Simulated Interactive Correction Rescues Suboptimal Segmentation Predictions of Military Wound Images

Daniel C Thompson, MBBS[1,2]; Felipe A Lisboa, MD[1,3,5]; Cameron Leonard, BS[1]; Katherine Weathersby, BA[1]; Katherine Nagel, BA[1]; Fausto A Bustos Carrillo, PhD[1,3]; Christopher J Dente, MD[1,4]; Eric A Elster, MD[1,5]; Seth A Schobel, PhD[1]

[1]Uniformed Services University (USU) Surgical Critical Care Initiative (SC2i), Bethesda, MD; [2]Royal Centre for Defence Medicine, Birmingham, UK;
[3]The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD; [4]Emory University, Atlanta GA; [5]Walter Reed National Military Medical Center, Bethesda, MD

INTRODUCTION

- **Wound boundary segmentation matters:** Automatically delineating wound boundaries in clinical photographs is a foundational step toward objective, image-based trauma assessment and longitudinal tracking.
- **Combat wounds challenge conventional models:** Traumatic wounds are highly heterogeneous and often caused by penetrating, blunt, and blast mechanisms. A conventional segmentation model would likely require a large, diverse training dataset to cover this variability, yet military wound imaging datasets are scarce.
- **Improved performance using a wound adapted foundation model:** Using civilian traumatic wound images, we developed a segmentation pipeline (Figure 1) that pairs a trained wound detector model with an adapted general-purpose segmentation model (SAM2.1[1]). Despite improved performance compared to standalone conventional models, the developed pipeline still makes suboptimal (Dice <0.95) segmentation predictions in 9-22% of externally tested images.
- **Human-in-the-loop refinement:** SAM 2.1 allows a user to refine a prediction by marking areas of error with correction point prompts.
- **Hypothesis:** Simulated user-correction points improve the SAM 2.1 segmentation pipeline predictions of military traumatic extremity wounds

METHODS

- **Two-stage pipeline:** Three trained wound detectors (EfficientNet-B0, FastViT-T8, and YOLOv8n) each pass a wound bounding box to a wound-adapted SAM2.1 Small model (LoRA[2] fine-tuning); see Figure 1.
- **Interactive correction:** A user-correction point is simulated at the deepest interior point of the largest remaining error region (maximum distance transform), over five iterative rounds, with the prediction updating in between rounds[3].
- **Data:** Training - 599 civilian traumatic wound images (101 patients); internal testing - 116 civilian images (21 patients); external testing - 100 military (27 patients) wound images from a separate institution. Ground-truth wound boundaries were manually annotated by a clinician and trained medical students.
- **Performance metric:** Dice coefficient (overlap between predicted and true wound boundaries, 1.0 = perfect agreement) with patient-level bootstrap 95% confidence intervals.

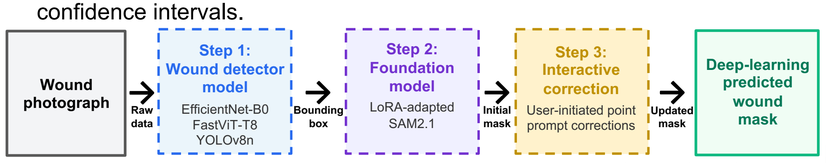

Figure 1. A lightweight trained wound detector (step 1) prompts a wound-adapted SAM2.1 (step 2), which is then refined by optional user-initiated correction (step 3).

RESULTS

FastViT-T8 | Over five correction points | External testing

Improved performance	Reduced suboptimal predictions	Reduced Dice spread
0.973 → 0.984	12% → 3%	0.025 → 0.008
Mean Dice	Dice < 0.95	Interquartile range

Table 1. Patient-level Dice [95% CI] at each correction point for each wound detector pipeline.

Correction points	EfficientNet-B0	FastViT-T8	YOLOv8n
Internal testing — civilian (n=116)			
0	0.971 [0.932–0.982]	0.967 [0.916–0.981]	0.969 [0.923–0.982]
1	0.982 [0.968–0.987]	0.982 [0.972–0.986]	0.982 [0.967–0.987]
2	0.985 [0.972–0.988]	0.985 [0.977–0.988]	0.985 [0.976–0.988]
3	0.986 [0.977–0.989]	0.987 [0.980–0.989]	0.987 [0.980–0.989]
4	0.987 [0.978–0.989]	0.987 [0.979–0.989]	0.987 [0.979–0.989]
5	**0.987 [0.980–0.990]**	**0.988 [0.983–0.989]**	**0.988 [0.982–0.990]**
External testing — military (n=100)			
0	0.975 [0.967–0.980]	0.973 [0.964–0.978]	0.939 [0.898–0.962]
1	0.980 [0.971–0.984]	0.980 [0.971–0.983]	0.954 [0.925–0.970]
2	0.982 [0.973–0.986]	0.982 [0.974–0.985]	0.968 [0.948–0.977]
3	0.983 [0.974–0.986]	0.983 [0.974–0.986]	0.975 [0.953–0.982]
4	0.984 [0.975–0.987]	0.984 [0.975–0.987]	0.978 [0.955–0.984]
5	**0.985 [0.977–0.987]**	**0.984 [0.975–0.987]**	**0.979 [0.957–0.985]**

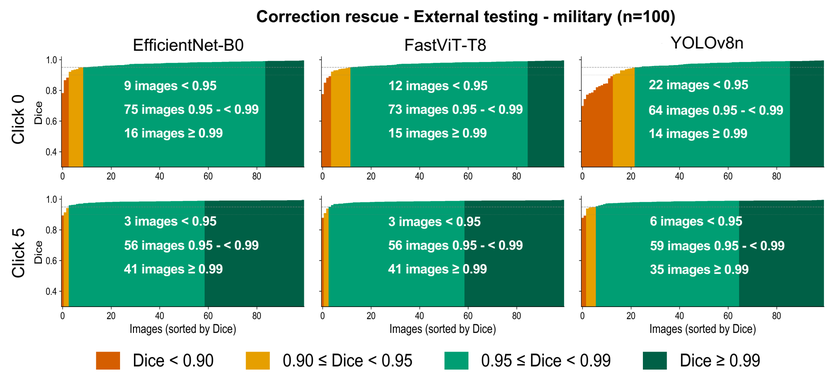

Figure 2. Five correction points increase the proportion of predictions above 0.95 Dice; images below 0.95 fall from 9→3 (B0), 12→3 (FastViT), and 23→7 (YOLO).

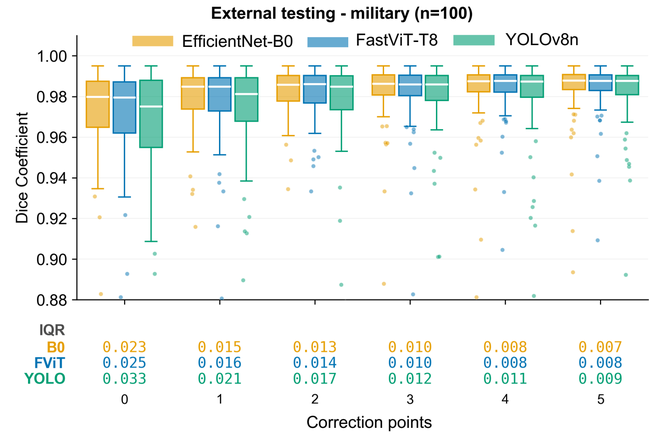

Figure 3. The interquartile range of Dice falls roughly 3-fold over five correction points showing reduced variability of the segmentation quality. IQR, interquartile range.

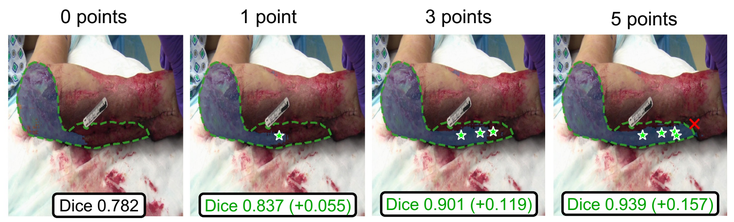

Figure 4. Clinical example showing segmentation prediction improvement over 5 correction points with Dice improvement from 0.782 to 0.939

DISCUSSION

- **Largest gains in the weakest predictions:** Correction points preferentially improve the lowest-Dice cases, reducing the proportion of predictions below 0.95 and decreasing the interquartile range of Dice ~3-fold.
- **Consistent across detectors:** The external testing performance gap between the smallest (YOLOv8n, 3.2M parameters) and largest (EfficientNet-B0, 5.3M) models shrank from 0.036 Dice at baseline to just 0.006 after five correction points (0.979–0.985).
- **Optimal choice:** Considering the simulated correction points equalize the accuracy of the initial wound detectors, model selection can balance other factors like size, inference speed, and power demand rather than accuracy. On this basis, we favor FastViT-T8.
- **Generalization to military wounds:** A lightweight, mobile-compatible segmentation framework, trained on readily available civilian data, can generalize to military traumatic wounds. User-initiated corrections can improve suboptimal predictions to enable downstream computer vision tasks.

REFERENCES

1. Ravi N, et al. SAM 2: Segment Anything in Images and Videos. arXiv:2408.00714, 2024.; 2. Hu EJ, et al. LoRA: Low-Rank Adaptation of Large Language Models. ICLR 2022 (arXiv:2106.09685).; 3. Sofiiuk K, et al. Reviving Iterative Training with Mask Guidance for Interactive Segmentation. ICIP 2022 (arXiv:2102.06583).

*SC2i is a consortium of Uniformed Services University of the Health Sciences, Walter Reed National Military Medical Center, Duke Health, Emory University School of Medicine, and Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc.