MS-Former: Multi-Scale Self-Guided Transformer for Medical Image Segmentation

Sanaz Karimijafarbigloo, Reza Azad, Amirhossein Kazerouni and Dorit Merhof
University of Regensburg and RWTH Aachen University
https://github.com/mindflow-institue/MS-Former

1. Introduction

Multi-scale representations have proven to be a powerful tool since they can take into account both the fine-grained details of objects in an image as well as the broader context. Inspired by this, we propose a novel dual-branch transformer network that operates on two different scales to encode global contextual dependencies while preserving local information. To learn in a self-supervised fashion, our approach considers the semantic dependency that exists between different scales to generate a supervisory signal for inter-scale consistency and also imposes a spatial stability loss within the scale for self-supervised image segmentation.

2. Data Set and Tasks

Skin Lesion Segmentation: Automatic skin lesion segmentation is one of the most demanding tasks in medical image analysis for accurate diagnosis and treatment.

Multiple Myeloma Segmentation: lung segmentation in CT images for accurate organ separation.

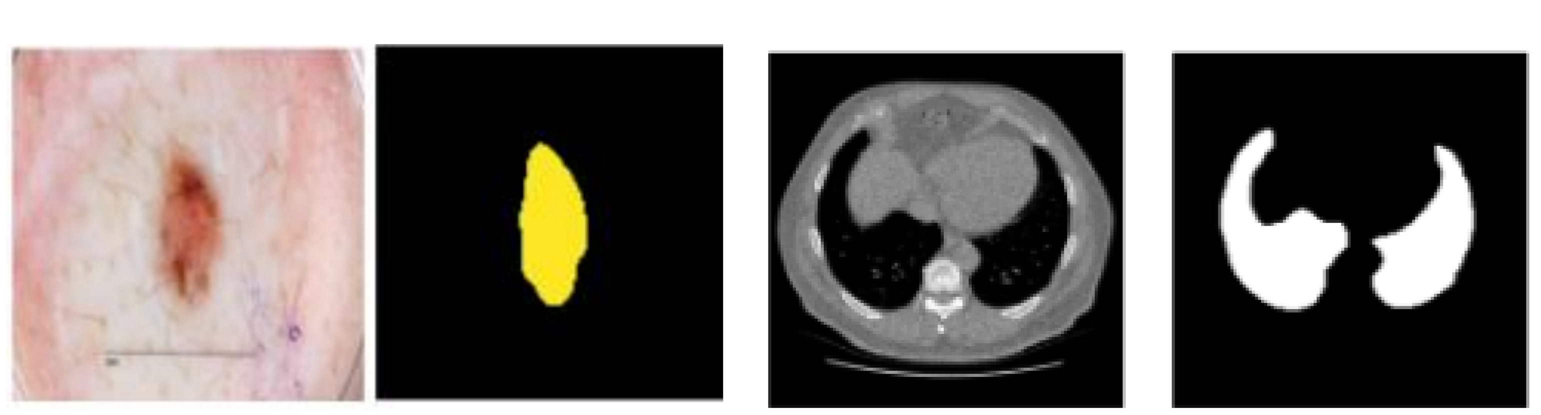

Fig. 1. Sample images from (left): skin lesion and (right): lung mask along with their grand truth masks

3. Proposed Method

The MMCFormer applies two vision trans-former models in parallel to capture multi-scale representation. Next by imposing inter-scale and intra-scale consistency loss it guides the network to learn content clustering based on the semantic information.

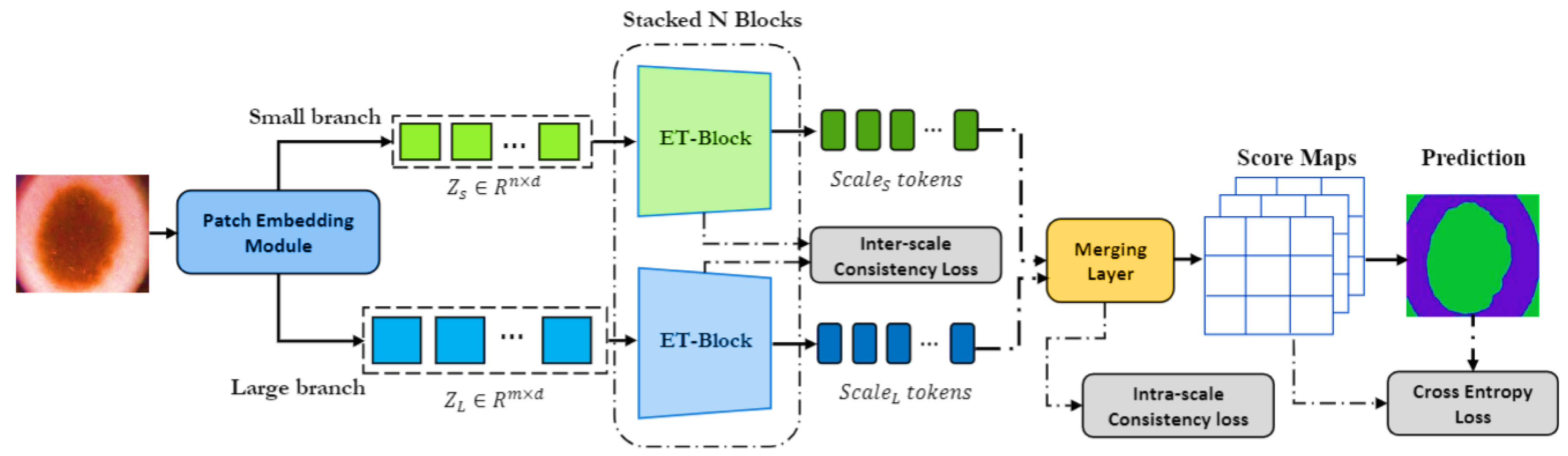

Fig. 2. The overview of the proposed MMCFormer

4. Visualization of the Proposed Modules

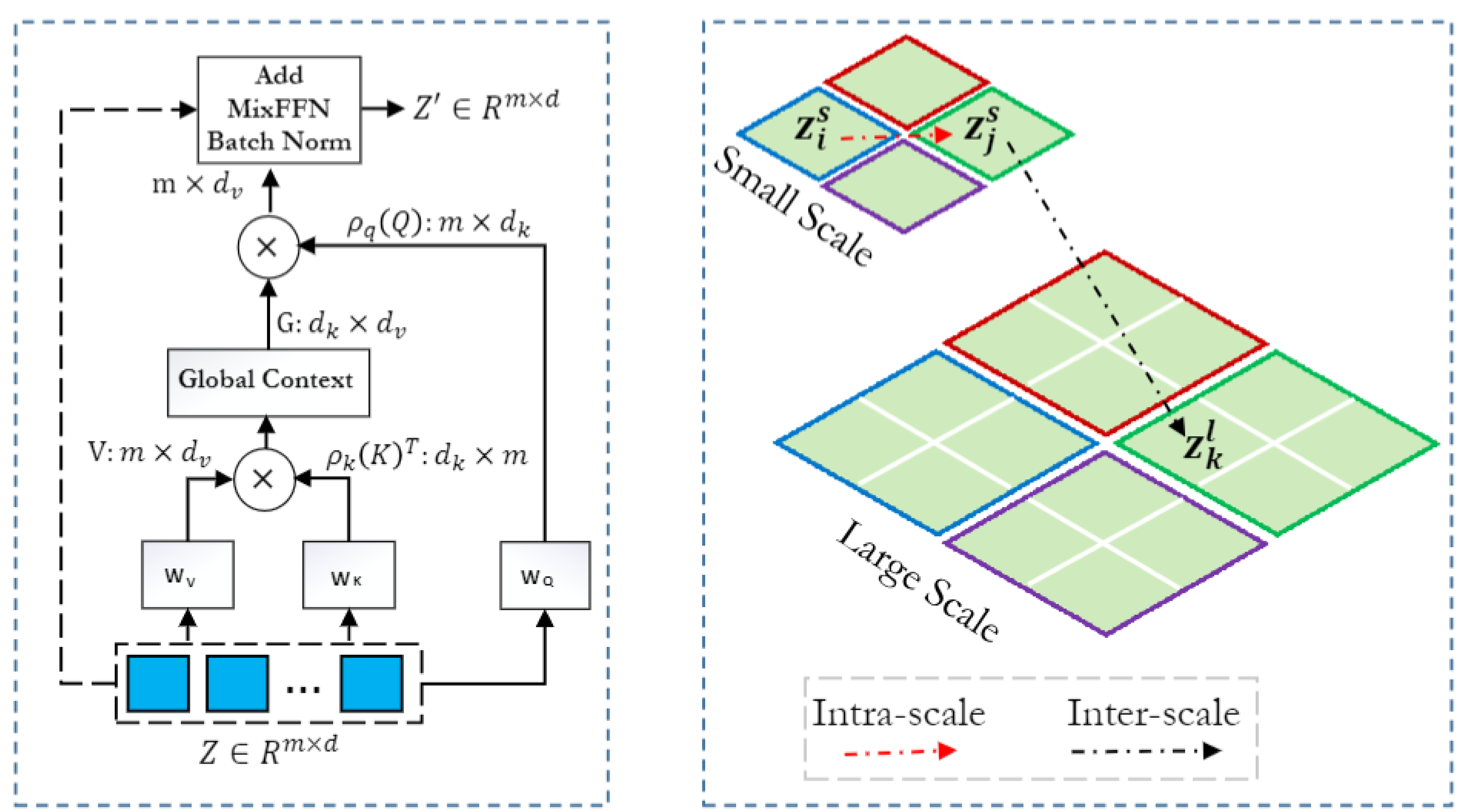

Fig. 3. (a): structure of the efficient Transformer block. (b): intra-scale and inter-scale dependencies.

5. Results on PH2 and Lung dataset

Table 1. Comparative results

Methods	PH²			Lung Segmentation		
	DSC ↑	HM ↓	XOR ↓	DSC ↑	HM ↓	XOR ↓
k-means	71.3	130.8	41.3	92.7	10.6	12.6
DeepCluster (Caron et al., 2018)	79.6	35.8	31.3	87.5	16.1	18.8
IIC (Ji et al., 2019)	81.2	35.3	29.8	-	-	-
SGSCN(Ahn et al., 2021)	83.4	32.3	28.2	89.1	16.1	34.3
Our Method	86.0	23.1	25.9	94.6	8.1	14.8

Table 2. Quantitative effect of our suggested modules on PH2

$\mathcal{L}_{ce}$	$\mathcal{L}_{intra}$	$\mathcal{L}_{inter}$	DSC ↑	HM ↓	XOR ↓
✓	✗	✗	83.6	25.8	30.2
✓	✓	✗	84.1	25.4	29.4
✓	✗	✓	84.3	25.3	28.4
✓	✓	✓	86.0	23.1	25.9

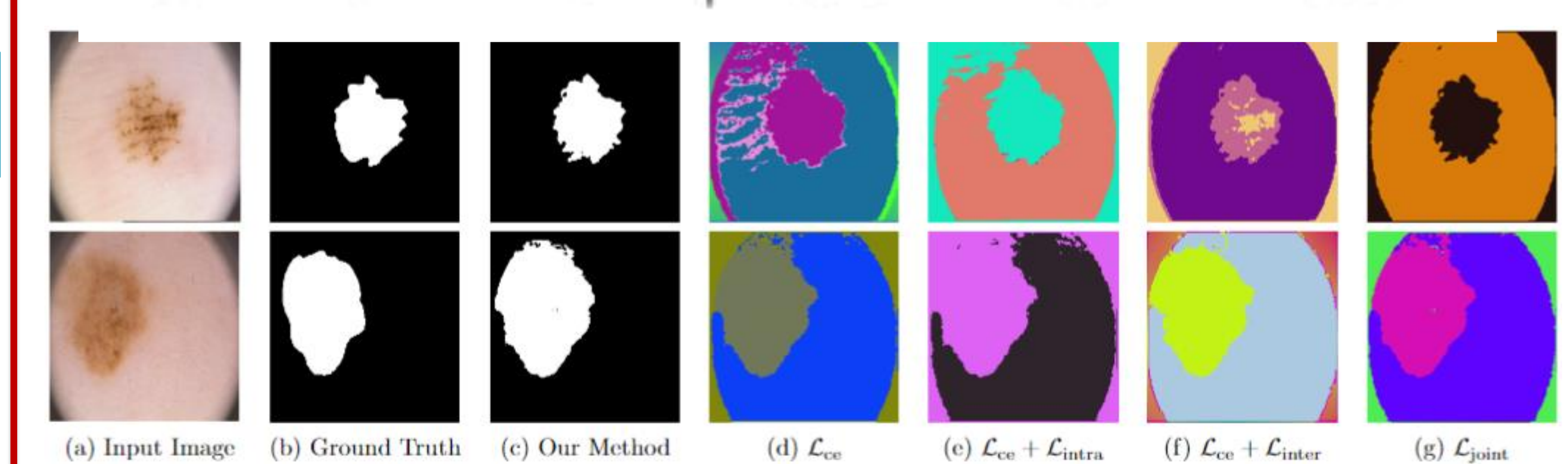

Fig. 4. Effect of our suggested auxiliary loss functions

6. Conclusion

This paper presents a self-supervised approach for medical image segmentation that eliminates the need for annotation masks. Our method utilizes a dual-branch strategy with an efficient self-attention mechanism, ensuring both intra-scale and inter-scale consistency to cluster pixels based on shared characteristics. Through iterative refinement, our algorithm generates highly accurate and semantically meaningful segmentation maps, surpassing SOTA methods in performance.