

Supplementary Material: Voxel Scene Graph for Intracranial Hemorrhage

1 Segmentation Results

Dataset	Bleeding	Ventricle System	Midline
INSTANCE2022	79.2±25.2	80.4±13.4	77.6±0.35
Private Cohort	66.5±37.5	66.5±15.8	66.3±10.0

Table 1. Segmentation performance (mean patient Dice score) of our method on both datasets.

2 Relation Prediction: Ablation Study

Replace with GT			INSTANCE2022			Private Cohort		
Box	No FP	Mask	R@8 \uparrow	mR@8 \uparrow	mAP@8 \uparrow	R@8 \uparrow	mR@8 \uparrow	mAP@8 \uparrow
			64.2 \pm 8.9	71.4 \pm 5.1	40.6 \pm 4.9	19.8 \pm 6.4	22.1 \pm 6.9	21.9 \pm 7.6
		✓	63.0 \pm 7.4	70.7 \pm 4.4	40.6 \pm 5.6	20.7 \pm 3.9	21.8 \pm 5.0	18.7 \pm 5.6
	✓		69.6 \pm 6.2	75.1 \pm 3.0	54.8 \pm 4.9	20.0 \pm 2.5	21.3 \pm 4.7	29.2 \pm 4.4
	✓	✓	69.4 \pm 8.7	75.8 \pm 5.9	56.4 \pm 8.2	20.8 \pm 4.2	22.3 \pm 6.5	30.4 \pm 4.0
✓	✓		72.5 \pm 4.2	77.0 \pm 3.5	54.6 \pm 3.7	57.2 \pm 13.5	45.7 \pm 11.3	48.8 \pm 6.9
✓	✓	✓	72.2 \pm 4.4	76.7 \pm 3.6	54.7 \pm 4.2	56.9 \pm 13.7	45.6 \pm 11.4	47.4 \pm 6.0

Table 2. Ablation study: we reuse the **segmentation-grounded V-MOTIF** models from the **Scene Graph Prediction** task with no further training. We use the ground truth annotation (GT) to replace or correct parts of the prediction. Box) GT object localization and label. No FP) Remove false positives from detected objects. Mask) GT segmentation mask.

Replace with GT			INSTANCE2022			Private Cohort		
Box	No FP	Mask	R@8 \uparrow	mR@8 \uparrow	mAP@8 \uparrow	R@8 \uparrow	mR@8 \uparrow	mAP@8 \uparrow
			74.0 \pm 6.1	78.1 \pm 4.6	34.0 \pm 2.8	22.7 \pm 5.0	22.0 \pm 4.6	16.5 \pm 3.1
		✓	73.1 \pm 8.7	77.7 \pm 6.3	34.7 \pm 3.7	23.8 \pm 5.6	23.2 \pm 5.9	16.2 \pm 3.8
	✓		76.3 \pm 4.4	80.2 \pm 3.5	50.8 \pm 2.7	29.2 \pm 1.4	27.6 \pm 2.3	24.7 \pm 4.1
	✓	✓	76.3 \pm 4.4	80.2 \pm 3.5	51.1 \pm 2.8	28.8 \pm 2.2	27.3 \pm 2.6	24.8 \pm 3.9
✓	✓		77.2 \pm 6.1	80.0 \pm 5.7	53.9 \pm 4.0	56.3 \pm 6.5	50.5 \pm 4.7	38.0 \pm 6.8
✓	✓	✓	76.5 \pm 7.3	79.7 \pm 6.4	53.6 \pm 4.5	55.7 \pm 6.0	50.2 \pm 4.5	39.1 \pm 6.3

Table 3. Ablation study: we reuse the **segmentation-grounded V-IMP** models from the **Scene Graph Prediction** task with no further training. We use the ground truth annotation (GT) to replace or correct parts of the prediction. Box) GT object localization and label. No FP) Remove false positives from detected objects. Mask) GT segmentation mask.