DeNVeR: Deformable Neural Vessel Representations for Unsupervised Video Vessel Segmentation

Supplementary Material

6. Additional Visualization Results

Figs. 11 to 13 demonstrate a comprehensive comparison where we consider supervised learning [21] as the upper bound for the vessel segmentation task, as well as all baseline methods mentioned in the main paper. For the supervised learning approach, both image-based and video-based inputs were considered. The image-based input utilized only the annotated image, while the video-based input involved using the annotated image along with two preceding and two subsequent frames, totaling five frames, as input. The results show that although supervised learning theoretically offers the best performance, our method achieves close to those of supervised learning methods without ground truth. Additionally, we found that using five consecutive images as input for nn-UNet [21] was only slightly better than using a single image as input. In contrast, our method exhibits significant improvement compared to both the traditional Hessian-based filter and self-supervised methods, demonstrating that the robust performance of our approach is not solely attributed to the increase in input images. We showcase some examples at the following anonymous URL: https://colab.research.google.com/drive/1IYGiJECwAaoLPq7KGHQE_dvtrdHz9fUA?authuser=2&hl=zh-tw#scrollTo=n1ppvOhqbRkV. Additionally, we include 5 out of 111 sequences of our XACV dataset and the source code in the zip file.

7. Temporal Coherency

Our method takes an entire X-ray video as input, thus producing segmentation results with better temporal coherency. Temporal coherency is essential for making medical diagnoses, especially when dealing with blood flow in vessels. Therefore, we conduct visual comparisons between our method and other compared methods by slicing horizontally or vertically and stacking the segmentation results. The results in Fig. 14 show our method strikes a better balance between segmentation accuracy and temporal coherency. While other baseline methods either produce false segmentation results or do not maintain consistent prediction along the temporal dimension.

8. Impact of prior

We add experiments demonstrating how the Hessian prior affects subsequent results, including ablation studies with different prior qualities. In our experiments, We replace the Hessian prior mask with a better mask (FreeCOS prediction) and observe a 2.5% improvement in dice score. We also provide visual results in Fig. 15.

9. Model and training details

We elaborate on the architectural details and training methodologies for all neural network components.

9.1. Stage1: Layer Separation on bootstrapping

This MLP (Multi-Layer Perceptron) model consists of these main components:

- Input Layer: Input dimension is 3 (color channels).
- Hidden Layer 1: Takes input of dimension 3 and outputs a dimension of 2. This layer has 256 neurons, with 4 hidden layers and an outermost linear layer.
- Hidden Layer 2: Takes input of dimension 2 and outputs a dimension of 3. This layer also has 256 neurons, with 4 hidden layers and an outermost linear layer.
- Output Layer: Takes input of dimension 3 and outputs a dimension of 4. This layer has 256 neurons, with 4 hidden layers and an outermost linear layer.

Important hyperparameters:

- λ_{smooth}: Controls the weight of the smoothness term in the bootstrapping loss. We set it to 0.001.
- λ_{limit}: Regularizes the foreground MLP in the bootstrapping loss. We set it to 0.02.

9.2. Stage 2: Vessel decomposition

In stage 2, We employ different standard U-Nets with three down and three up layers to predict masks and foreground canonical images. Both models utilize CNNs with 3×3 kernels, strides of 1, and padding of 1. We use batch norm and bilinear downsampling or upsampling after each layer in the U-Nets.

Training setting. We set the batch size to 16, including 512×512 image resolution, and trained on a 4090 GPU. Training on 80-90 cardiac images takes approximately 20 minutes.

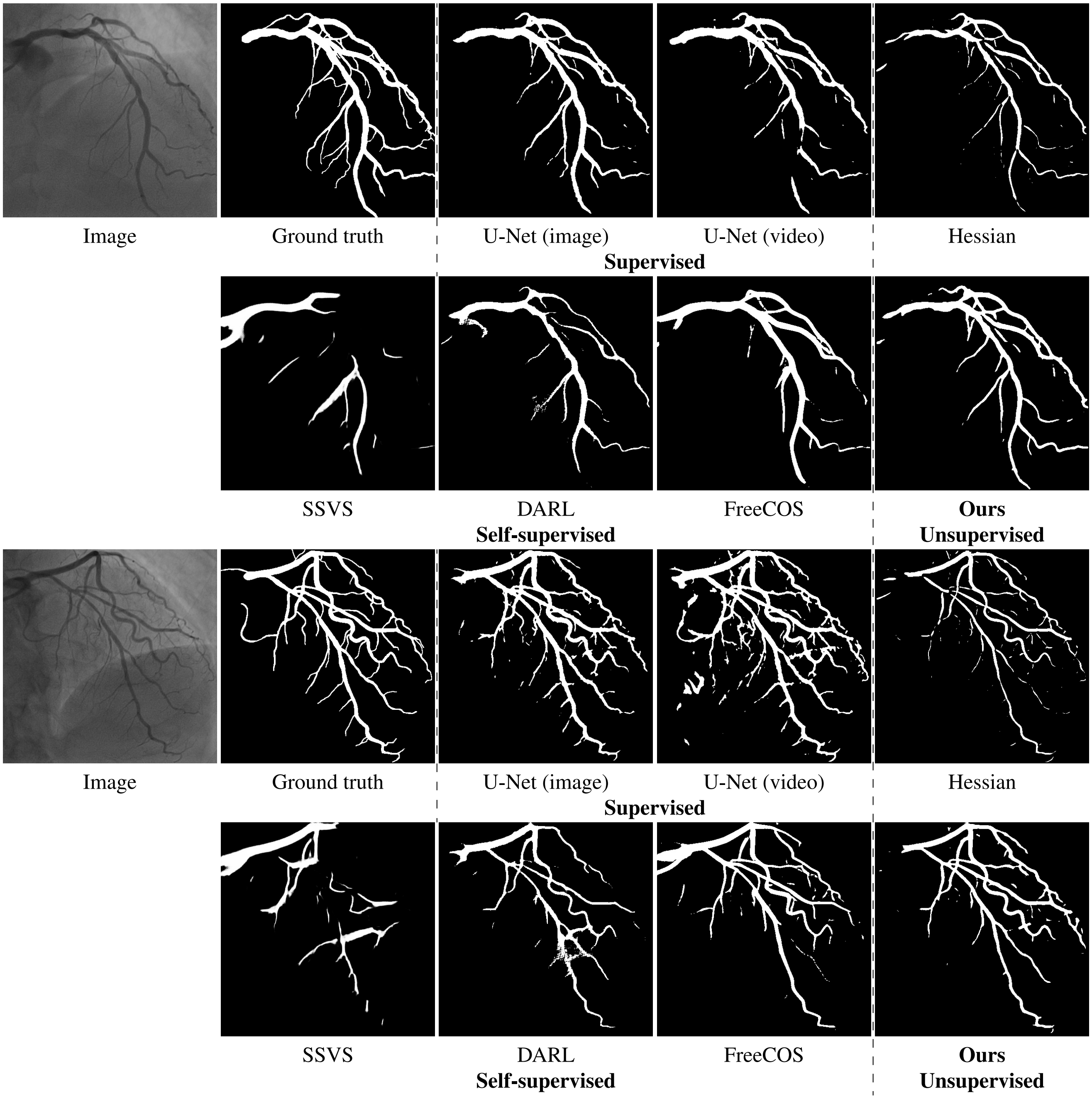

Figure 11. **Additional visualization results on the vessel segmentation.**

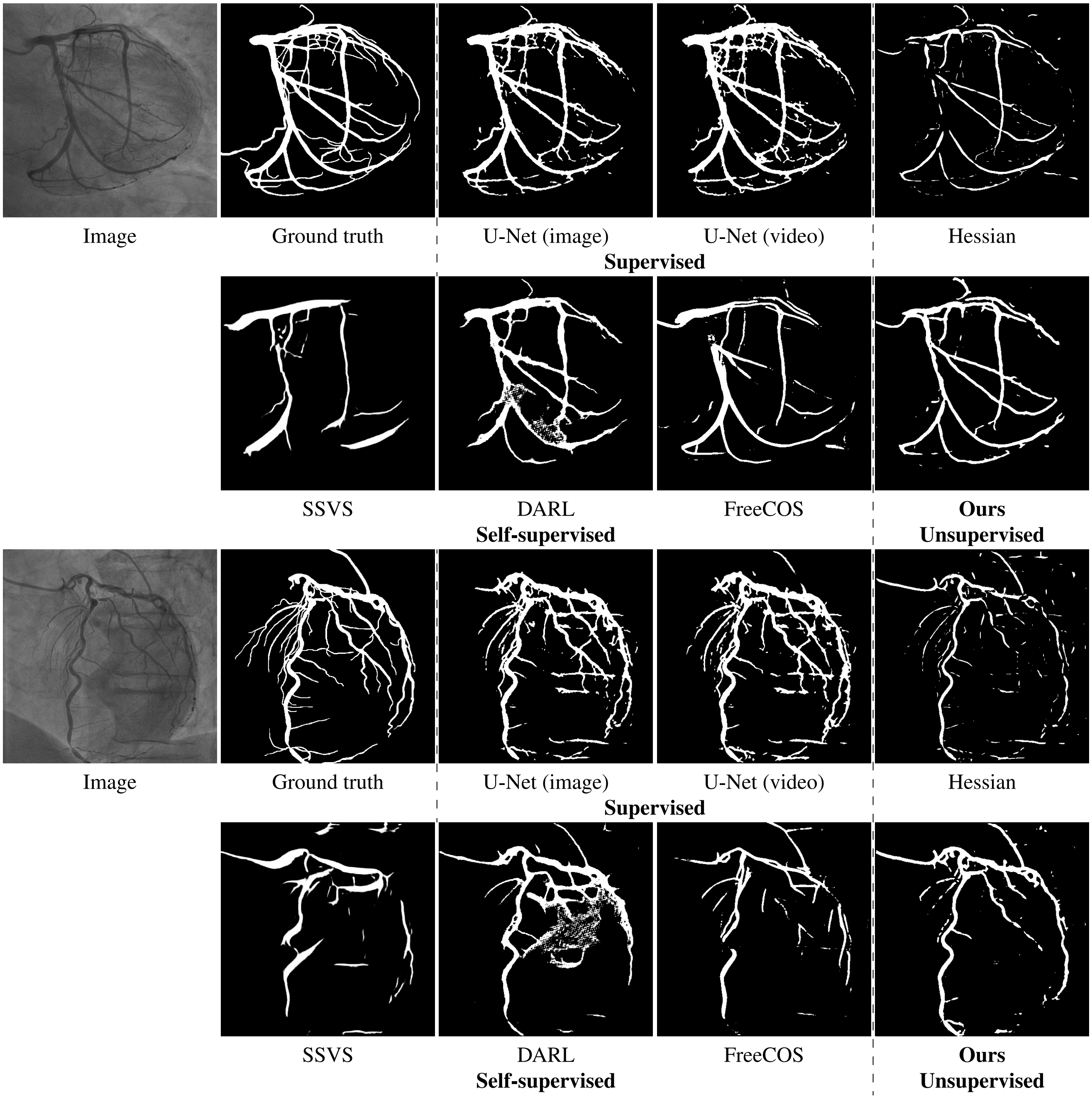

Image | Ground truth | U-Net (image) | U-Net (video) | Hessian

Supervised

SSVS | DARL | FreeCOS | **Ours**

Self-supervised | **Unsupervised**

Image | Ground truth | U-Net (image) | U-Net (video) | Hessian

Supervised

SSVS | DARL | FreeCOS | **Ours**

Self-supervised | **Unsupervised**

Figure 12. **Additional visualization results on the vessel segmentation.**

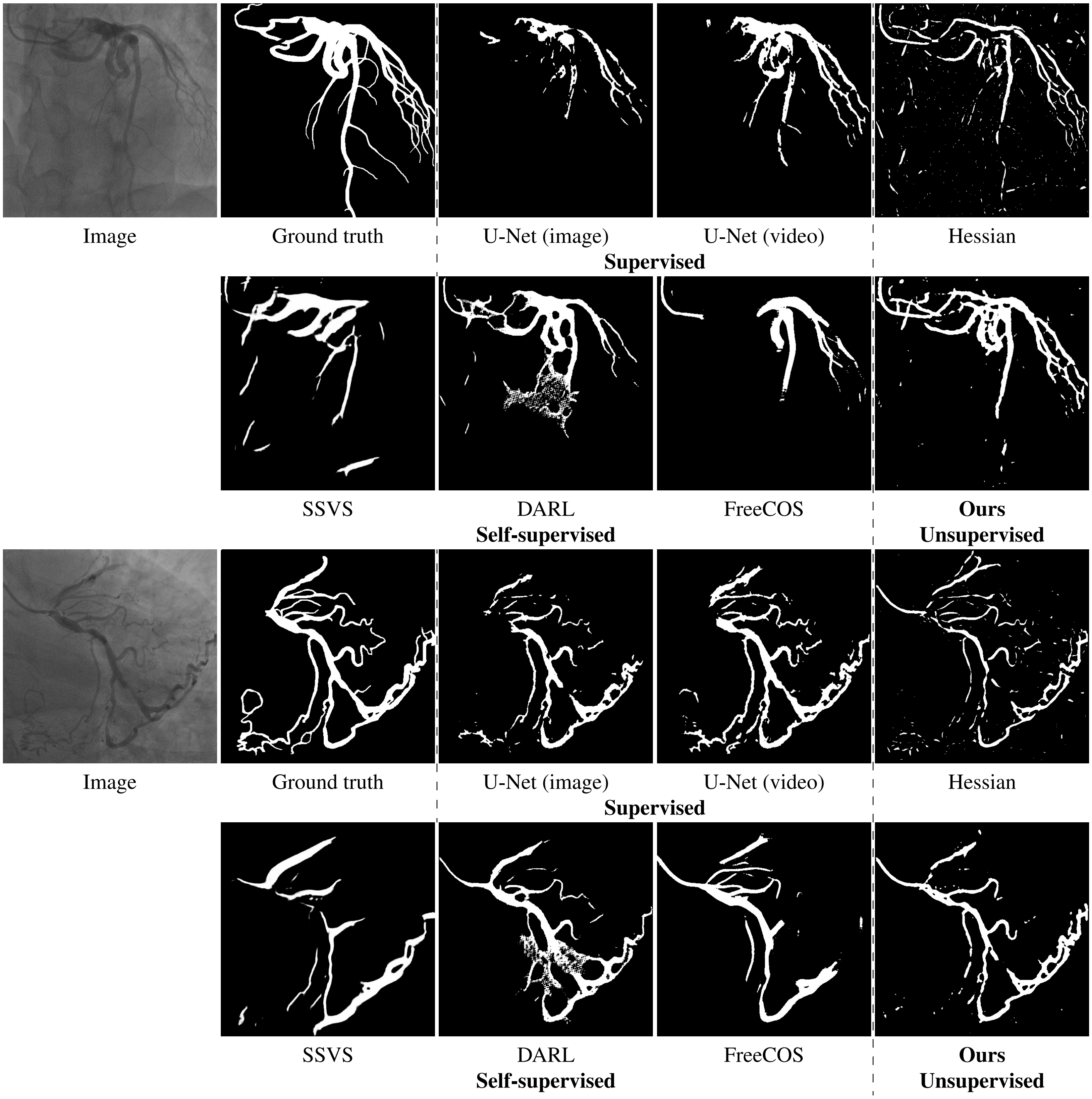

Figure 13. **Additional visualization results on the vessel segmentation.**

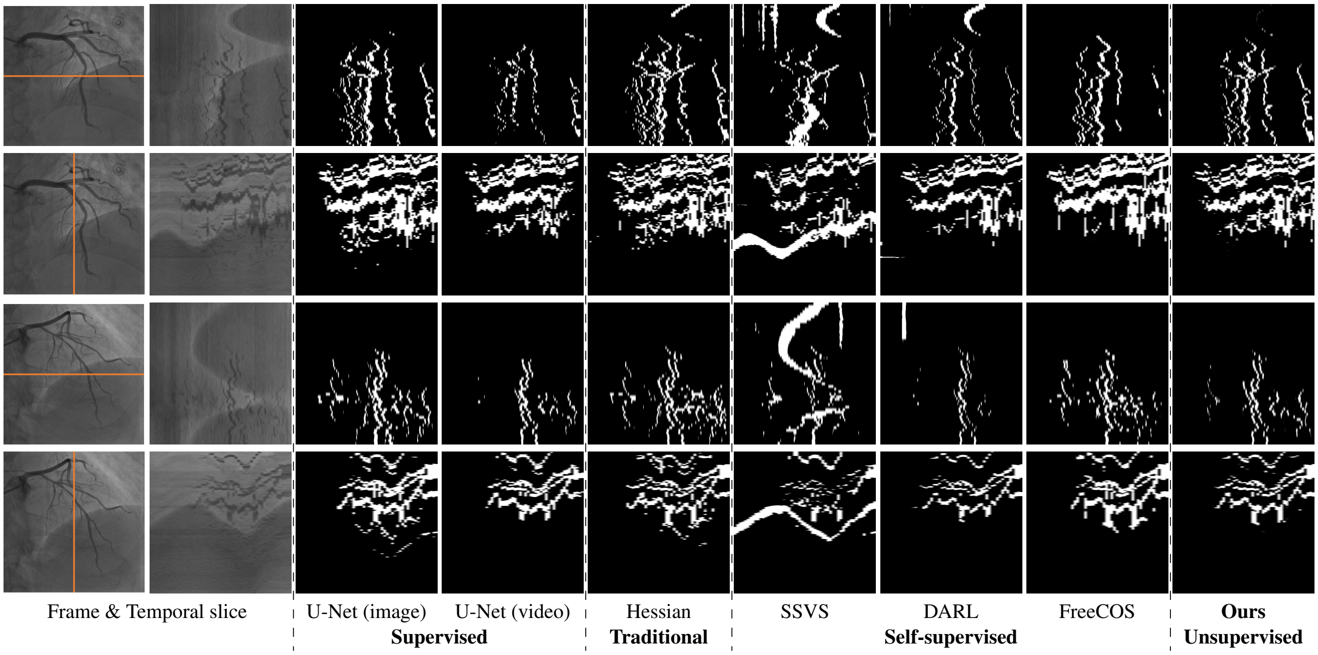

Figure 14. **Temporal coherency.**

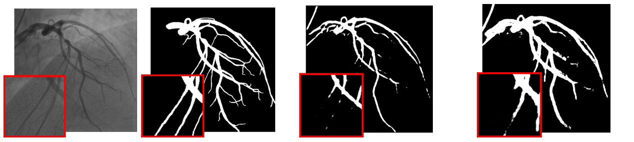

Figure 15. **Impact of prior.** We test the impact of the prior on our model. Replacing the original Hessian prior with the FreeCOS prediction results in a 2.5% improvement in dice score. Red zoom-in patches show that the FreeCOS-based prior has better predictive capabilities.