IWCNet: Intervention-based Weight Calibration Network for Multi-center Multi-modality MRI Lesion Segmentation

Ronghui Qi[1] , **Wenlong Song**[1] , **Wanyin Shi**[2,*] , **Chenchu Xu**[1,*]
[1]Key Laboratory of Intelligent Computing and Signal Processing of Ministry of Education, School of Computer Science and Technology, Anhui University, Hefei, China
[2]Department of Radiology, the First Affiliated Hospital of Anhui Medical University, Hefei, China
yfy1462112@fy.ahmu.edu.cn, cxu332@gmail.com

Abstract

Multi-modality MRI lesion segmentation leverages complementary contrasts across modalities to better characterize lesion morphology and tissue alterations, thereby improving lesion delineation and reducing misdiagnosis. Existing correlation-driven methods estimate fusion weights by converting cross-modality correlation scores into adaptive fusion weights. However, in multi-center settings, domain shift changes cross-modal correlation patterns, which breaks the learned correlation-to-weight mapping and leads to modality-weight bias that amplifies center-specific shortcuts. To address this issue, we propose an Intervention-based Weight Calibration Network (IWCNet) that models domain shift as a controllable embedding and leverages explicit interventions to mitigate its impact on fusion weight estimation. IWCNet constructs original and intervened branches by performing interventions on the domain embedding, measuring inter-branch prediction effects to calibrate fusion weights. IWCNet further introduces Operator-level Domain Modulation (OLDM) to construct structure-consistent domain-variant feature pairs via operator-level domain intervention, and Effect-Driven Consistency Calibration (EDCC) to jointly leverage intervention effects and cross-branch prediction consistency to calibrate weights. Extensive experiments on multi-center multi-modality MRI datasets demonstrate that IWCNet substantially improves segmentation accuracy and robustness under domain shift.

1 Introduction

Multi-modality MRI lesion segmentation leverages complementary contrasts across modalities to jointly characterize lesion morphology, microstructural alterations, and hemodynamic properties, thereby improving lesion delineation, reducing misdiagnosis, and supporting staging and treatment-response assessment [Gurney-Champion *et al.*, 2020]. In liver tumor MRI, T1- and T2-weighted images better delineate lesion boundaries and suggest compositional differences [Albiin, 2012], whereas diffusion-weighted imaging

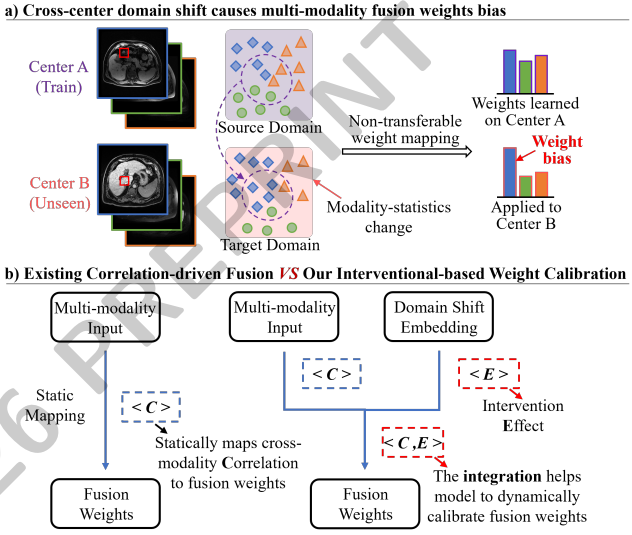

Figure 1: (a) Cross-center domain shift induces biased fusion weights in multi-center multi-modality MRI lesion segmentation. (b) Our method mitigates weight bias by calibrating fusion weights using prediction effects under domain-embedding interventions.

(DWI) is more sensitive to diffusion restriction and related microstructural changes [Parikh *et al.*, 2008], helping further differentiate lesion type and activity. Therefore, accurate multi-modality MRI lesion segmentation can reliably yield key quantitative measures such as lesion volume, the spatial extent of lesion margins, and the proportion of diffusion-restricted subregions, providing direct evidence for treatment planning and response evaluation [Bharwani and Koh, 2013].

Meanwhile, correlation-driven fusion methods have been widely adopted for multi-modality MRI lesion segmentation [Xu *et al.*, 2024; Chen *et al.*, 2025; Cao *et al.*, 2025]. These methods measure pairwise similarity between modality features or attention affinity scores to evaluate cross-modality correlation at each spatial location [Zhou *et al.*, 2022]. They then transform the resulting correlation scores into adaptive fusion weights via normalization or gating, where the weights determine each modality's contribution to the fused representation [Ding *et al.*, 2021]. By learning an adaptive mapping from cross-modality correlations to fusion weights, these methods effectively exploit complementary contrasts to

enhance lesion delineation and achieve superior performance under within-domain settings.

However, correlation-driven fusion methods [Zhou *et al.*, 2022; Xu *et al.*, 2024] are prone to fusion weight bias under domain shift in multi-center settings. Joint feature statistics vary across centers, altering the cross-modal correlations these methods rely on and changing their relative strength. Consequently, the correlation-to-weight mapping becomes miscalibrated at unseen centers, leading to systematically over- or under-weighting specific modalities and amplifying center-specific shortcuts rather than lesion-relevant evidence.

In this paper, we propose an Intervention-based Weight Calibration Network (IWCNet) for accurate multi-modality MRI lesion segmentation. Compared with correlation-driven fusion methods, IWCNet models domain shift as a controllable embedding and leverages explicit interventions to mitigate its impact on fusion weight estimation, thereby improving robustness under domain shift. Specifically, IWCNet constructs original and intervened branches by performing interventions on the domain embedding. It measures inter-branch prediction effects to estimate each modality's sensitivity to domain variation and uses these sensitivities to calibrate fusion weights. Additionally, to explicitly manipulate domain variations while preserving anatomical structure, IWCNet introduces Operator-level Domain Modulation (OLDM), which treats domain variation as a controllable variable via operator-level domain intervention, thereby constructing structure-consistent domain-variant feature pairs for effect estimation. Furthermore, to prevent fusion weights from over-relying on domain-sensitive shortcuts, IWCNet designs Effect-Driven Consistency Calibration (EDCC), which jointly leverages intervention effects and cross-branch prediction consistency to calibrate weights, thereby anchoring the weights to domain-invariant evidence.

The paper introduces the following contributions:

- A novel multi-center multi-modality MRI lesion segmentation method is proposed to calibrate modality fusion weights using prediction effects under domain-embedding interventions, thereby improving robustness to domain shift.
- An Operator-level Domain Modulation module is proposed to construct structure-consistent, domain-variant feature pairs via operator-level intervention, thereby explicitly modulating domain variations while preserving anatomical structure.
- An Effect-Driven Consistency Calibration module is proposed to jointly leverage intervention effects and cross-branch prediction consistency to calibrate fusion weights, thereby preventing the weights from over-relying on domain-sensitive shortcuts.
- Extensive experiments on multi-center multi-modality MRI datasets demonstrate that IWCNet achieves consistent improvements in both accuracy and robustness under domain shift.

2 Related Work

2.1 Multi-modality MRI Segmentation

Multi-modality MRI segmentation exploits complementary tissue contrasts across modalities to improve lesion delineation. Early approaches [Kamnitsas *et al.*, 2017] stacked modalities as channels and used a single encoder, which can blur modality-specific cues. Later works adopted multi-branch encoders [Havaei *et al.*, 2016; Dolz *et al.*, 2018] to preserve modality-specific representations and fuse them in a shared latent space. Multi-level fusion [Kamnitsas *et al.*, 2017] further injects interactions across resolutions to leverage both boundary detail and high-level semantics. Recently, correlation-driven fusion paradigms [Wang *et al.*, 2021; Zhang *et al.*, 2022] have become predominant, using cross-modal similarity/affinity as a proxy for inter-modality agreement and converting it into channel-/spatial-wise reweighting or cross-attention for adaptive fusion, while remaining computationally tractable [Zhao *et al.*, 2022]. Despite strong within-domain performance, these strategies implicitly anchor the correlation-to-weight mapping to training-domain co-occurrence statistics, making them sensitive to domain-induced appearance variations and leaving robust fusion under domain shifts in real-world multi-center settings insufficiently addressed [Yan *et al.*, 2019; Guan and Liu, 2021].

2.2 Controlled Interventions in Medical Image Analysis

Controlled interventions have been explored in medical imaging to probe model behavior and reduce shortcut reliance by creating paired examples that preserve clinically meaningful structure while inducing controlled appearance changes. Black-box explainers often generate progressively edited images, typically with generative models, to identify decision-critical evidence and audit predictions [Mertes *et al.*, 2022; Singla *et al.*, 2023; Moreira *et al.*, 2025]. Diffusion-based representations further enable smooth manipulations in semantically structured latent spaces, improving the plausibility and continuity of intervention trajectories for prediction analysis [Atad *et al.*, 2024]. Related ideas have also been applied to segmentation and adaptation, where disentangling causal content from domain-dependent factors improves robustness under domain shift [Lv *et al.*, 2025]. Different from these works that focus on explanation or single-input robustness, our goal is to use intervention effects to calibrate modality contributions for stable cross-modality fusion.

2.3 Domain Generalization in Medical Image Segmentation

Robust medical segmentation under domain shift is often studied through domain generalization and style-aware representation learning [Zhou *et al.*, 2021], with the goal of reducing reliance on domain-dependent appearance factors. One representative direction improves generalization by augmenting style or feature statistics, such as frequency-space mixing [Li *et al.*, 2023] and generative feature-style augmentation with regularization for more stable training [Huang *et al.*, 2025]. Related work also models the distribution of style statistics to support randomized sampling or mixing during

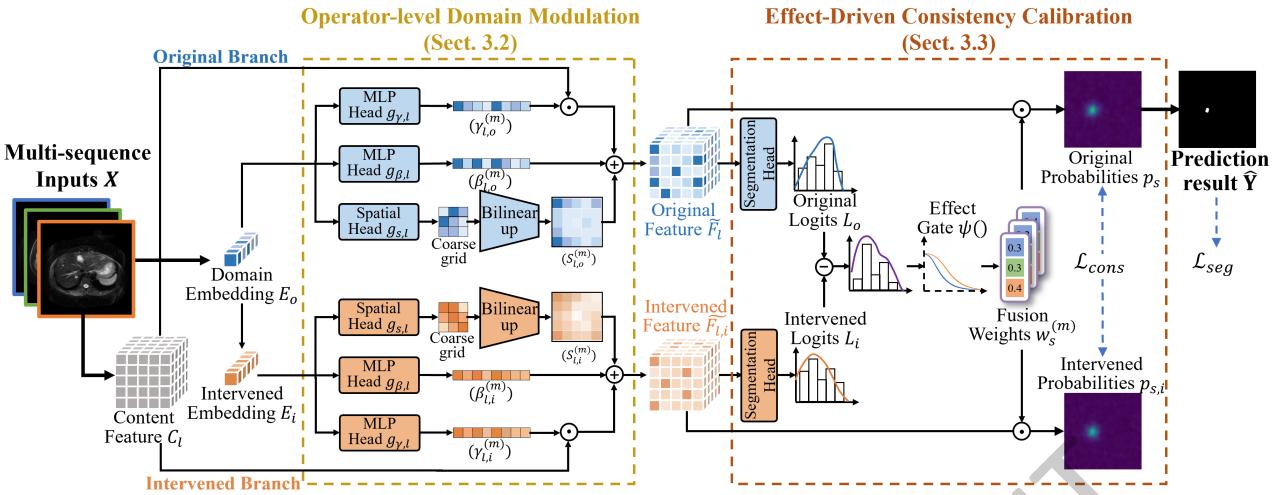

Figure 2: Overview of IWCNet. The proposed framework calibrates modality fusion weights using prediction effects under domain-embedding interventions for robust multi-modality MRI lesion segmentation.

training [Safdari *et al.*, 2025]. Diffusion-based formulations have been explored to improve segmentation robustness under domain shifts [Yang *et al.*, 2025]. While effective for single-stream robustness, these methods do not target fusion-weight calibration in multi-modality segmentation, nor do they use perturbation effects to quantify modality contributions for stable evidence aggregation under domain shift.

3 Methodology

Our IWCNet calibrates fusion weights for multi-modality MRI lesion segmentation using prediction effects under domain-embedding interventions. It keeps content features fixed, intervenes on the domain embedding to form an original and an intervened branch, and uses finite-difference logit changes to derive effect-based signals for sample-wise weighting and prediction regularization, thereby mitigating the impact of domain shift on cross-modality fusion. In addition, IWCNet introduces Operator-level Domain Modulation (OLDM), which injects the domain embedding into content features through channel-wise modulation and a low-frequency spatial field in a residual form, enabling fine-grained appearance modulation. Furthermore, IWCNet introduces Effect-Driven Consistency Calibration (EDCC), which maps effect scores to sample-wise fusion weights via Monotonic Effect Gating and enforces cross-branch prediction consistency to stabilize calibrated weighting.

3.1 IWCNet Network Architecture

IWCNet constructs an original branch and an intervened branch by keeping content features fixed and applying OLDM with an original domain embedding and its intervened counterpart. It measures prediction effects as finite-difference changes in segmentation logits between the two branches to compute per-modality effect scores, and maps these scores to sample-wise fusion weights to stabilize cross-modality aggregation under domain shift.

For each sample, the input is a set of aligned modalities $\mathbf{X} = \{X^{(m)}\}_{m=1}^{M}$ with a corresponding lesion mask Y, and the goal is to predict $\hat{Y}$ by aggregating complementary evidence across modalities while maintaining stable fusion behavior under domain shift.

Given each modality $X^{(m)}$, a shared encoder $E(\cdot)$ produces a feature pyramid $\{F_l^{(m)}\}_{l=1}^{L}$, where l indexes resolution levels. IWCNet factorizes the representation into a content feature map and a compact domain embedding via Spatial and Domain Dual Projection (SADP). The content feature preserves spatial structure relevant to anatomy and lesions

$$C_l^{(m)} = h_c\left(F_l^{(m)}\right), \tag{1}$$

while the domain embedding encodes global appearance statistics that characterize domain factors

$$E_o^{(m)} = h_d\left(\mathrm{AttnPool}\left(F_L^{(m)}\right)\right). \tag{2}$$

Here $\mathrm{AttnPool}(\cdot)$ aggregates global information into a compact token, and $h_c(\cdot)$ and $h_d(\cdot)$ are lightweight projection heads.

To generate the intervened embedding, IWCNet applies a controlled perturbation to the original domain embedding during training, while keeping $C_l^{(m)}$ fixed. Specifically, for each modality m, we construct

$$E_i^{(m)} = E_o^{(m)} + \kappa\,\Delta^{(m)}, \qquad \|\Delta^{(m)}\|_2 = 1, \tag{3}$$

where $\Delta^{(m)}$ is a unit-norm perturbation direction in the embedding space (e.g., randomly sampled or derived from mini-batch statistics), and κ controls the intervention strength.

On top of this factorization, OLDM turns the domain embedding into intervenable operators and injects them into $C_l^{(m)}$ in a residual manner. In the original branch, OLDM uses $E_o^{(m)}$ to obtain domain-conditioned features $\tilde{F}_l^{(m)}$. In the intervened branch, OLDM replaces $E_o^{(m)}$ with the intervened embedding $E_i^{(m)}$ while keeping $C_l^{(m)}$ fixed, yielding intervened features $\tilde{F}_{l,\mathrm{i}}^{(m)}$.

Based on the paired features from the two branches, IWCNet performs weight calibration via EDCC. Concretely, a shared prediction head $H(\cdot)$ produces per-modality logits for the original and intervened branches, denoted as $\{L_o^{(m)}\}_{m=1}^M$ and $\{L_i^{(m)}\}_{m=1}^M$, respectively. EDCC then quantifies an intervention effect score $L_r^{(m)}$ for each modality by measuring the finite-difference change between $L_o^{(m)}$ and $L_i^{(m)}$, which reflects the sensitivity of modality-specific predictions to domain-embedding interventions.

These effect scores are subsequently converted into sample-wise fusion weights $\{w^{(m)}\}_{m=1}^M$ through a Monotonic Effect Gating, such that modalities with smaller intervention effects receive larger weights. Using the calibrated weights, IWCNet fuses modality-specific features to obtain the fused prediction $\hat{Y}$ for the original branch, and analogously forms the fused prediction for the intervened branch. Finally, EDCC enforces prediction consistency between the fused outputs of the original and intervened branches.

During training, we optimize IWCNet with a supervised segmentation loss on the fused prediction of the original branch, together with the EDCC consistency regularizer. The overall objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{BCE}}\Big(\hat{Y}, Y\Big) + \mathcal{L}_{\mathrm{IoU}}\Big(\hat{Y}, Y\Big) + \lambda_{\mathrm{cons}}\,\mathcal{R}_{\mathrm{cons}}, \quad (4)$$

where $\mathcal{R}_{\mathrm{cons}}$ is defined in Eq. 14.

3.2 Operator-level Domain Modulation

OLDM instantiates a structure-preserving domain modulation mechanism that turns the domain embedding into intervenable operators acting on content features. Given the factorized representation, OLDM parameterizes domain factors as a composition of a channel-wise appearance operator and a smooth spatial bias field, and injects them into the content stream in a residual form. This operator-based reparameterization makes domain-dependent appearance variation explicit in feature space and enables controlled embedding interventions to induce appearance changes while keeping anatomy and lesion structure invariant.

OLDM operates on the content feature maps $\{C_l^{(m)}\}_{l=1}^L$ and the domain embeddings $\{E^{(m)}\}_{m=1}^M$ produced by SADP. For each modality m and level l, OLDM maps $E^{(m)}$ to channel-wise affine modulation parameters

$$\gamma_l^{(m)} = g_{\gamma,l}\Big(E^{(m)}\Big), \qquad \beta_l^{(m)} = g_{\beta,l}\Big(E^{(m)}\Big), \quad (5)$$

where $g_{\gamma,l}(\cdot)$ and $g_{\beta,l}(\cdot)$ are lightweight MLP heads producing vectors in $\mathbb{R}^{C_l}$ with C_l channels.

In addition, OLDM predicts a low-frequency spatial field from $E^{(m)}$ to capture smooth domain-dependent bias patterns. Concretely, it outputs a coarse grid and upsamples it to the feature resolution

$$S_l^{(m)} = \mathrm{Up}\Big(g_{s,l}\Big(E^{(m)}\Big)\Big), \quad (6)$$

where $g_{s,l}(\cdot)$ outputs a low-resolution field and $\mathrm{Up}(\cdot)$ denotes bilinear upsampling to match the spatial size of $C_l^{(m)}$.

OLDM injects these signals into the content feature map in a residual manner to obtain the domain-conditioned feature

$$\tilde{F}_l^{(m)} = C_l^{(m)} + \gamma_l^{(m)} \odot C_l^{(m)} + \beta_l^{(m)} + S_l^{(m)}, \quad (7)$$

where $\odot$ denotes channel-wise multiplication with broadcasting. This formulation separates channel-wise appearance modulation from spatially smooth bias modulation, enabling fine-grained appearance control without changing the spatial layout encoded in $C_l^{(m)}$.

OLDM is applied to both the original and intervened branches. For the intervened branch, OLDM replaces the original embedding $E_o^{(m)}$ with the intervened embedding $E_i^{(m)}$ while keeping $C_l^{(m)}$ fixed, yielding intervened features $\tilde{F}_{l,\mathrm{i}}^{(m)}$. This design allows controlled embedding interventions to generate appearance-varying yet structure-consistent feature pairs, which EDCC uses to obtain paired branch logits for effect estimation and weight calibration.

3.3 Effect-Driven Consistency Calibration

EDCC calibrates fusion weights by converting prediction effects under controlled domain-embedding interventions into sample-wise weighting signals. It forms paired predictions under an original and an intervened embedding while keeping content features fixed, measures per-modality effects via finite-difference logit changes, and maps these effects to fusion weights through Monotonic Effect Gating. In parallel, EDCC enforces cross-branch prediction consistency to stabilize calibrated weighting and suppress domain-dependent shortcuts under domain shift.

EDCC takes as input the domain-conditioned features from OLDM for both branches, namely $\{\tilde{F}_{s,l}^{(m)}\}$ and $\{\tilde{F}_{s,l,\mathrm{i}}^{(m)}\}$, where s indexes samples in a mini-batch, m indexes modalities, and l indexes resolution levels.

EDCC estimates a per-modality effect score for each sample via the finite-difference change of segmentation logits under embedding interventions. Using a shared lightweight prediction head $H(\cdot)$, EDCC obtains per-modality logits for the original and intervened branches

$$L_{o,s}^{(m)} = H\Big(\tilde{F}_{s,L}^{(m)}\Big), \qquad L_{i,s}^{(m)} = H\Big(\tilde{F}_{s,L,\mathrm{i}}^{(m)}\Big), \quad (8)$$

where L denotes the coarsest level used for effect estimation. The intervention effect is quantified as the mean absolute logit change

$$L_{r,s}^{(m)} = \frac{1}{K|\Omega|}\sum_{k=1}^{K}\sum_{u\in\Omega}\Big|L_{o,s,k}^{(m)}(u) - L_{i,s,k}^{(m)}(u)\Big|, \quad (9)$$

where Ω denotes spatial locations and K is the number of classes. Intuitively, $L_{r,s}^{(m)}$ measures how sensitive modality-specific predictions are to domain-embedding interventions.

EDCC maps $\{L_{r,s}^{(m)}\}_{m=1}^M$ to sample-wise fusion weights via Monotonic Effect Gating. Specifically, a monotonic decreasing mapping $\psi(\cdot)$ converts effect scores into unnormalized evidence

$$e_s^{(m)} = \psi\Big(L_{r,s}^{(m)}\Big), \quad (10)$$

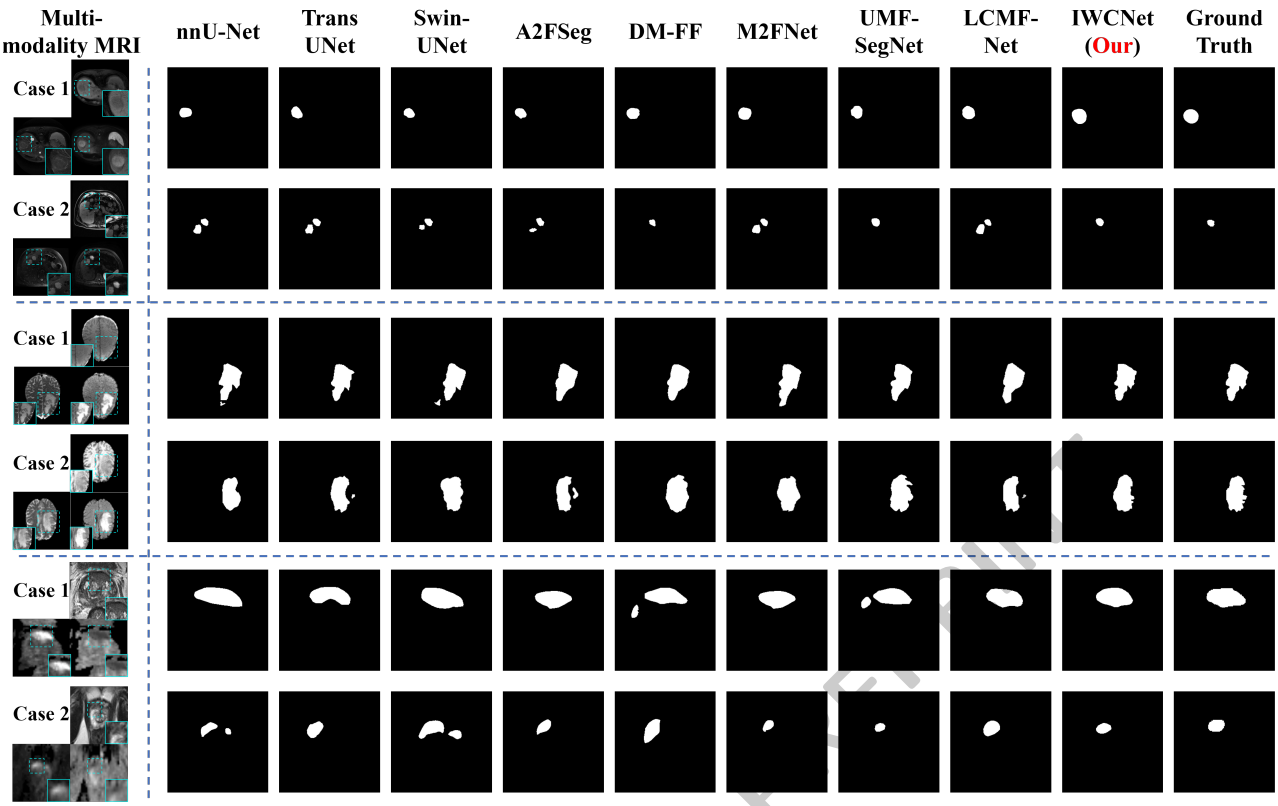

Figure 3: Our IWCNet accurately segments lesions from multi-modality MR images and outperforms all eight comparative methods across the liver, brain, and prostate datasets. Case 1 and Case 2 are from two different external test centers.

where we instantiate $\psi(\cdot)$ as a temperature-controlled monotonic mapping

$$e_s^{(m)} = \exp\left(-\frac{L_{r,s}^{(m)}}{\tau}\right), \tag{11}$$

and τ controls the sharpness of effect-to-evidence conversion. The fusion weights are then obtained by normalization

$$w_s^{(m)} = \frac{e_s^{(m)}}{\sum_{k=1}^{M} e_s^{(k)}}, \qquad w_s^{(m)} \geq 0, \ \sum_{m=1}^{M} w_s^{(m)} = 1. \tag{12}$$

This mapping preserves interpretability and stability by ensuring that modalities with smaller intervention effects receive larger weights through a strictly monotonic rule.

Using the calibrated weights, EDCC fuses the original features at each scale by weighted summation

$$F_{s,l}^{\text{fuse}} = \sum_{m=1}^{M} w_s^{(m)} \, \tilde{F}_{s,l}^{(m)}, \qquad l = 1, \ldots, L, \tag{13}$$

and feeds $\{F_{s,l}^{\text{fuse}}\}_{l=1}^{L}$ to a shared decoder $D(\cdot)$ to obtain the final probabilities $p_s = \sigma(L_{o,s})$. In the same manner, EDCC fuses the intervened features $\{\tilde{F}_{s,l,\text{i}}^{(m)}\}$ to obtain intervened probabilities $p_{s,\text{i}} = \sigma(L_{i,s})$.

Finally, EDCC enforces prediction consistency between the original and intervened branches to stabilize both weight calibration and final predictions under domain shift. We apply a symmetric Bernoulli KL consistency regularizer between the foreground probability maps $p_s = \sigma(L_{o,s})$ and $p_{s,\text{i}} = \sigma(L_{i,s})$:

$$\mathcal{R}_{\text{cons}} = \frac{1}{B} \sum_{s=1}^{B} \Big(\text{KL}(p_s \,\|\, p_{s,\text{i}}) + \text{KL}(p_{s,\text{i}} \,\|\, p_s) \Big), \tag{14}$$

which suppresses domain-dependent shortcuts that vary under domain-embedding interventions and encourages pathology-consistent aggregation across modalities.

4 Experiment

4.1 Datasets and Evaluation Metrics

The proposed IWCNet is evaluated on three multi-center, multi-modality MRI datasets for liver, brain, and prostate lesion segmentation. The **Liver MRI dataset** contains 450 patients from three clinical centers (250/120/80), including 280 hemangioma and 170 hepatocellular carcinoma cases, with three contrast-free modalities (T1/T2/DWI). The **Brain MRI dataset** contains 551 glioma patients from three centers (230/176/145), with three contrast-free modalities (T1/T2/FLAIR). The **Prostate MRI dataset** contains 370 prostate cancer patients from three centers (180/115/75), with three contrast-free modalities (T2W/DWI/ADC). All modalities and annotations are spatially aligned and resampled to a unified resolution for multi-modality fusion.

Method	Liver (T1/T2/DWI)				Brain (T1/T2/FLAIR)				Prostate (T2/DWI/ADC)			
	External C2		External C3		External C2		External C3		External C2		External C3	
	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓
nnU-Net [Isensee *et al.*, 2021]	67.43	16.56	64.27	18.47	69.75	13.67	70.34	14.03	51.63	11.26	49.45	12.65
TransUNet [Chen *et al.*, 2021]	71.34	14.46	67.12	15.27	72.34	10.45	74.78	9.23	53.75	10.63	49.73	12.16
Swin-UNet [Cao *et al.*, 2022]	68.38	16.48	63.74	17.43	70.74	11.27	73.26	9.83	52.94	10.86	50.72	12.08
A2FSeg [Wang and Hong, 2023]	71.82	15.36	73.24	13.05	74.23	8.45	70.23	13.26	54.07	9.62	57.36	8.37
DM-FF [Xu *et al.*, 2024]	74.26	12.54	72.34	12.26	73.05	9.36	71.27	10.85	56.86	9.41	58.26	9.12
M2FNet [Chen *et al.*, 2025]	75.23	9.26	71.52	13.25	75.76	7.07	70.34	11.53	58.36	9.25	56.37	11.23
UMF-SegNet [Lu *et al.*, 2025]	76.34	8.45	73.26	11.37	75.85	7.72	72.04	9.47	54.83	10.39	52.83	11.55
LCMF-Net [Cao *et al.*, 2025]	75.34	9.58	73.48	10.78	71.07	8.31	74.65	7.47	56.91	8.56	54.26	8.72
IWCNet (Ours)	**78.15**	**7.16**	**77.52**	**7.95**	**78.56**	**5.75**	**76.87**	**6.36**	**62.79**	**7.05**	**63.27**	**6.74**

Table 1: External-center evaluation under single-center training on three multi-center multi-modality MRI datasets. Models are trained and selected on the largest center (C1), and evaluated on the remaining two external centers (C2 and C3).

Method	In-center Dice↑	Cross-center Dice↑	Drop↓
nnU-Net	73.23	65.34	7.89
TransUNet	74.62	66.26	8.36
Swin-UNet	73.55	64.65	8.90
A2FSeg	79.17	72.38	6.79
DM-FF	79.51	74.63	4.88
M2FNet	76.65	70.39	6.26
UMF-SegNet	78.27	72.38	5.89
LCMF-Net	78.58	72.95	5.63
IWCNet	**81.67**	**79.36**	**2.31**

Table 2: Leave-one-center-out robustness on the liver MRI dataset.

To reflect a realistic deployment setting, we train and select models using only the largest center as the source domain (C1), and evaluate generalization on the remaining two centers (C2 and C3) as external test domains. We utilized the Dice Similarity Coefficient (Dice) and 95th percentile Hausdorff distance (HD95) [Huttenlocher *et al.*, 2002] for evaluating segmentation performance.

4.2 Implementation

All experiments are implemented in PyTorch. Hyperparameters are tuned on the validation set to achieve optimal performance. We use the Adam optimizer [Zhang, 2018] with a batch size of 1, and the initial learning rate is set to 1×10^{-4}, selected from $\{10^{-3}, 10^{-4}, 10^{-5}\}$ with an exponential decay factor of 0.95. Models are trained for 300 epochs. Training takes approximately 16 hours, and inference requires 0.38 seconds on average per test sample, where each sample consists of aligned multi-modality inputs of spatial size 256×256, using two NVIDIA RTX 3090 Ti GPUs.

4.3 Comparisons with the State-of-the-art

To comprehensively evaluate the effectiveness of IWCNet, we compare it with eight representative multi-modality MRI segmentation baselines, including nnU-Net, TransUNet, Swin-UNet, A2FSeg, DM-FF, M2FNet, UMF-SegNet, and LCMF-Net. These methods represent a broad spectrum of widely used fusion paradigms, ranging from early fusion and multi-stream multi-level fusion to more recent adaptive fusion and Transformer-based cross-modality interaction mechanisms. To ensure a rigorous and fair comparison, all methods are trained and evaluated under an identical single-center training protocol, preprocessing pipeline, and optimization schedule.

As reported in Tab.1, IWCNet consistently achieves the best external-center performance across all three datasets and both external centers, yielding superior Dice and lower HD95. On the liver dataset, IWCNet improves external-center Dice by 1.81–4.04 points over the second-best method while reducing HD95 by 1.29–2.83. On the brain dataset, it boosts external Dice by 2.09–2.71 and reduces HD95 by 1.11–1.32. On the prostate dataset, it delivers larger gains, improving external Dice by 4.43–5.01 and reducing HD95 by 1.51–1.63. These results indicate that IWCNet provides more accurate lesion delineation and better boundary quality under cross-center domain shift. Fig.3 further provides qualitative comparisons, where IWCNet produces cleaner lesion boundaries and fewer false positives in challenging cases.

To further assess robustness under domain shift, Tab.2 reports the leave-one-center-out (LOCO) evaluation on the liver dataset, where In-center Dice is measured on the seen training centers, Cross-center Dice on the held-out unseen center, and Drop denotes the Dice decrease. Notably, IWCNet yields the smallest performance drop (2.31), substantially outperforming the competing fusion baselines. This narrowed generalization gap confirms that our approach learns more transferable cross-modality fusion behavior, effectively mitigating the performance degradation caused by domain shift.

4.4 Ablation Study

Ablation Study of Proposed Components. We conduct progressive module-level ablations on three datasets to quantify how each design in IWCNet contributes to robust cross-center fusion. Tab. 3 reports Dice/HD95 averaged over the two external centers. All variants share the same backbone

Variant	Modules				Liver		Brain		Prostate	
	INT	OLDM	EDCC	Cons	Dice↑	HD95↓	Dice↑	HD95↓	Dice↑	HD95↓
A0: Base-Concat	×	×	×	×	66.52	17.36	68.32	13.26	49.86	13.03
A1: +Attention Fusion	×	×	×	×	68.45	16.02	70.47	11.16	50.34	12.36
A2: +INT	✓	×	×	×	72.58	12.73	72.37	9.59	53.26	10.85
A3: +INT +Cons	✓	×	×	✓	73.27	11.83	72.92	8.47	54.73	9.49
A4: +INT +OLDM +EDCC	✓	✓	✓	×	76.74	8.95	75.49	7.58	60.81	8.41
A5: Full (IWCNet)	✓	✓	✓	✓	**77.84**	**7.56**	**77.72**	**6.06**	**63.03**	**6.90**

Table 3: Progressive module-level ablation on three multi-center multi-modality MRI datasets under single-center training. Dice/HD95 are averaged over the two unseen test centers. INT denotes embedding intervention; Cons denotes the cross-branch consistency regularization.

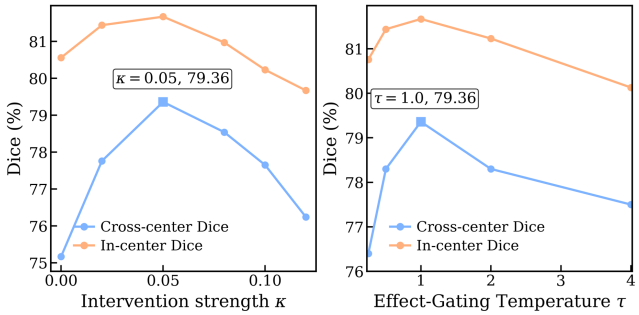

Figure 4: Ablation on the key hyperparameters in the liver multi-center setting (LOCO). (a) intervention strength κ and (b) effect-gating temperature τ.

and training settings. A0 is the concatenation baseline. A1 replaces concatenation with a standard attention fusion block. A2 activates INT, which constructs an intervened branch by applying embedding intervention while keeping content features fixed, enabling the model to observe prediction effects. A3 further adds the cross-branch consistency regularizer (Cons), enforcing agreement between fused predictions from the original and intervened branches. A4 enables OLDM together with EDCC to perform effect-driven weight calibration, while disabling Cons to decouple effect-based weighting from regularization. A5 is the full IWCNet.

Overall, A1 provides only marginal improvements over A0 across all datasets, indicating that correlation-to-weight mappings learned on the source center transfer poorly under domain shift. Introducing INT alone (A2) yields clear gains on both Dice and HD95, improving external Dice by 4.13/1.90/2.92 and reducing HD95 by 3.29/1.57/1.51, respectively, suggesting that embedding intervention exposes the model to plausible domain variations and mitigates domain-dependent shortcuts. Adding Cons on top of INT (A3) further stabilizes predictions, leading to additional HD95 reductions and consistent Dice improvements, demonstrating the benefit of enforcing invariance between the original and intervened fused outputs. When OLDM and EDCC are activated (A4), performance improves substantially, especially on boundary quality, with HD95 dropping to 8.95/7.58/8.41 and Dice increasing to 76.74/75.49/60.81, confirming that effect-driven calibration yields more reliable sample-wise fusion under domain shift. Finally, the full model (A5) achieves the best overall performance on all datasets, indicating that effect-driven calibration and cross-branch consistency act complementarily to produce transferable cross-modality aggregation.

Sensitivity to the Intervention Strength κ. We investigate how the intervention strength affects robustness under domain shift. Specifically, we construct the intervened embedding by applying a controlled intervention to the original domain embedding with different magnitudes κ, and report cross-center performance on the liver dataset. As shown in Fig. 4(a), increasing κ from 0 to 0.05 consistently improves Cross-center Dice (from 75.17% to 79.36%) and reduces the cross-center drop, indicating that moderate embedding interventions expose the model to realistic center-induced appearance variations and yield more reliable effect signals. When κ becomes large, performance decreases, suggesting that overly strong interventions deviate from plausible domain variations and destabilize the intervention effect.

Sensitivity to the Effect-Gating Temperature τ. We further study the temperature τ in Monotonic Effect Gating within EDCC, which controls the *sharpness* of mapping effect scores to fusion weights (Eq. 11). As shown in Fig. 4(b), very small τ leads to overly peaky, near-hard gating, causing the fusion to be dominated by a single modality and making the weights sensitive to estimation noise. Consequently, cross-center performance degrades. In contrast, large τ yields overly smooth (nearly uniform) weights, which diminishes the benefit of effect-driven calibration and approaches unweighted fusion. The best trade-off is achieved at $\tau = 1$, where IWCNet attains the highest Cross-center Dice (79.36%) with the lowest Drop, demonstrating that an appropriately calibrated gating sharpness is crucial for translating effect evidence into robust and stable sample-wise fusion.

5 Conclusion

In this paper, we propose IWCNet for multi-center multi-modality MRI lesion segmentation. IWCNet models domain shift as a controllable embedding and leverages explicit interventions to mitigate its impact on fusion weight estimation, enabling reliable cross-modality aggregation under domain shift. We evaluate IWCNet on three multi-center multi-modality datasets and compare it with eight state-of-the-art methods. The results consistently show that IWCNet achieves the best overall performance, demonstrating robust and accurate lesion segmentation under domain shift.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (62476001, 62406111), Anhui Province Excellent Scientific Research and Innovation Team (2024AH010004), the Anhui Province Clinical Medicine Research Transformation Project (202304295107020031), the Research Fund of Anhui Institute of Translational Medicine (2022zhyx-C72), and the Research Fund of Health Commission of Anhui Province (AHWJ2023A20115).

Contribution Statement

Ronghui Qi and Wenlong Song contributed equally to this work. Wanyin Shi and Chenchu Xu served as the corresponding authors.

References

[Albiin, 2012] Nils Albiin. Mri of focal liver lesions. *Current Medical Imaging*, 8(2):107–116, 2012.

[Atad *et al.*, 2024] Matan Atad, David Schinz, Hendrik Moeller, Robert Graf, Benedikt Wiestler, Daniel Rueckert, Nassir Navab, Jan S Kirschke, and Matthias Keicher. Counterfactual explanations for medical image classification and regression using diffusion autoencoder. *arXiv preprint arXiv:2408.01571*, 2024.

[Bharwani and Koh, 2013] N Bharwani and DM Koh. Diffusion-weighted imaging of the liver: an update. *Cancer imaging*, 13(2):171, 2013.

[Cao *et al.*, 2022] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[Cao *et al.*, 2025] Guogang Cao, Zhaojun Yang, Wanying Liang, Sai Zhang, Tao Zhong, Hongdong Mao, Dong Wang, and Ming Zong. Lcmf-net: A lightweight collaborative multimodal fusion network for brain tumor segmentation. *Neural Networks*, page 108257, 2025.

[Chen *et al.*, 2021] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[Chen *et al.*, 2025] Shannan Chen, Xuanhe Zhao, Yang Duan, Ronghui Ju, Peizhuo Zang, and Shouliang Qi. M2fnet: multi-modality multi-level fusion network for segmentation of acute and sub-acute ischemic stroke. *Complex & Intelligent Systems*, 11(6):1–24, 2025.

[Ding *et al.*, 2021] Yuhang Ding, Xin Yu, and Yi Yang. Rfnet: Region-aware fusion network for incomplete multi-modal brain tumor segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3975–3984, 2021.

[Dolz *et al.*, 2018] Jose Dolz, Karthik Gopinath, Jing Yuan, Herve Lombaert, Christian Desrosiers, and Ismail Ben Ayed. Hyperdense-net: A hyper-densely connected cnn for multi-modal image segmentation. *IEEE transactions on medical imaging*, 38(5):1116–1126, 2018.

[Guan and Liu, 2021] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis: a survey. *IEEE Transactions on Biomedical Engineering*, 69(3):1173–1185, 2021.

[Gurney-Champion *et al.*, 2020] Oliver J Gurney-Champion, Faisal Mahmood, Marcel van Schie, Robert Julian, Ben George, Marielle EP Philippens, Uulke A van der Heide, Daniela Thorwarth, and Kathrine R Redalen. Quantitative imaging for radiotherapy purposes. *Radiotherapy and Oncology*, 146:66–75, 2020.

[Havaei *et al.*, 2016] Mohammad Havaei, Nicolas Guizard, Nicolas Chapados, and Yoshua Bengio. Hemis: Hetero-modal image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 469–477. Springer, 2016.

[Huang *et al.*, 2025] Yunzhi Huang, Luyi Han, and Haoran Dou. Generative feature style augmentation for domain generalization in medical image segmentation. *Pattern Recognition*, 162:111416, 2025.

[Huttenlocher *et al.*, 2002] Daniel P Huttenlocher, Gregory A. Klanderman, and William J Rucklidge. Comparing images using the hausdorff distance. *IEEE Transactions on pattern analysis and machine intelligence*, 15(9):850–863, 2002.

[Isensee *et al.*, 2021] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[Kamnitsas *et al.*, 2017] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3d cnn with fully connected crf for accurate brain lesion segmentation. *Medical image analysis*, 36:61–78, 2017.

[Li *et al.*, 2023] Heng Li, Haojin Li, Wei Zhao, Huazhu Fu, Xiuyun Su, Yan Hu, and Jiang Liu. Frequency-mixed single-source domain generalization for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 127–136. Springer, 2023.

[Lu *et al.*, 2025] Haotian Lu, Mingyang Yu, Xinjian Wei, Xiaoxuan Xu, and Jing Xu. Unbiased multimodal fusion for medical image segmentation based on dual-stream adapter. *Knowledge-Based Systems*, page 114653, 2025.

[Lv *et al.*, 2025] Peiqing Lv, Yaonan Wang, Min Liu, Zhe Zhang, Yunfeng Ma, Licheng Liu, and Erik Meijering. Ciseg: Unsupervised cross-modality adaptation for 3d medical image segmentation via causal intervention. *IEEE Transactions on Medical Imaging*, 2025.

[Mertes *et al.*, 2022] Silvan Mertes, Tobias Huber, Katharina Weitz, Alexander Heimerl, and Elisabeth André. Ganterfactual—counterfactual explanations for medical non-

experts using generative adversarial learning. *Frontiers in artificial intelligence*, 5:825565, 2022.

[Moreira *et al.*, 2025] Catarina Moreira, Yu-Liang Chou, Chihcheng Hsieh, Chun Ouyang, João Pereira, and Joaquim Jorge. Benchmarking instance-centric counterfactual algorithms for xai: from white box to black box. *ACM Computing Surveys*, 57(6):1–37, 2025.

[Parikh *et al.*, 2008] Tejas Parikh, Stephen J Drew, Vivian S Lee, Samson Wong, Elizabeth M Hecht, James S Babb, and Bachir Taouli. Focal liver lesion detection and characterization with diffusion-weighted mr imaging: comparison with standard breath-hold t2-weighted imaging. *Radiology*, 246(3):812–822, 2008.

[Safdari *et al.*, 2025] Reza Safdari, Mohammad-Ali Nikouei Mahani, Mohamad Koohi-Moghadam, and Kyongtae Tyler Bae. Mixstyleflow: Domain generalization in medical image segmentation using normalizing flows. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 376–385. Springer, 2025.

[Singla *et al.*, 2023] Sumedha Singla, Motahhare Eslami, Brian Pollack, Stephen Wallace, and Kayhan Batmanghelich. Explaining the black-box smoothly—a counterfactual approach. *Medical Image Analysis*, 84:102721, 2023.

[Wang and Hong, 2023] Zirui Wang and Yi Hong. A2fseg: Adaptive multi-modal fusion network for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 673–681. Springer, 2023.

[Wang *et al.*, 2021] Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, and Jiangyun Li. Transbts: Multimodal brain tumor segmentation using transformer. In *International conference on medical image computing and computer-assisted intervention*, pages 109–119. Springer, 2021.

[Xu *et al.*, 2024] Chenchu Xu, Xue Wu, Boyan Wang, Jie Chen, Zhifan Gao, Xiujian Liu, and Heye Zhang. Accurate segmentation of liver tumor from multi-modality non-contrast images using a dual-stream multi-level fusion framework. *Computerized Medical Imaging and Graphics*, 116:102414, 2024.

[Yan *et al.*, 2019] Wenjun Yan, Yuanyuan Wang, Shengjia Gu, Lu Huang, Fuhua Yan, Liming Xia, and Qian Tao. The domain shift problem of medical image segmentation and vendor-adaptation by unet-gan. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 623–631. Springer, 2019.

[Yang *et al.*, 2025] Heran Yang, Wenbo Hua, Zongben Xu, and Jian Sun. Domain-generalized discrete diffusion model for cross-domain medical image segmentation. *IEEE Transactions on Medical Imaging*, 2025.

[Zhang *et al.*, 2022] Yao Zhang, Nanjun He, Jiawei Yang, Yuexiang Li, Dong Wei, Yawen Huang, Yang Zhang, Zhiqiang He, and Yefeng Zheng. mmformer: Multimodal medical transformer for incomplete multimodal learning of brain tumor segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 107–117. Springer, 2022.

[Zhang, 2018] Zijun Zhang. Improved adam optimizer for deep neural networks. In *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)*, pages 1–2. Ieee, 2018.

[Zhao *et al.*, 2022] Zechen Zhao, Heran Yang, and Jian Sun. Modality-adaptive feature interaction for brain tumor segmentation with missing modalities. In *International conference on medical image computing and computer-assisted intervention*, pages 183–192. Springer, 2022.

[Zhou *et al.*, 2021] Kaiyang Zhou, Yongxin Yang, Yu Qiao, and Tao Xiang. Domain generalization with mixstyle. *arXiv preprint arXiv:2104.02008*, 2021.

[Zhou *et al.*, 2022] Tongxue Zhou, Su Ruan, Pierre Vera, and Stéphane Canu. A tri-attention fusion guided multi-modal segmentation network. *Pattern Recognition*, 124:108417, 2022.