Empirical Analysis of a Segmentation Foundation Model in Prostate Imaging

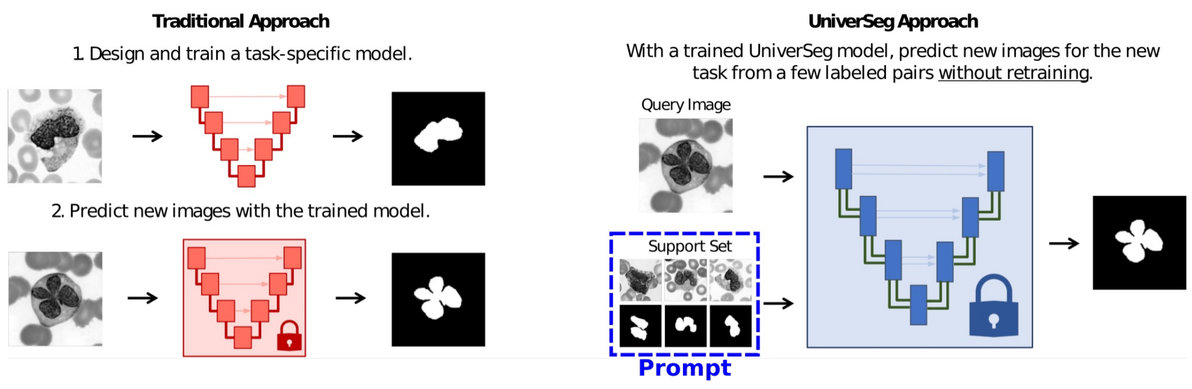
Heejong Kim[1], Victor Ion Butoi[3], Adrian V. Dalca[3,4], Daniel J.A. Margolis[1], Mert R. Sabuncu[1,2]

[1]Department of Radiology, Weill Cornell Medicine
[2]School of Electrical & Computer Engineering, Cornell University and Cornell Tech
[3]Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School
[4]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Weill Cornell Medicine
Cornell Tech

Personal page: https://heejongkim.com
Email: hek4004@med.cornell.edu
Paper link

MOTIVATION



Traditional Approach vs. Foundational Model Approach

- Traditionally models, e.g. nnUNet [1] are trained for each segmentation task, often from scratch.
- FMs like UniverSeg [2] and Segment Anything Model (SAM) [3] are pre-trained general-purpose models.
- Instead of retraining, prompts like support sets are used for UniverSeg [2] and points and masks for SAM [3].
- We conduct an empirical evaluation study in the context of prostate imaging and compare UniverSeg [2] and nnUNet [1].

METHOD

Different support set selection strategies for UniverSeg

- **Support Set Selection Strategies:**
  - Slice-index-aware Support Set Selection:
    Randomly select with a probability proportional to the weight.
    - Weights for query image = $1/(|Z_{I_t} - Z_{I_q}| + 1)$
      (Z: slice index, I_t: available images, I_q: query image)
  - Random Support Set Selection:
    Randomly draw from available labeled slices.
- **ROI-inclusive vs. ROI-agnostic:** Support set selection techniques can be restricted to slices.
  - ROI-inclusive: slices where the ROI is present
  - ROI-agnostic: all possible slices (i.e., be agnostic to whether the ROI is present or absent in the slice)

Note: this poster includes the results of prostate gland segmentation only.
For transitional and peripheral zone segmentation, please refer to the paper.

RESULT

Computational Resource Comparison.

	nnUNet–orig.	nnUNet–small	UniverSeg
#Params	20.6 M × 5 folds	1.3 M × 5 folds	**1.2 M**
Training time (ms)	1.6×10^8	1.2×10^8	–
Inference time (ms)	9.7×10^3	7.5×10^3	$\mathbf{6.9 \times 10^2}$

Table 1. Computational resource comparison. The values are averaged across ROIs and calculated for N=1 case for all methods. Tested on Nvidia TITAN Xp GPU (12 GB vRAM).

- Smaller number of parameters and faster inference time.
- Does not require task-specific training.
  - Saving substantial computational requirement.
  - Obviating the need for a GPU.

Quantitative Segmentation Performance

ROI	Method	$N = 1$	$N = 2$	$N = 5$	$N = 10$
Prostate-Gland	nnUNet-Orig	0.592 ± 0.088	0.714 ± 0.045	0.810 ± 0.007	0.817 ± 0.016
	nnUNet-Small	0.520 ± 0.076	0.698 ± 0.057	0.802 ± 0.008	0.808 ± 0.019
	UniverSeg	$\mathbf{0.711 \pm 0.008}$	$\mathbf{0.769 \pm 0.009}$	0.780 ± 0.003	0.802 ± 0.005

Table 2. 2D Dice scores results. The scores are averaged across 5 support/test splits.

- Overall comparable results to the nnUNet models.
- UniverSeg achieved good performance given extremely limited annotated data, e.g., N = 1, outperforming the nnUNet models.

Support Set Selection	N	Prostate
ROI-agnostic	1	0.596 ± 0.047
	2	0.690 ± 0.035
	5	0.716 ± 0.006
	10	0.778 ± 0.006

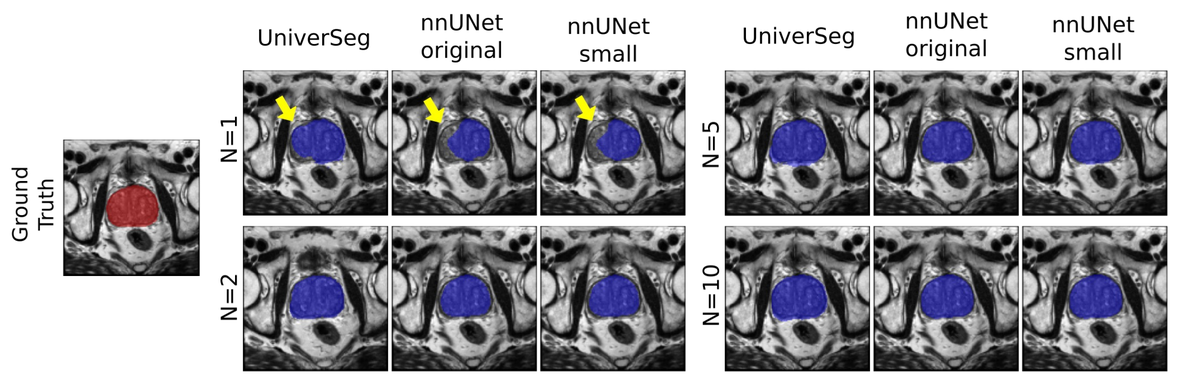
>

Support Set Selection	N	Prostate
ROI-inclusive	1	0.481 ± 0.035
	2	0.488 ± 0.034
	5	0.513 ± 0.027
	10	0.543 ± 0.013

Table 3. 3D Dice scores for UniverSeg models with two different support set selection strategies.

- Incorporating ROI-excluded slices improves results.
  - Need to account for ROI absence in query images in 3D contexts.

Qualitative Segmentation Performance



Ablation

ROI	Method	$N = 1$	$N = 2$	$N = 5$	$N = 10$
nnUNet-Orig	w/o augmentation	0.590 ± 0.085	0.712 ± 0.046	0.809 ± 0.007	0.815 ± 0.016
	fold-1	0.581 ± 0.086	0.681 ± 0.060	0.798 ± 0.011	0.808 ± 0.017
	fold-2	0.564 ± 0.095	0.710 ± 0.039	0.797 ± 0.010	0.798 ± 0.023
	fold-3	0.590 ± 0.092	0.691 ± 0.044	0.795 ± 0.014	0.807 ± 0.025
	fold-4	0.599 ± 0.088	0.708 ± 0.043	0.785 ± 0.006	0.804 ± 0.006
	fold-5	0.553 ± 0.046	0.692 ± 0.046	0.790 ± 0.006	0.810 ± 0.008
	default	$\mathbf{0.592 \pm 0.088}$	$\mathbf{0.714 \pm 0.045}$	$\mathbf{0.810 \pm 0.007}$	$\mathbf{0.817 \pm 0.016}$
nnUNet-Small	w/o augmentation	0.519 ± 0.072	0.696 ± 0.056	0.801 ± 0.007	0.807 ± 0.018
	fold-1	0.537 ± 0.047	0.668 ± 0.074	0.784 ± 0.014	0.801 ± 0.021
	fold-2	0.518 ± 0.068	0.686 ± 0.051	0.793 ± 0.012	0.792 ± 0.023
	fold-3	0.512 ± 0.091	0.689 ± 0.057	0.784 ± 0.011	0.803 ± 0.011
	fold-4	0.508 ± 0.076	0.705 ± 0.046	0.787 ± 0.015	0.792 ± 0.022
	fold-5	0.530 ± 0.089	0.680 ± 0.045	0.782 ± 0.014	0.798 ± 0.020
	default	$\mathbf{0.520 \pm 0.076}$	$\mathbf{0.698 \pm 0.057}$	$\mathbf{0.802 \pm 0.008}$	$\mathbf{0.808 \pm 0.019}$
UniverSeg	all	$\mathbf{0.711 \pm 0.008}$	$\mathbf{0.769 \pm 0.009}$	0.778 ± 0.006	0.799 ± 0.005
	random	–	–	0.777 ± 0.002	0.798 ± 0.005
	random+5 ensemble	–	–	0.779 ± 0.004	0.800 ± 0.006
	z-weighted	–	–	0.777 ± 0.002	0.798 ± 0.005
	z-weighted +5 ensemble	–	–	$\mathbf{0.780 \pm 0.003}$	$\mathbf{0.802 \pm 0.005}$
Average # of images available for support set		14.0 ± 2.1	31.4 ± 6.5	83.4 ± 2.9	148.0 ± 3.7

Table 4. 2D Dice scores from the ablation study conducted for the prostate segmentation task.

- Ensembling gave all models a boost.
- Support set quality is important for UniverSeg.
- Test time augmentation gives boost to nnUNet results.

CONCLUSION

- Our results and discussion highlight several important factors that will likely be important in the development and adoption of foundation models for medical image segmentation.
- **Based on the successful employment of FMs in multiple domains, we believe FMs will instigate a paradigm shift for medical imaging.**
- **Future directions:**
  - FMs for 3D MIS are needed, and promise to be impactful.
  - Second, adaptation of FMs should be further studied.
  - Third, clinical practitioners can easily adapt FMs in their workflows, as it obviates the need to fine-tune.

REFERENCE

[1] Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature methods 18(2), 203–211 (2021)
[2] Butoi, V.I., Ortiz, J.J.G., Ma, T., Sabuncu, M.R., Guttag, J., Dalca, A.V.: Uni- verseg: Universal medical image segmentation. arXiv preprint arXiv:2304.06131 (2023)
[3] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., et al.: Segment anything. arXiv preprint arXiv:2304.02643 (2023)

ACKNOWLEDGMENTS: This work was supported by NIH, United States grant R01AG053949 and 1R01AG064027, the NSF, United States NeuroNex grant 1707312, and the NSF, United States CAREER 1748377 grant.