#99 MedPrompt: Cross-Modal Prompting for Multi-Task Medical Image Translation

Xuhang Chen, Shenghong Luo, Chi-Man Pun and Shuqiang Wang
sq.wang@siat.ac.cn

Abstract

Translating medical images across different modalities is crucial for synthesizing missing data and aiding clinical diagnosis, but existing techniques often fail to capture cross-modal and global features and are typically tailored to specific modality pairs, limiting their practical utility. To address these challenges, we introduce **MedPrompt**, a multi-task framework designed for efficient translation among diverse modalities. MedPrompt incorporates a **Self-adaptive Prompt Block** that dynamically guides the translation network to handle various modalities effectively. We also introduce the **Prompt Extraction Block** and **Prompt Fusion Block** to efficiently encode cross-modal prompts, and leverage the **Transformer model** to enhance global feature extraction across modalities. Extensive experiments on five datasets covering four modality pairs demonstrate that MedPrompt achieves state-of-the-art visual quality and exhibits excellent generalization capability, highlighting its effectiveness and versatility in cross-modal medical image translation.

Methodology

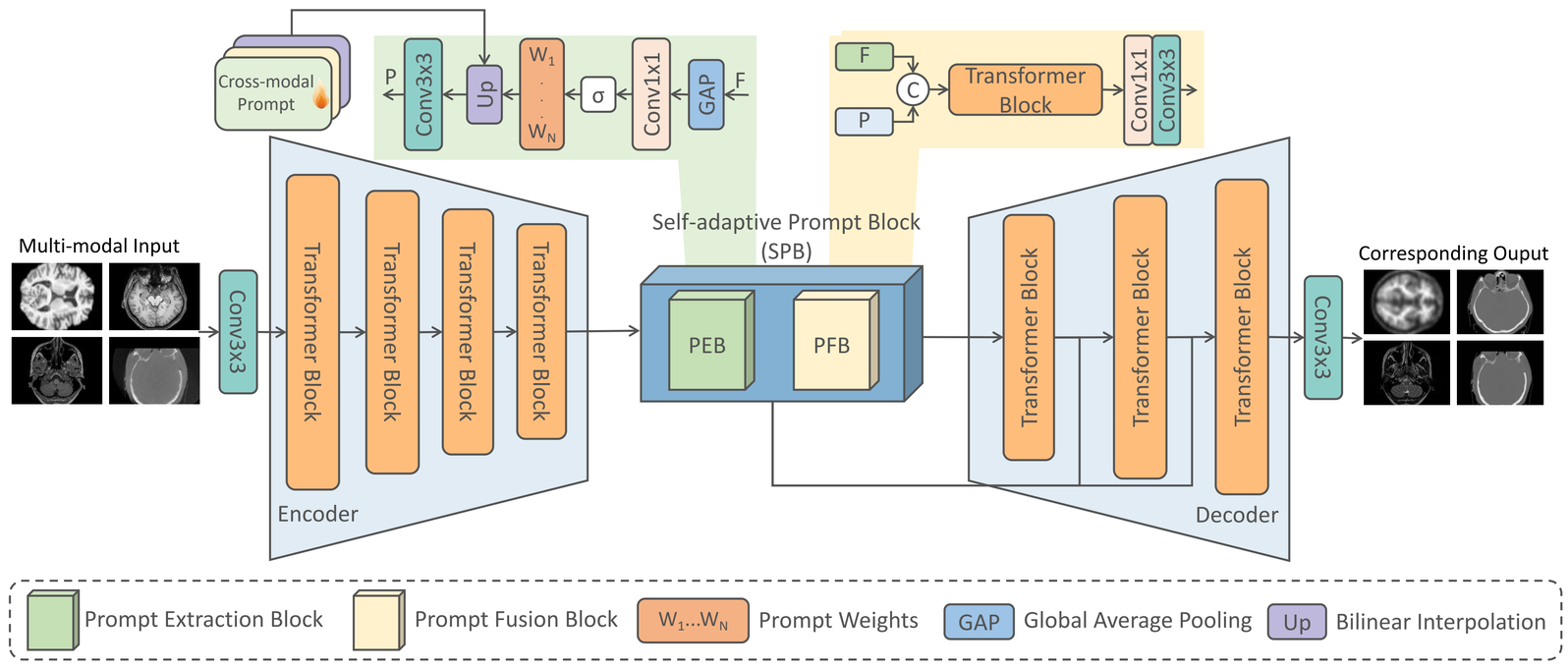


Figure 1: The MedPrompt pipeline follows a conventional encoder-decoder framework. During training, we input all the distinct modalities of the cross-modal dataset. To enhance the model's performance, we introduce the Self-adaptive Prompt Block (SPB) after the 4-level encoder. Additionally, we propose the Prompt Extraction Block (PEB) and Prompt Fusion Block (PFB) to effectively encode and aggregate prompt information from multiple modalities. Each SPB connects the transformer blocks in the encoder, allowing prompt information to propagate between decoders. This process ultimately leads to the generation of high quality results.

Conclusion

We present MedPrompt, a simple yet highly effective multi-task framework for medical image translation that achieves state-of-the-art performance across various modalities. Leveraging Transformer models and efficient cross-modal feature extraction through prompting, MedPrompt demonstrates excellent generalization capability. Key components include the Prompt Extraction Block (PEB), which generates modality-specific prompt weights for selective information extraction, and the Prompt Fusion Block (PFB), which dynamically fuses prompts relevant to the target modality. These innovations enhance multi-task learning, allowing MedPrompt to outperform existing methods with a single training process. Our future work will explore more effective prompt methods to further boost performance and applicability in multi-task medical image translation.

Acknowledgement

This work was supported in part by the National Natural Science Foundations of China under Grant 62172403 and 12326614, the Distinguished Young Scholars Fund of Guangdong under Grant 2021B1515020019, the Excellent Young Scholars of Shenzhen under Grant RCYX20200714114641211, in part by the Science and Technology Development Fund, Macau SAR, under Grant 0141/2023/RIA2 and 0193/2023/RIA3. This research has been conducted using the UK Biobank Resource under Application Number No.75310. This work was performed at SICC which is supported by SKL-IOTSC, University of Macau.

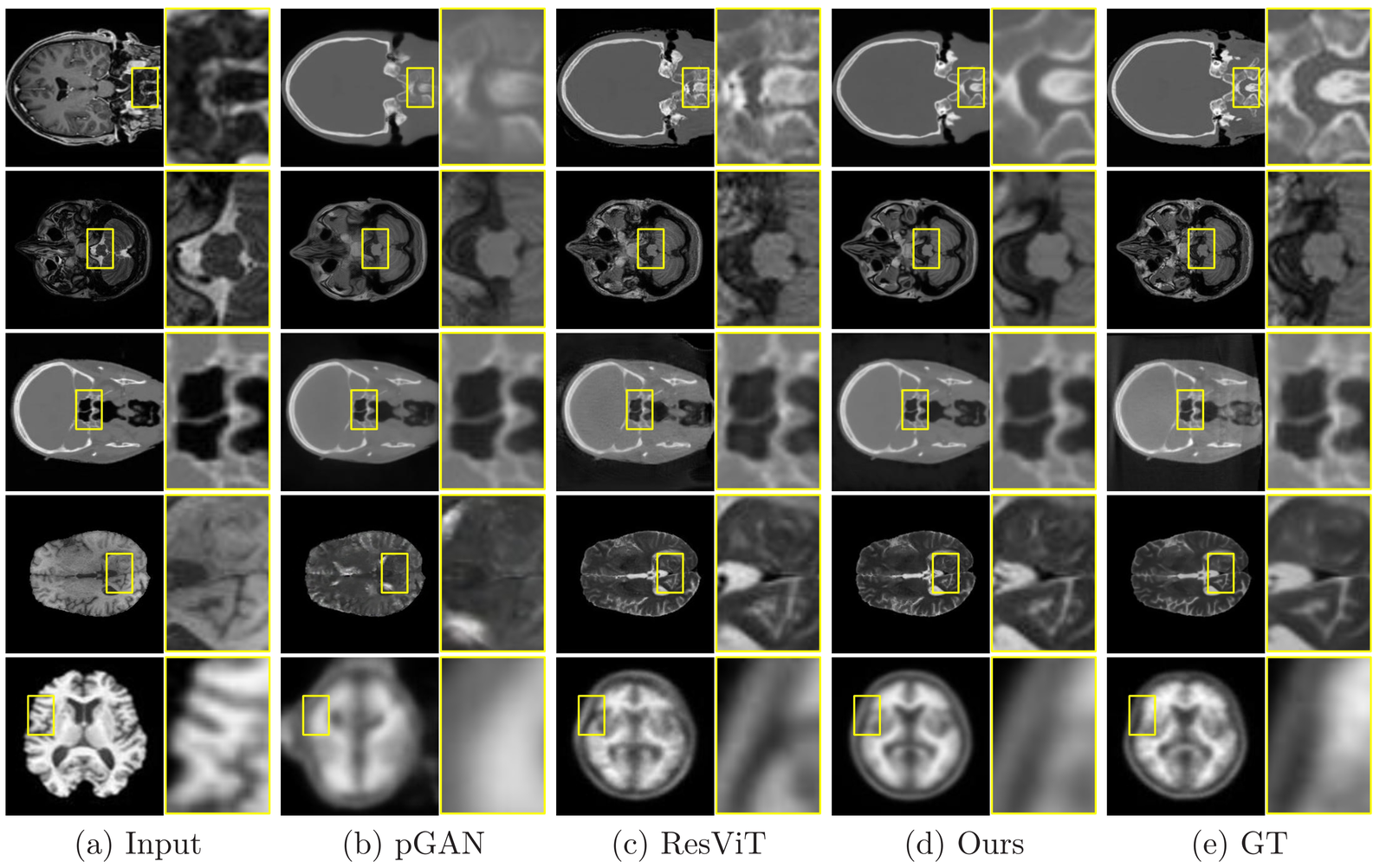
References

[1] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. *Advances in Neural Information Processing Systems*, 36, 2024.

Analysis

Table 1: Quantitative evaluation on SynthRAD2023 dataset.

Method	SynthRAD2023 Task1						SynthRAD2023 Task2					
	MRI → CT			CT → MRI			CBCT → CT			CT → CBCT		
	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓	PSNR↑	SSIM↑	MAE↓
CycleGAN	9.40	0.17	69.07	11.58	0.28	48.34	10.36	0.21	64.19	10.77	0.23	63.54
Pix2Pix	10.04	0.19	65.80	12.36	0.32	43.91	10.42	0.21	63.45	11.29	0.28	58.20
UNIT	9.65	0.17	67.91	12.50	0.31	42.17	10.29	0.20	64.74	10.94	0.24	62.60
MUNIT	10.12	0.20	64.97	12.49	0.31	42.65	10.46	0.21	62.98	11.24	0.27	57.88
FUNIT	8.93	0.04	75.22	8.47	0.11	74.19	10.28	0.46	64.58	9.88	0.44	70.57
U-GAT-IT	21.45	0.76	13.07	16.68	0.51	26.10	23.37	0.81	12.45	20.83	0.69	21.13
CUT	14.14	0.42	44.87	11.43	0.29	45.21	21.03	0.74	18.27	20.37	0.70	21.48
LPTN	16.79	0.56	24.33	13.37	0.34	36.74	22.15	0.79	14.31	21.28	0.71	18.59
medSynth	15.11	0.34	32.79	15.81	0.40	29.52	20.52	0.71	20.70	19.90	0.68	21.62
pGAN	20.78	0.73	15.04	18.19	0.55	20.10	21.49	0.75	14.93	20.71	0.68	19.78
RIED-Net	22.70	0.80	10.84	16.99	0.54	22.93	22.46	0.82	12.56	20.50	0.72	19.81
ResViT	22.98	0.79	**10.39**	18.88	0.58	17.59	24.15	0.82	**9.80**	22.87	0.72	15.58
Ours	**23.33**	**0.83**	10.63	**19.99**	**0.66**	**15.91**	**24.67**	**0.85**	9.83	**23.95**	**0.79**	**13.35**



(a) Input (b) pGAN (c) ResViT (d) Ours (e) GT

Figure 2: We conducted a visual comparison of different image enhancement methods using five distinct datasets. It is apparent that, in comparison to pGAN (b) and ResViT (c), our proposed method (d) demonstrates successful reconstruction of the target with remarkable fidelity.