

Rethinking Decoder Design: Improving Biomarker Segmentation Using Depth-to-Space Restoration and Residual Linear Attention

Saad Wazir, Daeyoung Kim

School of Computing, Korea Advanced Institute of Science & Technology (KAIST)



GOAL: To improve biomarker segmentation in medical images by addressing limitations in decoder design, enabling better feature integration and spatial restoration especially in datasets with staining and morphological variability and limited training samples.

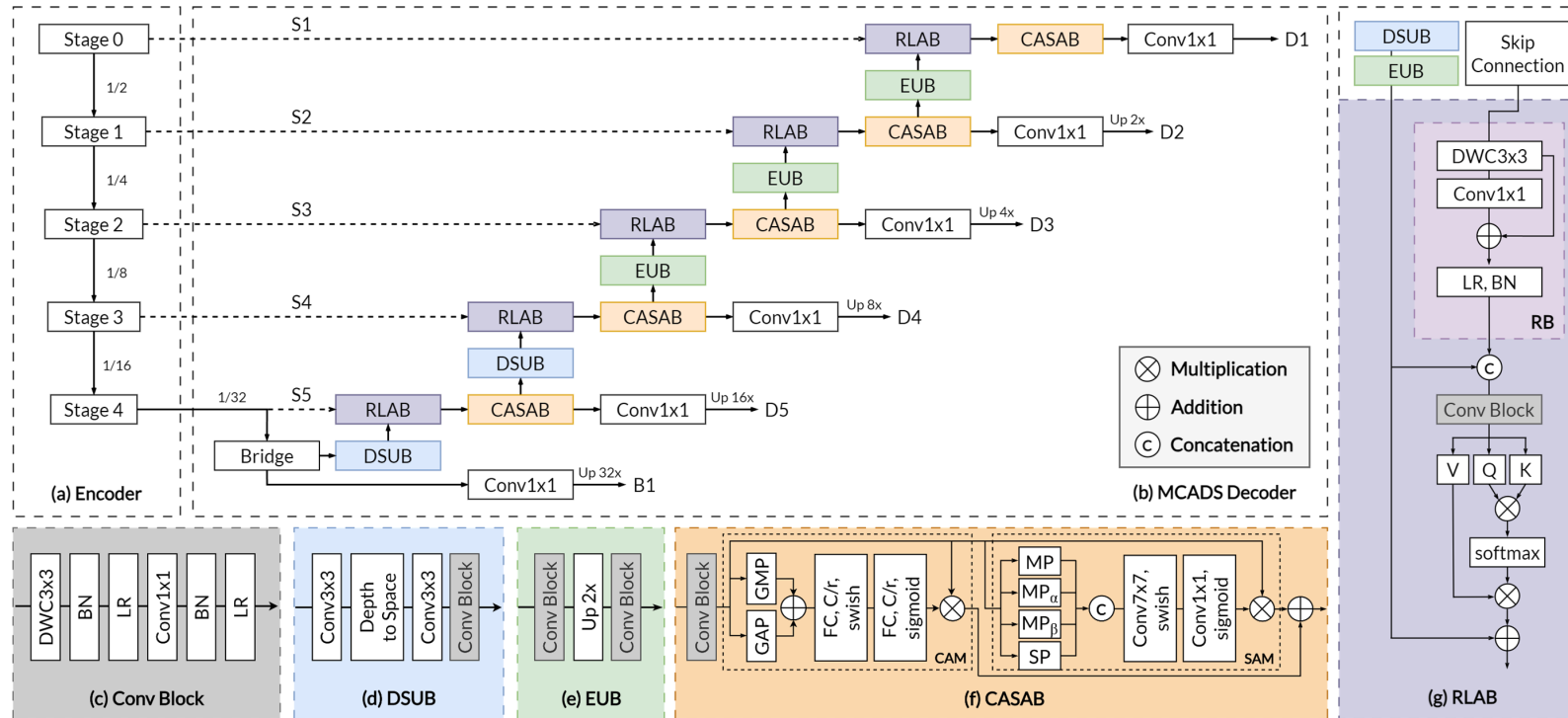
CHALLENGES:

- Poor transfer of multiscale features from encoder to decoder in existing models.
- Limited decoder capability to preserve spatial detail and context.
- Existing decoders suffer from coarse outputs and weak boundary refinement.

KEY CONTRIBUTIONS:

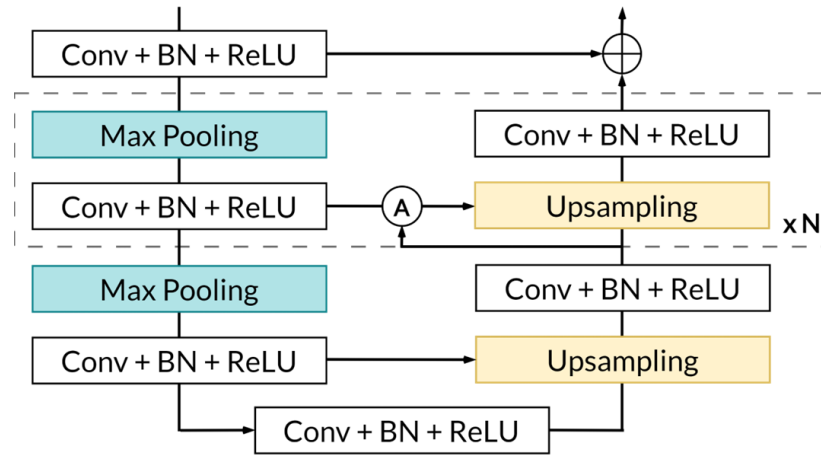
- **MCADS Decoder:** A novel Multiscale Convolution Attention with Depth-to-Space decoder that integrates features effectively, suppresses noise, and reconstructs spatial detail using custom upsampling blocks (DSUB, EUB).
- **Novel Attention Modules RLAB AND CASAB.**
- **Enhanced U2-Net Encoder:** Modified to capture richer multiscale local and global features, achieving superior performance when combined with MCADS.
- **Extensive Validation:** Demonstrated consistent outperformance over SOTA methods across four benchmark datasets (MoNuSeg, DSB, EM, TNBC) with up to 4.03% gain.

METHOD

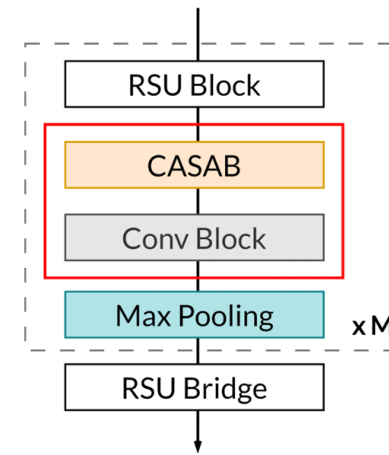


OVERVIEW OF THE PROPOSED ARCHITECTURE: (A) ENCODER FEATURE LEVELS AT DIFFERENT SCALES, (B) MULTISCALE CONVOLUTION ATTENTION WITH DEPTH-TO-SPACE (MCADS) DECODER, (C) CONVOLUTION BLOCK, (D) DEPTH-TO-SPACE UPSAMPLING BLOCK (DSUB), (E) EFFECTIVE UPSAMPLING BLOCK (EUB), (F) CHANNEL AND SPATIAL ATTENTION BLOCK (CASAB), AND (G) RESIDUAL LINEAR ATTENTION BLOCK (RLAB) WITH RESIDUAL BLOCK (RB).

METHOD



(a) RSU Block / Bridge



(b) Proposed Encoder

WE ENHANCE THE U2-NET ENCODER BY INTEGRATING ATTENTION MECHANISMS BETWEEN SKIP CONNECTIONS AND UPSAMPLING LAYERS TO BETTER CAPTURE LONG-RANGE, MULTI-SCALE FEATURES. CASAB IS ADDED AFTER EACH RESIDUAL U-BLOCK TO EMPHASIZE IMPORTANT CHANNELS AND SPATIAL REGIONS FOR RICHER FEATURE REPRESENTATION.

EVALUATION

Performance comparison on the MoNuSeg Dataset

Model	IoU ↑	Dice ↑	Prec. ↑	Rec. ↑	FOR ↓	HD95 ↓	ASD ↓
U-Net [44]	59.62	73.42	74.87	73.04	0.0766	3.6609	0.8109
Hover-Net [17]	62.90	77.26	77.67	77.54	0.0689	3.8687	0.8114
UNet++ [61]	69.34	81.83	75.31	90.29	0.0422	<u>2.7901</u>	0.6458
U2-Net [38]	68.89	81.33	77.20	86.88	0.0444	13.445	3.0647
nnU-Net [20]	67.60	80.42	81.63	80.88	0.0555	3.3357	0.6852
FusionU-Net [26]	66.57	79.32	73.29	87.26	0.0417	3.9401	0.8236
SegResNet [32]	69.02	81.00	<u>82.50</u>	80.97	0.0587	13.029	2.8829
UCTransNet [48]	65.56	79.20	75.83	82.88	0.0561	18.366	4.2000
TransUNet [8]	67.89	80.22	81.58	79.17	0.0610	3.5602	0.7379
Swin-Unet [6]	65.38	79.01	72.38	87.88	0.0403	3.7871	0.7893
PVT-CASCADE [40]	70.74	82.79	78.36	<u>88.24</u>	<u>0.0362</u>	3.0110	<u>0.6163</u>
EMCAD [41]	71.28	82.89	82.09	83.86	0.0488	3.1818	0.6285
InstaSAM [34]*	59.60	78.60	-	-	-	-	-
All-in-SAM [12]*	69.76	82.44	-	-	-	-	-
UN-SAM [10]*	70.82	82.86	-	-	-	-	-
SAC 0-expert [33]*	<u>72.61</u>	<u>84.03</u>	-	-	-	-	-
Ours	74.04	85.04	83.78	86.43	0.0308	1.8550	0.4810

Performance comparison on the DSB Dataset

Model	IoU ↑	Dice ↑	Prec. ↑	Rec. ↑	FOR ↓	HD95 ↓	ASD ↓
U-Net [44]	83.83	90.17	89.09	93.78	0.0128	7.6797	1.9192
Hover-Net [17]	81.29	89.04	88.73	91.01	0.0246	7.6289	2.1298
UNet++ [61]	83.95	91.10	88.49	<u>94.26</u>	<u>0.0115</u>	10.045	2.4359
U2-Net [38]	82.96	90.03	90.28	91.08	0.0163	10.869	2.3375
nnU-Net [20]	77.24	86.09	90.18	84.33	0.0337	9.9094	2.7911
FusionU-Net [26]	80.47	88.34	93.76	85.12	0.0332	7.6148	1.8682
SegResNet [32]	82.91	90.01	91.54	89.75	0.0262	7.5580	2.0236
UCTransNet [48]	85.28	91.25	<u>93.87</u>	90.02	0.0158	11.514	2.5552
TransUNet [8]	84.72	90.90	93.37	89.86	0.0162	11.213	2.7168
Swin-Unet [6]	84.49	91.03	90.77	92.72	0.0132	8.0415	1.9232
PVT-CASCADE [40]	85.05	91.07	90.55	93.02	0.0130	7.5427	1.7861
EMCAD [41]	86.36	92.18	90.53	94.89	0.0544	<u>6.4516</u>	<u>1.4400</u>
UN-SAM [10]*	86.93	92.84	-	-	-	-	-
SAC 0-expert [33]*	<u>87.32</u>	<u>93.04</u>	-	-	-	-	-
Ours	89.48	94.42	94.94	93.94	0.0036	2.6655	0.8559

EVALUATION

Performance comparison on the Electron Microscopy Dataset

Model	IoU \uparrow	Dice \uparrow	Prec. \uparrow	Rec. \uparrow	FOR \downarrow	HD95 \downarrow	ASD \downarrow
U-Net [44]	53.41	67.55	88.79	58.44	0.0548	21.101	5.0472
Hover-Net [17]	54.89	69.20	<u>88.68</u>	59.57	0.0539	23.017	5.0716
UNet++ [61]	60.94	74.53	77.89	75.20	0.0370	17.106	3.9056
U2-Net [38]	61.86	75.63	72.02	<u>82.43</u>	<u>0.0349</u>	14.659	3.4034
nnU-Net [20]	59.35	73.06	84.57	67.63	0.0511	16.768	3.3843
FusionU-Net [26]	56.80	71.44	87.84	62.21	0.0557	18.705	3.6262
SegResNet [32]	61.63	75.15	79.59	75.42	0.0362	15.314	3.1723
UCTransNet [48]	56.62	71.29	86.83	62.90	0.0549	18.340	3.5854
TransUNet [8]	59.71	73.31	85.69	66.90	0.0518	15.273	3.2620
Swin-Unet [6]	55.87	70.20	86.78	62.13	0.0515	19.421	4.3773
PVT-CASCADE [40]	62.46	76.11	82.49	74.41	0.0379	<u>13.506</u>	<u>2.7764</u>
EMCAD [41]	65.13	78.60	82.59	75.88	0.0501	15.953	3.3419
UN-SAM [10]*	72.27	83.89	-	-	-	-	-
Ours	<u>69.16</u>	<u>81.48</u>	79.63	84.51	0.0276	12.590	1.9611

Performance comparison on the TNBC Dataset

Model	IoU \uparrow	Dice \uparrow	Prec. \uparrow	Rec. \uparrow	FOR \downarrow	HD95 \downarrow	ASD \downarrow
U-Net [44]	77.45	85.37	88.82	84.95	0.0100	46.803	10.658
Hover-Net [17]	79.18	88.31	91.43	85.55	0.0084	12.624	2.1152
UNet++ [61]	<u>87.01</u>	<u>92.49</u>	92.53	<u>93.38</u>	0.0042	4.7633	1.1553
U2-Net [38]	86.31	92.01	91.92	92.88	<u>0.0041</u>	20.817	4.3750
nnU-Net [20]	85.02	91.86	<u>94.75</u>	89.23	0.0064	5.2735	1.2231
FusionU-Net [26]	85.92	92.37	93.99	90.87	0.0053	<u>4.1360</u>	<u>1.1485</u>
SegResNet [32]	84.51	90.87	92.79	90.13	0.0061	29.482	6.5738
UCTransNet [48]	85.99	92.32	91.55	93.24	0.0046	6.7678	1.5115
TransUNet [8]	84.86	91.77	94.33	89.42	0.0062	6.1746	1.4101
Swin-Unet [6]	83.72	91.08	92.18	90.10	0.0059	9.7707	1.8287
PVT-CASCADE [40]	86.28	92.05	91.64	93.28	0.0044	26.840	5.6181
EMCAD [41]	86.02	92.43	93.97	91.02	0.0055	19.857	4.1666
Ours	88.89	94.09	94.81	93.43	0.0040	2.6934	0.8552

FUTURE WORK:

This work can be extended by evaluating the proposed method on a broader range of medical imaging datasets, including organs, polyps, lesions, and retinal structures, across various modalities such as CT, MRI, fundus photography, and X-ray. Additionally, we plan to adapt the architecture for panoptic segmentation tasks to further enhance its applicability.

