A Novel Framework Based on Mask R-CNN and Histogram Thresholding for Scalable Segmentation of New and Old Rural Buildings

Supplementary Materials

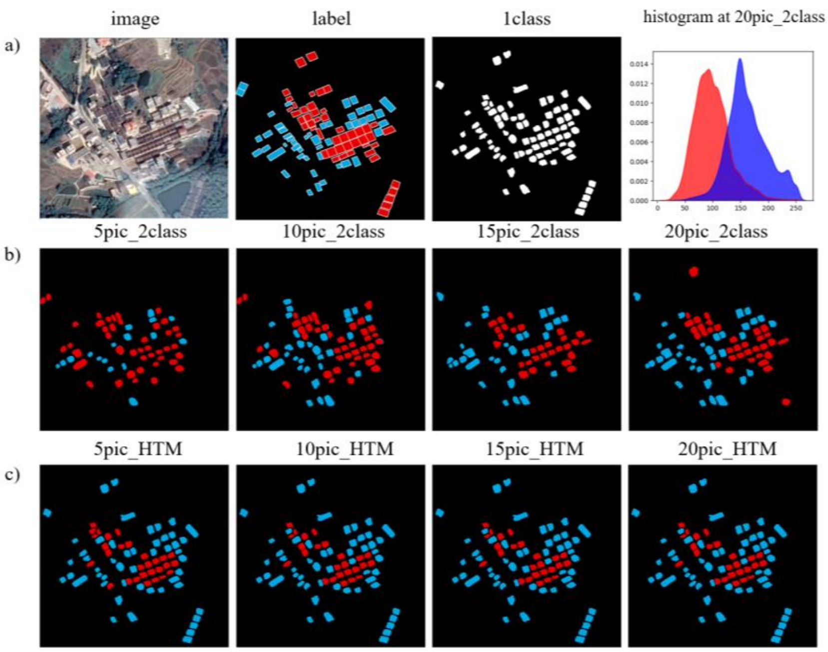

Figure S1: The result comparison of Site 2 between the baseline Mask R-CNN model and the HTMask R-CNN framework. (a) the satellite image from the test set, the second column shows the annotations, the third column shows the result of one-class model (R1), and the fourth column shows the grayscale histograms of the new and old building footprints from R2 when training images = 20. (b) the result of the two-class model (R2), with incremental training samples from 5 to 20. (c) the result of the HTMask R-CNN framework (R3) with incremental training samples.

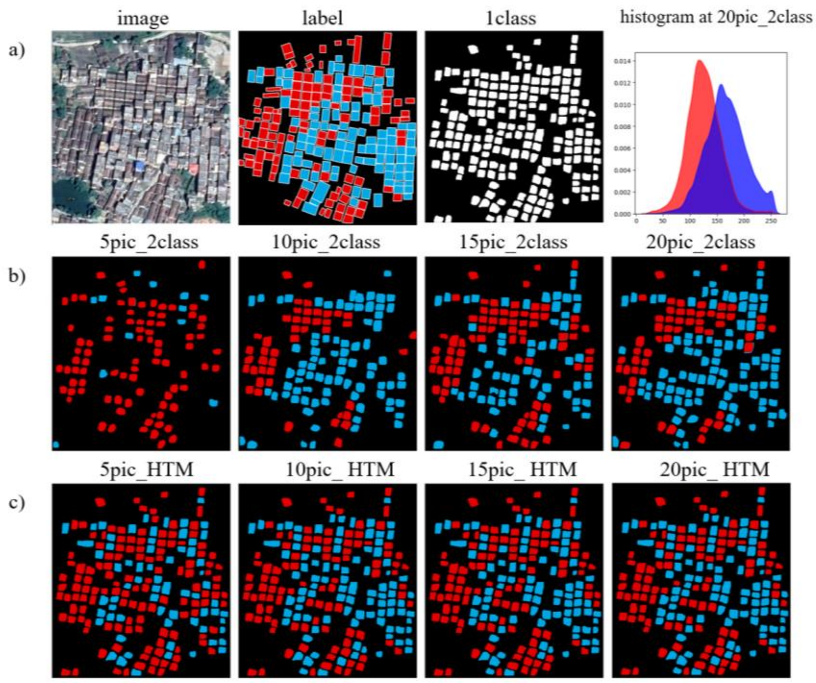

Figure S2: The result comparison of Site 3 between the baseline Mask R-CNN model and the HTMask R-CNN framework. (a) the satellite image from the test set, the second column shows the annotations, the third column shows the result of one-class model R1, and the fourth column shows the grayscale histograms of the new and old building footprints from R2 when training images = 20. (b) the result of the two-class model, R2, with incremental training samples from 5 to 20. (c) the result of the HTMask R-CNN framework R3 with incremental training samples.

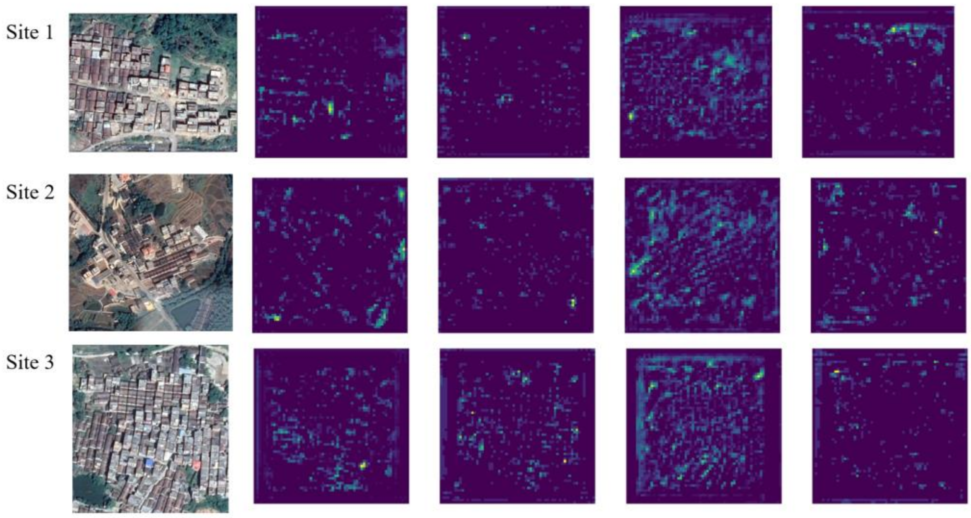

Figure S3: The feature maps of the three sites in the two-class model at the 50th epoch. The first column is the satellite image from the test set. The second to fifth columns show the feature maps of the res4w layer (64 * 64 * 1024) of the ResNet101 backbone.

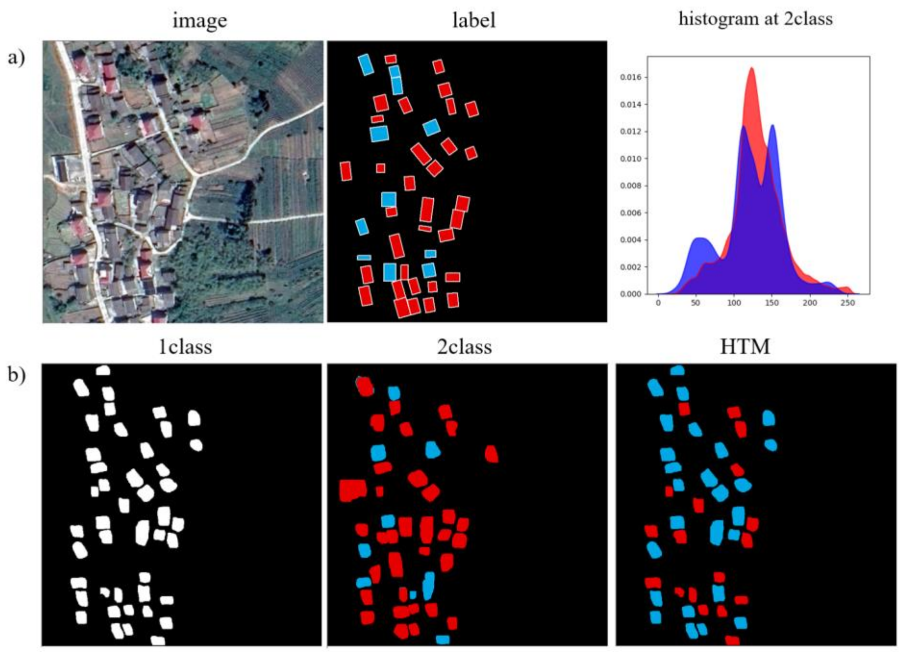

Figure S4: The baseline Mask R-CNN model and the HTMask R-CNN framework were tested in Fuliang County, Jiangxi Province, China. (a) the satellite image from the test set. The second column shows the annotations, and the third column shows the gray distribution of new and old buildings from the two-class model R2. (b) the first column shows the result of the one-class model, R1. The second column shows the result of the two-class model, R2. The third column shows the result of the HTMask R-CNN framework, R3.

Table S1: The mAP_{75} between R2 and R3 at all levels of training

models\epoch	1	5	10	15	20	25	30	35	40	45	50
5_2class	0.00	0.00	0.00	0.00	0.03	0.01	0.01	0.02	0.02	0.02	0.02
5_HTM	0.00	0.05	0.01	0.07	0.08	0.09	0.12	0.09	0.09	0.10	0.09
10_2class	0.00	0.00	0.02	0.03	0.03	0.05	0.07	0.05	0.06	0.06	0.07
10pic_ HTM	0.00	0.06	0.02	0.07	0.08	0.09	0.12	0.09	0.09	0.10	0.10
15_2class	0.00	0.00	0.01	0.01	0.01	0.02	0.04	0.04	0.06	0.05	0.07
15pic_HTM	0.00	0.06	0.02	0.07	0.08	0.09	0.12	0.09	0.09	0.10	0.10
20_2class	0.00	0.00	0.04	0.05	0.08	0.07	0.07	0.10	0.11	0.09	0.10
20_HTM	0.00	0.06	0.02	0.07	0.08	0.09	0.12	0.09	0.09	0.10	0.10

Table S2: Tuning hyperparameters.

We used the settings below to tune the hyperparameters.

Hyperparameters	Range	Optimal value
Iteration	50	50
Batch size	2	2
Learning rate	[1e-5, 1e-2]	1e-5
Optimizer	SGD, Adam	SGD
Weight decay	[1e-5, 1e-3]	1e-4
Learning momentum	[0.7, 0.9]	0.9

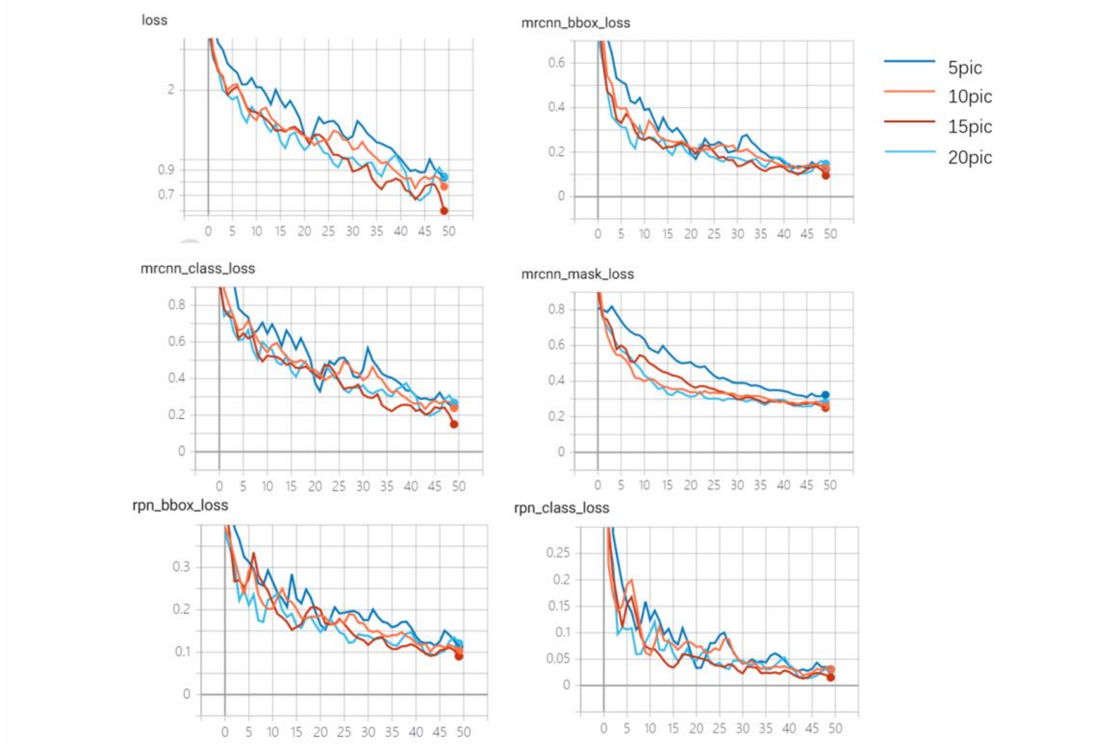

Figure S5: The loss function per training iteration. It shows the multi-task loss, the mrcnn_bbox_loss, the mrcnn_class_loss, the mrcnn_mask_loss, the rpn_bbox_loss, and the rpn_class_loss.

Equation S1: Loss function of the two-class model R2.

$$L = L_{rpn\_bbox} + L_{rpn\_cls} + L_{cls} + L_{bbox} + L_{mask} \qquad (1)$$

Mask R-CNN utilizes a multi-task loss function(L) that combines the loss of classification, localization and segmentation mask as illustrated in the equation below. We optimized the total loss L during the training stage. The variable L_{rpn_bbox} represents the rpn_bbox_loss, the variable L_{rpn_cls} represents the rpn_class_loss, the variable L_{cls} represents the mrcnn_class_loss, the variable L_{bbox} represents the mrcnn_bbox_loss, the variable L_{mask} represents the mrcnn_mask_loss.