An Attention-based Multi-Scale Feature Learning Network for Multimodal Medical Image Fusion

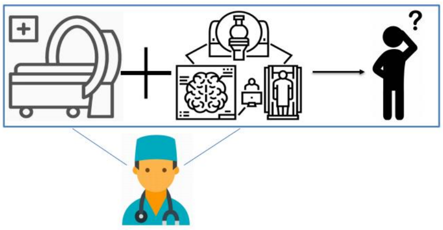
Meng Zhou[1,2], Xiaolan Xu[1], Yuxuan Zhang[1]

[1]Department of Computer Science, University of Toronto, Canada

[2]The Hospital for Sick Children, Toronto, Canada

Motivation

- Multi-modal medical images can provide doctors with a wealth of information about a patient's condition. This can help them make an accurate diagnosis and determine the best course of treatment.
- Analyzing multiple images can be time-consuming and tedious.



- Image fusion is a way to combine important features from several images into a single image. In multimodal image fusion, doctors can get gain more information from the fused image that is more comprehensive and contains more details.

OBJECTIVE: ***propose a feature learning framework for multimodal medical image fusion using Convolutional Neural Networks with Attention and Residual mechanism to improve the fusion performance***

Related Work

- The traditional method of medical image fusion focuses on the transform domain and can be applied at the pixel, feature, and decision levels. Neural networks with **multi-scale decomposition** are effective in dealing with uncertainty and improving fusion efficiency.
- Hermessi et al. [1] proposed a **CNN-based** method in the shearlet domain for extracting high-frequency fusion feature maps. Lahoud et al. [2] proposed a real-time image fusion method using only the pretrained VGG19 model.
- Fu et al. [3] introduced a residual pyramid attention structure: **MSRPAN** framework, which combines the residual attention and pyramid attention mechanisms in the fusion process
- Li et al. [4] proposed the **MSDRA** network, which combines a residual network and attention to capture important detailed features without causing gradient issues in the network.

References

[1] H. Hermessi, O. Mourali, and E. Zagrouba, "Multimodal medical image fusion review: Theoretical background and recent advances," Signal Processing, vol. 183, p. 108036, 2021

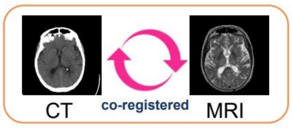
[2] Lahoud, F., & Süsstrunk, S. (2019, July). Zero-learning fast medical image fusion. In *2019 22th International Conference on Information Fusion (FUSION)* (pp. 1-8). IEEE.

[3] J. Fu, W. Li, J. Du, and Y. Huang, "A multiscale residual pyramid attention network for medical image fusion," Biomedical Signal Processing and Control, vol. 66, p. 102488, 2021

[4] W. Li, X. Peng, J. Fu, G. Wang, Y. Huang, and F. Chao, "A multi-scale double-branch residual attention network for anatomical–functional medical image fusion," Computers in Biology and Medicine, vol. 141, p. 105005, 2022.

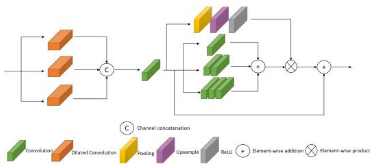
Materials and Method

- **Data**: The Whole Brain Atlas  → **184 pairs**



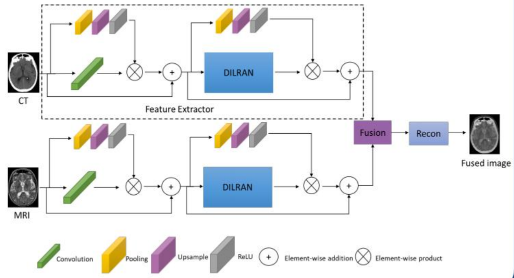
Data pre-processing:

- **Model Architecture**



DILRAN

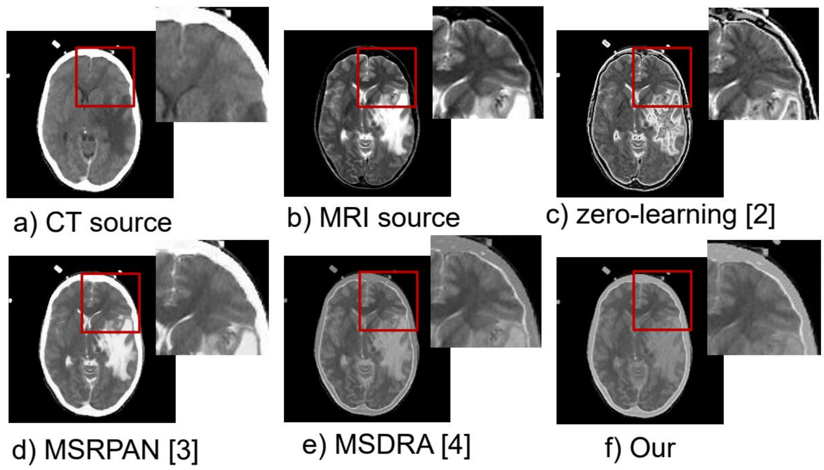
- The attention mechanism improves the feature expression capability by producing weights to each individual pixel according to their importance
- Three different losses are incorporated in the training phase: L2 distance, image gradient differences, and perceptual differences



Pipeline Overview

Experimental Results

- ***Qualitative Results***



- ***Quantitative Results***

Between algorithms:

	PSNR	SSIM	MI	FMI	FSIM	EN
Zero-learning	13.525	0.681	4.633	0.836	0.738	4.279
MSRPAN	15.693	0.697	4.586	0.867	0.797	8.167
MSDRA	14.528	**0.741**	**4.652**	0.874	0.808	8.969
Ours	**18.258**	0.740	4.558	**0.891**	**0.829**	**9.887**

- ***Ablation study:***

	PSNR	SSIM	MI	FMI	FSIM	EN
Only MSE	**18.610**	0.739	4.428	0.890	**0.829**	8.634
All loss	18.258	**0.740**	**4.558**	**0.891**	**0.829**	**9.887**