FreMIM: Fourier Transform Meets Masked Image Modeling for Medical Image Segmentation – Supplementary Material

Wenxuan Wang[1,*] Jing Wang[1,*] Chen Chen[2] Jianbo Jiao[3]
Yuanxiu Cai[1] Shanshan Song[1] Jiangyun Li[1†]
[1]School of Automation and Electrical Engineering, University of Science and Technology Beijing
[2]Center for Research in Computer Vision, University of Central Florida
[3]University of Birmingham
s20200579@xs.ustb.edu.cn, chen.chen@crcv.ucf.edu, leejy@ustb.edu.cn

Overview

In this supplementary material, we provide the following items:

1. (Sec. 1) More detailed information about the adopted three benchmark datasets (*i.e.* BraTS 2019, ISIC 2018 and ACDC 2017).
2. (Sec. 2) Implementation details on the utilized three benchmark datasets (*i.e.* BraTS 2019, ISIC 2018 and ACDC 2017).
3. (Sec. 3) More quantitative results about ablation studies of decoder structure and pre-training loss, as well as more experimental comparison on 3D baselines.
4. (Sec. 4) Visual comparison of reconstruction results and brain tumor segmentation results on BraTS 2019 dataset [1, 2, 6], and skin lesion segmentation on ISIC 2018 dataset [4, 9] for qualitative analysis.

Our code will be publicly available.

1. More Details about the Benchmark Datasets

Our proposed method is evaluated on three benchmark datasets for medical segmentation. The Brain Tumor Segmentation 2019 challenge (**BraTS 2019**) dataset [1, 2, 6] is composed of multi-institutional pre-operative MRI sequences, including 335 patient cases for training and 125 cases for validation. Each sample contains four modalities (FLAIR, T1, T1c, T2) with the size of $240 \times 240 \times 155$, and the corresponding ground truth consists of 4 classes: background (label 0), necrotic and non-enhancing tumor (label 1), peritumoral edema (label 2) and GD-enhancing tumor (label 4). The Dice score and the Hausdorff distance (95%) metrics are used for evaluating the segmentation accuracy of different regions, including enhancing tumor region (ET, label 4), regions of the tumor core (TC, labels 1 and 4), and the whole tumor region (WT, labels 1,2 and 4). The International Skin Imaging Collaboration 2018 (**ISIC 2018**) dataset [4, 9] is a collection of 2594 RGB images of skin lesion for training, around 100 samples for validation, and 1000 samples for testing. Five metrics are specifically employed for the quantitative assessment of model performance, including Dice, Jaccard Index (JI), Accuracy, Recall, and Precision. The Automated Cardiac Diagnosis Challenge 2017 (**ACDC 2017**) dataset [3] is collected from different patient cases using MRI scanners, including 3D cardiac MRI cine for both end-diastolic (ED) and end-systolic (ES) phases instances. The publicly available training dataset consists of 100 patient scans, which are split into 80 training samples and 20 testing samples. The ground truth contains 3 classes: right ventricle (RV), myocardium (Myo) and left ventricle (LV).

2. Implementation Details

Config	Pre-training	Fine-tuning
optimizer	Adam	Adam
base learning rate	10^{-4}	10^{-4}
weight decay	10^{-5}	10^{-5}
batch size	64	64
lr decay schedule	cosine decay	cosine decay
training epochs	250	500

Table 1. Training settings on BraTS 2019 dataset.

The proposed method is implemented in PyTorch [7] and trained with two NVIDIA Geforce RTX 3090 GPUs. The specific training hyper-parameter configurations of our FreMIM on BraTS 2019, ISIC 2018 and ACDC 2017 can be found in Table 1, 2, 3 respectively.

[*]Equal Contribution.[†]Corresponding author.

Config	Pre-training	Fine-tuning
optimizer	SGD	SGD
base learning rate	10^{-3}	5×10^{-4}
weight decay	10^{-8}	10^{-8}
batch size	12	12
lr decay schedule	poly	poly
training epochs	125	300

Table 2. Training settings on ISIC 2018 dataset.

Config	Pre-training	Fine-tuning
optimizer	SGD	SGD
base learning rate	10^{-2}	10^{-2}
weight decay	10^{-4}	10^{-4}
batch size	16	16
lr decay schedule	poly	poly
training epochs	300	1200

Table 3. Training settings on ACDC 2017 dataset.

3. More Quantitative Results.

Importance of the bilateral aggregation decoder (BAD) and focal loss: We also conduct supplementary ablation studies to validate the effectiveness of BAD and focal loss, in Table 4, which clearly justifies the importance and effectiveness of our design choices.

Decoder	Loss	Dice Score (%) ↑			
		ET	WT	TC	Average
Single	Focal	77.88	90.31	82.01	83.40
BAD	L1	78.75	90.83	82.19	83.92(+0.48)
BAD	MSE	79.18	90.47	82.79	84.15(+0.71)
BAD	Focal	**79.65**	**90.80**	**83.33**	**84.59**(+1.15)

Table 4. Ablation study on the type of decoder and loss function for self-supervised pre-training.

Evaluations on 3D baselines: Noticeably, our framework is easily extendable to 3D version, enhancing 3D baseline's performance. To convince this point, we also conduct experiments on a commonly used 3D benchmark dataset BTCV [5], with 3D UNet and 3D Swin UNETR [8] as 3D baselines for comparison. The employed pre-training methods (*i.e.* Model genesis [10] and Swin UNETR [8]) are both previous efforts on SSL for medical image analysis. We follow the same pre-training and fine-tuning settings as in Swin UNETR for a fair comparison. Besides, we evaluate the effectiveness of our approach in terms of five-fold cross-validation on the training set and the evaluation metrics stay the same as in Swin UNETR. Results in Table 5 provide substantial evidence of our method's generalization ability and potential.

Method	Scratch	Models genesis [10]	Swin UNETR [8]	Ours
3D UNet	80.41	81.25	-	**81.72**
Improvement↑	-	(+0.84)	-	(+1.31)
Swin UNETR 3D	81.06	-	82.25	**82.80**
Improvement↑	-	-	(+1.19)	(+1.74)

Table 5. Comparison with previous SSL works on BTCV dataset.

4. Visual Comparison for Qualitative Analysis

Segmentation Results. Firstly, the skin lesion segmentation results on ISIC 2018 dataset is presented in Fig. 1. It can be obviously seen that the model can generate much more accurate and fine-grained segmentation masks compared with baseline with the benefit of employing our proposed FreMIM. Simultaneously, we compare the segmentation performance of different self-supervised methods, including MAE, DINO, and FreMIM on the BraTS 2019 dataset with visualization results. As shown in Fig. 2, our method promotes the detailed pixel delineation of brain tumors and obtains more accurate predictions.

Reconstruction Results. To convincingly prove the superiority of our FreMIM, we further supplement more visual comparison of reconstruction results on BraTS 2019 dataset for qualitative analysis. As is shown in Fig. 3, our method can nicely achieve the reconstruction task of Fourier spectrum and generate the corresponding reconstruction spectrum approximately the same as original image. To be mentioned, for each image slice, the first row is the original image and the second row is our reconstruction results of the Fourier spectrum.

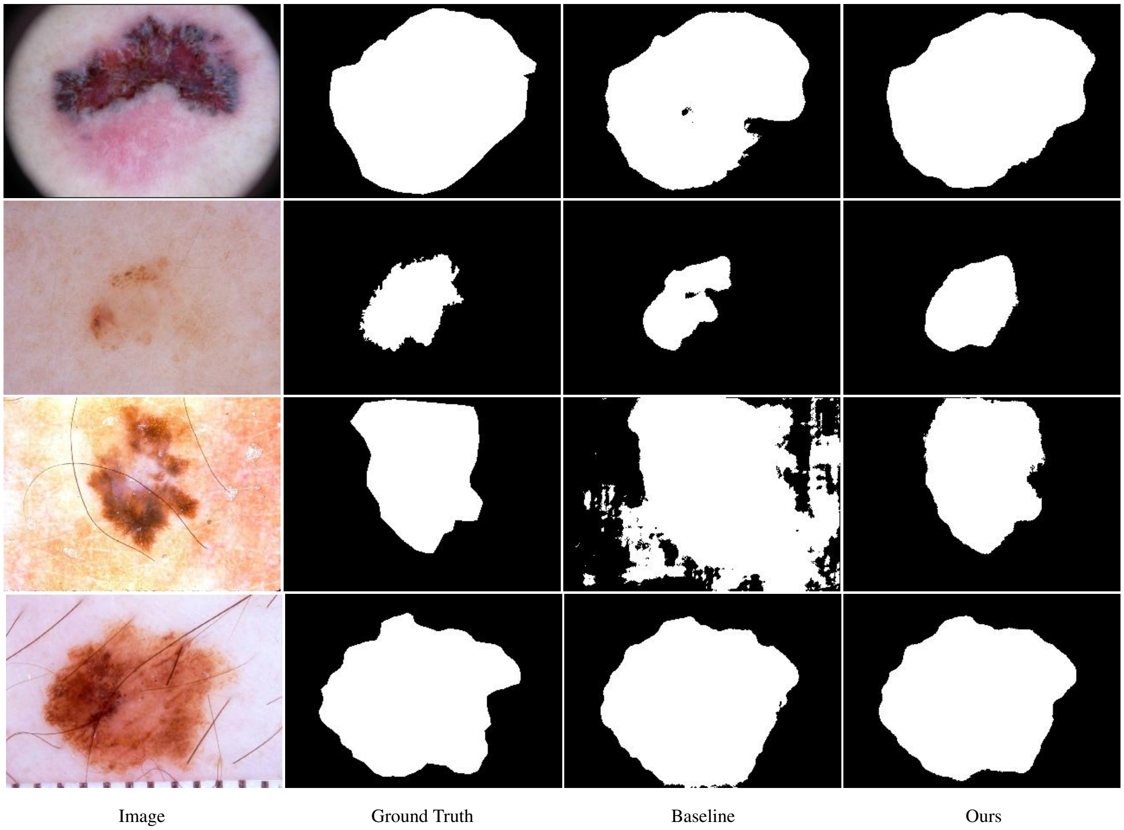

Figure 1. The visual comparison of skin lesion segmentation results on ISIC 2018 dataset with TransBTSV2 as the baseline.

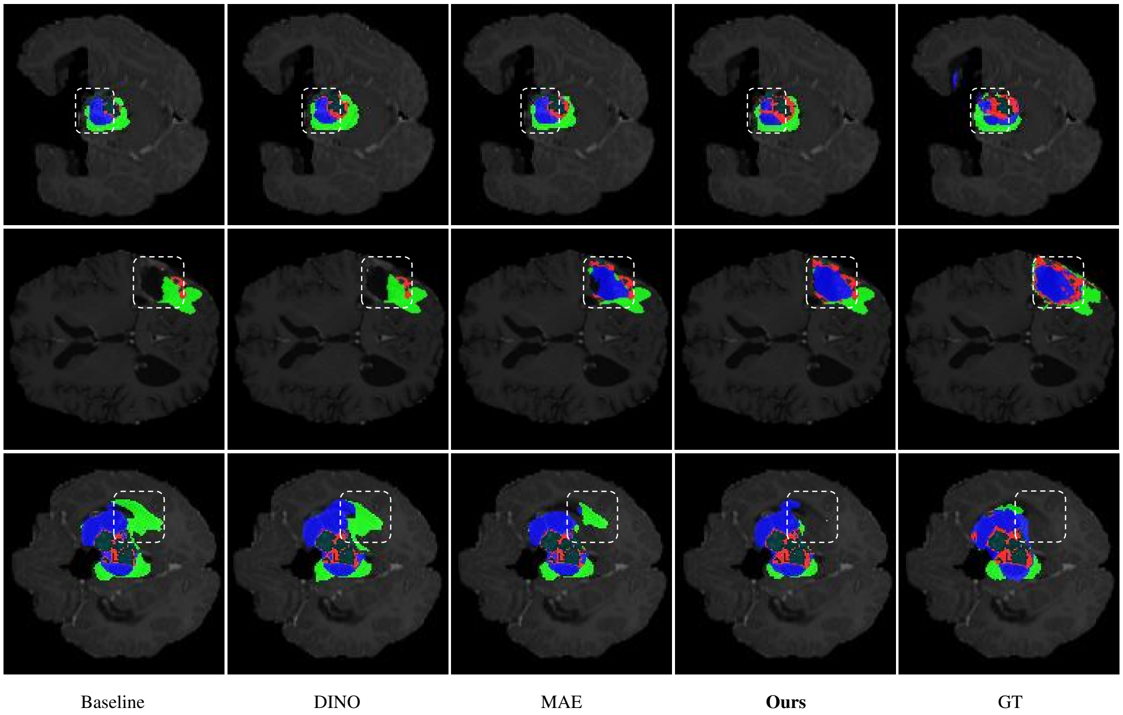

Figure 2. The visual comparison of MRI brain tumor segmentation results with UNETR as baseline. The blue regions denote the enhancing tumors, the red regions denote the non-enhancing tumors and the green ones denote the peritumoral edema.

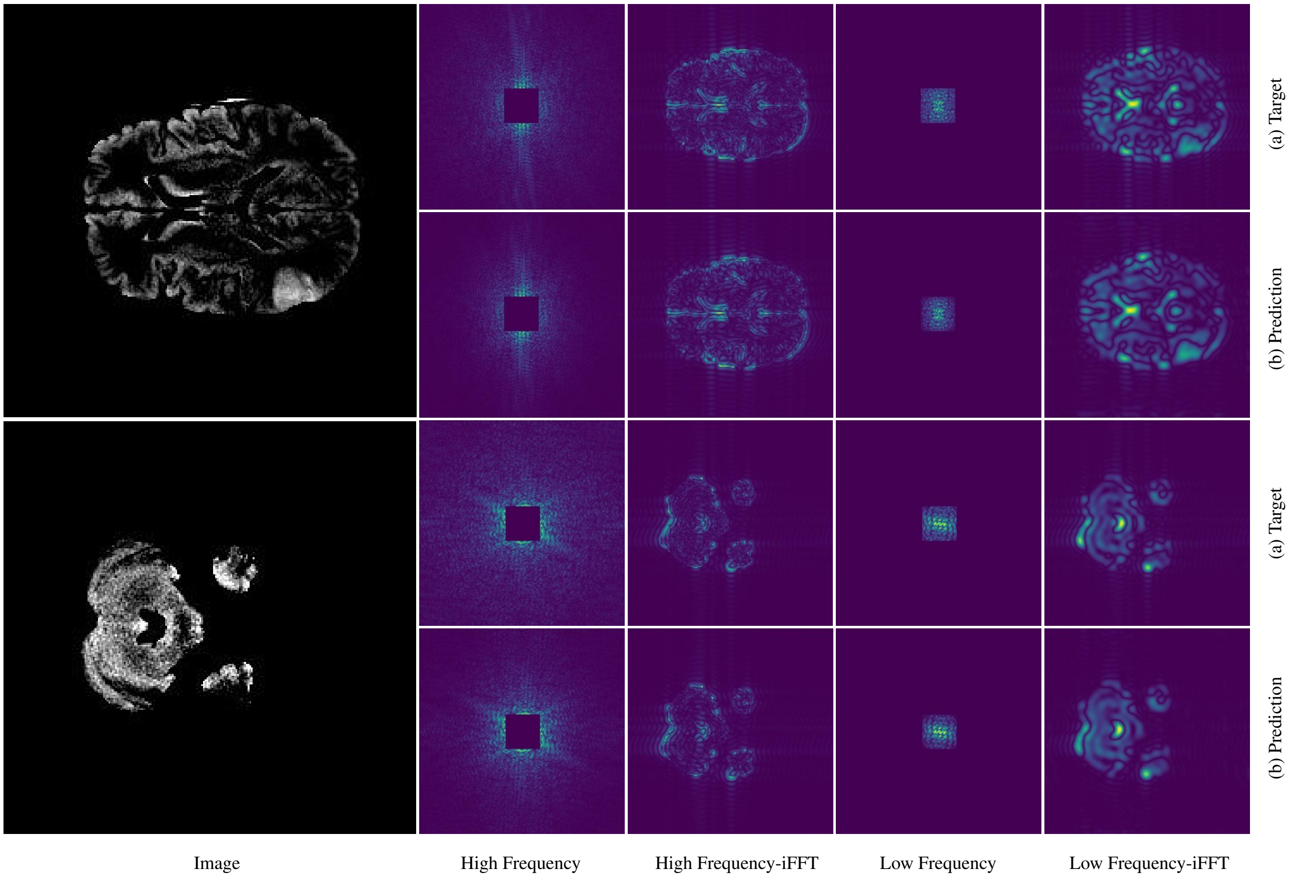

Figure 3. The visualization of reconstruction results by our FreMIM in the frequency domain.

References

[1] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4:170117, 2017. 1

[2] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018. 1

[3] Olivier Bernard, Alain Lalande, Clement Zotti, Frederick Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: is the problem solved? *IEEE transactions on medical imaging*, 37(11):2514–2525, 2018. 1

[4] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 1

[5] Bennett Landman, Zhoubing Xu, J Igelsias, Martin Styner, T Langerak, and Arno Klein. Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge. In *Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge*, volume 5, page 12, 2015. 2

[6] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014. 1

[7] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019. 1

[8] Yucheng Tang, Dong Yang, Wenqi Li, Holger R Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20730–20740, 2022. 2

[9] Philipp Tschandl, Cliff Rosendahl, and Harald Kittler. The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific data*, 5(1):1–9, 2018. 1

[10] Zongwei Zhou, Vatsal Sodha, Md Mahfuzur Rahman Siddiquee, Ruibin Feng, Nima Tajbakhsh, Michael B Gotway, and Jianming Liang. Models genesis: Generic autodidactic models for 3d medical image analysis. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22*, pages 384–393. Springer, 2019. 2