MaSA-UNet: Manhattan Self-Attention UNet with Weighted Composed Loss for Skin Lesion Segmentation

Abel A. Reyes[1], Sidike Paheding[2].

[1]Dept. of Applied Computing, Michigan Technological University, Houghton, MI, USA

[2]Dept. of Computer Science and Engineering, Fairfield University, Fairfield, CT, USA

Michigan Tech

Fairfield UNIVERSITY

CVPR SEATTLE, WA JUNE 17-21, 2024

Highlights

- Introduced **MaSA-UNet**, a novel DL architecture for skin lesion segmentation.
- Combines Manhattan Self-Attention UNet with **Weighted Composed Loss (WCL).**
- Integrates **bidirectional Manhattan Self-Attention** and varied loss functions to improve **segmentation performance.**

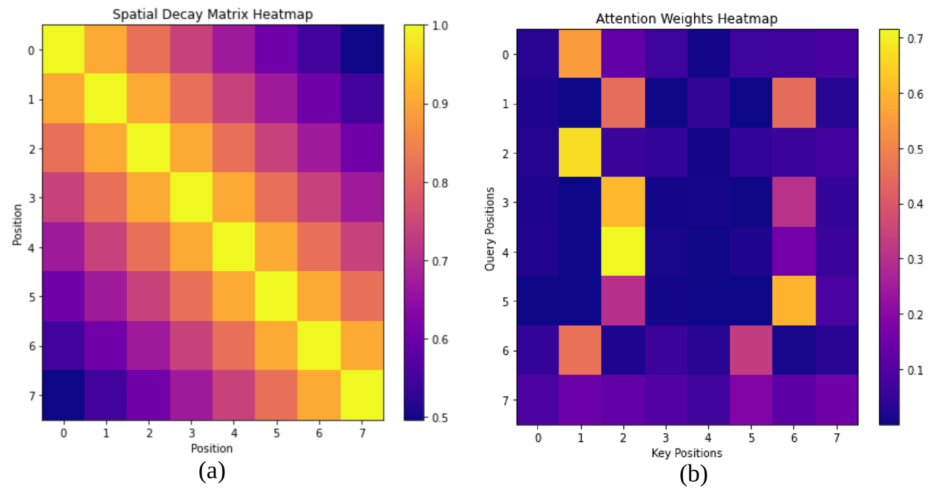
Manhattan Self-Attention

- Introduced by Fan et al. and inspired by the Retentive Network.
- Novel approach to enhance efficiency and performance of attention-based models.
- Reduces computational demands without sacrificing accuracy.
- Simplifies attention computation for efficient resource use.

$$BiRetention(X) = (QK^T \odot D^{Bi})V$$

where $D_{nm}^{Bi} = \gamma^{|n-m|}$

$$MaSA(X) = (Softmax(QK^T) \odot D^{2d})V$$

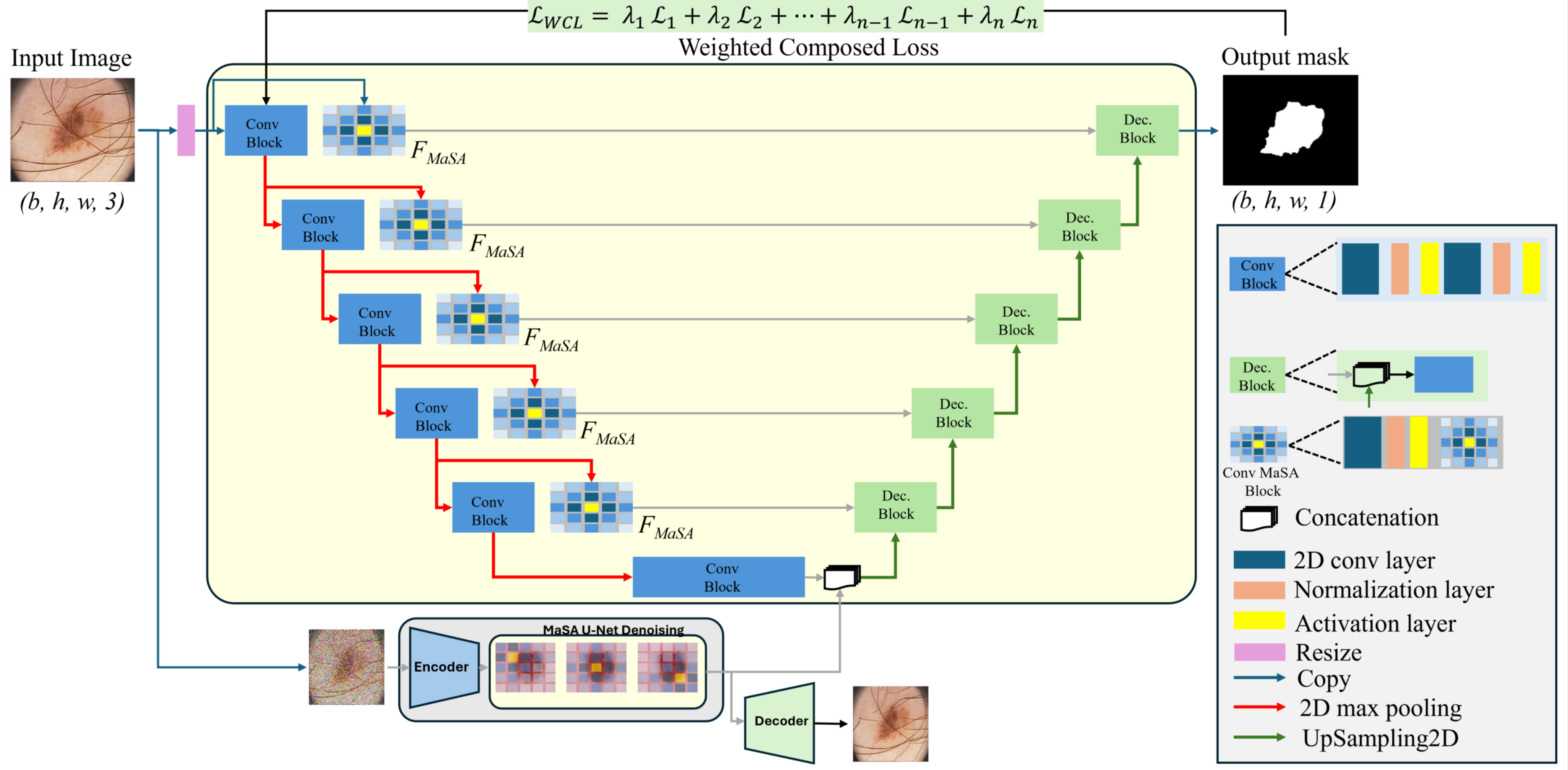


Weighted Composed Loss

- Optimization approach for improving biomedical image segmentation.
- Combines distinct loss functions with specific weights (λ) to effectively navigate the loss landscape.
- General expression:

$$WCL_n = \sum_{i=1}^{n} \lambda_i \mathcal{L}_i$$

Methodology



- Integrates MaSA, self-supervised encoder pre-training, and WCL for skin lesion segmentation.
- MaSA enhances feature focus through spatial attention.
- Encoder pre-training refines pattern discernment in biomedical images.
- WCL combines loss functions with weighted optimization for the segmentation of medical images.
- Achieves competitive performance in skin lesion segmentation.
- The figure above illustrates the **MaSA-UNet** architecture design.

$$MaSA-UNet(X) = F_{MaSA}(E_{pre-trained}(X))$$

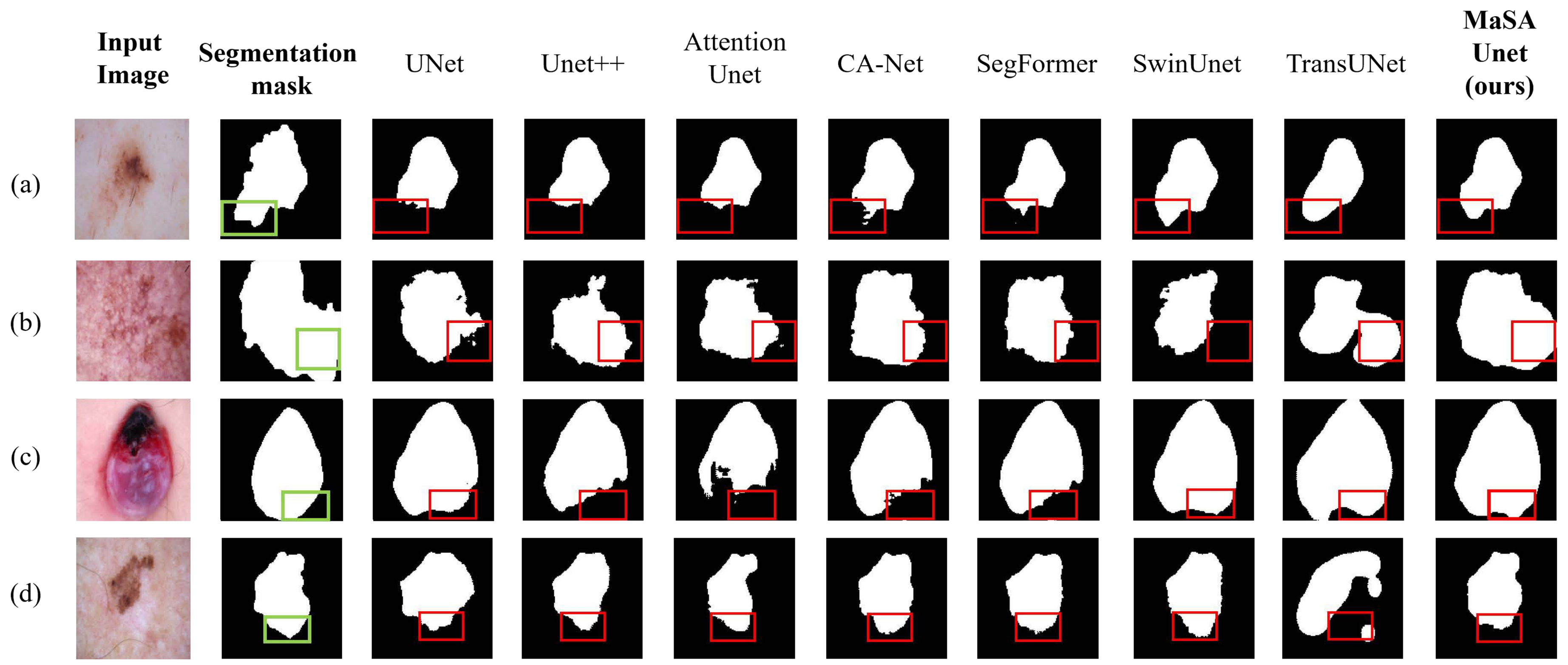
X: Input image, F_{MaSA} : MaSA within UNet, $E_{pre-trained}$: Pre-trained encoder.

$$\mathcal{L}_{WCL} = \sum_i \lambda_i \mathcal{L}_i(MaSA-UNet(X), Y)$$

Y : Ground truth mask, $\mathcal{L}_i$: Individual loss functions, λ_i : Weight for each loss function.

Benchmark Results

(a) ISIC 2016, (b) ISIC 2017, (c) ISIC 2018, and (d)PH[2]



Comparing with other models

Method	ISIC 2016		ISIC 2017		ISIC 2018		PH2	
	JS	DC	JS	DC	JS	DC	JS	DC
UNet	0.812	0.887	0.658	0.794	0.685	0.813	0.794	0.873
UNet++	0.818	0.889	0.686	0.814	0.714	0.833	0.813	0.890
Attention Unet	0.811	0.886	0.688	0.815	0.730	0.844	0.802	0.880
CA-Net	0.807	0.881	-	-	-	-	0.752	0.847
SegFormer	-	-	0.825	0.904	0.842	0.914	-	-
Swin Unet	-	-	0.741	0.851	0.768	0.869	-	-
TransUNet	0.849	0.913	0.775	0.873	0.825	0.904	0.840	0.910
MaSA-UNet (ours)	0.860	0.925	0.830	0.907	0.836	0.911	0.879	0.936

Conclusion

- Demonstrated improved accuracy in Jaccard score and Dice coefficient metrics.
- MaSA enables to capture of long-range feature relationships of the image structure.
- Future work includes enhancing weight-search efficiency, exploring broader applications of Manhattan Self-Attention, and improving the interpretability of segmentation outcomes.

PRESENTER NAME:

Abel A. Reyes

WORKSHOP DATE:

June 18, 2024

WORKSHOP ABBREVIATION:

LXCV

TAG / NOTES