A Two-Stage Deep Learning Framework Achieves Generalizable Military Wound Segmentation Using Civilian Training Data

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]

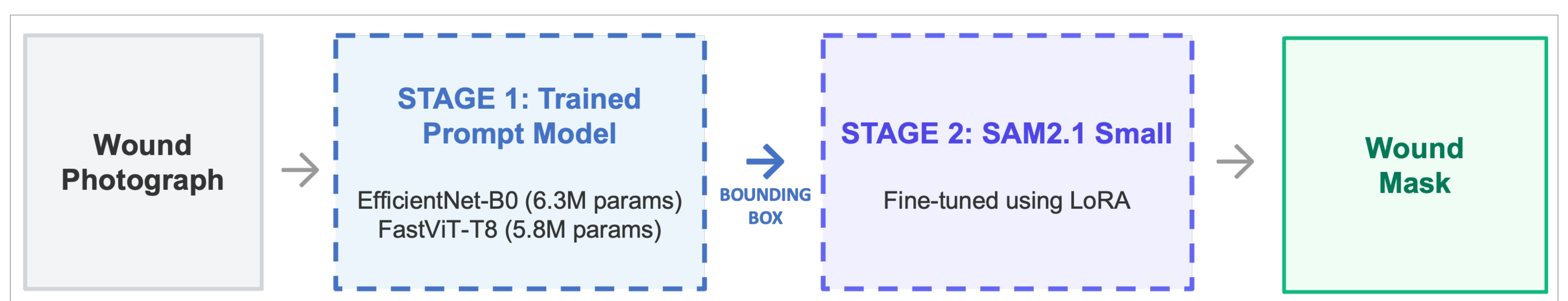
[1]Uniformed Services University (USU) Surgical Critical Care Initiative (SC2i), Bethesda, MD; [2]Royal Centre for Defence Medicine, Birmingham, UK; [3]The Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Bethesda, MD; [4]Emory University, Atlanta GA; [5]Walter Reed National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

INTRODUCTION

- **Wound boundary segmentation** — automatically identifying wound regions in clinical photographs using deep learning — is a foundational step toward objective, image-based trauma assessment
- **Traumatic extremity wounds span diverse injury mechanisms**, producing a wide range of wound appearances. Military wounds, often caused by blast fragmentation and high-velocity projectiles, can differ substantially from civilian trauma
- **A single segmentation model would likely require a large, diverse training dataset** to cover this variability, yet military wound imaging datasets are scarce. No segmentation approach has been validated to generalize from civilian to military trauma populations
- **Hypothesis:** A two-stage approach, pairing a lightweight wound-detection model with a general-purpose segmentation model (SAM2.1) pre-trained on millions of natural images and adapted to wounds, could generalize from civilian to military populations without requiring an exhaustive combat-wound training dataset

METHODS

- **Two-stage pipeline:** Trained first-stage model detects wound → bounding box → second-stage SAM2.1 (fine-tuned to wound images using Low-Rank Adaptation, LoRA) refines segmentation (Figure 1)
- **First-stage models compared:** EfficientNet-B0 and FastViT-T8
- **Training:** 599 wound images from 101 patients at a Level 1 civilian trauma center in the USA, annotated by a clinician and trained medical students; no patient appears in both training and testing sets
- **Internal test:** 116 images, 21 held-out civilian patients; **External test:** 100 military extremity trauma images, 27 patients (separate institution)
- **Performance:** Dice coefficient (overlap between predicted and true wound boundaries, 1.0 = perfect agreement) with patient-level bootstrap 95% confidence intervals



Why two stages?

A lightweight model localises the wound; a powerful foundation model refines the boundary. Divides the task by what each architecture does best.

What is the "prompt"?

SAM2 needs to be told where to look. The bounding box from Stage 1 tells it, "Segment what's inside this rectangle."

What is LoRA?

Low-Rank Adaptation — injects small trainable matrices into a frozen model. Adapts SAM2 to wounds using only ~0.4% of parameters (189K of 46M).

Figure 1. Two-stage wound segmentation pipeline. A lightweight prompt model generates a bounding box; SAM2.1 (LoRA-adapted) refines the wound boundary.

RESULTS

0.962
External Dice (Both Pipelines)

p=0.005
FastViT-T8 + SAM2.1 vs. Standalone

3:1
Benefit-to-Harm Ratio (External)

Pipeline	Internal Dice [95% CI]	External Dice [95% CI]
EfficientNet-B0 alone	0.961 [0.931–0.973]	0.956 [0.936–0.966]
EfficientNet-B0 + SAM2.1	**0.969 [0.937–0.978]**	**0.962 [0.950–0.970]**
FastViT-T8 alone	0.955 [0.921–0.968]	0.947 [0.923–0.960]
FastViT-T8 + SAM2.1	**0.965 [0.926–0.977]**	**0.962 [0.950–0.970]**

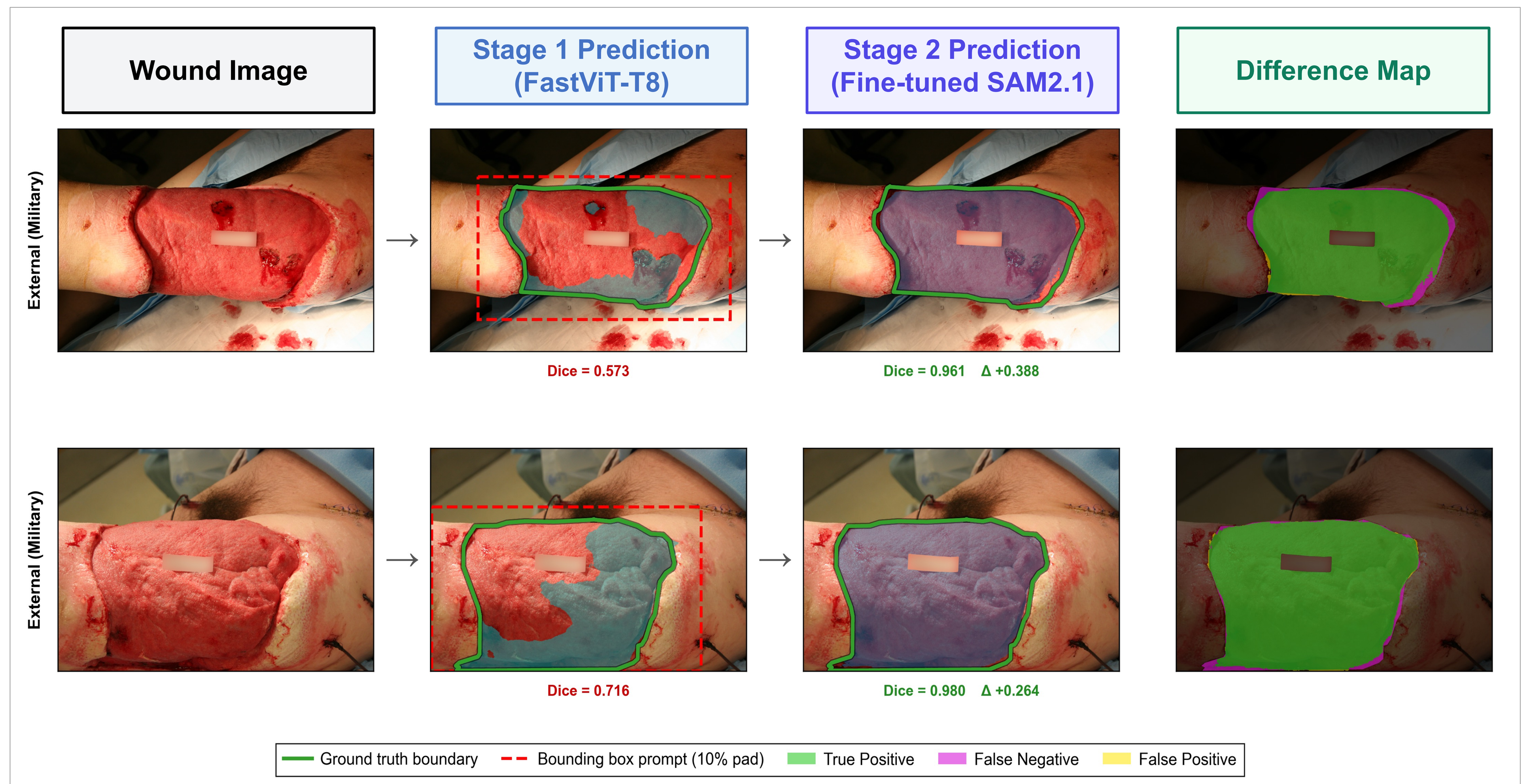


Figure 2. Two-stage pipeline rescue on external military data using FastViT-T8 first-stage prediction (green = ground truth boundary), improving Dice from 0.573→0.961 (Δ+0.388) and 0.716→0.980 (Δ+0.264).

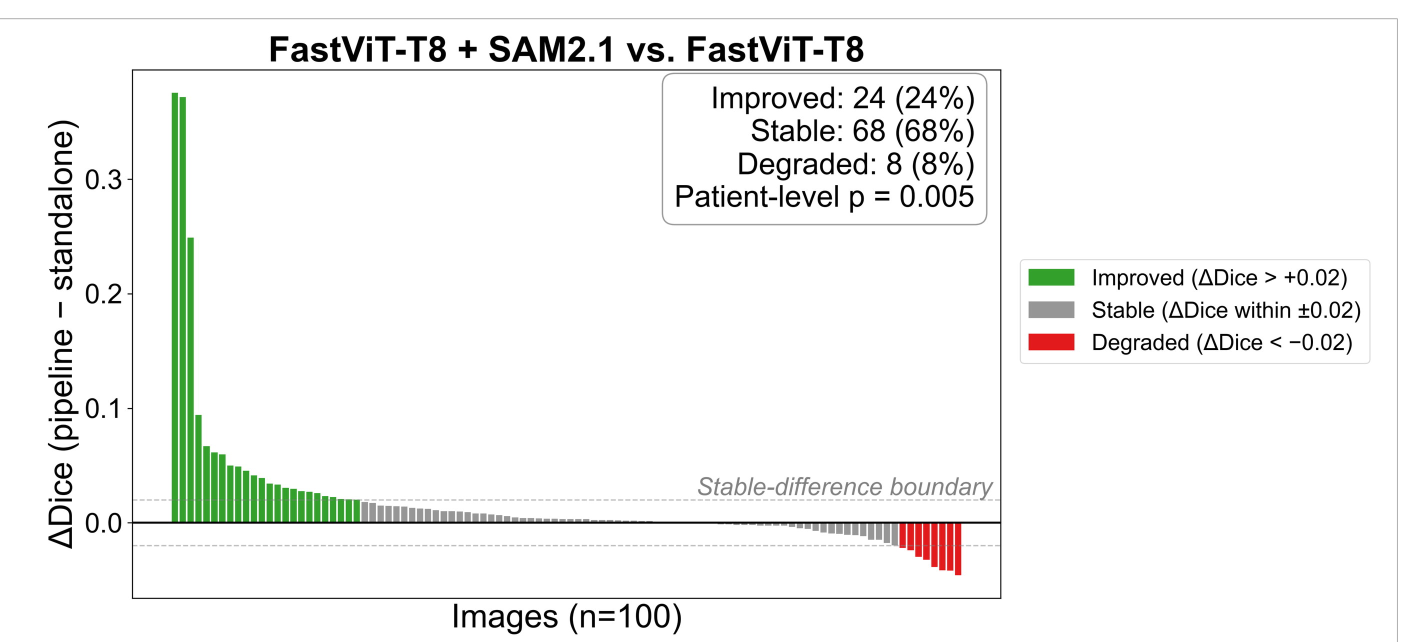


Figure 3. Per-image Dice difference on external testing dataset (green = improved, grey = stable, red = degraded). FastViT: 24% improved vs. 8% degraded (3:1 benefit ratio).

KEY FINDING

The two-stage pipeline preferentially rescues the weakest predictions. FastViT-T8 + SAM2.1 improved 24% of images while degrading only 8% on external military data (Wilcoxon p=0.005, Figure 3).

DISCUSSION

- This is the first demonstration that a wound segmentation framework trained on civilian trauma data generalizes to military trauma, in a domain where SAM-based approaches have been largely unexplored for acute wound assessment
- Both pipelines converge to equivalent external Dice (0.962) after adding SAM2.1, suggesting the two-stage pipeline acts as a performance equalizer across architectures (Figure 4)
- The two-stage pipeline adds resilience by rescuing suboptimal predictions while infrequently degrading strong ones (Figure 3). The already-stronger EfficientNet-B0 benefited less (p=0.269), consistent with selective rescue of the weakest predictions
- Lightweight first-stage architectures (~6M parameters) support potential mobile deployment for use across all echelons of care

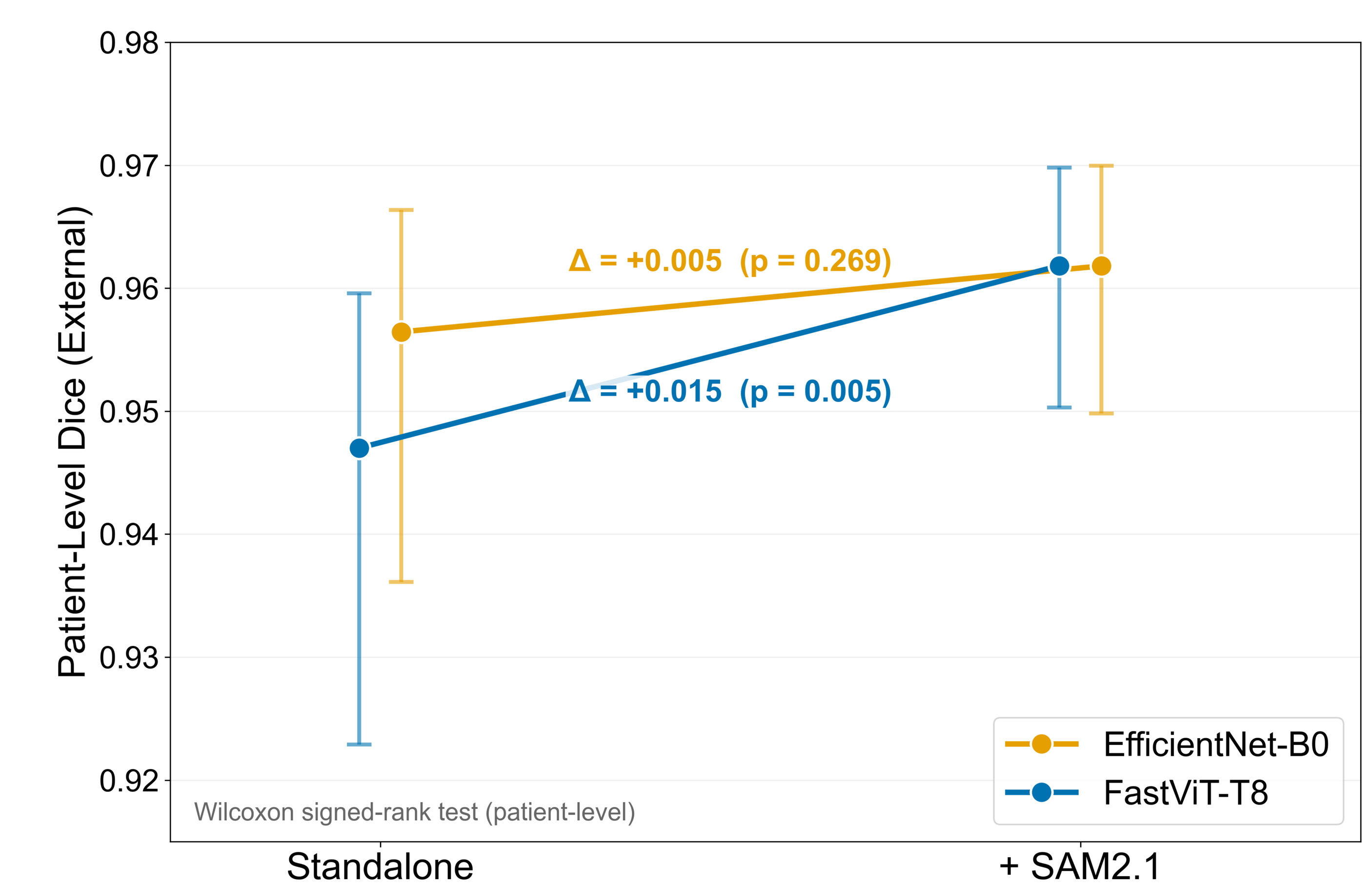


Figure 4. The two-stage pipeline equalizes performance across architectures on external data. Both first-stage models converge to equivalent Dice (0.962) despite different standalone performance (EfficientNet-B0: 0.956, FastViT-T8: 0.947). Error bars show patient-level bootstrap 95% CIs.

CONCLUSIONS

- Despite our two-stage method for automated detection of wound boundaries being trained only on civilian data, it generalizes to military extremity trauma.
- Our two-stage pipeline typically improves weak first-stage predictions while rarely degrading strong ones, resulting in more high-quality predictions overall.
- All tested models are mobile-compatible, enabling their potential use in a field deployable device and supporting image-based objective wound assessment across all echelons of care.

*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.

REFERENCES

1. Ravi N et al. SAM 2: Segment Anything in Images and Videos. arXiv:2408.00714 (2024).; 2. Tan M & Le QV. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. ICML (2019); 3. Vasu PKA et al. FastViT: A Fast Hybrid Vision Transformer using Structural Reparameterization. ICCV (2023); 4. Hu EJ et al. LoRA: Low-Rank Adaptation of Large Language Models. arXiv:2106.09685 (2021).

The contents of this presentation are the sole responsibility of the author(s) and do not necessarily reflect the views, opinions or policies of Uniformed Services University of the Health Sciences (USUHS), Henry M. Jackson Foundation for the Advancement of Military Medicine, Inc., Department of War (DoW), or Departments of the Army, Navy, or Air Force. Mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.