ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 24-03-23
j.inst.Korean.electr.electron.eng.Vol.28,No.3,405~411,September 2024
https://doi.org/10.7471/ikeee.2024.28.3.405
163

피부병변 영상 분할의 성능향상을 위한 VmCUnet

VmCUnet for Improving the Performance of Skin lesion Image Segmentation

김 홍 진*, 이 태 희**, 황 우 성**, 최 명 렬*★

Hong-Jin Kim*, Tae-Hee Lee**, Woo-Sung Hwang**, Myung-Ryul Choi*★

Abstract

In this paper, we have proposed VmCUnet, a deep learning model designed to enhance image segmentation performance in skin lesion image. VmCUnet has combined Vm-UnetV2 with the CIM(Cross-Scale Interaction Module), and the features extracted from each layer of the encoder have been integrated through CIM to accurately recognize the boundaries of various patterns and objects. VmCUnet has performed image segmentation of skin lesions using ISIC-2017 and ISIC-2018 datasets and has outperformed Unet, TransUnet, SwinUnet, Vm-Unet, and Vm-UnetV2 on the performance metrics IoU and Dice Score. In future work, we will conduct additional experiments on different medical imaging datasets to validate the generalization performance of the VmCUnet model.

요 약

본 논문에서는 피부병변 영상에서 이미지 분할 성능을 향상시키기 위해 설계된 딥러닝 모델인 VmCUnet을 제안한다. VmCUnet은 Vm-UnetV2와 CIM(Cross-Scale Interaction Module)을 결합하여 인코더의 각 계층에서 추출한 특징들을 CIM으로 통합하여 다양한 패턴과 경계를 정확하게 인식할 수 있다. VmCUnet은 ISIC-2017와 ISIC-2018 데이터 세트를 사용하여 피부 병변의 이미지 분할을 수행하였고 Unet, TransUnet, SwinUnet Vm-Unet, Vm-UnetV2와 비교하여 성능 지표인 IoU, Dice Score에서 더 높은 성능을 보였다. 향후 작업에서는 다양한 의료 영상 데이터 세트에 대한 추가 실험을 수행하여 VmCUnet 모델의 일반화 성능을 검증할 예정이다.

Key words : CNN, U-net, Medical Image Segmentation, Vmamba, VM-Unet, VM-UnetV2

Ⅰ. 서론

최근 의료 전문가들은 CT, X-ray, MRI와 같은 첨단 장비를 활용하여 환자의 내부 구조를 자세히 본다. 이러한 중요한 정보를 통해 정확한 진단을 내리고 환자의 상태를 파악하고 최선의 치료 방법을 결정한다. 음성 인식, 자율주행, 의료 분야등 다양한 분야에서 인공지능 기술들이 활용되고 연구가 활발히 진행되고 있다. 특히, 의료 분야에서는 딥러닝 기반의 인공지능 기술의 발전으로 대용량의 데이터를 빠르게 처리할 수 있고 인간의 눈으로

* Graduate Student, Dept. of Applied Artificial Intelligence, Hanyang University.
★ Corresponding author
E-mail : choimy@hanyang.ac.kr, Tel : +82) 31-400-4036
※ Acknowledgment
This work was supported by the Korea Innovation Foundation(INNOPOLIS) grant funded by the Korea government (MSIT) 2024-IT-RD-0122
Manuscript received Sep.13, 2024; revised Sep. 24, 2024; accepted Sep. 26, 2024.
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(http://creativecommons.org/licenses/by-nc/3.0) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

감지하기 어려운 패턴까지 학습하여 의료 전문가들이 더욱 정확하고 빠르게 진단을 내릴 수 있도록 도와준다.

컴퓨터 비전에서 크게 이미지 분류, 객체 감지, 이미지 분할 세 가지 방법이 있다. 이미지 분류는 입력되는 이미지를 특정한 클래스로 분류해주고 객체 감지는 객체를 구분하고 위치 정보도 표시해준다. 그러나 이미지 분류나 객체 감지와 달리 영역의 경계를 정확하게 구분하고 분석하기 위해서는 이미지 분할이 주로 활용된다[1]. 이미지 분할이란 이미지 내에서 관심 있는 영역이나 객체를 픽셀 단위로 구분하는 과정으로 의료 분야에서 병변의 부위를 추출하는데 주로 사용되고 있다.

본 논문에서는 CIM(Cross-Scale Interaction Module)과 Vm-UnetV2 모델을 결합한 VmCUnet을 제안한다[2], [3]. 제안한 모델을 테스트하기 위해 피부 병변 데이터 세트인 ISIC(International Skin Imaging Collaboration)-2017과 ISIC-2018을 사용하여 실험을 진행하였다[4], [5].

II. 본론

2.1. 관련된 연구

이미지 분할을 위한 신경망 구조로 자주 사용되는 U-Net은 2015년 Olaf Ronneberger 등이 제안한 모델이다[6]. 입력 이미지의 세부 정보를 보존하는 특징이 있는 U-Net은 주로 의료 영상에서 객체의 경계를 정확하게 분할하는데 사용된다. 점차 해상도를 높이는 방식으로 분할을 수행하는 U-net은 크게 인코더와 디코더 두 가지로 나눌 수 있다. 인코더는 입력 이미지를 계층이 깊어질수록 작은 해상도로 축소하여 주요 특징을 추출한다. 디코더는 이를 다시 원래 해상도로 복원하면서 객체의 경계를 정확하게 구성한다. 또한, 인코더와 디코더 사이의 Skip-Connection을 배치하여 이미지의 세부 정보가 손실되지 않도록 한다.

Vm-UnetV2은 Vmamba의 VSS Block 및 SDI (Semantics and Detail Infusion)를 활용한 U-net 구조로 기존의 U-net과 비교해 성능평가 지표인 Dice와 IoU가 증가하는 결과를 보였다[7], [8].

2024년 원격 감지 이미지 세분화 문제를 해결하기 위해 제안된 논문에서 활용된 CIM(Cross-Scale Interaction Module)은 이미지 분할에 있어 복잡한 공간 정보를 처리하고 다양한 크기의 특징들을 분석함으로써 세밀한 정보를 얻어낼 수 있었다.

2.2. 제안하는 모델

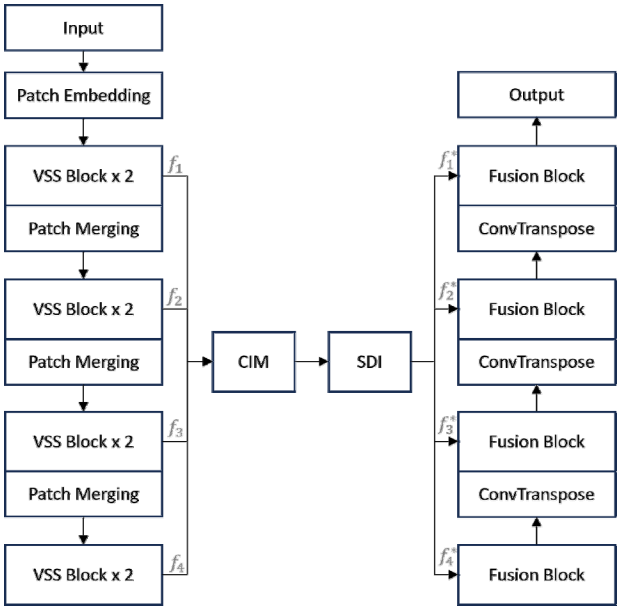


Fig. 1. Architecture of VmCUnet.
그림 1. VmCUnet 구조도

본 논문에서는 VM-UnetV2에 CIM을 추가한VmCUnet을 제안한다. VmCUnet은 크게 Encoder, CIM, SDI, Decoder로 구성되어 있다. 그림 1은 VmCUnet 모델의 구조도이다.

Encoder는 입력된 이미지를 고차원 벡터로 변환하기 위해 이미지의 정보를 작은 패치 단위로 분할한다. 이 과정은 모델이 이미지의 중요한 특징을 효과적으로 학습할 수 있도록 돕는다. 다음으로 VSS block과 Patch Merging 단계가 반복된다. VSS Block은 서로 다른 방향으로 스캔하여 이미지 내의 중요한 패턴과 특징을 학습하고, Patch Merging은 공간적 차원인 높이와 너비를 줄이는 동시에 채널의 수를 늘려 더 추상적이고 고수준의 특징을 볼 수 있게 한다. 각 단계의 VSS block에서 나온 특징들은 CIM에 들어가기 위해 동일한 크기로 다운샘플링을 한다. 식 (1)은 다운샘플링 과정을 수식으로 나타낸 것이다.

$$f_i^d = Downsample(f_i),\ \ i \in \{1,2,3,4\} \qquad (1)$$

CIM은 Rotated Multi-Scale Interaction Networ에서 핵심 구성요소로 인코더의 각 레이어에서 나온 특징들을 처리함으로써 크고 작은 객체들을 효과적으로 다룰 수 있다.

$$h^m = \left\lfloor \frac{h-1}{m}+1 \right\rfloor, w^m = \left\lfloor \frac{w-1}{m}+1 \right\rfloor, m \in \{1,\ldots,M\} \qquad (2)$$

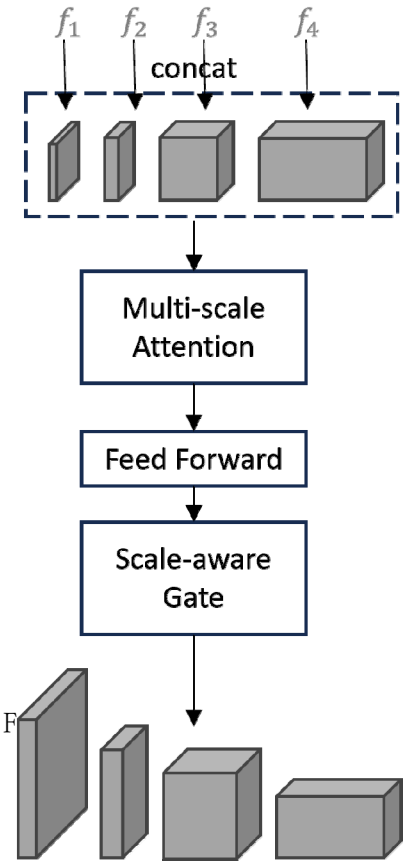


Fig. 2. Cross-scale Interaction Module.
그림 2. Cross-scale Interaction 모듈 구조

CIM은 그림 2와 같이 다운샘플링된 특징들을 통합한 후 Multi-scale Attention Layer에 입력으로 들어가 depth-wise convolution을 사용해 식 (2)와 같은 스케일로 변환하여 정보를 통합한다. 식 (2)에서 M의 값은 변환된 스케일의 수이다. 통합된 특징은 Attention 매커니즘내에서 Key와 Value 역할을 하고, Multi-scale Attention Layer에 들어가기 전의 특징은 Query 역할을 한다. 식 (3)는 위의 내용을 식으로 나타낸 것으로 MA는 Multi-scale Attention Layer이다.

$$f_i^c = MA\left(Concat\left(f_1^d, f_2^d, f_3^d, f_4^d\right)\right),\ i \in \{1,2,3,4\} \quad (3)$$

이후 Feed Forward와 Scale-aware Gate를 진행한다. Feed Forward는 Transformer와 동일한 Multi-scale Attention Layer를 따른다. Scale-aware Gate는 식 (4)와 같이 CIM에 들어오기 전의 특징들과 Feed Forward를 통해 나온 특징들을 업샘플링한 값을 입력으로 넣어 다양한 스케일에서 수집된 특징들 간의 오차를 줄여준다.

$$f_i^o = SG(f_i, FFN(f_i^c)),\ i \in \{1,2,3,4\} \quad (4)$$

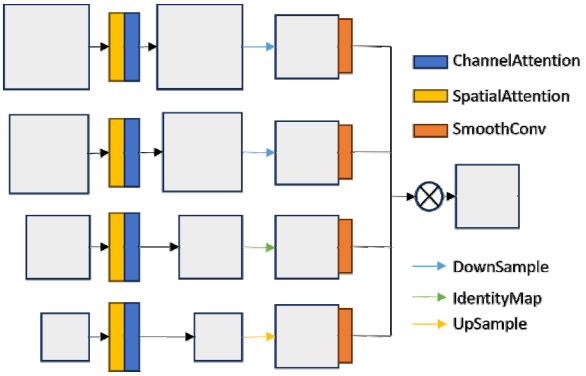


Fig. 3. Structure of SDI modules.
그림 3. SDI 모듈의 구조

다음으로 CIM에서 나온 특징들은 SDI에 입력으로 들어간다. 그림 x는 SDI 모듈의 구조이다. SDI에 들어온 특징들은 Spatial attention과 channel attention을 통해 공간 정보와 채널 정보를 강조하고 다운샘플링과 SmoothConv를 거친 특징들은 Hadamard 곱을 통해 결합되어 의미 정보와 디테일한 정보가 강화된 특징이 된다. 위의 과정으로 저수준과 고수준의 특징 정보를 통합하여 풍부한 의미를 가지게 한다. 식 (5)는 CIM과 마찬가지로 입력값과 출력값의 크기가 동일하다. 따라서 CIM의 입력값 f_i는 SDI의 출력값 f_i^*의 크기와 같기 때문에 정보 손실 없이 중요한 정보를 유지할 수 있다. SDI에서 나온 특징들은 Decoder에서 복원될 때 전달된다.

$$f_i^* = SDI(f_i^o),\ i \in \{1,2,3,4\} \quad (5)$$

Decoder는 두 가지 블록으로 구성되어 있다. 첫 번째는 Fusion Block으로, 복원된 특징과 SDI에서 넘어온 특징을 결합하는 역할을 한다. Fusion Block은 네트워크의 Layer가 깊어짐에 따라 정보 손실이 발생하는 문제를 해결하기 위해 설계된 블록이다. 두 번째는 ConvTranspose Block으로, 축소된 해상도를 원래 해상도로 복원하는 역할을 한다. 이 블록은 역컨볼루션 연산을 통해 축소된 특징을 입력 이미지와 동일한 해상도로 복원하는 데 사용된다. 이러한 과정을 거쳐 최종적으로 입력 이미지와 같은 해상도로 복원된 결과를 얻게 된다.

기존의 Vm-UnetV2 구조에 CIM을 추가하여 인코더에서 추출된 다중 스케일 특징들을 통합하고, 특징들 간의 상호작용 관계를 강화한다. 이를 통해 복잡한 패턴, 객체의 경계, 그리고 다양한 크기의 객체를 더 정확하게 인식할 수 있게 되었다. 실험 결과, 제안된 방법은 이미지 분할 작업에서 성능 향상을 보였다.

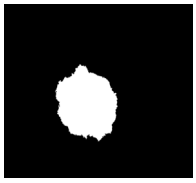
2.3. ISIC-2017 & ISIC-2018 데이터 세트

본 논문에서 사용한 데이터는 피부 병변 진단을 위한 피부암 이미지 데이터 세트이다. ISIC 데이터 세트는 피부병변의 이미지 분할을 위한 데이터세트로 가장 많이 사용되고 있다. ISIC-2017의 Train, Valid, Test 데이터 수는 각각 2,000장, 150장, 2,000장으로 총 4,150장이며, ISIC-2018의 Train, Valid, Test 데이터 수는 각각 2,594장, 100장, 1,000장으로 총 3,694장으로 구성되어 있다. 표 2는 각 데이터 세트의 이미지와 마스크의 샘플을 보여준다.

Table 1. Number of ISIC-2017 & ISIC-2018 datasets.
표 1. ISIC-2017 & ISIC-2018 데이터 세트 수

	ISIC-2017	ISIC-2018
Train	2,000	2,594
Validation	150	100
Test	2,000	1,000
Total	**4,150**	**3,694**

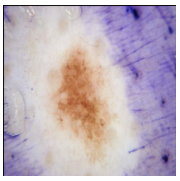
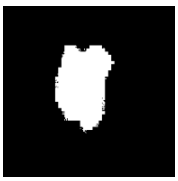
Table 2. Sample ISIC-2017 & ISIC-2018 datasets.
표 2. ISIC-2017 & ISIC-2018 데이터 세트 샘플

	Image	Mask
ISIC-2017		
ISIC-2018		

2.4. VmCUnet 학습 및 결과

본 논문에서는 표 1의 ISIC 데이터 세트를 사용하며, 2.2절에서 설계한 VmCUnet의 성능을 평가하기 위해 가장 기본적인 모델인 Unet과 최신 모델인 TransUnet, SwinUnet, Vm-Unet, Vm-UnetV2를 대조군으로 설정하여 결과를 비교한다[9],[10]. ISIC 데이터 세트를 사용하여 이미지를 256×256 크기로 조정하고, 학습률은 0.001로 설정한다. 배치 사이즈는 16으로 하고, 옵티마이저로 AdamW를 사용한다. 학습률은 CosineAnnealingLR 스케줄러를 통해 조정되며, 총 300개의 에포크 동안 학습이 진행된다. 또한, 성능이 향상되지 않으면 최대 50

Table 3. Qualitative Comparison of ISIC-2017 & ISIC-2018 datasets by model.
표 3. ISIC-2017 & ISIC-2018 데이터 세트의 모델별 정성적 비교

	ISIC-2017	ISIC-2018
Image		
GT		
Unet		
TransUnet		
SwinUnet		
Vm-Unet		
Vm-UnetV2		
VmCUnet		

회까지 Early Stopping을 적용하여 학습을 조기에 종료한다.

모델의 성능을 평가하기 위해 정성적 실험과 정량적 실험을 수행하였다. 정성적 실험의 결과인 표 3은 각 모델이 예측한 결과와 데이터세트의 Ground Truth를 시각적으로 비교한 것이다. ISIC-2017와 ISIC-2018 데이터세트에서 VmCUnet과 TransUnet은 Unet, SwinUnet, Vm-Unet, Vm-UnetV2와 비교해서 병변의 경계면이 Ground Truth와 가장 비슷한 결과를 보여준다.

정량적 실험에서는 IoU(Intersection over Union)와 Dice Score를 사용하였다. 모델의 예측 성능을 정량적으로 평가하는데 중요한 역할을 하는 IoU와 Dice Score는 이미지 분할과 객체 탐지에서 널리 사용되는 지표이다. IoU는 모델이 예측한 영역과 실제 영역 사이의 겹치는 정도를 측정하는 지표이다. 식 (6)은 IoU의 수식으로, A는 모델이 예측한 영역을, B는 실제 영역을 나타낸다.

$$IoU = \frac{|A \cap B|}{|A \cup B|} \tag{6}$$

Dice Score는 두 집합 간의 유사성을 측정한다. 식 (7)는 Dice Score의 수식으로 IoU와 유사하지만 교집합 영역을 두 배로 고려하여 계산된다.

$$Dice\ Score = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{7}$$

0에서 1 사이의 값을 가지는 IoU와 Dice Score는 1에 가까워질수록 두 영역이 더 비슷하다는 것을 의미한다. Dice Score는 교집합의 2배를 두 집합의 크기 합으로 나누기 때문에 IoU에 비해 작은 오차에도 더 민감하게 반응하며 작은 객체에 대해 IoU보다 더 높은 값을 가진다.

표 4는 각 모델들의 성능평가 지표 결과를 보여준다. 본 논문에서 제시한 VmCUnet은 ISIC-2017에서 Vm-UnetV2에 비해 IoU가 0.055, Dice Score가 0.04 상승했으며, ISIC-2018에서는 IoU가 0.031, Dice Score가 0.018 증가한 성과를 보였다. 또한, Unet, SwinUnet, Vm-Unet보다 IoU와 Dice Score가 모두 향상된 결과를 확인할 수 있다. TransUnet은 ISIC-2017에서 IoU와 Dice Score 값이 소폭 높지만 모델의 파라미터 수가 4배 가까이 크기 때문에 VmCUnet이 더 효율적이다.

표 5는 각 모델들의 IoU와 Dice의 표준편차를 나타내는 것이다. 표준편차가 낮을수록 새로운 데이터가 들어와도 일관된 성능을 보여준다. 표 5에서 VmCUnet은 ISIC-2017에서 TransUnet을 제외하고 IoU와 Dice의 표준편차가 모두 낮은 값이 나온 것을 보여준다. TransUnet이 ISIC-2017에서 IoU와 Dice Score도 높게 나오고 표준편차도 낮게 나타났지만 모델의 파라미터수가 VmCUnet에 비해 3배 이상 차이나기 때문에 제안한 모델이 더 효율적이다.

Table 4. Performance Metrics of Different Models on ISIC-2017 and ISIC-2018 Datasets.
표 4. ISIC-2017 & ISIC-2018 데이터 세트의 모델별 성능지표 결과

	Params (M)	ISIC-2017		ISIC-2018	
		IoU	Dice	IoU	Dice
Unet	20.7M	0.822	0.891	0.76	0.847
Trans Unet	104M	0.919	0.957	0.805	0.881
Swin Unet	28.1M	0.846	0.913	0.799	0.878
Vm-Unet	35.8M	0.844	0.91	0.795	0.875
Vm-UnetV2	22.7M	0.845	0.909	0.783	0.87
VmC Unet	**27.3M**	**0.906**	**0.949**	**0.814**	**0.888**

Table 5. Standard deviation by model for ISIC-2017 & ISIC-2018 datasets.
표 5. ISIC-2017 & ISIC-2018 데이터 세트의 모델별 표준편차

	ISIC-2017		ISIC-2018	
	std_IoU	std_Dice	std_IoU	std_Dice
Unet	0.111	0.08	0.19	0.155
SwinUnet	0.091	0.061	0.147	0.119
TransUnet	0.045	0.025	0.154	0.123
Vm-Unet	0.116	0.086	0.151	0.125
Vm-UnetV2	0.128	0.097	0.14	0.111
VmCUnet	**0.061**	**0.038**	**0.144**	**0.111**

Ⅲ. 결론

본 논문에서는 의료 영상 분석의 중요한 기법 중 하나인 이미지 분할에 딥러닝 기술을 적용하여 피부 병변의 분할을 위한 모델인 VmCUnet을 제안했다. VmCUnet의 성능을 검증하기 위해 ISIC-2017 및 ISIC-2018 데이터 세트를 활용하여 학습 및 검증을 진행하였다. 실험 결과, VM-UnetV2와 CIM을 결합한 VmCUnet은 피부 병변의 경계를 더욱 정확히 분할하고 기존 모델인 VM-UnetV2와 비교해 IoU 및 Dice Score는 향상된

성능을 보여주었다.

향후 연구에서는 본 논문에서 제안한 VmCUnet 모델에 추가적인 모듈을 도입하여 모델의 성능을 더욱 향상시키고 기존에 사용한 데이터 세트 외에도 Kvasir-SEG, ClinicDB, ColonDB, ETIS, CVC-300와 같은 의료 영상 데이터 세트를 활용하여 일반화 성능을 검증할 예정이다. 이를 통해 본 논문에서 제안된 모델이 의료 분야에서 다양하게 활용되어 의료 전문가들의 오진을 예방하고 진단 및 치료에 많은 기여를 할 것으로 기대된다.

References

[1] Minaee, Shervin, et al, "Image segmentation using deep learning: A survey", *IEEE transactions on pattern analysis and machine intelligence,* vol.44, no.7, pp.3523-3542, 2021.
DOI: 10.1109/TPAMI.2021.3059968

[2] Liu, Sihan, et al, "Rotated multi-scale interaction network for referring remote sensing image segmentation", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 2024. DOI: 10.48550/arXiv.2312.12470

[3] Zhang, Mingya, et al, "VM-UNET-V2: Rethinking Vision Mamba UNet for Medical Image Segmentation," *International Symposium on Bioinformatics Research and Applications, Singapore: Springer Nature Singapore,* 2024
DOI: 10.48550/arXiv.2403.09157

[4] Berseth, Matt, "ISIC 2017-skin lesion analysis towards melanoma detection," *arXiv preprint arXiv:*1703.00523 (2017).
DOI: 10.48550/arXiv.1703.00523

[5] Codella, Noel, et al, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," a*rXiv preprint arXiv:*1902.03368 (2019).
DOI: 10.48550/arXiv.1902.03368

[6] Ronneberger, Olaf, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," *Medical image computing and computer-assisted intervention-MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18, Springer International Publishing*, 2015.
DOI: 10.48550/arXiv.1505.04597

[7] Zhu, Lianghui, et al, "Vision mamba: Efficient visual representation learning with bidirectional state space model," *arXiv preprint arXiv:*2401.09417 (2024). DOI: 10.48550/arXiv.2401.09417

[8] Peng, Yaopeng, Milan Sonka, and Danny Z. Chen, "U-Net v2: Rethinking the skip connections of U-Net for medical image segmentation," *arXiv preprint arXiv:*2311.17791 (2023).
DOI: 10.48550/arXiv.2311.17791

[9] Chen, Jieneng, et al, "Transunet: Transformers make strong encoders for medical image segmentation," *arXiv preprint arXiv:*2102.04306 (2021).
DOI: 10.48550/arXiv.2102.04306

[10] Cao, Hu, et al, "Swin-unet: Unet-like pure transformer for medical image segmentation," *European conference on computer vision. Cham: Springer Nature Switzerlan*d, 2022.
DOI: 10.48550/arXiv.2105.05537

BIOGRAPHY

Hong-Jin Kim (Member)

2023 : BS degree in smart Information and Communication Engineering, Sangmyung University.
2023 : MS candidate in Applied Artificial Intelligence, Hanyang University.

Tae-Hee Lee (Member)

2021 : BS degree in Electronics Engineering, Hanyang University.
2023 : MS degree in Electronics Engineering, Hanyang University.
2023 : Ph.D candidate in Electronics Engineering, Hanyang University.

Woo-Sung Hwang (Member)

2004 : BS degree, School of Electronics and Computer Engineering, Hanyang University.
2006 : MS degree, Dept. of Electronic, Electrical, Control & Instrumentation Engineering, Hanyang University.

2024 : Ph.D degree in Electronics Engineering, Hanyang University.
2006~2007 : Research Engineer, Hi-tek

Myung-Ryul Choi (Member)

1983 : BS degree in Electronics Engineering, Hanyang University.
1985 : MS degree in Computer Engineering, Michigan State University.
1991 : Ph.D. degree in Computer Engineering, Michigan State University.
1991~1992 : Research Engineer and Assistant Prof. KITECH.
1992~ : Professor, Electrical Engineering in Hanyang University.