Edge-Enhanced Dilated Residual Attention Network for Multimodal Medical Image Fusion

IEEE BIBM 2024 Short Paper

Meng (Simon) Zhou[*1], Yuxuan Zhang [*1], Xiaolan Xu [1], Jiayi Wang [2], and Farzad Khalvati [1,3]

[1] Department of Computer Science, University of Toronto

[2] Desautels Faculty of Management, McGill University

[3] Department of Medical Imaging, University of Toronto

UNIVERSITY OF TORONTO

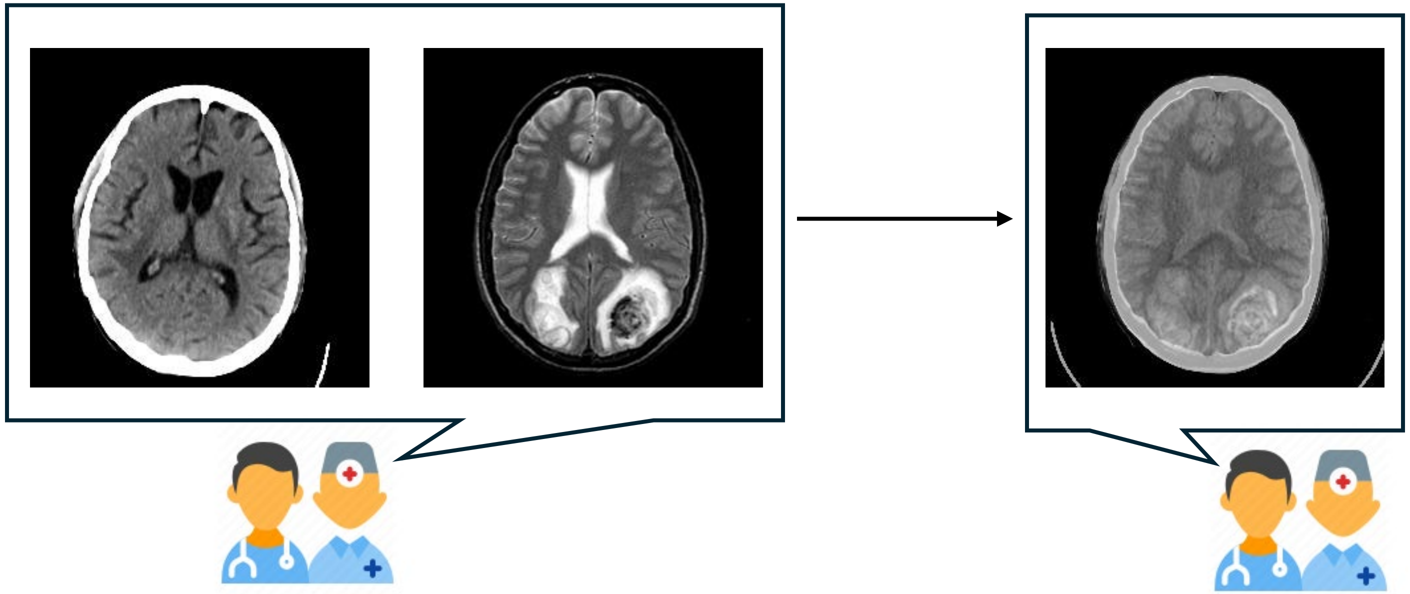
McGill UNIVERSITY

Outline

- Background
- Related Works
- Problem
- Proposed Method
- Experiments
- Conclusions and Future work

Background

- Multimodal image fusion plays an increasingly prominent role in clinical diagnosis.
- It aims to aggregate common and complementary information from different image modalities as well as integrate the information to generate more clearer and informative images
- Physicians must analyze multiple images to make informed decisions, a process that is both time-consuming and laborious.



Related works

- CNN-based:
  - IFCNN [1]: CNN-based image fusion framework for multi-focus, infrared-visible, and multimodal medical image fusion. Elementwise fusion rules to combine feature maps directly.
  - MSRPAN [2]: Residual pyramid attention network for multimodal medical image fusion and Feature Energy Ratio Strategy to fuse feature maps
  - MSDRA [3]: Double residual attention network for multimodal medical image fusion and uses weighted L1 Norm to fuse feature maps

Limitations...

1. Losing the structural information and edge details, which are crucial for medical images
2. Lack of multiscale learning capabilities

[1] Zhang, Y., Liu, Y., Sun, P., Yan, H., Zhao, X., & Zhang, L. (2020). IFCNN: A general image fusion framework based on convolutional neural network. *Information Fusion*, *54*, 99-118.
[2] Fu, J., Li, W., Du, J., & Huang, Y. (2021). A multiscale residual pyramid attention network for medical image fusion. *Biomedical Signal Processing and Control*, *66*, 102488.
[3] Li, W., Peng, X., Fu, J., Wang, G., Huang, Y., & Chao, F. (2022). A multiscale double-branch residual attention network for anatomical–functional medical image fusion. *Computers in biology and medicine*, *141*, 105005.

Related works

- Transformer-based:
  - SwinFusion [3]: Combining a CNN feature extractor with a cross-domain transformer model to fuse local and global information.
  - MRSCFusion [4]: Combining a multiscale CNN model and applied residual Swin Transformer layers to fuse cross-domain information.
  - MACTFusion [5]:Light-weight cross modality transformer with window and grid attention.

Limitations…

1. Computational cost is high and hinders the clinical application.
2. Applying the fusion method to real clinical diagnosis tasks is not fully explored.

[3] Ma, J., Tang, L., Fan, F., Huang, J., Mei, X., & Ma, Y. (2022). SwinFusion: Cross-domain long-range learning for general image fusion via swin transformer. *IEEE/CAA Journal of Automatica Sinica*, *9*(7), 1200-1217.

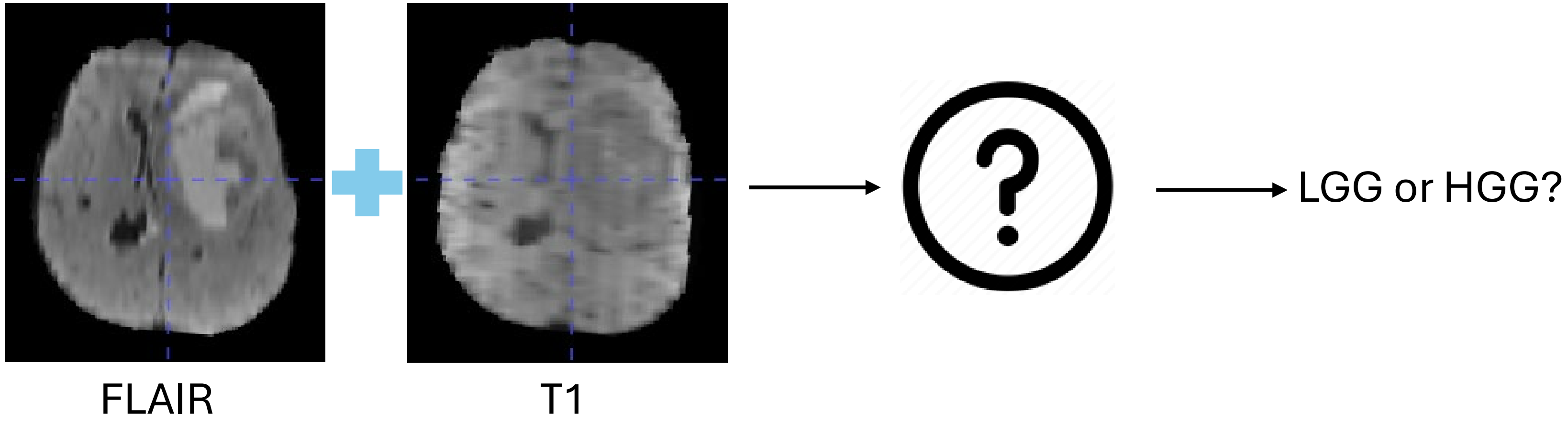
[4] Xie, X., Zhang, X., Ye, S., Xiong, D., Ouyang, L., Yang, B., ... & Wan, Y. (2023). MRSCFusion: Joint residual Swin transformer and multiscale CNN for unsupervised multimodal medical image fusion. *IEEE Transactions on Instrumentation and Measurement*.

[5] Xie, X., Zhang, X., Tang, X., Zhao, J., Xiong, D., Ouyang, L., ... & Teo, K. L. (2024). MACTFusion: Lightweight Cross Transformer for Adaptive Multimodal Medical Image Fusion. *IEEE Journal of Biomedical and Health Informatics*.

Problem

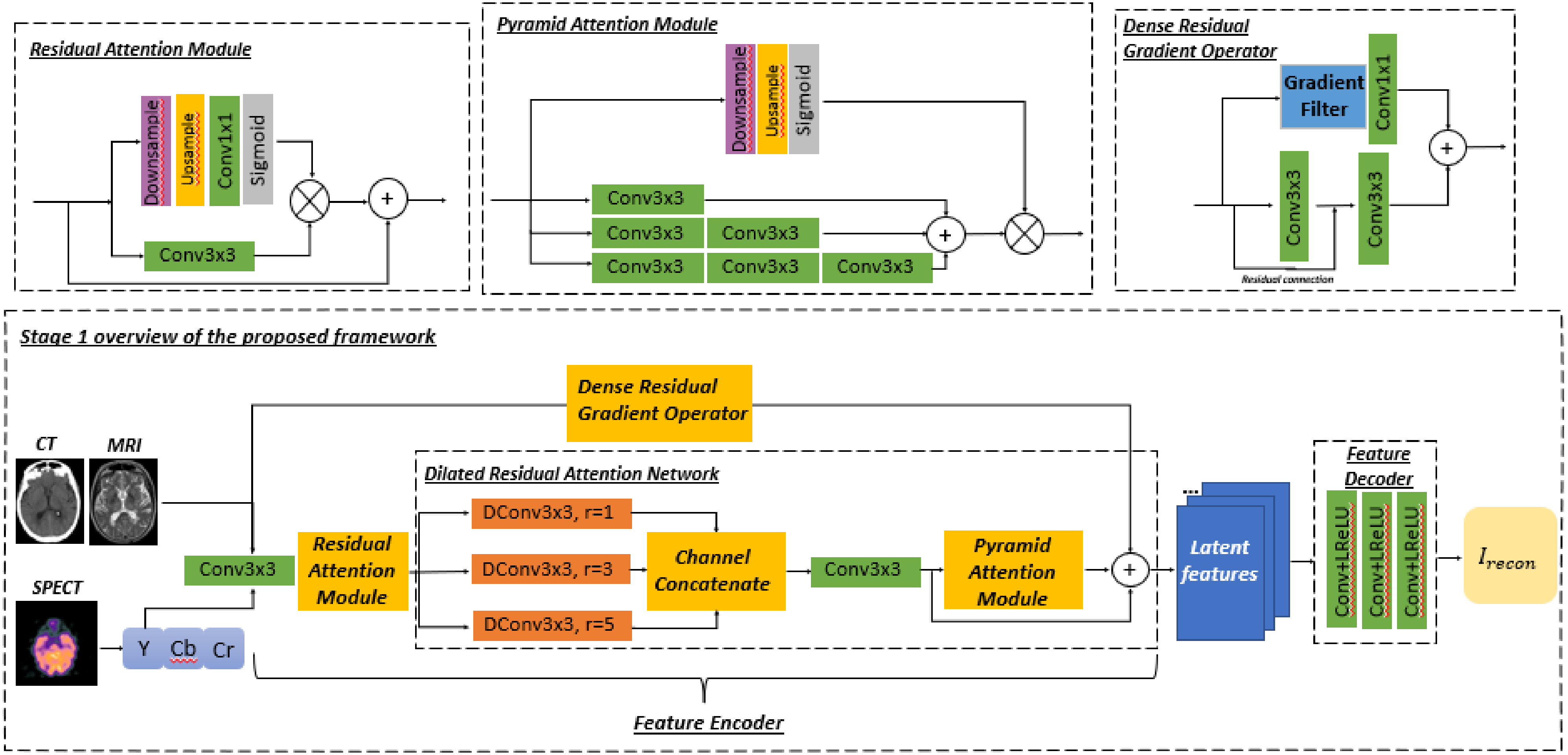
- Two most common fusion tasks in medical imaging: MRI-CT and MRI-SPECT fusion tasks

- To further evaluate the effectiveness of the fusion method, we apply it to a downstream clinical brain tumor pathology classification task between Low-Grade and High-Grade Gliomas.



Proposed Method

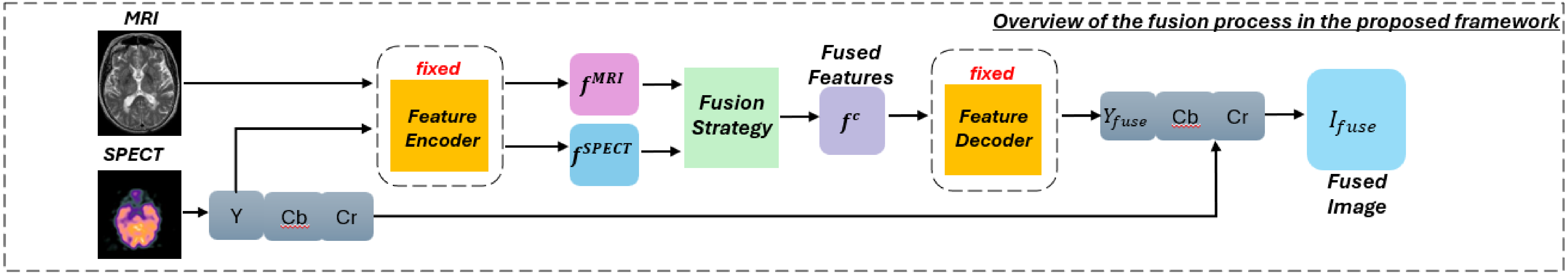
Stage 1



Asymmetric Autoencoder (Stage 1 model) in the proposed framework

Proposed Method

Stage 2



SFNN Strategy

$$W_k = \frac{\phi(\|S(x_i)^k\|_*)}{\sum_{k=1}^{C} \phi(\|S(x_i)^k\|_*)}$$

Loss function

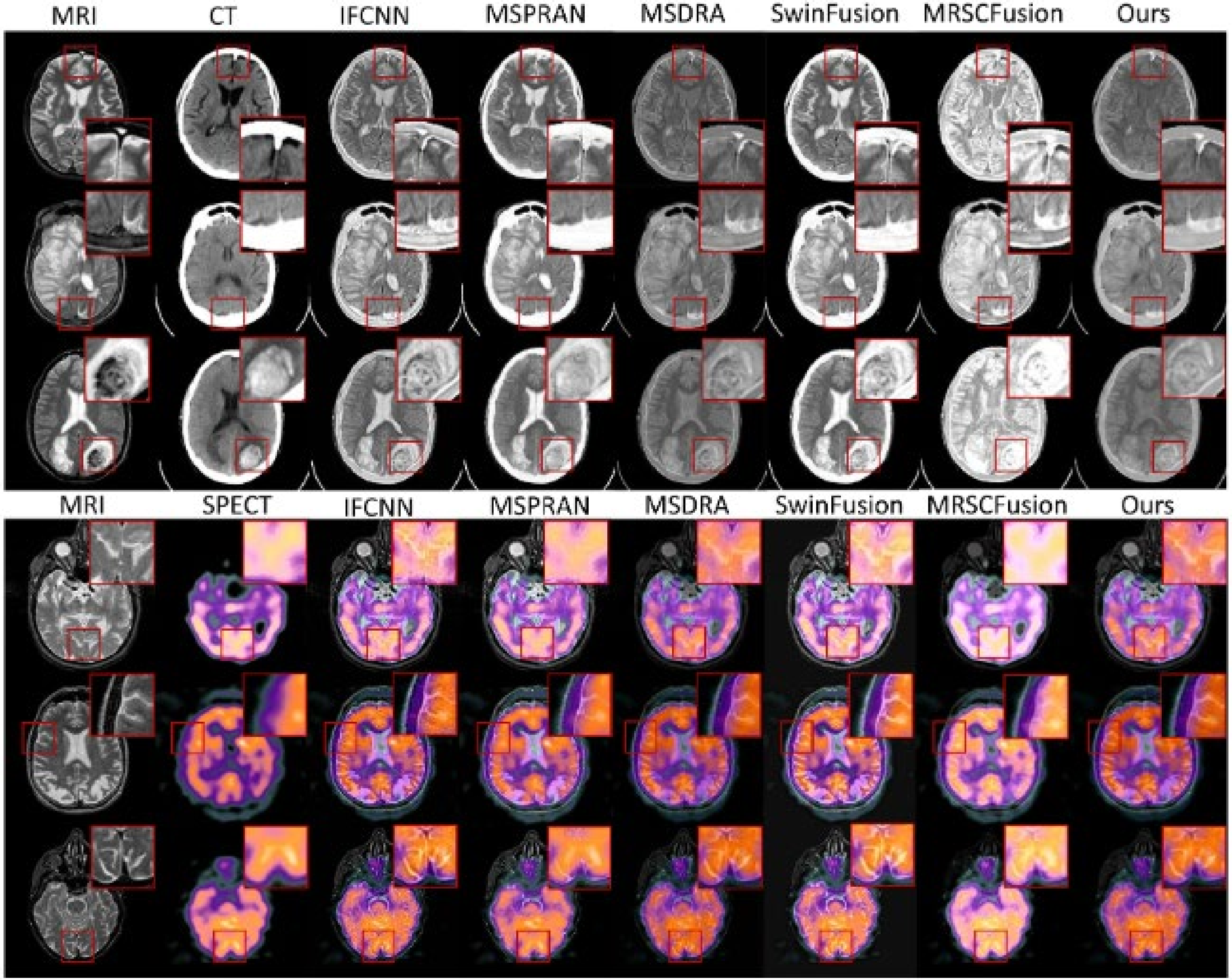
$$\mathcal{L}_{pixel} = \|x - \hat{x}\|_2^2, \ \mathcal{L}_{grad} = \|\nabla x - \nabla \hat{x}\|_2^2,$$

$$\mathcal{L}_{perp} = \sum_{k=1}^{C} \|f_i^k(x) - f_i^k(\hat{x})\|_2^2$$

$$\mathcal{L}(\theta) = \mathcal{L}_{pixel} + \lambda_1 * \mathcal{L}_{grad} + \lambda_2 * \mathcal{L}_{perp}$$

Experiments

- Qualitative Results



Experiments

- Quantitative Results

Dataset	Method	PSNR	SSIM	FMI	FSIM	EN
MRI-CT	IFCNN [7]	15.594±0.112	0.700±0.015	0.870±0.012	0.801±0.001	8.968±0.227
	MSRPAN [8]	14.790±0.233	0.749±0.003	0.744±0.001	0.804±0.001	7.773±0.273
	MSDRA [9]	15.308±0.437	0.742±0.037	0.872±0.002	0.788±0.005	9.554±0.767
	SwinFusion [13]	14.962±0.173	**0.768±0.007**	0.882±0.002	0.810±0.001	8.445±0.078
	MRSCFusion [1]	14.476±0.205	0.713±0.012	0.877±0.006	0.791±0.010	7.544±0.232
	EH-DRAN(Ours)	**16.830±0.490**	0.753±0.007	**0.883±0.005**	**0.820±0.003**	**10.727±0.531**
MRI-SPECT	IFCNN [7]	19.728±0.228	0.721±0.025	0.846±0.062	0.783±0.027	10.167±0.429
	MSRPAN [8]	19.174±0.046	0.732±0.002	0.838±0.003	0.793±0.002	9.737±0.202
	MSDRA [9]	19.662±0.165	0.725±0.003	0.839±0.003	0.794±0.003	10.784±0.447
	SwinFusion [13]	17.557±0.021	0.728±0.004	0.808±0.007	0.819±0.011	**13.066±0.428**
	MRSCFusion [1]	18.412±0.211	0.734±0.012	0.827±0.009	0.814±0.006	9.87±0.600
	EH-DRAN(Ours)	**21.455±0.071**	**0.736±0.002**	**0.876±0.004**	**0.843±0.003**	11.970±0.538

Bold and underline numbers represent the best and second-best results for each dataset, respectively

Experiments

- Ablation study

Dataset	Method	PSNR	SSIM	FMI	FSIM	Entropy
MRI-CT	Base Model	15.623±0.032	0.745±0.013	0.878±0.003	0.802±0.003	9.122±0.706
	Base Model+$\mathcal{L}_{grad}$	16.355±0.038	0.749±0.010	0.881±0.002	0.818±0.002	9.771±0.528
	Base Model+$\mathcal{L}_{grad}$+DRGO	**16.830±0.490**	**0.753±0.007**	**0.883±0.005**	**0.820±0.003**	**10.727±0.531**
MRI-SPECT	Base Model	20.698±0.002	**0.743±0.008**	0.833±0.006	0.836±0.005	10.010±0.563
	Base Model+$\mathcal{L}_{grad}$	20.738±0.026	0.740±0.011	0.837±0.002	0.838±0.004	10.454±0.426
	Base Model+$\mathcal{L}_{grad}$+DRGO	**21.455±0.071**	0.736±0.002	**0.876±0.004**	**0.843±0.003**	**11.970±0.538**

- Fusion time comparison

	IFCNN	MSRPAN	MSDRA	SwinFusion	MRSCFusion	Ours
Params(M)	0.08	0.10	0.20	0.97	23.00	0.50
Time(s)	0.89	0.79	0.81	1.31	2.85	1.26

Experiments

- Results on ROI-based LGG/HGG tumor pathology types classification

	AUC	F1-Score	Accuracy
T2 (1-channel)	0.722±0.021	0.703±0.018	0.604±0.037
FLAIR (1-channel)	0.727±0.024	0.701±0.008	0.611±0.017
T2+FLAIR (2-channel)	0.723±0.028	0.717±0.012	0.640±0.015
Fused (1-channel)	**0.769±0.003**	**0.723±0.006**	**0.640±0.011**

Conclusions and Future Work

- Novel asymmetric autoencoder architecture incorporating a Dilated Residual Attention Network (DRAN) for effective multi-scale feature extraction
- Integrated a Dense Residual Gradient Operator (DRGO) as an edge enhancer to capture fine-grained edge details
- Introduced a family of parameter-free fusion strategies for multimodal image fusion, designed to operate without requiring parameter computation during both training and inference phases
- Extensive evaluated on three datasets to valid the effectiveness of the proposed approach

- Future Work: Extend to 3D, Explore Mamba-based methods.

Thank you for your listening

- Code:

- Arxiv extended version: