

1 Supplementary Material

1.1 Metrics

The metrics employed to assess binary segmentation performance include:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (1)$$

$$PA = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

where TP, FP, TN, and FN denote true positives, false positives, true negatives, and false negatives, respectively.

To assess depth estimation performance, the following metrics are employed:

$$AbsRel = \frac{1}{|D|} \sum_{d \in D} |d^* - d|/d^* \quad (5)$$

$$SqRel = \frac{1}{|D|} \sum_{d \in D} |d^* - d|^2/d^* \quad (6)$$

$$RMSE = \sqrt{\frac{1}{|D|} \sum_{d \in D} |d^* - d|^2} \quad (7)$$

$$RMSElog = \sqrt{\frac{1}{|D|} \sum_{d \in D} |\log d^* - \log d|^2} \quad (8)$$

$$\delta = \frac{1}{|D|} \{d \in D | \max(\frac{d^*}{d}, \frac{d}{d^*}) < 1.25\} \times 100\% \quad (9)$$

where d and d^* denote the predicted and ground truth depth values, respectively.

1.2 Additional Qualitative Results

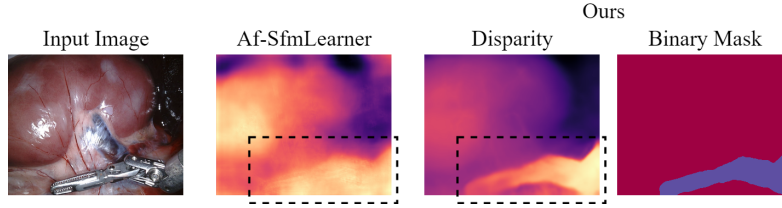


Figure 1 presents zero-shot qualitative results for depth estimation and binary tool segmentation on the Hamlyn Dataset.

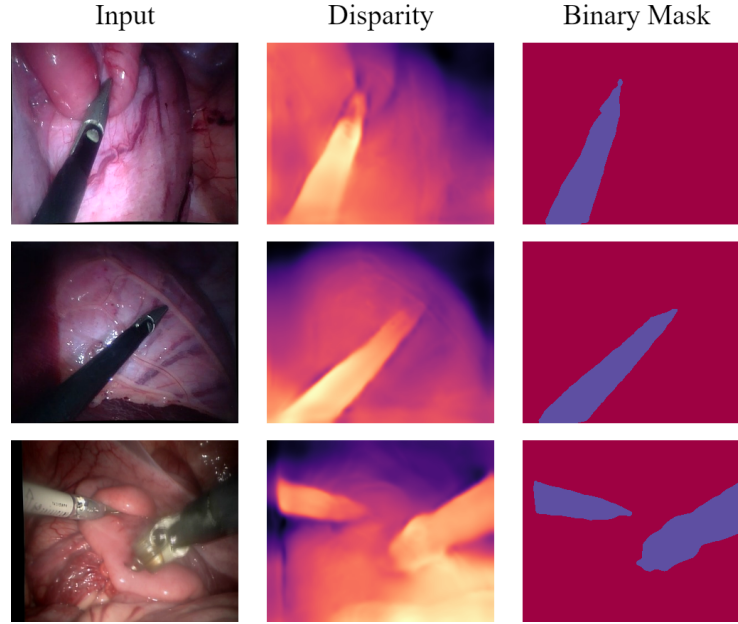


Fig. 1. Zero-shot qualitative results on the Hamlyn dataset.