Centre for Artificial Intelligence and Robotics
Hong Kong Institute of Science & Innovation, Chinese Academy of Sciences

IEEE International Conference on Image Processing
27 - 30 October 2024 | Abu Dhabi, UAE

PWISeg: Weakly-Supervised Surgical Instrument Instance Segmentation

Zhen Sun, Huan Xu, Jinlin Wu, Zhen Chen, Hongbin Liu, Zhen Lei

Background

- While surgical instrument localization is crucial for computer-assisted surgeries, existing research mainly focuses on endoscopic settings, neglecting the broader operating room environment where occlusion poses significant challenges.
- Instance segmentation methods could address occlusion issues, but the high cost of obtaining pixel-level annotations necessitates weakly supervised approaches that can achieve accurate results with less labor-intensive labeling.
- The lack of high-quality, open-source datasets for occluded surgical instrument localization hinders progress in this field, highlighting the need for both new datasets and weakly supervised methods to advance research in this area.

Methodology

- FCN-based architecture with box and mask prediction branches
- Supervised box branch training using bounding box annotations
- Weakly supervised mask branch training using:
  - Unsupervised projection loss: Leverages projection relation between masks and boxes

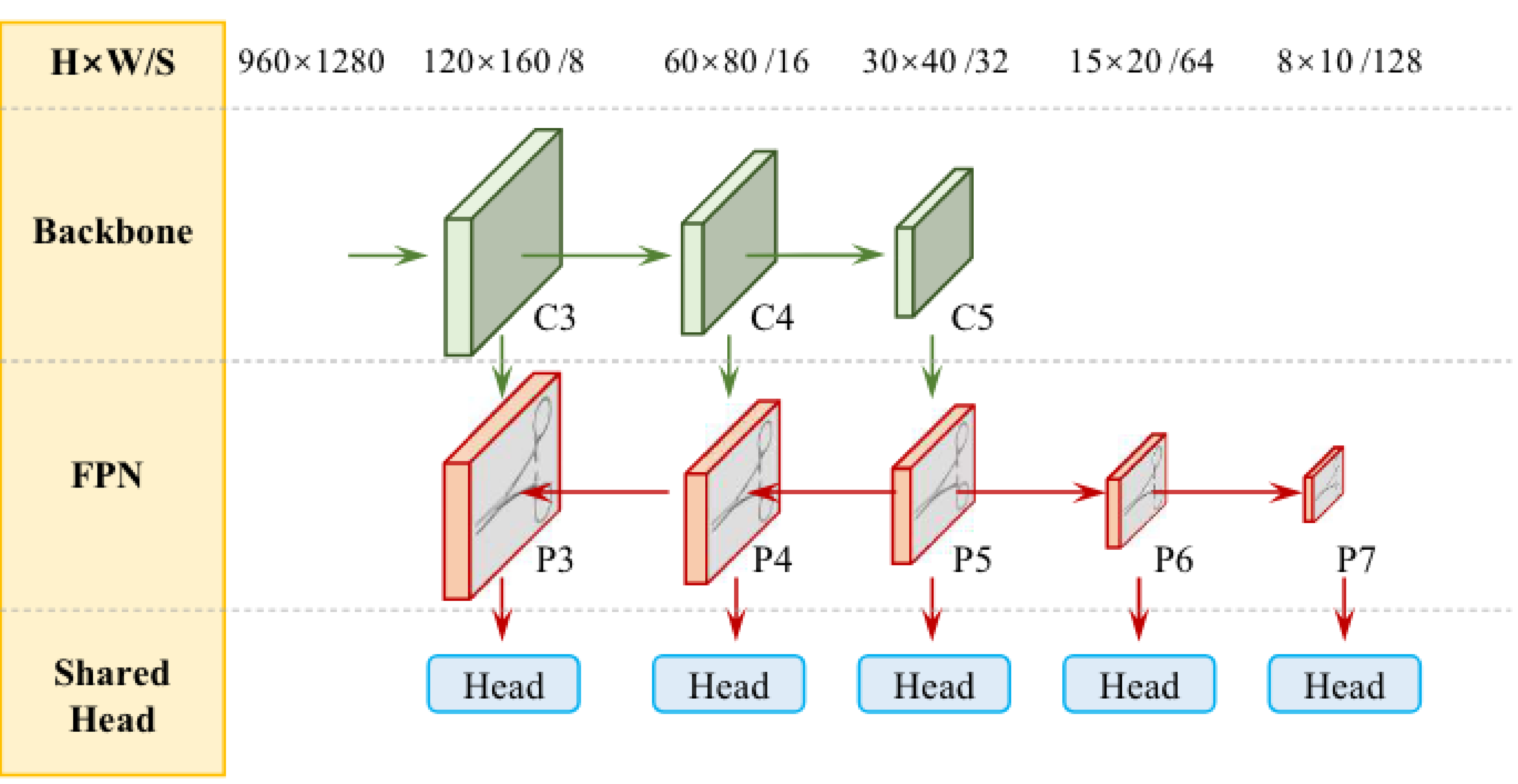
$$L_{\text{proj}} = Dice\left(\max_x(\hat{w}), \max_x(t)\right) + Dice\left(\max_y(\hat{w}), \max_y(t)\right) ,$$

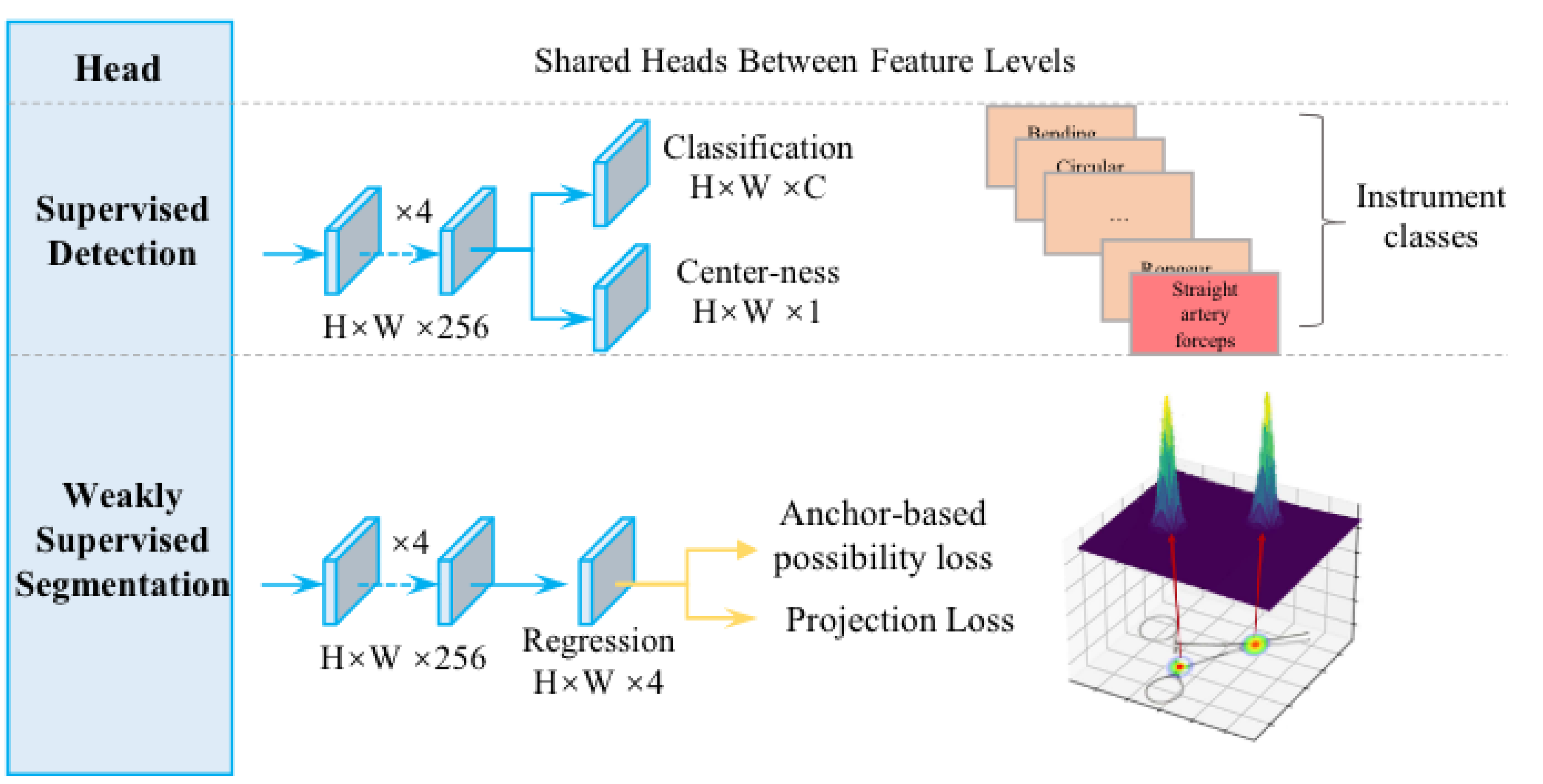
  - Key-pixels association loss: Diffuses labels of key points within bounding box

$$L_{\text{ass}} = -\frac{1}{N}\sum_{(x,y)\in bbox} \hat{y}_{(x,y)} \log p_{(x,y)} + (1-\hat{y}_{(x,y)}) \log p_{(x,y)} .$$

  - Key-pixels distribution loss: Optimizes distribution of key-pixels heatmap

$$L_{\text{dis}} = \sum_{(x,y)\in \text{bbox}} \| p_{(x,y)} - \Phi(h_{(x,y)}) \|_{L_1} ,$$





The dataset we collected.

Download here!

Results

- PWISeg outperforms state-of-the-art weakly supervised instance segmentation methods on both Surg-Inst and HOSPI-Tools datasets
- Strong performance on public HOSPI-Tools dataset demonstrates robustness

Method	Backbone	Detection			Segmentation		
		mAP	mAP_{50}	mAP_{75}	mAP	mAP_{50}	mAP_{75}
Discobox [19]	ResNet-50	62.50	90.40	73.40	13.70	36.40	8.90
BoxLevelSet [20]	ResNet-50	61.20	87.90	71.20	20.70	**69.40**	4.80
BoxInst [18]	ResNet-50	59.30	93.20	69.40	21.30	60.80	13.00
PWISeg (Ours)	ResNet-50	**64.20**	**96.80**	**75.70**	**23.90**	66.30	**13.80**

Table 2: Performance for object detection and segmentation on the Surg-Inst dataset.

Method	Backbone	Detection			Segmentation		
		mAP	mAP_{50}	mAP_{75}	mAP	mAP_{50}	mAP_{75}
Discobox [19]	ResNet-50	**74.20**	94.60	**87.90**	25.30	74.10	8.80
BoxLevelSet [20]	ResNet-50	72.60	94.30	80.10	28.10	80.10	10.30
BoxInst [18]	ResNet-50	66.00	88.20	74.90	29.10	77.10	15.00
PWISeg (Ours)	ResNet-50	73.20	**95.20**	84.40	**30.60**	**80.50**	**15.80**

Table 3: Performance for object detection and segmentation on the HOSPI-Tools dataset.

Loss Function			Detection			Segmentation		
L_{proj}	L_{ass}	L_{dis}	mAP	mAP_{50}	mAP_{75}	mAP	mAP_{50}	mAP_{75}
✓			51.60	84.40	57.50	15.20	53.91	9.70
	✓		52.40	83.60	57.70	18.50	51.90	9.80
		✓	40.90	70.10	44.30	11.10	35.80	6.30
✓	✓		60.10	94.40	71.50	20.40	55.20	11.10
✓	✓	✓	**64.20**	**96.80**	**75.70**	**23.90**	**66.30**	**13.80**

Table 4: The performance comparison on Surg-Inst dataset by using the different loss terms based on ResNet-50.

Conclusions

- Introduced Surg-Inst dataset to advance surgical instrument instance segmentation
- PWISeg effectively leverages weak annotations to produce strong segmentation
- Improves accuracy of occluded instrument segmentation
- Streamlines annotation and promises improvements in automated surgical tool recognition