Appendix

The additional binary metrics used in Table 2:

$$Accuracy = \frac{TP + TN}{P + N} \tag{3}$$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

$$F_1 = \frac{2TP}{2TP + FP + FN}, \tag{6}$$

where P and N denote positive and negative ground truth values and TP, TF, FP, and FN denote the corresponding true and false predictions.

Input	GT	ResNet-50	PIDNet-S	PIDNet-M	PIDNet-L

Table 9. Visualisation of outputs obtained using the BoxSnake supervised ResNet-50 and the corresponding distilled PIDNet models.

Input	GT	ResNet-101	PIDNet-S	PIDNet-M	PIDNet-L
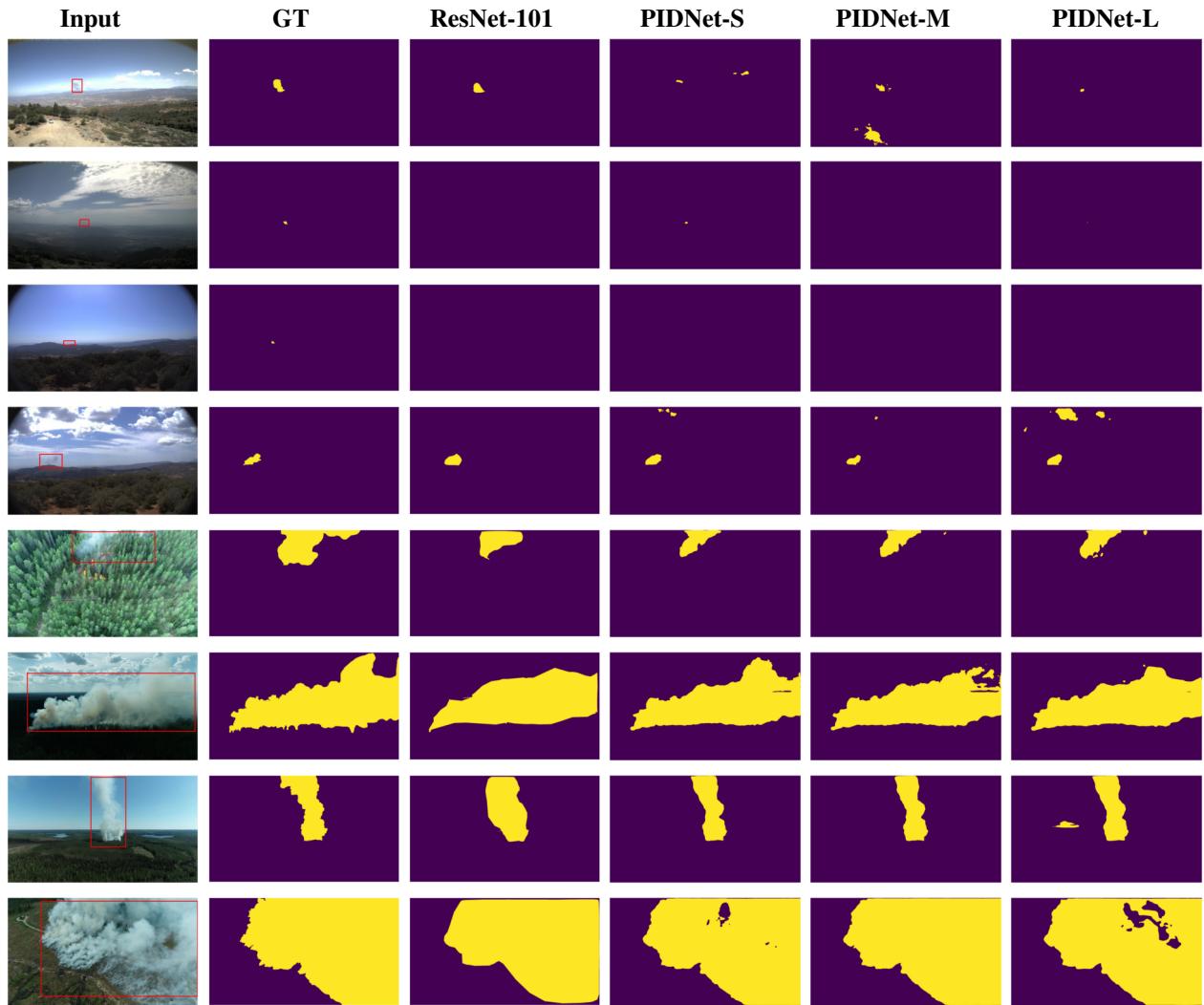
Table 10. Visualisation of outputs obtained using the BoxSnake supervised ResNet-101 and the corresponding distilled PIDNet models.
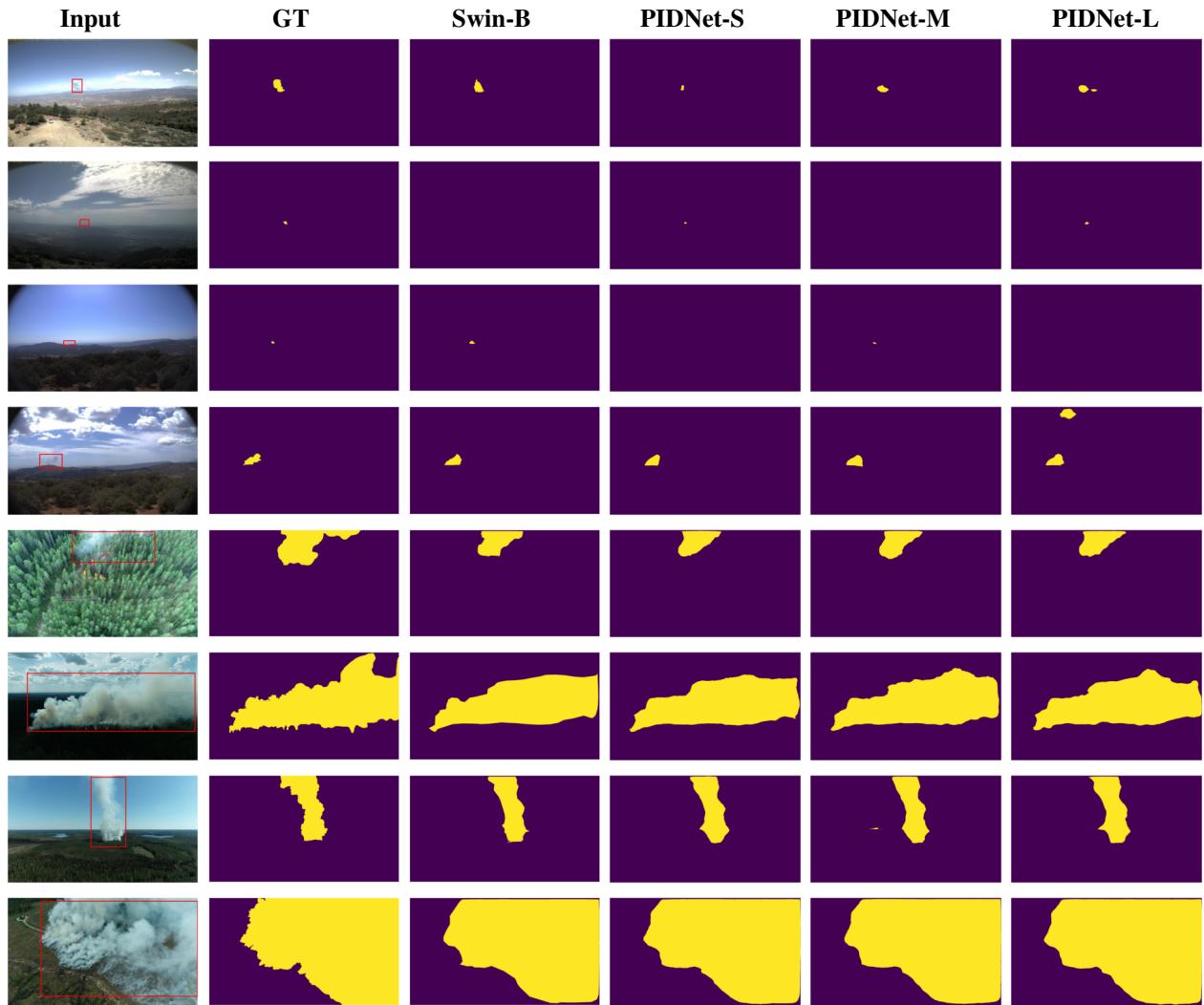
Table 11. Visualisation of outputs obtained using the BoxSnake supervised Swin-B and the corresponding distilled PIDNet models.
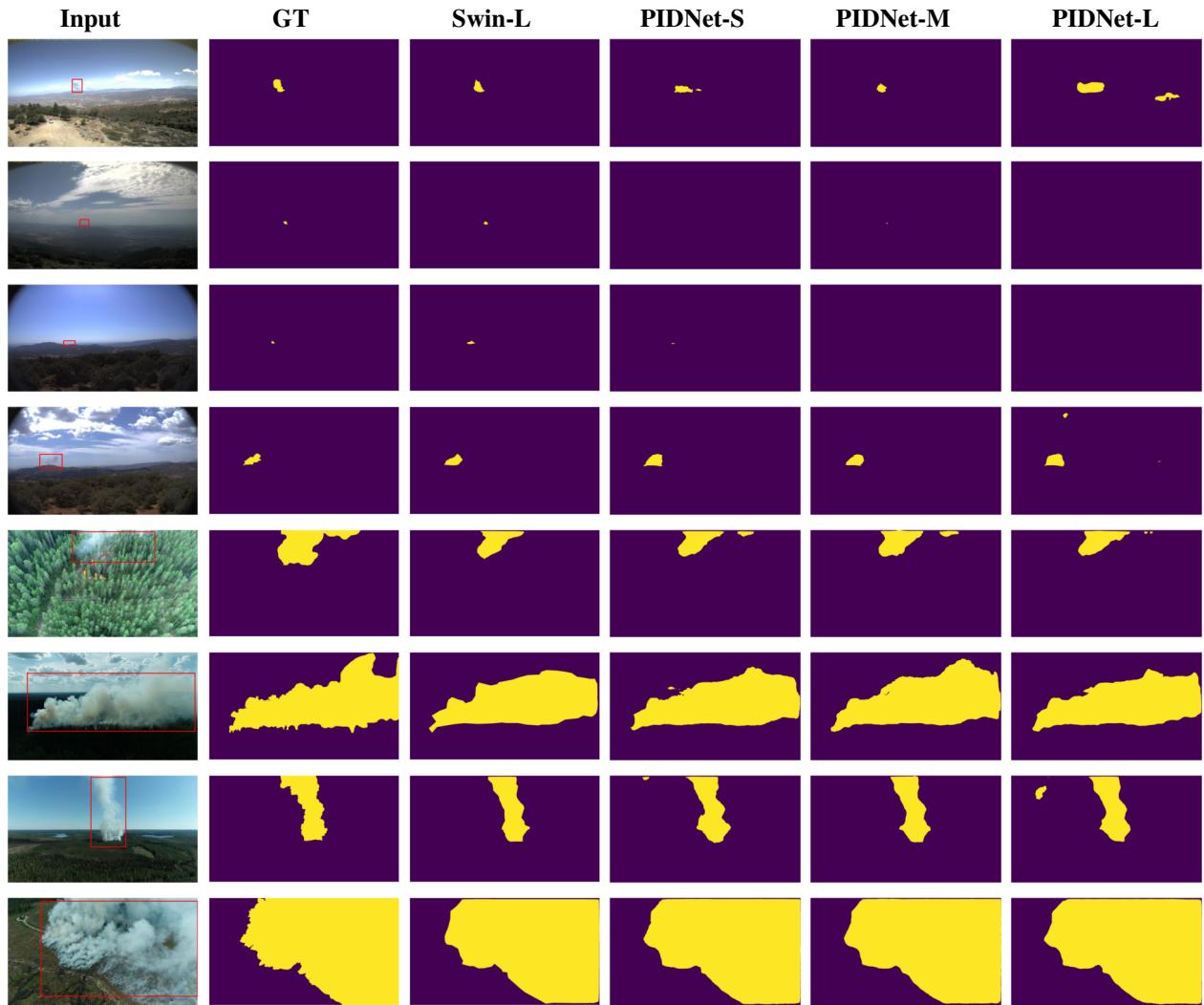
Table 12. Visualisation of outputs obtained using the BoxSnake supervised Swin-L and the corresponding distilled PIDNet models.
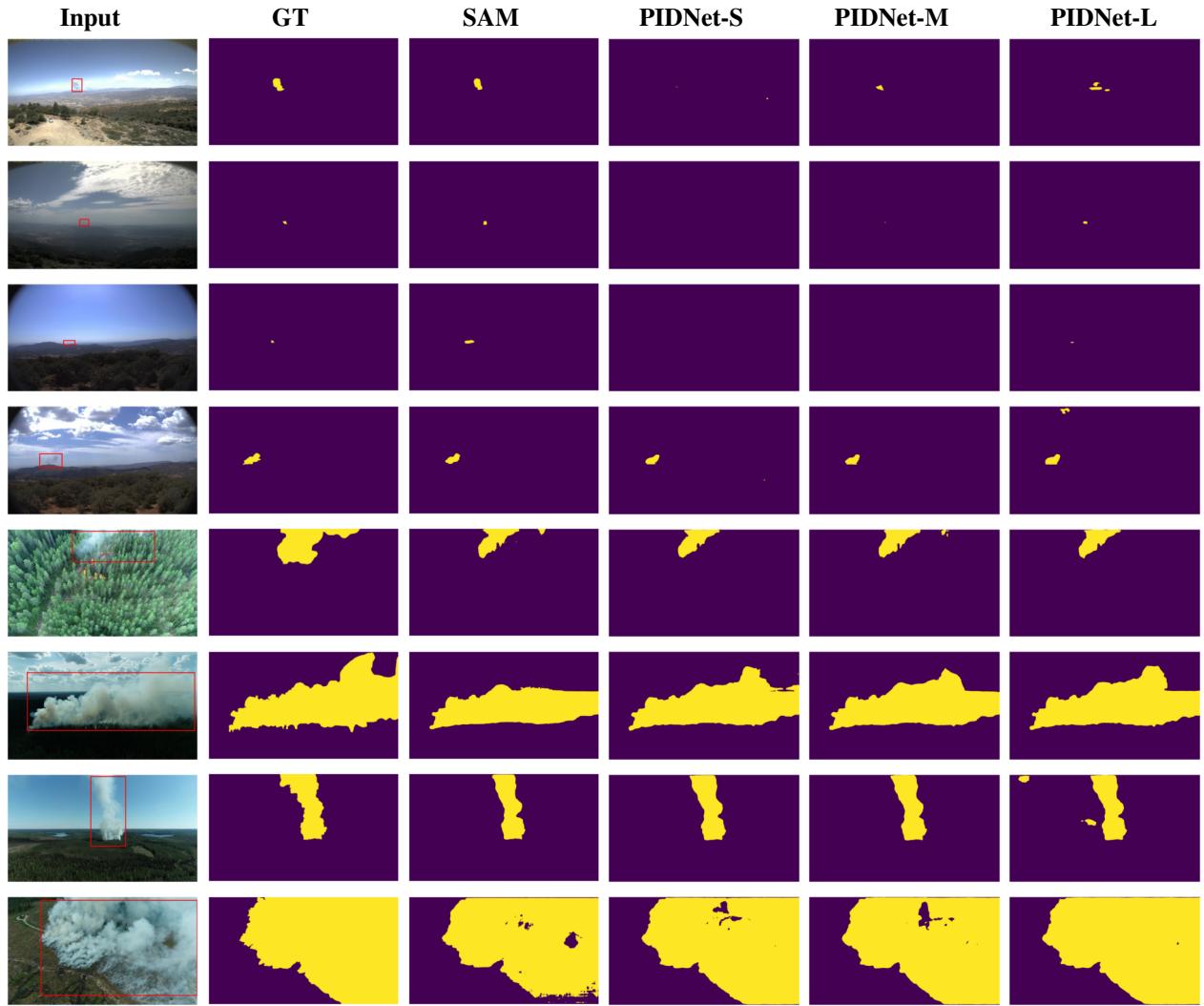
Table 13. Visualisation of outputs obtained using SAM and the corresponding distilled PIDNet models.

Device	CPU				GPU							
TensorRT	-				No				Yes			
Power setting	10W	15W	25W	MAX	10W	15W	25W	MAX	10W	15W	25W	MAX
Model	FPS											
PIDNet-S	1.22	1.40	1.32	1.71	6.31	6.25	6.19	12.47	10.90	10.77	11.97	25.88
PIDNet-M	0.47	0.54	0.70	0.88	2.89	2.79	2.83	5.44	4.65	4.39	5.25	9.86
PIDNet-L	0.32	0.37	0.53	0.63	2.17	2.18	2.17	4.29	3.35	3.37	3.96	7.47

Table 14. PIDNet model frame rate benchmark on the NVIDIA Jetson Orin NX computer with different power supply limitation, GPU, CPU and TensorRT options. With the GPU the inputs and outputs were 1080×1920 pixels and all operations were computed in full precision (32-bit float). For the CPU, images with a resolution of 576×1024 were used. With the 1080×1920 resolution the fps was always below 1. The specifications of the different power settings are shown in Table 15.

Power Budget	MAXN	10W	15W	25W
Online CPU	8	4	4	8
CPU Max Frequency (MHz)	1984	1190.4	1420.8	1497.6
GPU Max Frecuency (MHz)	918	612	612	408
DLA cores	2	1	1	2
DLA Core Max Frequency (MHz)	614.4	153.6	614.4	614.4
DLA Falcon Max Frequency (MHz)	614.4	153.6	614.4	614.4
PVA Cores	1	0	0	1
PVA VPS Max Frequency (MHz)	704	N/A	N/A	512
PVA AXI Max Frequency (MHz)	486.4	N/A	N/A	358.4

Table 15. The specifications of the different NVIDIA Jetson Orin NX power settings [40].

Teacher	Student	mIoU	Accuracy	Precision	Recall	F_1
ResNet-50-RCNN-FPN	PIDNet-S	0.444	0.855	0.690	0.585	0.561
ResNet-101-RCNN-FPN	PIDNet-S	0.437	0.857	**0.841**	0.499	0.572
Swin-B-FPN	PIDNet-S	<u>0.472</u>	**0.874**	0.735	<u>0.600</u>	<u>0.585</u>
Swin-L-FPN	PIDNet-S	**0.482**	<u>0.873</u>	0.749	**0.629**	**0.607**
SAM	PIDNet-S	0.410	0.848	<u>0.824</u>	0.481	0.534
ResNet-50-RCNN-FPN	PIDNet-M	**0.460**	**0.867**	0.750	**0.601**	**0.586**
ResNet-101-RCNN-FPN	PIDNet-M	0.452	0.855	0.774	<u>0.542</u>	0.575
Swin-B-FPN	PIDNet-M	0.406	0.857	0.737	0.512	0.525
Swin-L-FPN	PIDNet-M	<u>0.454</u>	**0.867**	<u>0.822</u>	0.509	<u>0.582</u>
SAM	PIDNet-M	0.374	0.832	**0.867**	0.406	0.502
ResNet-50-RCNN-FPN	PIDNet-L	**0.513**	**0.882**	<u>0.807</u>	<u>0.588</u>	**0.637**
ResNet-101-RCNN-FPN	PIDNet-L	0.410	0.845	0.799	0.486	0.538
Swin-B-FPN	PIDNet-L	<u>0.507</u>	<u>0.881</u>	0.796	0.593	<u>0.633</u>
Swin-L-FPN	PIDNet-L	0.463	0.862	0.696	**0.630**	0.584
SAM	PIDNet-L	0.479	0.873	**0.823**	0.552	0.614

Table 16. Test set metrics of the different student models on the Croatian Center for Wildfire Research data. In all metrics higher is better. All teacher models besides SAM were trained using BoxSnake. The best results for each student model are in bold and the second best are underlined.
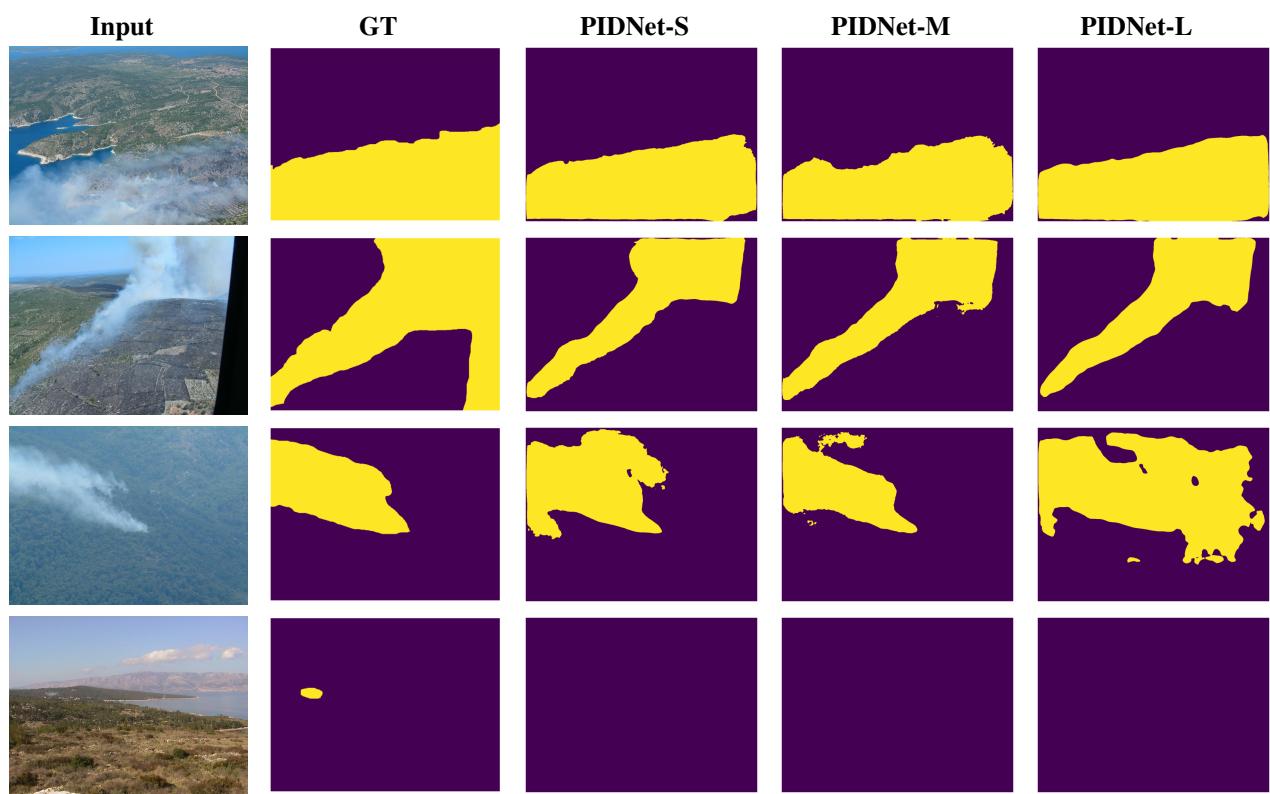
Table 17. Qualitative results of smoke predictions obtained on the Croatian Center for Wildfire Research data using the different SAM supervised models.