MeViT: Medium-Resolution Vision Transformer for Semantic Segmentation on Landsat Satellite Imagery for Agriculture in Thailand

Teerapong Panboonyuen[*1,2], Chaiyut Charoenphon[1], Chalermchon Satirapod[1]

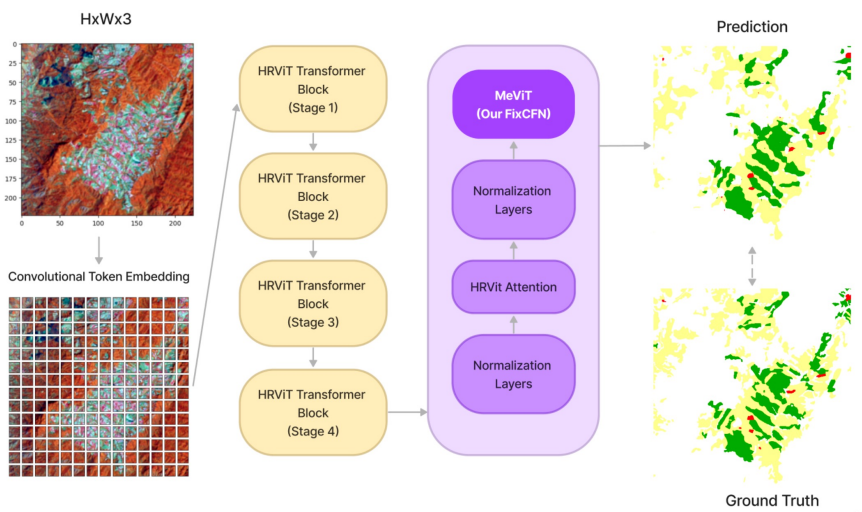
1 Faculty of Engineering, Chulalongkorn University

2 MARS (Motor AI Recognition Solution), Bangkok, Thailand

Contact: teerapong.panboonyuen@gmail.com, **Website:** https://kaopanboonyuen.github.io/

SCAN ME

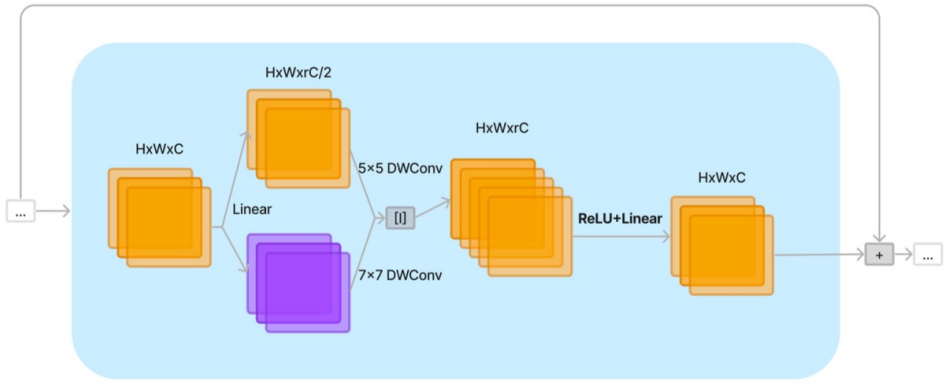
Overview



Our work presents **MeViT**, a Vision Transformer framework designed to perform semantic segmentation on Landsat satellite imagery for agriculture in Thailand, focusing on crops like **para rubber, corn, and pineapple.**

Improving Model Design with MeViT

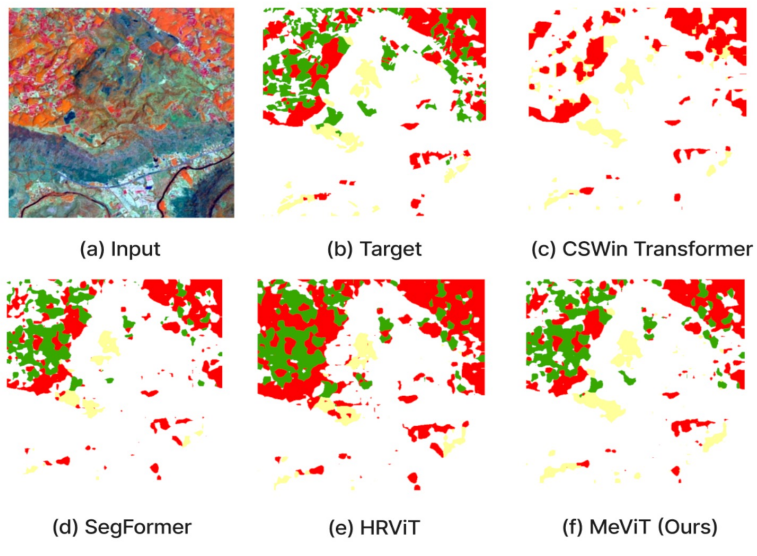
- **MeViT** incorporates multi-branch architecture for multi-scale learning, with depth-wise convolutions to effectively capture both local and global features.
- The **MixCFN** module, improved with ReLU activation, boosts precision while maintaining computational efficiency, and the model is trained and evaluated on the Thai Landsat-8 dataset.



Results

Model	Precision	Recall	Mean F1	Mean IoU
AutoDeeplab [34]	0.8946	0.8156	0.8533	0.7293
SwinTransformer [35,36]	0.9065	0.9055	0.906	0.8092
Twins [37]	0.8985	0.9168	0.9076	0.8112
CSWinTransformer [38]	0.8928	0.9313	0.9117	0.8168
SegFormer [39]	0.8979	0.9243	0.9109	0.8165
HRViT [23]	0.9111	0.9165	0.9138	0.823
MeViT (Ours)	0.9222	0.9469	0.9344	0.8363

Model	Para Rubber	Corn	Pineapple
AutoDeeplab [34]	0.8537	0.9379	0.8487
SwinTransformer [35,36]	0.921	0.966	0.811
Twins [37]	0.8953	0.8703	0.848
CSWinTransformer [38]	0.9127	0.9428	0.7546
SegFormer [39]	0.9021	0.8912	0.8222
HRViT [23]	0.8876	0.9419	0.8014
MeViT (Ours)	0.9239	0.9785	0.9087



- **MeViT Performance**: Outperformed other methods with **92.22% Precision**, **94.69% Recall**, **93.44% F1 Score**, and **83.63% IoU** on the dataset.
- **Visual Comparison**: MeViT achieves sharp object boundaries and effectively identifies rare classes, as illustrated in crop segmentation examples.
- **(para rubber, corn, pineapple).**

Conclusion

MeViT sets a new benchmark for transformer models in medium-resolution satellite imagery, combining multi-scale learning and high accuracy for agriculture-focused segmentation.

Reference: *Panboonyuen, Teerapong, Chaiyut Charoenphon, and Chalermchon Satirapod. "MeViT: A Medium-Resolution Vision Transformer for Semantic Segmentation on Landsat Satellite Imagery for Agriculture in Thailand." Remote Sensing 15.21 (2023): 5124.*