Segregation and Context Aggregation Network for Real-time Cloud Segmentation

Yijie Li[1], Hewei Wang[1], Jiayi Zhang[2], Jinjiang You[1], Jinfeng Xu[3], Puzhen Wu[4], Yunzhong Xiao[1], Soumyabrata Dev[5]

[1]Carnegie Mellon University [2]University of Nottingham [3]The University of Hong Kong [4]Cornell University [5]University College Dublin

Overview

Cloud segmentation from intensity images is a pivotal task in atmospheric science and computer vision, aiding weather forecasting and climate analysis. We introduce SCANet, a novel lightweight cloud segmentation model featuring Segregation and Context Aggregation Module (SCAM), which refines rough segmentation maps into weighted sky and cloud features processed separately. SCANet achieves state-of-the-art performance while drastically reducing computational complexity. SCANet-large (4.29M) achieves comparable accuracy to state-of-the-art methods with 70.9% fewer parameters.

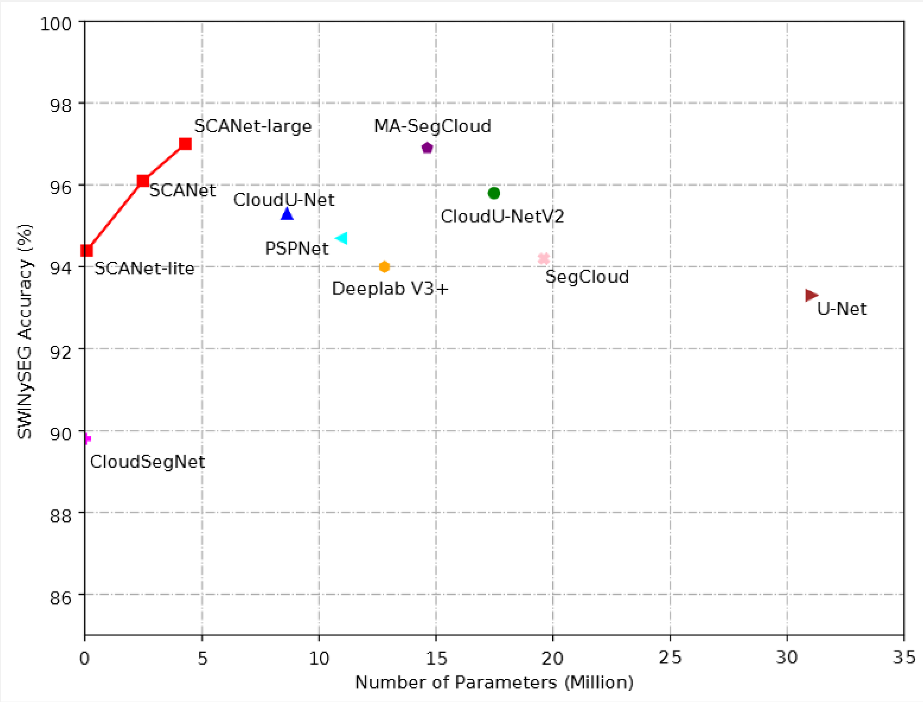

Figure 1: #Params vs. SWINySEG Accuracy.

Overall Pipeline

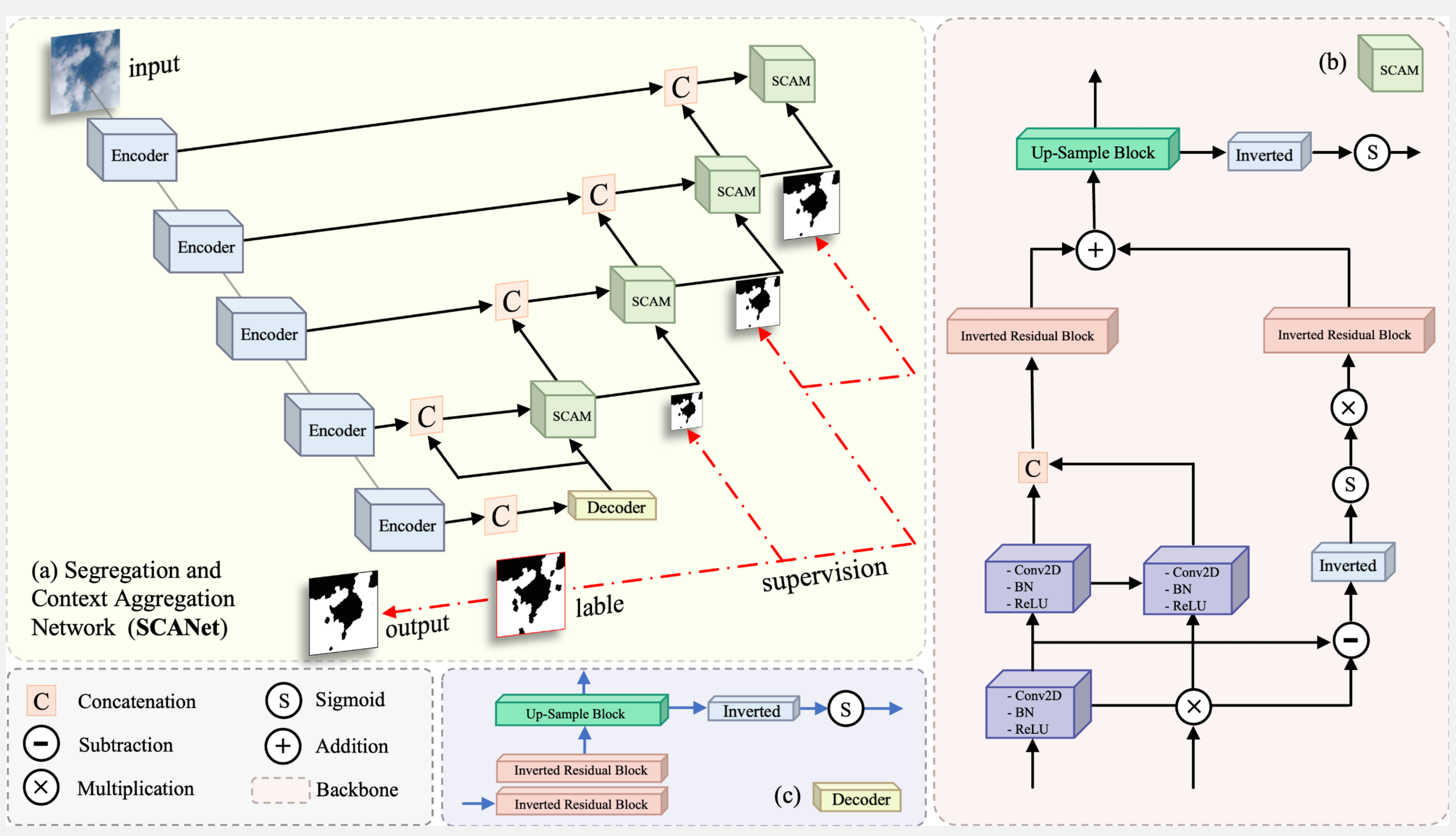

Figure 2: The overall architecture of SCANet and SCAM. (a) presents the pipeline of our SCANet; (b) details the design of our proposed SCAM module; (c) depicts the decoder structure preceding the SCAM modules.

In our research work, we use binary cross entropy (BCE) and Intersection of Union (IOU) loss as the loss function in the training of SCANet. These loss functions can be defined as follows:

$$\mathcal{L}_{\text{bce}}(p, y) = -\frac{1}{N} * \sum_{j=1}^{N} (y_j * \log p_j + (1 - y_j) * \log(1 - p_j)) \quad (1)$$

$$\mathcal{L}_{\text{iou}}(p, y) = 1 - \frac{1}{n} \sum_{j=1}^{N} \left(\frac{y_j \times p_j}{y_j + p_j - y_j \times p_j} \right) \quad (2)$$

then our total training loss under deep supervision can be formulated as:

$$\mathcal{L}(p, y) = \sum_{i=1}^{4} \alpha_i * (\mathcal{L}_{\text{bce}}(p_i, y_i) + \mathcal{L}_{\text{iou}}(p_i, y_i)) \quad (3)$$

in which α_i represents the coefficient of i th SCAM or decoder.

Qualitative Comparison

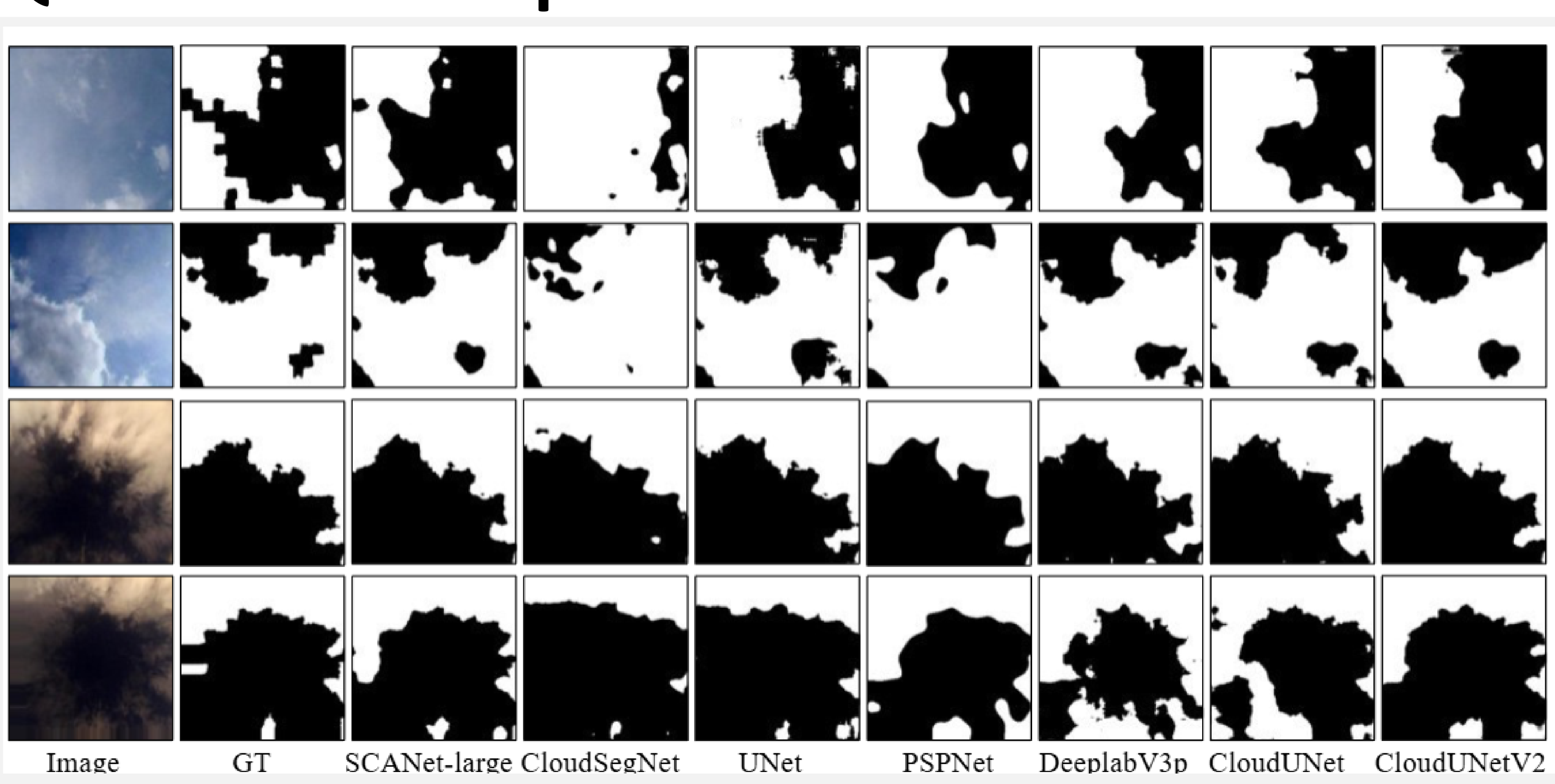

Figure 3: Qualitative comparison of SCANet-large with state-of-the-art approaches on day-time and night-time images from SWINySEG dataset.

Ablation Study

No.	Backbone	SCAM Configs		Loss Functions		Pre-training		#Params	SWINySEG				
		L Branch	R Branch	BCE	IOU	SWPT	INPT		Accuracy	Precision	Recall	F-score	MIoU
1	baseline	✗	✗	✓	✗	✗	✗	<u>0.32M</u>	0.927	0.934	0.910	0.922	0.832
2	MobileNetV2-lite	✓	✗	✓	✗	✗	✗	**0.09M**	0.936	0.940	0.925	0.932	0.850
3	MobileNetV2-lite	✓	✓	✓	✗	✗	✗	**0.09M**	0.943	0.944	0.932	0.938	0.863
4	MobileNetV2-lite	✓	✓	✓	✗	✓	✗	**0.09M**	0.944	0.943	0.933	0.938	0.863
5	MobileNetV2-lite	✓	✓	✗	✓	✓	✗	**0.09M**	0.940	0.941	0.934	0.938	0.856
6	MobileNetV2-lite	✓	✓	✓	✓	✓	✗	**0.09M**	0.944	0.936	0.944	0.940	0.865
7	MobileNetV2	✓	✓	✓	✗	✗	✓	2.49M	0.960	0.958	0.951	0.955	0.898
8	MobileNetV2	✓	✓	✓	✓	✗	✓	2.49M	0.961	0.955	<u>0.958</u>	0.957	0.902
9	EfficientNet-B0	✓	✓	✓	✗	✗	✓	4.29M	<u>0.969</u>	**0.972**	0.956	<u>0.964</u>	<u>0.919</u>
10	EfficientNet-B0	✓	✓	✓	✓	✗	✓	4.29M	**0.970**	<u>0.971</u>	**0.960**	**0.966**	**0.923**

Table 3: Ablation study on different model/training configurations.

Methods	MFlops	FP32		FP16	
		FPS	Latency	FPS	Latency
SCANet-lite	111.216	750	1.3 ms	1390	0.7 ms
SCANet	1451.31	465	2.1 ms	1124	0.8 ms
SCANet-large	365.85	299	3.3 ms	392	2.6 ms

Table 2: MFLops, inference latency and FPS of SCANets on an NVIDIA Tesla V100 GPU, deployed via NVIDIA TensorRT.

Quantitative Comparison

Methods	#Params	Day-time					Night-time					Day+Night time				
		Acc.	Prec.	Rec.	F1	MIoU	Acc.	Prec.	Rec.	F1	MIoU	Acc.	Prec.	Rec.	F1	MIoU
General Semantic Segmentation Models																
U-Net (Ronneberger et al., 2015)	31.05M	0.933	0.938	0.919	0.928	0.844	0.933	0.938	0.919	0.928	0.844	0.933	0.938	0.919	0.928	0.844
PSPNet (Zhao et al., 2017)	20.95M	0.947	0.949	0.935	0.942	0.873	0.947	0.950	0.935	0.942	0.873	0.947	0.950	0.935	0.942	0.873
DeeplabV3+ (Chen et al., 2017)	12.80M	0.940	0.941	0.920	0.936	0.860	0.940	0.941	0.932	0.936	0.860	0.940	0.941	0.932	0.936	0.860
Special Designed Sky/Cloud Segmentation Models																
CloudSegNet (Dev et al., 2019b)	0.005M	0.898	0.920	0.876	0.898	0.777	0.898	0.920	0.876	0.898	0.777	0.898	0.920	0.876	0.898	0.777
SegCloud (Xie et al., 2020)	19.61M	0.941	0.953	0.934	0.943	0.889	0.955	0.936	0.960	0.948	0.912	0.942	0.952	0.936	0.944	0.891
UCloudNet (Li et al., 2024b)	/	0.940	0.920	0.940	0.930	/	0.960	0.950	0.950	0.950	/	0.940	0.920	0.940	0.930	/
DDUNet (Li et al., 2025b)	0.33M	0.953	0.953	/	/	0.882	0.954	0.951	/	/	0.900	0.953	0.952	/	/	0.884
CloudU-Net (Shi et al., 2020)	8.64M	0.953	0.949	0.955	0.948	0.885	0.953	0.949	0.947	0.948	0.885	0.953	0.949	0.947	0.948	0.885
CloudU-NetV2 (Shi et al., 2021)	17.48M	0.958	0.955	0.952	0.953	0.895	0.958	0.955	0.952	0.953	0.895	0.958	0.955	0.952	0.953	0.900
MA-SegCloud (Zhang et al., 2022)	14.63M	<u>0.969</u>	<u>0.971</u>	**0.970**	**0.970**	**0.940**	<u>0.969</u>	<u>0.960</u>	**0.970**	<u>0.965</u>	**0.940**	<u>0.969</u>	<u>0.970</u>	**0.970**	**0.970**	**0.940**
SCANet-lite	0.09M	0.944	0.936	0.944	0.940	0.865	0.944	0.936	0.944	0.940	0.865	0.944	0.936	0.944	0.940	0.865
SCANet	2.49M	0.961	0.955	0.958	0.957	0.901	0.961	0.955	0.958	0.957	0.902	0.961	0.955	0.958	0.957	0.902
SCANet-large	4.29M	**0.970**	**0.971**	<u>0.960</u>	<u>0.966</u>	<u>0.923</u>	**0.970**	**0.971**	<u>0.960</u>	**0.966**	<u>0.923</u>	**0.970**	**0.971**	<u>0.960</u>	<u>0.966</u>	<u>0.923</u>

Table 1: Comparison with other state-of-the-art methods on day time, night time, and day+night time images. We highlight the optimal and suboptimal methods in **bold** and <u>underline</u>, respectively. "/" indicates that the corresponding metric was not reported in the original paper.