint optimization of adaptive semantic compression and resource allocation for semantic-enabled uav communications. Journal on Wireless Communications and Networking **2026**, 93 (2026). https://doi.org/10.1186/s13638-026-02637-2

[8] H. Wei, J. Lang, Adaptive cross-modal uav semantic communications via hierarchical drl. Journal on Wireless Communications and Networking (2026). https://doi.org/10.1186/s13638-026-02664-z

[9] C. Muñoz, et al., Latest advances in image acceleration: All dimensions are fair game. Journal of Magnetic Resonance Imaging **57**(2), 387–402 (2023). https://doi.org/10.1002/jmri.28462

[10] B. Shafieizargar, R. Byanju, J. Sijbers, Systematic review of reconstruction techniques for accelerated quantitative mri. Magnetic Resonance in Medicine **90**(3), 1172–1208 (2023). https://doi.org/10.1002/mrm.29721

[11] S. Kim, H. Park, S.H. Park, A review of deep learning-based reconstruction methods for accelerated mri using spatiotemporal and multi-contrast redundancies. Biomedical Engineering Letters **14**(6), 1221–1242 (2024). https://doi.org/10.1007/s13534-024-00425-9

[12] Q. Chen, X. Xing, Z. Chen, Z. Xiong, *Accelerated Multi-Contrast MRI Reconstruction via Frequency and Spatial Mutual Learning*, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. 15007

(Springer Nature Switzerland, 2024), pp. 56–66. https://doi.org/10.1007/978-3-031-72104-5_6

[13] G. Elmas, S.U.H. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Ozbey, S. Avestimehr, T. Cukur, Federated learning of generative image priors for mri reconstruction. IEEE Transactions on Medical Imaging **42**(7), 1996–2009 (2023). https://doi.org/10.1109/TMI.2022.3220757

[14] O. Dalmaz, M.U. Mirza, G. Elmas, M. Ozbey, S.U.H. Dar, E. Ceyani, K.K. Oguz, S. Avestimehr, T. Cukur, One model to unite them all: Personalized federated learning of multi-contrast mri synthesis. Medical Image Analysis **94**, 103121 (2024). https://doi.org/10.1016/j.media.2024.103121

[15] Y. Yan, H. Wang, Y. Huang, N. He, L. Zhu, Y. Xu, Y. Li, Y. Zheng, Cross-modal vertical federated learning for mri reconstruction. IEEE Journal of Biomedical and Health Informatics **28**(11), 6384–6394 (2024). https://doi.org/10.1109/JBHI.2024.3360720

[16] L. Xiang, Y. Chen, W. Chang, Y. Zhan, W. Lin, Q. Wang, D. Shen, Deep-learning-based multi-modal fusion for fast mr reconstruction. IEEE Transactions on Biomedical Engineering **66**(7), 2105–2114 (2019). https://doi.org/10.1109/TBME.2018.2883958

[17] S.U.H. Dar, M. Yurt, M. Shahdloo, M.E. Ildiz, B. Tinaz, T. Cukur, Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. IEEE Journal of Selected Topics in Signal Processing **14**(6), 1072–1087 (2020). https://doi.org/10.1109/JSTSP.2020.3001737

[18] L. Sun, Z. Fan, X. Fu, Y. Huang, X. Ding, J. Paisley, A deep information sharing network for multi-contrast compressed sensing mri reconstruction. IEEE Transactions on Image Processing **28**(12), 6141–6153 (2019). https://doi.org/10.1109/TIP.2019.2925288

[19] X. Liu, J. Wang, H. Sun, S.S. Chandra, S. Crozier, F. Liu, On the regularization of feature fusion and mapping for fast mr multi-contrast imaging via iterative networks. Magnetic Resonance Imaging **77**, 159–168 (2021). https://doi.org/10.1016/j.mri.2020.12.019

[20] C.M. Feng, Y. Yan, G. Chen, Y. Xu, Y. Hu, L. Shao, H. Fu, Multimodal transformer for accelerated mr imaging. IEEE Transactions on Medical Imaging **42**(10), 2804–2816 (2023). https://doi.org/10.1109/TMI.2022.3180228

[21] H. Zhang, Q. Wang, J. Shi, S. Ying, Z. Wen, Deep unfolding network with spatial alignment for multi-modal mri reconstruction. Medical Image Analysis **99**, 103331 (2025). https://doi.org/10.1016/j.media.2024.103331

[22] X. Chen, L. Ma, S. Ying, D. Shen, T. Zeng, Fefa: Frequency enhanced multi-modal mri reconstruction with deep feature alignment. IEEE Journal of Biomedical and Health Informatics **28**(11), 6751–6763 (2024). https://doi.org/10.1109/JBHI.2024.3432139

[23] J. Zou, L. Liu, Q. Chen, S. Wang, Z. Hu, X. Xing, J. Qin, Mmr-mamba: Multi-modal mri reconstruction with mamba and spatial-frequency information fusion. Medical Image Analysis **102**, 103549 (2025). https://doi.org/10.1016/j.media.2025.103549

[24] Y. He, C. Lian, R. Xiao, F. Ju, C. Zou, Z. Xu, J. Ma, Anatomy-aware deep unrolling for task-oriented acceleration of multi-contrast mri. IEEE Transactions

on Medical Imaging **44**(9), 3832–3844 (2025). https://doi.org/10.1109/TMI.2025.3568157

[25] Y. Shi, L. Yu, Q. Zhu, Swinjscc-gan: Enhanced joint source-channel coding for semantic communication with swin transformer and generative adversarial networks. Journal on Wireless Communications and Networking **2026**, 8 (2026). https://doi.org/10.1186/s13638-025-02547-9

[26] S. Inokuma, G. Sawada, D. Hisano, K. Maruta, Image value-based dynamic spatial stream allocation in deep joint source-channel-coded mimo transmission for extreme environment. Journal on Wireless Communications and Networking **2026**, 92 (2026). https://doi.org/10.1186/s13638-026-02633-6

[27] H. Xue, S. Inati, T.S. Sørensen, P. Kellman, M.S. Hansen, Distributed mri reconstruction using gadgetron-based cloud computing. Magnetic Resonance in Medicine **73**(3), 1015–1025 (2015). https://doi.org/10.1002/mrm.25213

[28] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, *Image-to-Image Translation with Conditional Adversarial Networks*, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 1125–1134. https://doi.org/10.1109/CVPR.2017.632

[29] S.U.H. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, T. Cukur, Image synthesis in multi-contrast mri with conditional generative adversarial networks. IEEE Transactions on Medical Imaging **38**(10), 2375–2388 (2019). https://doi.org/10.1109/TMI.2019.2901750

[30] T. Zhou, H. Fu, G. Chen, J. Shen, L. Shao, Hi-net: Hybrid-fusion network for multi-modal mr image synthesis. IEEE Transactions on Medical Imaging **39**(9), 2772–2781 (2020). https://doi.org/10.1109/TMI.2020.2975344

[31] G. Yang, S. Yu, H. Dong, G. Slabaugh, P.L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, D. Firmin, Dagan: Deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. IEEE Transactions on Medical Imaging **37**(6), 1310–1321 (2018). https://doi.org/10.1109/TMI.2017.2785879

[32] C.M. Feng, Y. Yan, K. Yu, Y. Xu, H. Fu, J. Yang, L. Shao, Exploring separable attention for multi-contrast mr image super-resolution. IEEE Transactions on Neural Networks and Learning Systems **35**(9), 12251–12262 (2024). https://doi.org/10.1109/TNNLS.2023.3253557

[33] Y. Yang, N. Wang, H. Yang, J. Sun, Z. Xu, *Model-Driven Deep Attention Network for Ultra-Fast Compressive Sensing MRI Guided by Cross-Contrast MR Image*, in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (Springer, 2020), pp. 188–198. https://doi.org/10.1007/978-3-030-59713-9_19

[34] Z. Wang, A. Liu, J. Wei, G. Yang, Y. Yu, W. Wei, X. Chen, Dainet: Dual-alignment implicit network for accelerated multi-contrast mri reconstruction. Biomedical Signal Processing and Control **123**, 110571 (2026). https://doi.org/10.1016/j.bspc.2026.110571

[35] S. Zheng, S. Du, C. Qin, Sgf-mri: Structure guided fusion for multi-contrast mri super-resolution and reconstruction. Pattern Recognition **180**, 114188 (2026). https://doi.org/10.1016/j.patcog.2026.114188