SwIPE: Efficient and Robust Medical Image Segmentation with Implicit Patch Embeddings

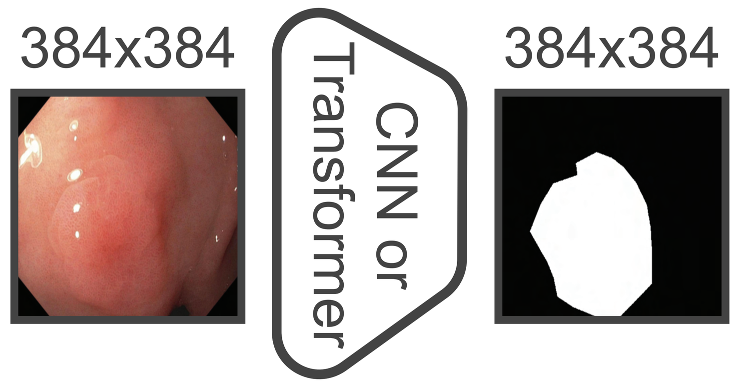
Code </>

Paper

Charley Yejia Zhang, Pengfei Gu, Nishchal Sapkota, Danny Z. Chen
University of Notre Dame, Computer Science and Engineering

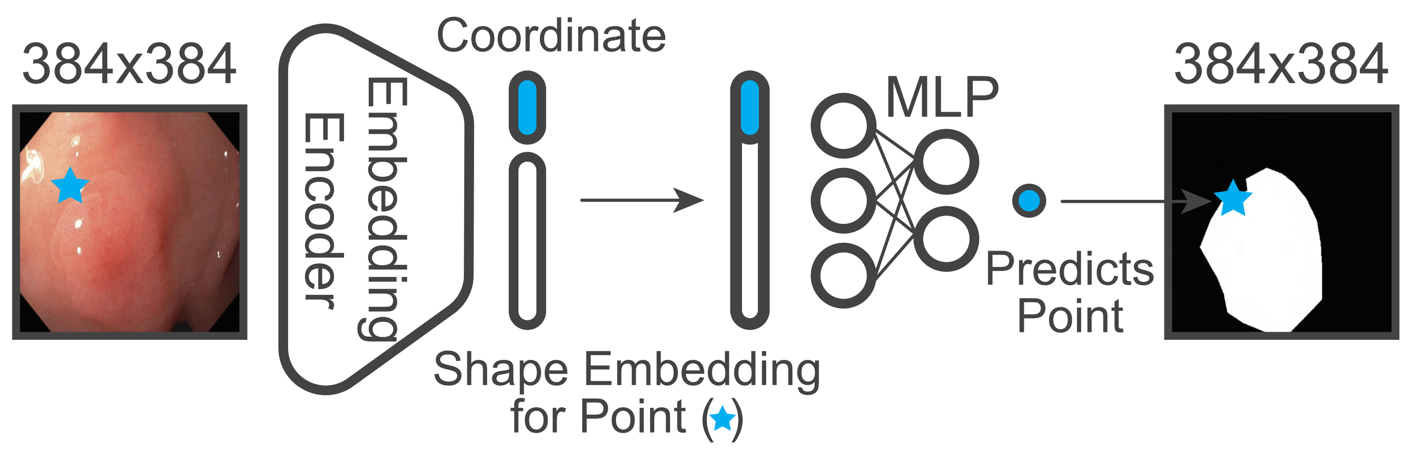
Motivation

I Drawbacks in Traditional Segmentation



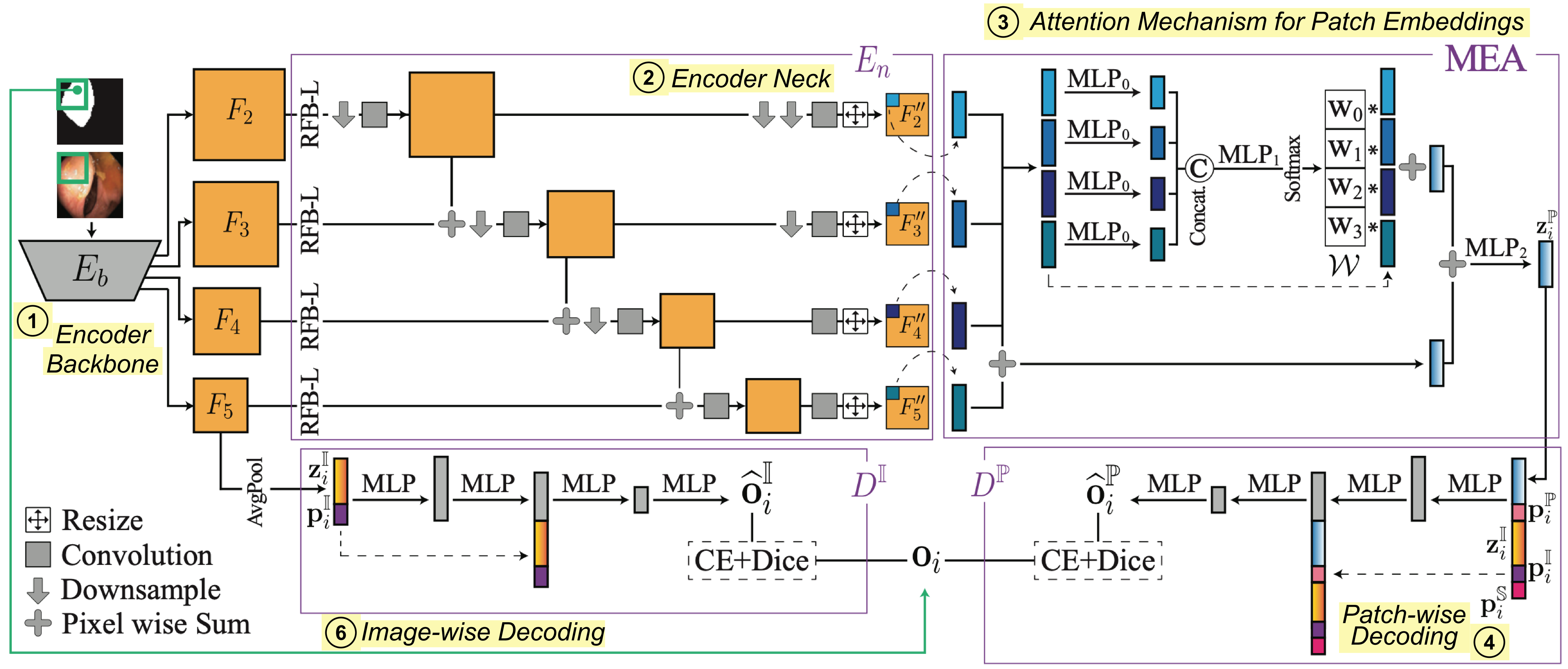
- Medical image segmentation (seg.) is a critical task where masks of pertinent objects are predicted
- Traditional seg. [1] uses **discrete representations** (masks as pixel grids) which has several disadvantages:
  - ✗ Limited spatial flexibility
  - ✗ Poor computational scaling
  - ✗ No direct shape modeling
  - ✗ Unconstrained predictions (unrealistic objects) esp. with limited labels and inbalanced classes

II Seg. w/ Implicit Neural Representations (INRs)



- INR-based seg. [2] uses **continuous representations** with an MLP decoder that maps an image coordinate and a shape embedding (i.e., vector with shape info.) to a class score. With INRs, fix-sized image inputs can:
  - ✓ predict masks of arbitrary sizes by modulating the density of continuous coordinates
  - ✓ attain constant memory scaling w.r.t. image size
  - ✓ directly model shapes (an object's boundary is the MLP decoder's decision boundary)
  - ✓ train better and more efficiently with less labels
- We divide objects into parts (i.e., images into patches) and use INR segmentation to predict local shapes, unlike others that do point-wise or per-image prediction

SwIPE Framework



SwIPE (Segmentation with Implicit Patch Embeddings) uses patch and image embeddings to predict point-wise occupancies:

1. Encoder Backbone extracts 4 multi-scale feature maps
2. Encoder Neck enriches features in a bottom-up cascade
3. MEA dynamically weighs global/abstract to local/fine-grained info. and gets final patch embeddings
4. Patch-wise decoding is performed at an image coordinate using 5 inputs
5. Image-wise decoding is performed given the image embedding and an image coordinate

Results

I. Main Results

+ Beats discrete SOTAs w/ 10x fewer params.
+ Outperforms INR SOTAs (+2.5% F1 on Sessile, +4.5% F1 on BCV)

2D Polyp Sessile				3D CT BCV			
Method	Params (M)	FLOPs (G)	Dice (%)	Method	Params (M)	FLOPs (G)	Dice (%)
Discrete Approaches							
U-Net$_{15}$	7.9	83.3	63.89±1.30	U-Net$_{15}$	16.3	800.9	74.47±1.57
PraNet$^{*}_{20}$	30.5	15.7	82.56±1.08	UNETR$^{*}_{21}$	92.6	72.6	81.14±0.85
Res2UNet$_{21}$	25.4	17.8	81.62±0.97	Res2UNet$_{21}$	38.3	**44.2**	79.23±0.66
Implicit Approaches							
OSSNet$_{21}$	5.2	6.4	76.11±1.14	OSSNet$_{21}$	7.6	55.1	73.38±1.65
IOSNet$_{22}$	4.1	**5.9**	78.37±0.76	IOSNet$_{22}$	6.2	46.2	76.75±1.37
SwIPE (*ours*)	**2.7**	10.2	**85.05**±0.82	SwIPE (*ours*)	**4.4**	71.6	**81.21**±0.94

II. Robustness (datasets & resolutions)

+ Superior inference dice for both segmentation tasks (left table)
+ Improved prediction accuracy across resolutions (right table)

Across Datasets

Method	Dice
Polyp Sessile → CVC	
1 PraNet	68.37
2 IOSNet	59.42
3 SwIPE	70.10
CT BCV → CT AMOS (liver class only)	
4 UNETR	81.75
5 IOSNet	79.48
6 SwIPE	82.81

Across Resolutions

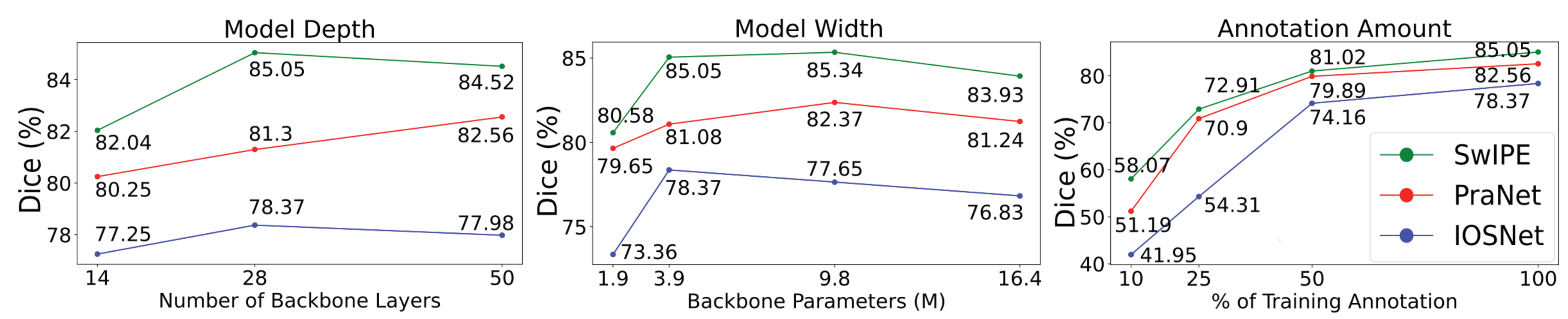
Method	Size	Dice
Varying Output Size		
1 PraNet	128↓	72.64
2 IOSNet	128↓	76.18
3 SwIPE	128↓	81.26
4 PraNet	896↑	74.95
5 IOSNet	896↑	78.01
6 SwIPE	896↑	84.33
Varying Input Size		
7 PraNet	128↓	68.79
8 PraNet	896↑	43.92

Evaluated on 2 tasks, and 4 datasets

Task 1 - 2D polyp seg. (colonoscopy): Sessile & CVC
Task 2 - 3D organ seg. (CT): BCV & AMOS

III. Model Parameter and Annotation Efficiency

+ On 2D polyp segmentation, SwIPE outperforms competitors across model depths, widths, and annotation availabilities



References

[1] Salpea et al. "Medical image segmentation: A review of modern architectures." ECCV Workshops. Cham: Springer Nature Switzerland, 2022.
[2] Khan et al. "Implicit Neural Representations for Medical Imaging Segmentation." MICCAI. Cham: Springer Nature Switzerland, 2022.