Supplementary Materials: Diff-VPS: Video Polyp Segmentation via a Multi-task Diffusion Network with Adversarial Temporal Reasoning

Yingling Lu, Yijun Yang, Zhaohu Xing, Qiong Wang, Lei Zhu

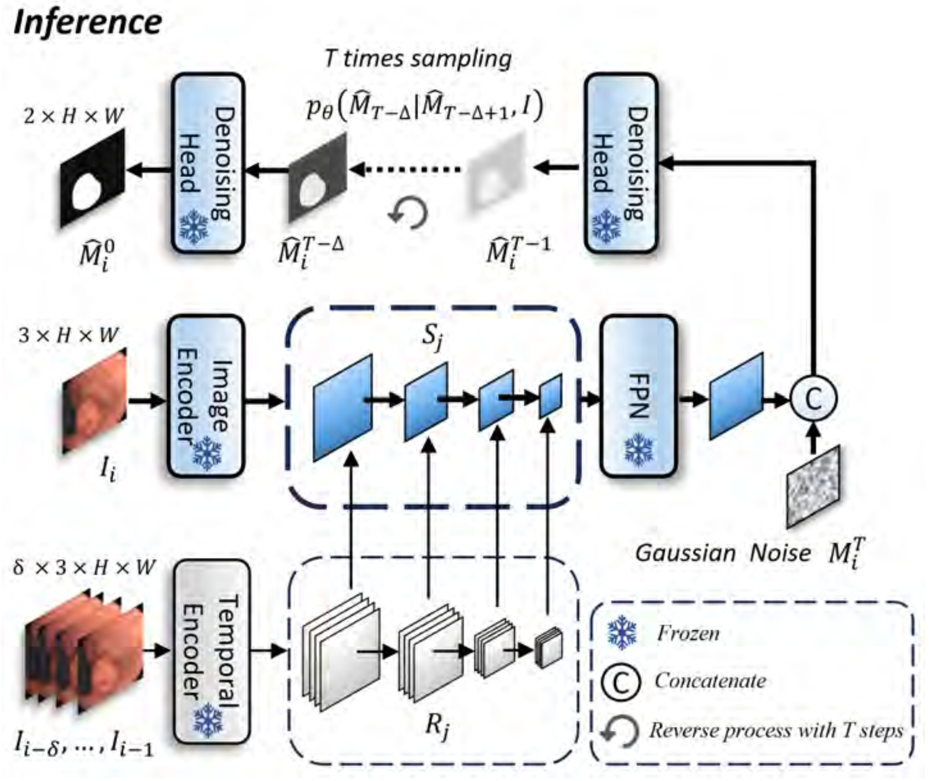

Fig. 1. Overview of Diff-VPS inference phase. Given a test clip of $\delta + 1$ frames, the model starts with a random noise map sampled from a Gaussian distribution and gradually refines the prediction. To speed up the inference, we adopt the paradigm of DDIM. In each sampling step t, the random noise M_i^T or the predicted noisy map $\hat{M}_i^{T-\varDelta}$ from the last step is fused with the conditional feature map and sent to the frozen denoising head for mask prediction.

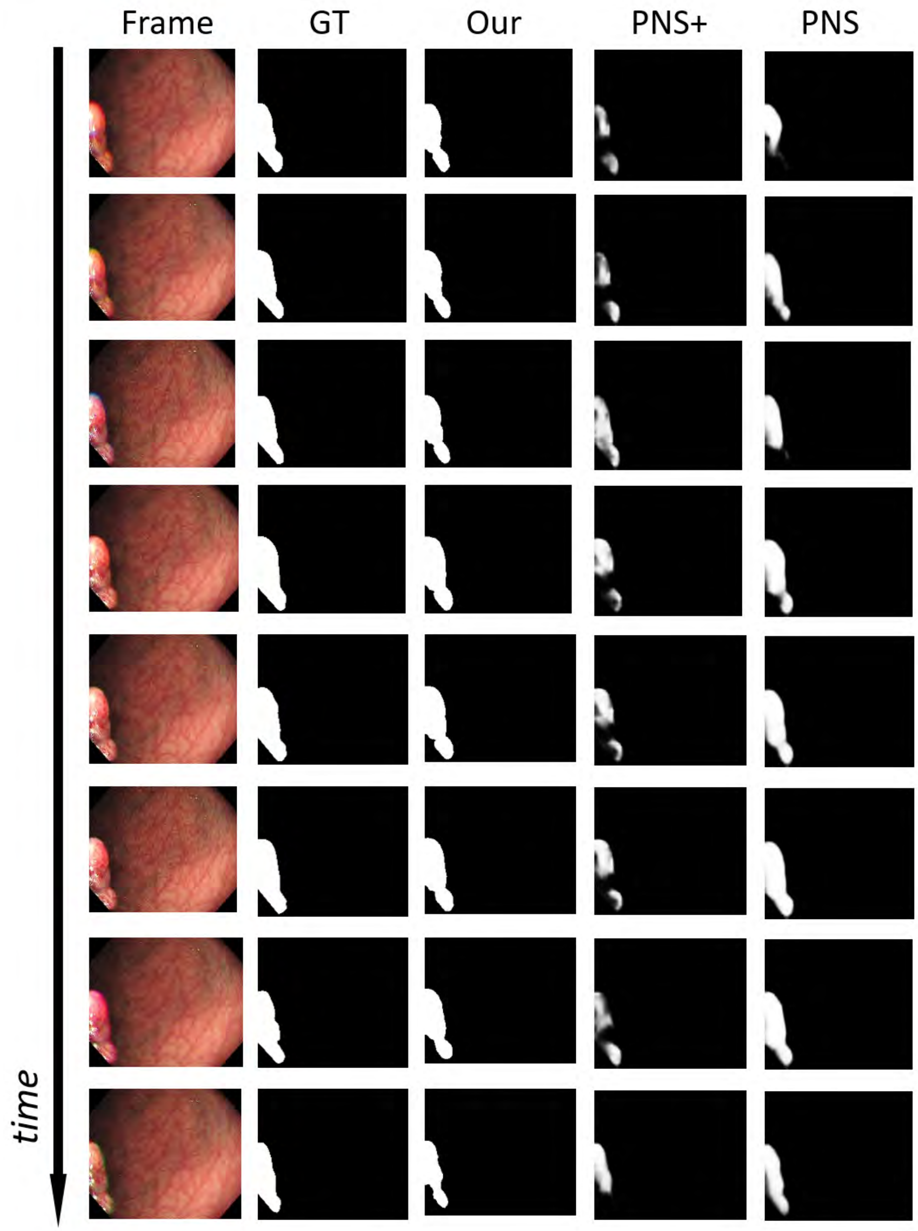

Fig. 2. Qualitative results on SUN-SEG hard-seen dataset case 83 .