

# Multi-modal MRI Translation via Evidential Regression and Distribution Calibration

Jiyao Liu<sup>1</sup>, Shangqi Gao<sup>2✉</sup>, Yuxin Li<sup>1</sup>, Lihao Liu<sup>2,3</sup>, Xin Gao<sup>1</sup>, Zhaohu Xing<sup>4</sup>, Junzhi Ning<sup>3</sup>, Yanzhou Su<sup>5</sup>, Xiao-Yong Zhang<sup>6</sup>, Junjun He<sup>3,7</sup>, Ningsheng Xu<sup>1</sup>,  
and Xiahai Zhuang<sup>1✉</sup>

<sup>1</sup> Fudan University, Shanghai, China

<sup>2</sup> University of Cambridge, Cambridge, United Kingdom

<sup>3</sup> Shanghai Artificial Intelligence Laboratory, Shanghai, China

<sup>4</sup> The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

<sup>5</sup> Fuzhou University, Fujian, China

<sup>6</sup> Shanghai Jiao Tong University, Shanghai, China

<sup>7</sup> Shanghai Innovation Institute, Shanghai, China

**Abstract.** Multi-modal Magnetic Resonance Imaging (MRI) translation leverages information from source MRI sequences to generate target modalities, enabling comprehensive diagnosis while overcoming the limitations of acquiring all sequences. While existing deep-learning-based multi-modal MRI translation methods have shown promising potential, they still face two key challenges: 1) lack of reliable uncertainty quantification for synthesized images, and 2) limited robustness when deployed across different medical centers. To address these challenges, we propose a novel framework that reformulates multi-modal MRI translation as a multi-modal evidential regression problem with distribution calibration. Our approach incorporates two key components: 1) an evidential regression module that estimates uncertainties from different source modalities and an explicit distribution mixture strategy for transparent multi-modal fusion, and 2) a distribution calibration mechanism that adapts to source-target mapping shifts to ensure consistent performance across different medical centers. Extensive experiments on three datasets from the BraTS2023 challenge demonstrate that our framework achieves superior performance and robustness across domains.

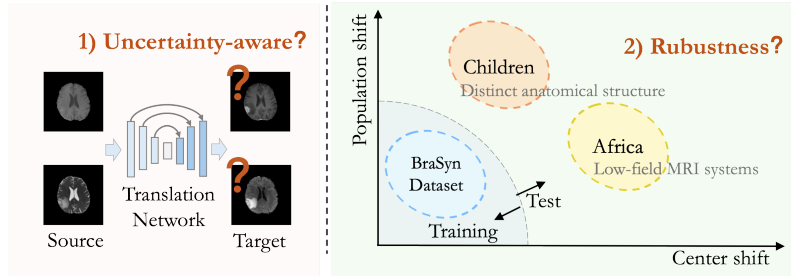
**Keywords:** Medical image translation · Uncertainty · Evidential regression · Calibration

## 1 Introduction

Magnetic resonance imaging (MRI) has become indispensable in modern clinical practice for its ability to provide detailed soft tissue visualization. Multi-modal

---

✉ corresponding authors: Shangqi Gao (sg2162@cam.ac.uk) and Xiahai Zhuang (zxh@fudan.edu.cn)

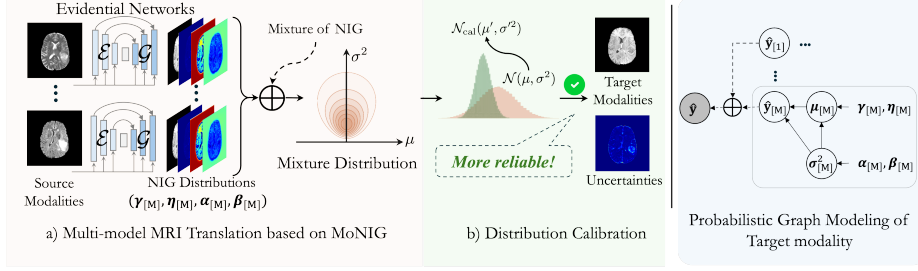


**Fig. 1.** Two challenges in multi-modal MRI translation: (1) lack of reliability assessment for synthesized images, and (2) limited robustness across domains.

MRI sequences, such as native T1-weighted (T1n), T2-weighted (T2w), T2-FLAIR (T2f), and post-contrast T1-weighted (T1c), each capture distinct tissue characteristics that together enable comprehensive diagnosis and treatment planning [2]. However, acquiring a complete set of sequences is often impractical due to various constraints including prolonged scanning time, high costs, and patient-specific limitations. Deep learning-based medical image translation [10, 12] has emerged as a promising solution by enabling the synthesis of missing modalities from available sequences. Recent advances in this field have demonstrated remarkable progress by developing multi-modal fusion techniques [4, 15]. Generative adversarial networks [23, 28] or diffusion models [20, 25] were employed to produce high-fidelity synthetic images. To achieve multi-modal integration, three key strategies: image-level concatenation [20], feature-level fusion [27, 28], and attention-based fusion mechanisms [7], were developed.

However, despite extensive studies, multi-modal MRI translation still faces two practical challenges (Fig. 1). First, most multi-modal methods lack mechanisms to assess the reliability of synthesized images, making it difficult to determine whether the generated results can be trusted by clinicians. Second, these methods are often sensitive to domain shifts when deployed across centers, where variations in scanning protocols and population characteristics can significantly impact model performance.

To evaluate the reliability of multi-modal learning, few works have focused on the uncertainty-guided fusion strategy [6, 8, 26]. Existing uncertainty estimation methods include Bayesian neural networks [5, 19] and deep ensembles [3], which often require complex modifications to network architectures or multiple forward passes. In contrast, evidential learning [1, 22] provides a more efficient and direct approach by learning to predict the parameters of probability distributions. To enhance the reliability of multi-modal fusion, the Mixture of Normal-Inverse Gamma (MoNIG) distribution [18] combined predictions from each source through an explicit uncertainty-aware fusion rule. To improve model robustness, recent studies have investigated calibration approaches [9, 13] in machine learning. However, their applicability in multi-modal MRI translation is still underexplored.



**Fig. 2.** Overview of the proposed framework. Through MoNIG-based image translation, we synthesize a target sequence from existing multi-modal source modalities based on explicit distribution fusion. Then, to adapt to the new source-target mapping, distribution calibration is performed to compute the calibrated generated target images and prediction uncertainty.

In this paper, we propose an uncertainty-aware multi-modal MRI translation framework based on evidential regression and distribution calibration. As shown in Fig. 2, we estimate the translation uncertainty of each modality through evidential networks. Based on uncertainties, we apply fusion rules derived from MoNIG to obtain the final generated results, thereby fully leveraging the multi-modal information in a reliable way. Additionally, we improve the robustness across domains through an efficient distribution calibration mechanism based on quantile regression, which enables rapid adaptation to new clinical environments with minimal data requirements.

Our main contributions include three aspects. First, we propose a novel probabilistic framework that reformulates multi-modal MRI translation as an evidential regression problem guided by uncertainty. Second, we develop a distribution calibration mechanism to adapt different domains for multi-modal fusion. Third, extensive experiments on three BraTS2023 datasets demonstrate that our approach delivers superior robustness across different centers.

## 2 Methodology

The aim of our method is to derive a distribution of target modality with uncertainty and utilize this uncertainty for explicit multi-modal fusion. As shown in Fig. 2, our framework mainly consists of two parts, *i.e.*, multi-modal MRI translation with uncertainty quantification and distribution calibration. In the multi-modal MRI translation part, each source modality first estimates the target modality distribution through an evidential network and then obtains a mixture distribution through the explicit MoNIG fusion strategy. Distribution calibration employs quantile regression to obtain the calibrated target distribution, from which the final generated target modality and uncertainty are derived. The details of our method are discussed in the following sections.

## 2.1 Uncertainty Estimation and Multi-modal Fusion

**Evidential Regression Module** Consider multi-modal medical image translation as a regression problem: given a dataset  $\mathcal{D} = \{\{\mathbf{x}_{[m]}\}_{m=1}^M, \mathbf{y}\}$ , where  $\mathbf{x}_{[m]}$  represents the  $m$ -th source modality and  $\mathbf{y}$  corresponds to the target modality. Traditional approaches typically train a model by minimizing the  $L_1$  or  $L_2$  loss between prediction and target. In contrast, we model the synthesized target  $\hat{\mathbf{y}}$  as being drawn from a Gaussian distribution  $\hat{\mathbf{y}} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ .

To incorporate uncertainty into the model, following [1], we further model the mean and variance using a Normal-Inverse Gamma (NIG) distribution, where  $\boldsymbol{\mu} \sim \mathcal{N}(\boldsymbol{\gamma}, \boldsymbol{\sigma}^2 \boldsymbol{\eta}^{-1})$  and  $\boldsymbol{\sigma}^2 \sim \Gamma^{-1}(\boldsymbol{\alpha}, \boldsymbol{\beta})$ . To train this evidential regression model, we minimize the negative log-likelihood loss:

$$\mathcal{L}^{\text{NLL}} = \frac{1}{2} \log\left(\frac{\pi}{\boldsymbol{\eta}}\right) - \boldsymbol{\alpha} \log(\boldsymbol{\omega}) + \left(\boldsymbol{\alpha} + \frac{1}{2}\right) \log((\mathbf{y} - \boldsymbol{\gamma})^2 \boldsymbol{\eta} + \boldsymbol{\omega}) + \log(\boldsymbol{\phi}), \quad (1)$$

where  $\boldsymbol{\omega} = 2\boldsymbol{\beta}(1 + \boldsymbol{\eta})$  and  $\boldsymbol{\phi} = \frac{\Gamma(\boldsymbol{\alpha})}{\Gamma(\boldsymbol{\alpha} + \frac{1}{2})}$ . We incorporate an additional regularization term:

$$\mathcal{L}^{\text{NIG}} = \mathcal{L}^{\text{NLL}} + \lambda_{\text{R}} \mathcal{L}^{\text{R}}, \quad (2)$$

where  $\mathcal{L}^{\text{R}} = \|\mathbf{y} - \boldsymbol{\gamma}\|_1 \cdot (2\boldsymbol{\eta} + \boldsymbol{\alpha})$ , and  $\lambda_{\text{R}} > 0$  balances these two terms.

Given the trained model, we can compute aleatoric uncertainty ( $\text{AU} = \frac{\boldsymbol{\beta}}{(\boldsymbol{\alpha}-1)}$ ) and epistemic uncertainty ( $\text{EU} = \frac{\boldsymbol{\beta}}{\boldsymbol{\eta}(\boldsymbol{\alpha}-1)}$ ). After obtaining these NIG distribution parameters, the synthesized target image is derived as  $\hat{\mathbf{y}} = \boldsymbol{\gamma}$ .

**Uncertainty-guided Multi-modal Fusion** Given predictions from different source modalities, each characterized by a NIG distribution, they are combined through the NIG summation operation based on MoNIG [18]:

$$\text{NIG}(\boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \bigoplus_{m=1}^M \text{NIG}_{[m]}(\boldsymbol{\gamma}_{[m]}, \boldsymbol{\eta}_{[m]}, \boldsymbol{\alpha}_{[m]}, \boldsymbol{\beta}_{[m]}), \quad (3)$$

where the  $\oplus$  operator follows the fusion rule:

$$\begin{aligned} \boldsymbol{\gamma} &= (\boldsymbol{\eta}_1 + \boldsymbol{\eta}_2)^{-1}(\boldsymbol{\eta}_1 \boldsymbol{\gamma}_1 + \boldsymbol{\eta}_2 \boldsymbol{\gamma}_2), \quad \boldsymbol{\eta} = \boldsymbol{\eta}_1 + \boldsymbol{\eta}_2, \quad \boldsymbol{\alpha} = \boldsymbol{\alpha}_1 + \boldsymbol{\alpha}_2 + \frac{1}{2}, \\ \boldsymbol{\beta} &= \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 + \frac{1}{2} \boldsymbol{\eta}_1 (\boldsymbol{\gamma}_1 - \boldsymbol{\gamma})^2 + \frac{1}{2} \boldsymbol{\eta}_2 (\boldsymbol{\gamma}_2 - \boldsymbol{\gamma})^2. \end{aligned} \quad (4)$$

## 2.2 Distribution Calibration

To address the distribution shifts across different medical centers, we propose a non-parametric calibration scheme based on quantile regression [13]. For each target center, we first collect a small pixel-wise calibration dataset  $\mathcal{D}$ . We use the evidential model to estimate its predictive cumulative distribution function (CDF)  $\hat{F}_i(y) = P(\hat{y}_i \leq y)$ . The calibration dataset is then constructed as:

$$\mathcal{D} = \left\{ \left( F_n(y_n), \hat{P}(F_n(y_n)) \right) \right\}_{n=1}^N, \quad (5)$$

where  $\hat{P}(p) = \frac{|\{y_n | F_n(y_n) \leq p\}|}{N}$  represents the empirical CDF, quantifying the actual coverage proportion of predicted quantiles.

To ensure monotonicity of the calibrated distribution, we train an isotonic regression model [13]  $R : [0, 1] \rightarrow [0, 1]$  on the calibration dataset to learn the center-specific distribution mapping. The calibrated distribution is then obtained through composition of the learned isotonic regression with the original CDF:

$$\tilde{F}(y|x) = R \circ \hat{F}(y|x) \quad (6)$$

### 2.3 Training Paradigm

This section shows the network architecture in our framework and explains the training strategy. Our framework consists of a conditional encoder  $\mathcal{E}$  and a conditional evidential generator  $\mathcal{G}$ . For each source modality  $\mathbf{x}_{[m]}$ , the prediction follows  $\text{NIG}_{[m]} = \mathcal{G}(\mathcal{E}(\mathbf{x}_{[m]}))$ , where  $\mathcal{G}$  outputs four parameter maps  $\{\gamma, \eta, \alpha, \beta\}$ . We modify an alias-free GAN [11, 24] to implement these modules. A softplus activation is applied to ensure positivity constraints on  $\{\eta, \alpha, \beta\}$ , with  $\alpha$  additionally incremented by 1 to satisfy  $\alpha > 1$ .

The overall training objective combines source-specific losses and fusion loss:

$$\mathcal{L} = \sum_{m=1}^M \mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}} + \mathcal{L}^{\text{Fuse}}, \quad (7)$$

where for each source modality:

$$\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}} = \mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{Adv}} + \lambda_{\text{NIG}} \mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{NIG}} + \lambda_{\text{Pix}} (\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{Syn}} + \mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{x}_{[m]}}^{\text{Rec}}), \quad (8)$$

with  $\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{Syn}} = \|\gamma_{[m]} - \mathbf{y}\|_1$  for target synthesis,  $\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{x}_{[m]}}^{\text{Rec}} = \|\mathcal{G}(\mathbf{x}_{[m]}) - \mathbf{x}_{[m]}\|_1$  for self-reconstruction,  $\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{NIG}}$  for uncertainty estimation as defined in Equ. 2, and  $\mathcal{L}_{\mathbf{x}_{[m]} \rightarrow \mathbf{y}}^{\text{Adv}}$  for adversarial training to ensure realistic synthesis. The fusion-level loss after MoNIG is defined as:

$$\mathcal{L}^{\text{Fuse}} = \mathcal{L}^{\text{Adv}} + \lambda_{\text{NIG}} \mathcal{L}^{\text{NIG}} + \lambda_{\text{Pix}} \mathcal{L}^{\text{Syn}}, \quad (9)$$

where each term serves the same as Equ. 8 but operates on the fused prediction. The balancing weights  $\lambda_{\text{NIG}}$  and  $\lambda_{\text{Pix}}$  are set to 0.5 and 100 respectively.

## 3 Experiments

### 3.1 Dataset, Evaluation Metrics, and Implementation

**Datasets** We conduct experiments using three brain MRI datasets from the BraTS2023 challenge [2, 21]: BraSyn, BraTS-Africa, and BraTS-PED. Each dataset comprises four registered MRI sequences (T1n, T2w, T2f, and T1c sequences) and expert-annotated tumor segmentation masks.

*BraSyn dataset (1,251 cases)* serves as our primary training dataset. To evaluate cross-domain robustness, we employ two additional datasets with domain shifts for validation. *a) BraTS-Africa dataset (60 cases)* represents variations in imaging protocols, collected from African medical centers using lower-field MRI systems that have lower image contrast and resolution. *b) BraTS-PED dataset (90 cases)* represents variations in patient populations, focusing on pediatric high-grade glioma of children with distinct anatomical structures due to age-specific brain development.

For data preprocessing, we normalize the intensity values of each modality to  $[-1, 1]$ . The 3D volumes are processed slice-by-slice along the axial plane and resized to  $256 \times 256$ . To ensure sufficient foreground information, we discard the first and last 30 slices where the brain regions are typically minimal. For BraSyn dataset, we randomly split the subjects into training, validation, and testing sets with a ratio of 7:1:2.

**Evaluation Metrics** We employ multiple complementary metrics to comprehensively evaluate our method: 1) For synthesis quality, we use the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM); 2) For uncertainty estimation reliability, we calculate the expected uncertainty calibration error (UCE) [14], which quantifies the alignment between predicted uncertainties and actual errors through weighted  $L_1$ -norm; 3) For clinical utility assessment, we evaluate downstream task performance using the Dice score for multi-modal tumor segmentation. Specifically, we apply a top-performing BraTS 2021 algorithm [17] on both synthesized and original sequences. We conducted experiments using T1n and T2w as input to synthesize T2f and T1c modalities. The table results show the average performance across T2f and T1c synthesis.

**Implementation Details** The framework was implemented in PyTorch and trained on a single NVIDIA H100 GPU. We employed the Adam optimizer with an initial learning rate of  $2 \times 10^{-3}$ , which was modulated using a cosine annealing warm restart schedule [16] over  $2 \times 10^5$  iterations. The evidential loss balancing coefficient  $\lambda_R$  in Equ. 2 was gradually increased from 0 to 1 during training to ensure stable convergence. For the MoNIG fusion module, we initialized the concentration parameters  $\{\alpha, \beta\}$  to small positive values (0.1) to encourage uncertainty-aware learning from the start.

### 3.2 Results

**Intra-domain Comparison Study** To comprehensively evaluate our framework, we conducted extensive comparisons with other medical image translation methods, including: (1) CNN-based methods: Pix2pix [10], CycleGAN [29], RegGAN [12], and MM-GAN [23]; (2) Transformer-based method: ResViT [4]; (3) Diffusion-based method: M2DN [20].

As shown in Tab. 1, the top section presents the intra-center validation results on the BraSyn dataset. Our method outperforms the second-best approach with

**Table 1.** Comparison of different methods on target sequence synthesis using BraTS2023 Challenge datasets.

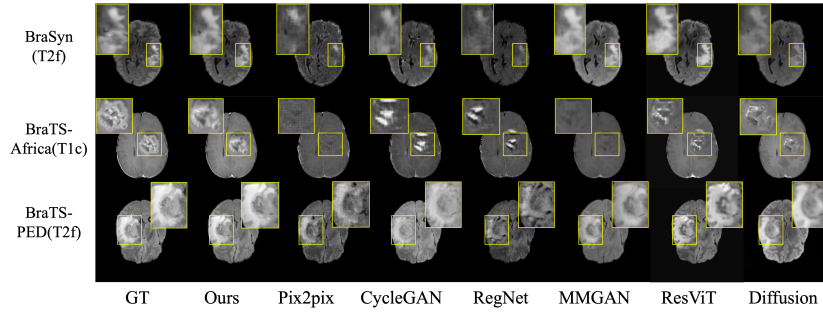
Dataset	Method	PSNR $\uparrow$	SSIM $\uparrow$	Dice $\uparrow$
BraSyn (in-distribution)	Pix2pix	27.12 $\pm$ 4.17	0.904 $\pm$ 0.031	0.807 $\pm$ 0.102
	CycleGan	26.71 $\pm$ 4.31	0.903 $\pm$ 0.033	0.807 $\pm$ 0.103
	MM-GAN	29.06 $\pm$ 3.39	0.924 $\pm$ 0.023	0.827 $\pm$ 0.096
	RegGAN	25.79 $\pm$ 3.95	0.888 $\pm$ 0.035	0.797 $\pm$ 0.106
	ResViT	26.46 $\pm$ 3.33	0.908 $\pm$ 0.033	0.814 $\pm$ 0.099
	M2DN	31.65 $\pm$ 3.27	0.929 $\pm$ 0.029	<b>0.854<math>\pm</math>0.093</b>
	Ours	<b>32.37<math>\pm</math>3.09</b>	<b>0.937<math>\pm</math>0.029</b>	0.852 $\pm$ 0.094
BraTS-Africa (cross-center)	Pix2pix	20.42 $\pm$ 3.07	0.817 $\pm$ 0.170	0.749 $\pm$ 0.174
	CycleGan	21.16 $\pm$ 2.68	0.835 $\pm$ 0.142	0.762 $\pm$ 0.165
	MM-GAN	21.42 $\pm$ 3.69	0.862 $\pm$ 0.165	0.782 $\pm$ 0.162
	RegGAN	19.22 $\pm$ 2.79	0.810 $\pm$ 0.171	0.743 $\pm$ 0.178
	ResViT	22.12 $\pm$ 3.37	0.830 $\pm$ 0.168	0.769 $\pm$ 0.168
	M2DN	24.71 $\pm$ 2.29	0.858 $\pm$ 0.095	0.795 $\pm$ 0.151
	Ours	<b>28.48<math>\pm</math>2.17</b>	<b>0.886<math>\pm</math>0.071</b>	<b>0.828<math>\pm</math>0.157</b>
BraTS-PED (cross-population)	Pix2pix	22.00 $\pm$ 3.66	0.875 $\pm$ 0.045	0.667 $\pm$ 0.153
	CycleGan	22.32 $\pm$ 3.45	0.884 $\pm$ 0.042	0.679 $\pm$ 0.144
	MM-GAN	22.84 $\pm$ 4.54	0.910 $\pm$ 0.037	0.700 $\pm$ 0.139
	RegGAN	20.91 $\pm$ 3.33	0.863 $\pm$ 0.044	0.662 $\pm$ 0.156
	ResViT	23.83 $\pm$ 3.68	0.887 $\pm$ 0.038	0.686 $\pm$ 0.146
	M2DN	26.39 $\pm$ 3.05	0.870 $\pm$ 0.110	0.712 $\pm$ 0.143
	Ours	<b>29.14<math>\pm</math>2.92</b>	<b>0.883<math>\pm</math>0.105</b>	<b>0.723<math>\pm</math>0.140</b>

improvements of 0.72dB in PSNR and 0.008 in SSIM while achieving competitive Dice scores with the best-performing method. This superior performance is due to our explicit MoNIG fusion strategy, where uncertainty-guided integration is more robust compared to other fusion strategies.

**Cross-domain Comparison Study** To demonstrate robustness, we conducted experiments on two cross-center datasets, BraTS-Africa and BraTS-PED, as shown in Tab. 1. To ensure a fair comparison, we randomly selected 10 slices from each cross-center dataset for fine-tuning each comparison model pre-trained on BraSyn as well as training our calibration model.

Our method achieved superior performance on both datasets compared to other approaches, demonstrating the strong robustness of our uncertainty-based framework. Moreover, through calibration training with very few samples, our model could quickly adapt to new source-target mappings.

**Ablation Study** We conducted an ablation study to evaluate the effectiveness of our key components. As shown in the right of Fig. 4, MoNIG achieves



**Fig. 3.** Visual comparisons of ours and other methods.

Hard case (T1c)				
	Synthesis	GT	GT+mask	Uncertainty
Easy case (T2f)				
	Synthesis	GT	GT+mask	Uncertainty

Method	PSNR $\uparrow$	SSIM $\uparrow$	Dice $\uparrow$	UCE $\downarrow$
Baseline	24.25 $\pm 2.42$	0.855 $\pm 0.121$	0.783 $\pm 0.156$	-
+Deep Ensemble	25.72 $\pm 2.31$	0.864 $\pm 0.108$	0.809 $\pm 0.152$	0.167 $\pm 0.023$
+MoNIG	26.39 $\pm 2.25$	0.871 $\pm 0.095$	0.819 $\pm 0.147$	0.142 $\pm 0.019$
+MoNIG+Calib.	<b>28.48</b> $\pm 2.17$	<b>0.886</b> $\pm 0.071$	<b>0.828</b> $\pm 0.157$	<b>0.089</b> $\pm 0.012$

**Fig. 4.** Left: Reliability visualization assessment, where purple mask shows enhanced tumor region, green mask shows oedema region, and red mask shows necrotic core. Right: Ablation study results evaluated on BraTS-Africa dataset.

lower UCE compared to Deep Ensemble [3], indicating more accurate uncertainty estimation. Through uncertainty-guided fusion, MoNIG further achieves higher-quality synthesis results. Furthermore, incorporating calibration on top of MoNIG leads to optimal performance in both image synthesis and uncertainty estimation, demonstrating the robustness of our MoNIG and calibration strategies.

Furthermore, the left of Fig. 4 demonstrates the reliability of our uncertainty estimation. We visualize uncertainty maps for hard and easy cases. In the first row (hard case), the yellow arrow indicates a risky region in the generated image, corresponding to the enhancing tumor area. This region shows high uncertainty as it is difficult to identify in T1n and T2w source images, correctly indicating potential errors. In contrast, the second row shows a case with lower uncertainty, suggesting a more reliable synthesis. Such uncertainty visualization can provide valuable guidance for clinical diagnosis by helping physicians identify regions requiring additional verification.

## 4 Conclusion

In this paper, we presented a framework for multi-modal MRI translation that addresses uncertainty quantification and cross-center adaptation. By reformulat-



ing the translation task as an evidential regression problem, our approach provides reliable uncertainty estimates that enhance clinical decision-making. The MoNIG fusion strategy effectively integrates information from multiple source modalities. Furthermore, our distribution calibration mechanism enables efficient adaptation to new clinical environments with minimal data requirements.

**Acknowledgments.** This work was supported by the National Natural Science Foundation of China (grant No. 62372115), the Shanghai Municipal Education Commission (grant No. 24KXZNA13), the National Natural Science Foundation of China (grant No. 62488101), and Shanghai Artificial Intelligence Laboratory.

**Disclosure of Interests.** The authors have no competing interests to declare that are relevant to the content of this article.

## References

1. Alexander Amini, Wilko Schwarting, Ava Soleimany, and Daniela Rus. Deep evidential regression. *Advances in Neural Information Processing Systems*, 33:14927–14937, 2020.
2. Ujjwal Baid, Satyam Ghodasara, Suyash Mohan, Michel Bilello, Evan Calabrese, Errol Colak, Keyvan Farahani, Jayashree Kalpathy-Cramer, Felipe C Kitamura, Sarthak Pati, et al. The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*, 2021.
3. Laëtitia Buisson, Wilfried Thuiller, Nicolas Casajus, Sovan Lek, and Gael Grenouillet. Uncertainty in ensemble forecasting of species distribution. *Global Change Biology*, 16(4):1145–1157, 2010.
4. Onat Dalmaz, Mahmut Yurt, and Tolga Çukur. Resvit: Residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, 41(10):2598–2614, 2022.
5. Erik Daxberger, Eric Nalisnick, James U Allingham, Javier Antorán, and José Miguel Hernández-Lobato. Bayesian deep learning via subnetwork inference. In *International Conference on Machine Learning*, pages 2510–2521. PMLR, 2021.
6. Zheyao Gao, Yuanye Liu, Fuping Wu, Nannan Shi, Yuxin Shi, and Xiahai Zhuang. A reliable and interpretable framework of multi-view learning for liver fibrosis staging. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 178–188. Springer, 2023.
7. Luyi Han, Tianyu Zhang, Yunzhi Huang, Haoran Dou, Xin Wang, Yuan Gao, Chunyao Lu, Tao Tan, and Ritse Mann. An explainable deep framework: Towards task-specific fusion for multi-to-one mri synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 45–55. Springer, 2023.
8. Zongbo Han, Changqing Zhang, Huazhu Fu, and Joey Tianyi Zhou. Trusted multi-view classification with dynamic evidential fusion. *IEEE transactions on pattern analysis and machine intelligence*, 45(2):2551–2566, 2022.
9. Jimmy Hickey, Jonathan P Williams, and Emily C Hector. Transfer learning with uncertainty quantification: Random effect calibration of source to target (recast). *Journal of Machine Learning Research*, 25(338):1–40, 2024.

10. Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.
11. Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021.
12. Lingke Kong, Chenyu Lian, Detian Huang, Yanle Hu, Qichao Zhou, et al. Breaking the dilemma of medical image-to-image translation. *Advances in Neural Information Processing Systems*, 34:1964–1978, 2021.
13. Volodymyr Kuleshov, Nathan Fenner, and Stefano Ermon. Accurate uncertainties for deep learning using calibrated regression. In *International conference on machine learning*, pages 2796–2804. PMLR, 2018.
14. Max-Heinrich Laves, Sontje Ihler, Jacob F Fast, Lüder A Kahrs, and Tobias Ortmaier. Well-calibrated regression uncertainty in medical imaging with deep learning. In *Medical Imaging with Deep Learning*, pages 393–412. PMLR, 2020.
15. Jiang Liu, Srivathsa Pasumarthi, Ben Duffy, Enhao Gong, Keshav Datta, and Greg Zaharchuk. One model to synthesize them all: Multi-contrast multi-scale transformer for missing data imputation. *IEEE transactions on medical imaging*, 42(9):2577–2591, 2023.
16. Ilya Loshchilov and Frank Hutter. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016.
17. Huan Minh Luu and Sung-Hong Park. Extending nn-unet for brain tumor segmentation. In *International MICCAI brainlesion workshop*, pages 173–186. Springer, 2021.
18. Huan Ma, Zongbo Han, Changqing Zhang, Huazhu Fu, Joey Tianyi Zhou, and Qinghua Hu. Trustworthy multimodal regression with mixture of normal-inverse gamma distributions. *Advances in Neural Information Processing Systems*, 34:6881–6893, 2021.
19. Wesley J Maddox, Pavel Izmailov, Timur Garipov, Dmitry P Vetrov, and Andrew Gordon Wilson. A simple baseline for bayesian uncertainty in deep learning. *Advances in neural information processing systems*, 32, 2019.
20. Xiangxi Meng, Kaicong Sun, Jun Xu, Xuming He, and Dinggang Shen. Multi-modal modality-masked diffusion network for brain mri synthesis with random modality missing. *IEEE Transactions on Medical Imaging*, 2024.
21. Sarthak Pati, Ujjwal Baid, Brandon Edwards, Micah Sheller, Shih-Han Wang, G Anthony Reina, Patrick Foley, Alexey Gruzdev, Deepthi Karkada, Christos Davatzikos, et al. Federated learning enables big data for rare cancer boundary detection. *Nature communications*, 13(1):7346, 2022.
22. Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classification uncertainty. *Advances in neural information processing systems*, 31, 2018.
23. Anmol Sharma and Ghassan Hamarneh. Missing mri pulse sequence synthesis using multi-modal generative adversarial network. *IEEE transactions on medical imaging*, 39(4):1170–1183, 2019.
24. Zhiyun Song, Xin Wang, Xiangyu Zhao, Sheng Wang, Zhenrong Shen, Zixu Zhuang, Mengjun Liu, Qian Wang, and Lichi Zhang. Alias-free co-modulated network for cross-modality synthesis and super-resolution of mr images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 66–76. Springer, 2023.

25. Zhaohu Xing, Sicheng Yang, Sixiang Chen, Tian Ye, Yijun Yang, Jing Qin, and Lei Zhu. Cross-conditioned diffusion model for medical image to image translation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 201–211. Springer, 2024.
26. Lipei Zhang, Yanqi Cheng, Lihao Liu, Carola-Bibiane Schönlieb, and Angelica I Aviles-Rivero. Biophysics informed pathological regularisation for brain tumour segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 3–13. Springer, 2024.
27. Yue Zhang, Chengtao Peng, Qiuli Wang, Dan Song, Kaiyan Li, and S Kevin Zhou. Unified multi-modal image synthesis for missing modality imputation. *IEEE Transactions on Medical Imaging*, 2024.
28. Tao Zhou, Huazhu Fu, Geng Chen, Jianbing Shen, and Ling Shao. Hi-net: hybrid-fusion network for multi-modal mr image synthesis. *IEEE transactions on medical imaging*, 39(9):2772–2781, 2020.
29. Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.