POLITECNICO DI TORINO
Repository ISTITUZIONALE

IMTOP, IMage-TO-Print Wound Dressings: From a Single Calibrated RGB Photograph to a Patient-Specific Patch

Original
IMTOP, IMage-TO-Print Wound Dressings: From a Single Calibrated RGB Photograph to a Patient-Specific Patch / Salerno, F., Ulrich, L., Secco, J., Azzena, G., Laurano, R., Moos, S., Vezzetti, E.. - In: COMPUTERS. - ISSN 2073-431X. - ELETTRONICO. - 15:10(2026), pp. 1-30. [10.3390/computers15100710]

Availability:
This version is available at: 11583/3016631 since: 2026-10-14T08:38:20Z

Publisher:
MDPI

Published
DOI:10.3390/computers15100710

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

25 October 2026

computers

MDPI

Article

IMTOP, IMage-TO-Print Wound Dressings: From a Single Calibrated RGB Photograph to a Patient-Specific Patch

Federico Salerno [1], Luca Ulrich [1,*], Jacopo Secco [2], Gabriele Azzena [2], Rossella Laurano [3], Sandro Moos [1] and Enrico Vezzetti [1]

[1] Department of Management and Production Engineering (DIGEP), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; federico.salerno@polito.it (F.S.); sandro.moos@polito.it (S.M.); enrico.vezzetti@polito.it (E.V.)

[2] Department of Electronics and Telecommunications (DET), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; jacopo.secco@polito.it (J.S.); gabriele.azzena@polito.it (G.A.)

[3] Department of Mechanical and Aerospace Engineering (DIMEAS), Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy; rossella.laurano@polito.it

* Correspondence: luca.ulrich@polito.it

Abstract

Patient-specific dressings for diabetic foot ulcers (DFUs) must follow the individual lesion, yet image-to-fabrication pipelines rely on depth sensing and seldom report the accuracy of the fabricated object. We present IMTOP, a low-cost pipeline that turns a single calibrated RGB photograph into a printable patient-specific patch, and we budget its area error stage by stage. Two terms were measured: on 18 polylactic acid phantoms printed by fused deposition modelling, geometry was reconstructed to 0.32% area error and printing deviated by 0.69%; on 149 public clinical DFU images (uncalibrated, single-operator reference), three reference-seeded segmenters gave median patch area errors of 3.9% to 6.0%. The scale term was simulated (assumed 3 px placement noise on a 117 px reference; standard deviation 7.3%). Monte Carlo propagation of the terms gave an estimated median absolute total error of 7.3% to 8.0%; segmentation and scale were comparable, and printing contributed under 0.1% of the variance. Mask overlap correlated with, but did not determine, the patch error. A normal perilesional offset of 14% to 21% of the wound equivalent diameter covered the wound in about 94% of cross-validated cases. The results support the feasibility of the approach as a proof of concept.

Keywords: diabetic foot ulcer; patient-specific wound dressing; image segmentation; Segment Anything Model; additive manufacturing; error budget; uncertainty propagation; decision support

Check for updates

Academic Editor: Atsushi Teramoto

Received: 27 August 2026
Revised: 5 October 2026
Accepted: 10 October 2026
Published: 13 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Chronic skin wounds, and diabetic foot ulcers (DFUs) in particular, are a growing clinical and economic burden [1]. The lifetime incidence of a foot ulcer in people with diabetes is high, and its consequences are severe, including lower-limb amputation and a post-ulceration mortality comparable to that of several common cancers [2,3]. A recurring cause of poor healing is non-personalized treatment: standardized protocols and pre-shaped dressings ignore wound-specific features such as contour, area and local tissue condition [1], and pre-shaped dressings conform poorly to irregular morphologies even though close contact with the wound bed is considered important for efficacy [4].

This has motivated personalized wound care. Hydrogel dressings are widely used because they keep the wound moist, a condition that accelerates epithelialization [5], while absorbing exudate, allowing gas exchange and releasing therapeutics locally [6]. To match the dressing to the individual lesion, additive manufacturing has been used to fabricate patient-specific shapes, from 3D-scanned antimicrobial patches to printed functional hydrogels [4,7,8]. The geometric step, turning a clinical image of the wound into a fabricable object, has so far relied mainly on colour-and-depth (RGB-D) pipelines that acquire depth and reconstruct a three-dimensional model before generating printable instructions [9]. Two limitations follow: (i) such pipelines raise the hardware requirement and are ill-suited to point-of-care acquisition, where only a consumer camera may be available [9]; (ii) accuracy is usually reported on the segmentation mask with region-overlap metrics, as throughout the wound-segmentation literature [10,11], not on the object that is finally fabricated, a gap shared with other patient-specific printed devices, such as surgical guides, whose fabricated geometry is rarely validated against the computer-aided design (CAD) model [12].

DFUs range from superficial partial- or full-thickness lesions to deep wounds that reach tendon or bone [13]. This work targets the superficial class treated with conformal dressings, for which a patch that reproduces the wound contour and carries a parametric thickness is a reasonable design choice; deep or undermined wounds fall outside its scope. For this class, we argue that a single RGB photograph is geometrically sufficient: absolute scale is recovered from a physical reference in the field of view, as in image-based wound planimetry [14], with no depth sensor and no reconstruction of the wound-bed topography. The contribution of such a low-cost, depth-free pipeline is not the patch itself, which is a simple geometric transformation, but the characterization of how the upstream errors reach the fabricated object and the conversion of that characterization into a design rule. Manufacturing metrology routinely decomposes the deviation of a produced part into contributions from successive process stages and recombines them into an error budget, following the Guide to the Expression of Uncertainty in Measurement (GUM) [15] and the dimensional-tolerance studies of additive manufacturing [16]; to our knowledge, this perspective has not been applied to image-to-print wound dressings.

We organize the study around three research questions:

RQ1 How does the fabricated-patch error distribute across the segmentation, scale and printing stages, in both magnitude and variance?

RQ2 Is mask overlap a sufficient predictor of fabricated-patch accuracy, or are boundary and coverage measures also required?

RQ3 Can the residual segmentation uncertainty be turned into a perilesional offset that covers the wound at a target level?

We address them through three contributions:

- a stage-wise error budget that splits the fabricated-patch error into segmentation (E_{Seg}), scale (E_{Scale}) and printing (E_{Pr}) terms, the first and the third measured in dedicated experiments and the second simulated, and that keeps the bias of each term apart from its dispersion (RQ1);
- a fabrication-aware evaluation that complements region overlap with boundary metrics and with clinically asymmetric coverage metrics, separating under-coverage, which leaves wound exposed, from over-coverage onto healthy skin, which is less harmful but not cost-free (RQ2);
- an uncertainty-driven sizing of the perilesional offset, yielding the margin that covers the wound at a target level, on the evaluated data, for a chosen segmentation back-end, and checked by cross-validation (RQ3).

Seen as an intelligent system for complex clinical data, IMTOP is a human-in-the-loop decision-support tool: for every candidate segmentation it presents the operator with overlap, boundary and coverage metrics together with the resulting patch, and it converts the uncertainty of the segmentation stage into an actionable design parameter, the perilesional offset, under an explicit and clinically asymmetric cost. The stage-wise error budget is an interpretable attribution of that uncertainty, indicating which stage is responsible for the error of the fabricated object and therefore where an intervention would pay off.

The study is a technical proof of concept of the geometric chain from the photograph to the printed part. The conditions of the evaluation (operator-seeded segmentation, a single-operator reference, uncalibrated clinical images and rigid phantoms) are discussed in Section 5.5.

The pipeline, including the interactive application used to produce the manual reference segmentations and the patches, will be released to support reproducibility (data availability statement). The rest of the paper presents related work (Section 2), the system and methods (Section 3), the results (Section 4), a discussion with the study limitations (Section 5) and the conclusions (Section 6); Appendix A reports a sensitivity analysis on the handling of outliers.

2. Related Work

As noted in Section 1, additive manufacturing has been used to give engineered dressings a patient-specific shape [4,7,8]; since conformal contact with the wound bed matters for efficacy [4], the geometric fidelity of the patch is itself a design target. In situ bioprinting deposits the biomaterial directly on the wound bed [17,18] and removes the patch altogether, at the cost of dedicated hardware at the bedside; the image-to-print route addressed here instead fabricates the dressing off-line from an image of the wound.

Extracting geometry from a clinical image follows two routes. The first stays on plain RGB: planimetry recovers wound area from a photograph with a physical scale reference [14], mobile apps standardize acquisition on a smartphone [19], deep models segment wound tissue on-device [20], and phone photographs predict ulcer healing [21]; these report area, tissue or prognosis but not a fabricable object. The second reconstructs three-dimensional geometry: robot-mounted RGB-D rigs reconstruct the wound surface [22] and derive measurements [23], Cavazzana et al. add a three-dimensional analysis for objective assessment [24], and Chae et al. [9] reconstruct an RGB-D surface to print bio-printable DFU patches. This route needs a depth sensor or a controlled rig, ill-suited to point-of-care capture, and validates a measurement or surface, not a fabricated object, including Chae et al. [9], who reach a printable patch but validate the segmentation and the reconstructed surface. IMTOP keeps the depth-free acquisition of the first route while reaching the patch, and is conceived as the fabrication step that can follow an assessment pipeline such as that of Cavazzana et al. [24].

The segmentation feeding these pipelines has a large literature. Classical region methods, Otsu thresholding [25], marker-controlled watershed [26] and GrabCut [27], were superseded by deep learning: fully convolutional networks [28], U-Net [29], applied to DFUs by Goyal et al. [11], to general wounds by Wang et al. [10] and through convolutional neural network (CNN) ensembles [30], with self-configuring nnU-Net [31]; public challenges rank clinician-delineated masks by overlap [32,33]. Promptable foundation models, the Segment Anything Model (SAM) [34], its medical adaptation MedSAM [35] and its successor SAM 2 [36], segment from a box or point prompt without task-specific training. In almost all of this work accuracy is scored on the mask. IMTOP uses three such segmenters only as interchangeable inputs to its geometric chain, the smallest SAM checkpoint

being chosen to keep the memory footprint compatible with ordinary hardware, and the two classical methods being retained because they run on a CPU without downloaded weights, which matters at the point of care.

That scoring relies mainly on region-overlap metrics, including the Dice coefficient [37] and the intersection over union (IoU), sometimes complemented by precision and recall or by false-negative and false-positive rates [10,32]. Overlap is an incomplete descriptor: two masks with identical Dice can differ in boundary placement, captured by the Hausdorff distance [38]; and although recall and precision separate missed wound from excess skin, to our knowledge no study has propagated this asymmetry to the fabricated dressing or used it to size the margin. Both limitations become concrete once the mask becomes a physical patch, which motivates the use of boundary and coverage measures and the measurement of the error on the fabricated object.

On the fabrication side, the deviation of a produced part is routinely split across process stages and recombined into an error budget, following the GUM [15] and experimental studies of dimensional accuracy in fused deposition modelling (FDM) [16,39]; similar metrological protocols address fabricated medical devices such as surgical guides [12]. When the input distributions are skewed or the errors are not small, the first-order law of propagation is complemented by the propagation of distributions through Monte Carlo simulation, described in Supplement 1 to the GUM [40]. To our knowledge this perspective has not been applied to image-to-print wound dressings, where the stages are segmentation, scale recovery and printing.

Table 1 places IMTOP among these methods. It propagates the segmentation, scale and printing errors to the fabricated patch, complements overlap with boundary and coverage metrics, and turns the residual segmentation uncertainty into a perilesional offset, thus extending mask-only segmentation studies [10,11,32] and depth-based reconstruction pipelines [9,22,23] toward a depth-free, fabrication-aware route in which the error is evaluated on the patch.

Table 1. Methods and pipelines relevant to image-based wound geometry, grouped from general-purpose segmentation to fabrication. Among the listed works, only IMTOP reaches a printable patch from a single RGB photograph without depth; it measures the accuracy of the fabricated object on phantoms of known geometry and propagates the segmentation and scale errors to the patch for clinical images. In the Depth column a filled circle (●) marks methods that require a depth sensor or a 3D rig, an open circle (○) image-only methods.

Work	Year	Depth	Output	Validated on
General-purpose segmentation				
Long et al. [28]	2015	○	mask	overlap
Ronneberger et al. [29]	2015	○	mask	overlap
Isensee et al. [31]	2021	○	mask	overlap
Kirillov et al. [34]	2023	○	promptable	overlap
Ma et al. [35]	2024	○	promptable	overlap
Wound and DFU segmentation				
Goyal et al. [11]	2017	○	mask	overlap
Wang et al. [10]	2020	○	mask	overlap
Mahbod et al. [30]	2022	○	mask	overlap
Ramachandram et al. [20]	2022	○	tissue	overlap
Yap et al. [32]	2024	○	mask	overlap
Planimetry, mobile and monitoring				
Yap et al. [19]	2018	○	std. image	repeatability
Foltynski [14]	2018	○	area	clinical area
Kim et al. [21]	2020	○	healing risk	clinical outcome

Table 1. *Cont.*

Work	Year	Depth	Output	Validated on
3D reconstruction and fabrication				
Filko et al. [22]	2021	●	3D surface	reconstruction
Chae et al. [9]	2022	●	patch (print)	3D surface
Filko and Nyarko [23]	2023	●	mask + 3D	measurement
Cavazzana et al. [24]	2025	●	mask + 3D	assessment
IMTOP (this work)	2026	○	patch (print)	printed phantoms; propagated patch

3. Materials and Methods

Figure 1 summarizes the IMTOP pipeline, which turns a single RGB photograph into a printable patch in six stages grouped in three phases. In the acquisition phase the image is loaded and the metric scale s (px/mm) is set from two points placed on a reference of known length. In the segmentation and analysis phase the operator delineates the reference contour, a back-end seeded from that contour (SAM, GrabCut or watershed) produces the mask under evaluation, and the two are compared with mask-level, patch-level and coverage metrics. In the output phase the selected contour is offset by the perilesional margin, extruded and finished at the edge into a watertight mesh exported as a stereolithography (STL) file. Each stage is unlocked by a completion event of the previous one. The following subsections describe the geometric and evaluation methods behind these stages (Sections 3.1–3.5), the experiments and the statistical analysis (Sections 3.6 and 3.7), and the software that implements the pipeline (Section 3.8).

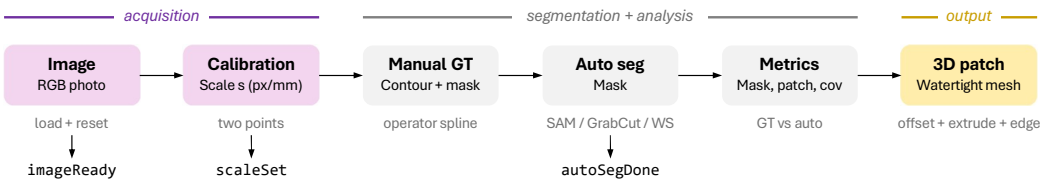

Figure 1. Six-stage pipeline from a single RGB photograph to a printable dressing, grouped in three colour-coded phases: image loading and two-point scale calibration (acquisition); manual reference delineation by the operator (GT, ground truth), reference-seeded segmentation with one of three back-ends (SAM, GrabCut or watershed, WS) and computation of the mask, patch and coverage metrics (segmentation and analysis); offset, extrusion and edge finishing of the patch (output). The names below the boxes are the completion events that unlock the next stage; the segmentation event (`autoSegDone`) unlocks the metric and patch stages.

3.1. Metric Scale Calibration from a Physical Reference

Absolute scale is recovered from two points placed by the operator on a reference of known length d visible in the photograph, such as a caliper jaw opening or a ruler segment. Given the pixel coordinates $\mathbf{q}_0$ and $\mathbf{q}_1$ of the endpoints, the scale factor, in pixels per millimetre, is

$$s = \frac{\|\mathbf{q}_1 - \mathbf{q}_0\|_2}{d} \quad [\text{px/mm}], \tag{1}$$

and every downstream geometric quantity is expressed in millimetres through s; $L = \|\mathbf{q}_1 - \mathbf{q}_0\|_2$ denotes the length of the reference in pixels. Because s is the only source of absolute dimension, its uncertainty is isolated and characterized separately (Section 4.3). The formulation needs neither camera intrinsics nor a depth sensor, which makes the method independent of the capture device; its single assumption is that the reference and the wound lie in a common plane approximately parallel to the image plane (a fronto-parallel view), discussed in Section 5.

3.2. From Contour to Printable Patch

The mask is converted to a printable patch through a deterministic geometric chain (Algorithm 1). The largest external contour is extracted and fitted with a closed periodic B-spline of degree five resampled to $N = 300$ points, which removes the pixel staircasing of the raster boundary while preserving the overall shape. The contour is converted to millimetres through s and offset outward, along the radial direction from its centroid, by a perilesional margin δ that extends the patch over the surrounding healthy skin; δ is not fixed a priori but derived from the coverage analysis of Section 3.5, where the radial construction is also compared with a uniform normal offset. The offset contour is extruded to a user-set thickness t (3 mm by default) with a flat top, and the bottom rim is optionally finished with a 45° chamfer or a circular fillet of radius r_e sampled over $N_{arc} = 20$ rings, with r_e clamped to the thickness and $r_e = 0$ leaving a square edge. Because the caps and the lateral wall share their boundary vertices, the mesh is watertight (closed and manifold, without holes), as required for STL export and slicing, and its volume follows from the divergence theorem,

$$V = \frac{1}{6}\left|\sum_{f\in\mathcal{F}} \mathbf{v}_1^{(f)} \cdot \left(\mathbf{v}_2^{(f)} \times \mathbf{v}_3^{(f)}\right)\right|, \tag{2}$$

with $\mathcal{F}$ the oriented triangular faces and $\mathbf{v}_{1,2,3}^{(f)}$ the vertices of face f. The patch is therefore a planar solid of parametric thickness with a finished edge: conformability to the wound bed is delegated to the flexibility of the thin dressing and not to a built-in curvature, and the wound-bed topography is not reconstructed, a scope limit stated in Section 5. For the manual reference, steps 1 and 2 of Algorithm 1 are replaced by a closed cubic interpolating spline through the operator control points (Section 3.3); the remaining steps are identical, so the two patches are built by the same operator and are directly comparable.

Algorithm 1: Mask-to-patch geometric chain

Data: binary mask M, scale s, thickness t, offset δ, edge radius r_e, edge mode m
Result: watertight triangle mesh T

1 $C \leftarrow$ `largest_contour`(M);
2 $\{\mathbf{p}_i\}_{i=1}^{N} \leftarrow$ `spline`(C) ; `/* closed degree-5 B-spline, N = 300 */`
3 $\mathbf{p}_i \leftarrow \mathbf{p}_i/s, \quad i = 1,\ldots,N$; `/* px → mm */`
4 $\bar{\mathbf{c}} \leftarrow \frac{1}{N}\sum_{i=1}^{N} \mathbf{p}_i$;
5 **for** $i \leftarrow 1$ **to** N **do**
6 $\quad \hat{\mathbf{u}}_i \leftarrow \frac{\mathbf{p}_i - \bar{\mathbf{c}}}{\|\mathbf{p}_i - \bar{\mathbf{c}}\|}$;
7 $\quad \mathbf{p}_i \leftarrow \mathbf{p}_i + \delta\,\hat{\mathbf{u}}_i$; `/* perilesional offset */`
8 $R^{\uparrow} \leftarrow \{(\mathbf{p}_i, t)\}, \quad R^{\downarrow} \leftarrow \{(\mathbf{p}_i, 0)\}$;
9 **if** $r_e > 0$ **then**
10 $\quad \mathcal{E} \leftarrow$ `edge_rings`$(\{\mathbf{p}_i\}, r_e, m)$;
11 **else**
12 $\quad \mathcal{E} \leftarrow \varnothing$;
13 $T \leftarrow$ `triangulate`$(R^{\uparrow}, \mathcal{E}, R^{\downarrow})$;
14 **return** T;

3.3. Segmentation Back-Ends and Reference

The segmentation stage offers three interchangeable back-ends, a promptable foundation model and two classical region methods, so the geometric pipeline can be exercised independently of any single segmenter. In IMTOP the back-ends are not run unaided:

each is seeded from the contour drawn by the operator and returns a refined boundary, so the workflow is semi-automatic by design. In the present evaluation the same operator contour is both the seed of the three back-ends and the reference against which their result is scored. The reported agreement therefore describes boundary refinement under a common, reference-derived initialization: it is an upper bound on what an unseeded workflow would achieve and not the accuracy of fully automatic wound segmentation. The dependence is strongest for the two classical methods, whose result is confined by construction between an eroded and a dilated copy of the reference (see below), and weaker for SAM, which receives only the bounding box of the reference and one interior point. When no manual delineation is available, a reference-free prompt is derived automatically by Otsu thresholding [25], a border-based polarity check and selection of the largest interior connected component while discarding border-touching blobs, so the tool also runs without an operator; this mode, which drives SAM only, was designed for high-contrast scenes, such as a phantom on a uniform background, and was run on the clinical images only as a check of the dependence on the seed (Section 4.1).

The reference is delineated by an operator who places a variable number of control points, depending on the wound complexity, interpolated by a closed periodic cubic interpolating spline resampled to $N = 300$ points, then rasterized by polygon filling. This is the same sample count and resampling as the automatic contour, so manual and automatic masks pass through the identical operator of Algorithm 1. The delineation was performed by a single operator, with a geometric and not a clinical background, a point revisited in Section 5.5.

The first back-end, and the default one, is SAM with the Vision Transformer base (ViT-B) image encoder [34], prompted with the axis-aligned bounding box of the reference contour and a single positive point placed at the maximum of the distance transform and not at the centroid, since for concave contours the centroid can fall outside the mask whereas the distance-transform maximum is interior by construction; multiple-mask output is disabled. The second is GrabCut [27], initialized with probable foreground inside the reference region, certain foreground on a core eroded by two iterations and certain background outside a band dilated by two iterations, run for five iterations and restricted to the dilated band. The third is the marker-controlled watershed of Meyer [26], with the eroded core as foreground markers and the region outside a dilated band as background, the intervening band resolved by flooding. Both classical results are therefore bounded between the eroded core and the dilated band of the reference. In all cases the morphological radius of the erosion and dilation is tied to the image size, $k_{\mathrm{m}} = \max(7,\, 0.02 \max(h, w))$ px with $h \times w$ the image size in pixels, and the raw output is cleaned by an opening and closing with an elliptical kernel of size $k_{\mathrm{c}} = \max(3,\, 0.012 \max(h, w))$ px forced odd, followed by retention of the largest connected component, before entering the geometric chain.

3.4. Evaluation Metrics

The manual and automatic results are compared at two levels, mask agreement and clinical coverage. Areas and distances are expressed in millimetres through s where a physical reference is available; the clinical images carry no such reference, so s is undefined there and the clinical metrics (Table 2, including the Hausdorff distance) are reported in pixels, while millimetre quantities are reported only for the calibrated phantom experiments. Let G and S be the reference and automatic masks (the latter being the output of the reference-seeded back-end), with boundaries $\partial G, \partial S$, and $|\cdot|$ the pixel count. Region

overlap is summarized by the Dice coefficient [37] and the IoU, a monotone transform of Dice reported for comparability with the literature,

$$\mathrm{Dice}(G,S) = \frac{2\,|G \cap S|}{|G| + |S|}, \tag{3}$$

$$\mathrm{IoU}(G,S) = \frac{|G \cap S|}{|G \cup S|}. \tag{4}$$

Table 2. Agreement of the three reference-seeded back-ends with the single-operator reference on the 149 clinical DFU images, all valid cases (no exclusion of outliers). n is the number of images on which the back-end returned a mask; it is lower for watershed, which returned no mask on three images. Dice and IoU: overlap of the automatic and reference masks (Equations (3) and (4)); patch area error: relative difference of their areas (Equation (6)); Hausdorff distance (Equation (8)) in pixels, because the images are uncalibrated. IQR: interquartile range; SD: standard deviation.

Back-End	n	Statistic	Dice	IoU	Patch Area Error (%)	Hausdorff (px)
SAM	149	Median [IQR]	0.954 [0.932–0.968]	0.911 [0.873–0.937]	5.4 [3.0–10.2]	4.6 [3.1–8.8]
		Mean ± SD	0.947 ± 0.031	0.901 ± 0.053	7.1 ± 5.8	8.7 ± 11.8
GrabCut	149	Median [IQR]	0.954 [0.932–0.972]	0.913 [0.874–0.946]	3.9 [1.7–8.9]	5.8 [3.7–8.8]
		Mean ± SD	0.943 ± 0.044	0.895 ± 0.071	7.7 ± 10.5	9.4 ± 10.6
Watershed	146	Median [IQR]	0.948 [0.918–0.961]	0.900 [0.848–0.925]	6.0 [2.8–11.4]	6.7 [4.8–9.4]
		Mean ± SD	0.928 ± 0.073	0.873 ± 0.096	9.0 ± 10.6	12.4 ± 22.3

Size agreement is the relative area error on the calibrated areas,

$$A_X = \frac{|X|}{s^2}, \tag{5}$$

$$\varepsilon_A = \frac{|A_G - A_S|}{A_G}. \tag{6}$$

The error ε_A is unsigned; its signed counterpart, $(A_S - A_G)/A_G$, negative when the automatic mask is smaller than the reference, is used in the error budget (Section 3.5). The worst-case boundary deviation is the symmetric Hausdorff distance [38], built from the directed distance $\vec{h}$,

$$\vec{h}(X,Y) = \max_{x \in X} \min_{y \in Y} \|x - y\|, \tag{7}$$

$$H(\partial G, \partial S) = \frac{1}{s} \max\{\vec{h}(\partial G, \partial S), \vec{h}(\partial S, \partial G)\}. \tag{8}$$

Finally, because the dressing must cover the wound, the disagreement is split into the clinically unsafe under-coverage A_u (wound left exposed) and the conservative over-coverage A_o (excess on healthy skin), with coverage fraction c,

$$A_u = \frac{|G \setminus S|}{s^2}, \tag{9}$$

$$A_o = \frac{|S \setminus G|}{s^2}, \tag{10}$$

$$c = 1 - \frac{A_u}{A_G} = \frac{|G \cap S|}{|G|}, \tag{11}$$

as illustrated in Figure 2d. In the results A_u and A_o are reported relative to the reference area, A_u/A_G and A_o/A_G, so that $c = 1 - A_u/A_G$; c coincides with the recall (sensitivity) of S with respect to G. For the uncalibrated clinical images $s = 1$ px/px in Equations (5)–(11), so the patch area error, a ratio, is scale-free and Algorithm 2 runs unchanged. Overlap alone is

an incomplete predictor of geometric fidelity, since a disagreement spread along the whole contour and one concentrated in a single region affect overlap and boundary deviation, and hence the fabricated patch, differently (Figure 2b,c); overlap and boundary measures are therefore reported jointly and their relationship to the patch error is examined in Section 4.4. The released tool also reports shape descriptors and further boundary measures (perimeter error and average symmetric surface distance), omitted here for compactness.

Algorithm 2: Dataset evaluation across segmentation back-ends

Data: image set $\mathcal{I}$, reference length d, back-end set $\mathcal{B}$, parameters (t, δ, r_e, m)
Result: per-image, per-back-end record set $\mathcal{R}$

$\mathcal{R} \leftarrow \varnothing$;
for *each image* $I \in \mathcal{I}$ **do**
  $s \leftarrow$ `calibrate`(I, d) ; `/* two-point scale, Equation (1); $s = 1$ if no reference */`
  $M_G \leftarrow$ `manual_reference`(I);
  $T_G \leftarrow$ `chain`$(M_G, s, t, \delta, r_e, m)$; `/* reference patch, Algorithm 1 */`
  for *each back-end* $b \in \mathcal{B}$ **do**
    $M_S \leftarrow$ `auto_seg`(I, M_G, b) ; `/* seeded from $M_G$ */`
    if $M_S = \varnothing$ **then** $\mathcal{R} \leftarrow \mathcal{R} \cup \{(I, b, \text{fail})\}$; **continue**;
    $\mathcal{M} \leftarrow$ `metrics`(M_G, M_S, s) ; `/* Equations (3)-(11) */`
    $T_S \leftarrow$ `chain`$(M_S, s, t, \delta, r_e, m)$;
    $|\varepsilon_{\text{Seg}}| \leftarrow \dfrac{|a(T_S) - a(T_G)|}{a(T_G)}$; `/* patch area error */`
    $\mathcal{R} \leftarrow \mathcal{R} \cup \{(I, b, \mathcal{M}, |\varepsilon_{\text{Seg}}|)\}$;
return $\mathcal{R}$;

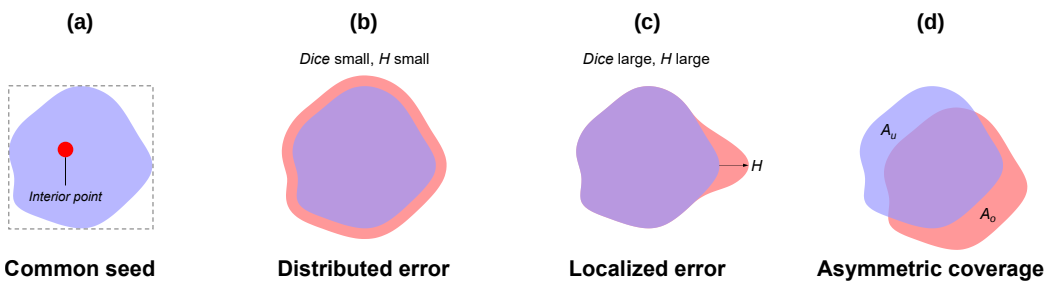

Figure 2. Concept of the fabrication-aware evaluation; in each panel the reference mask is drawn in violet and the automatic mask in red: (**a**) all back-ends are initialized from the operator's reference mask (dashed box: SAM box prompt; dot: positive point at the distance-transform maximum; the classical back-ends use an eroded core of the same mask); (**b**,**c**) a disagreement distributed along the whole contour and one concentrated in a single region give different combinations of overlap (Dice) and worst-case boundary deviation (Hausdorff distance H), so that neither metric alone describes the fabricated patch; (**d**) the coverage split into under-coverage A_u (reference not covered by the automatic mask, that is wound left exposed) and over-coverage A_o (automatic mask outside the reference, that is excess on healthy skin).

3.5. Error Budget and Offset Sizing

The error budget concerns one quantity, the relative error of the in-plane area of the fabricated patch, called the total area error below, and the three stages that act on it in sequence. Writing ε_k for the signed relative area error introduced at stage k, negative when the area is reduced, the patch area relative to its reference value is the product

$$1 + \varepsilon_{\text{patch}} = (1 + \varepsilon_{\text{Seg}})\,(1 + \varepsilon_{\text{Scale}})\,(1 + \varepsilon_{\text{Pr}}). \tag{12}$$

The segmentation term ε_{Seg} is the signed relative difference between the areas of the automatic and reference masks, $(A_S - A_G)/A_G$, whose magnitude is the patch area error ε_A of Equation (6); it is measured on the clinical images (Experiment A). The printing term ε_{Pr} is the relative in-plane area error of the printed part against its STL model and is measured on the phantoms (Experiment C). The scale term $\varepsilon_{\text{Scale}}$ is the relative area error caused by the placement of the two calibration points; it could not be measured on the clinical images, which contain no physical reference, and is obtained by simulation (Section 3.6).

Four descriptors that are often used interchangeably are distinguished. The error of a single case is the signed deviation ε_k. The bias B_k is the mean of ε_k over the cases, the systematic part, which a correction could in principle remove. The standard uncertainty E_k is the standard deviation (SD) of ε_k, the dispersion around the bias, and its square is the variance. The median of $|\varepsilon_k|$ is reported as the typical magnitude of a single case, because the distributions are skewed. For the product of Equation (12) and uncorrelated terms, the law of propagation of uncertainty of the GUM gives, to first order (clause 5.1.6 of [15]),

$$B_{\text{patch}} \approx B_{\text{Seg}} + B_{\text{Scale}} + B_{\text{Pr}}, \tag{13}$$

$$E_{\text{patch}} = \sqrt{E_{\text{Seg}}^2 + E_{\text{Scale}}^2 + E_{\text{Pr}}^2}. \tag{14}$$

The sum in quadrature is thus applied to standard uncertainties and not to mean absolute errors, and the biases are added algebraically and reported apart. Because the segmentation errors are skewed and not small, Equations (13) and (14) are complemented by a propagation of distributions by Monte Carlo simulation, the approach of Supplement 1 to the GUM [40], which relies neither on linearization nor on normality. For each back-end, 20,000 values of ε_{Seg} are drawn with replacement from the observed per-image signed errors, so that their bias, skewness and tails are retained without a parametric model; $\varepsilon_{\text{Scale}}$ is drawn from the two-point calibration model of Section 3.6; ε_{Pr} is drawn from a normal distribution centred on the mean in-plane area error measured in Experiment C (0.69%), with an assumed SD of 0.10%, of the order of the dispersion of the dimensional error (0.11%), since doubling it leaves the printing share of the variance below 0.1%; and $\varepsilon_{\text{patch}}$ follows from Equation (12). The simulation returns the mean and SD of $\varepsilon_{\text{patch}}$, the median and the 95th percentile of $|\varepsilon_{\text{patch}}|$ and the fraction of draws that exceed given thresholds. The share of the variance attributable to each term is E_k^2/E_{patch}^2, from Equation (14).

The three terms are treated as uncorrelated. In this study that holds by construction: ε_{Seg} is a ratio in which the scale cancels, because the reference and automatic patches are built with the same s (Algorithm 2), $\varepsilon_{\text{Scale}}$ is simulated independently, and printing acts on the final STL. In practice a blurred or low-contrast photograph may degrade both the segmentation and the placement of the calibration points; a positive correlation ρ between ε_{Seg} and $\varepsilon_{\text{Scale}}$ would add $2\rho E_{\text{Seg}} E_{\text{Scale}}$ under the square root of Equation (14). The budget is therefore an estimate propagated from separate experiments, each against its own reference, and not a measurement of one wound carried through the whole chain. It also has two omissions, which make it a lower bound of the error with respect to the true wound. A real wound has no physical ground-truth patch, so the manual reference is the proxy ground truth and ε_{Seg} quantifies the agreement with that delineation, not with the true wound boundary; and any departure from the planar, fronto-parallel geometry assumed in Section 3.1 is not included (Section 5.5).

The complete data flow is shown in Figure 3 and the evaluation loop over images and back-ends in Algorithm 2, in which METRICS returns Dice, IoU, ε_A, H, A_u and A_o (Equations (3)–(11)), $a(\cdot)$ is the in-plane area and CHAIN is Algorithm 1. In the batch

evaluation the reference contour was rebuilt from the saved reference mask with the same chain as the automatic one, and the patch area error of Algorithm 2, the magnitude of $\varepsilon_{\mathrm{Seg}}$, was computed from the pixel areas of the two masks; computing it from the areas of the two fitted contours changes the per-image value by a median of 0.2 percentage points.

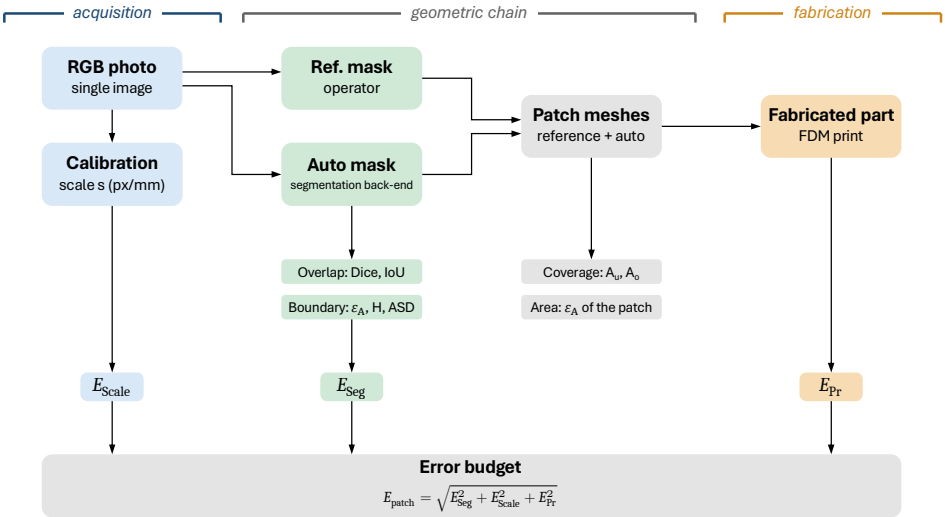

Figure 3. Data flow of the stage-wise error budget (schematic): the three terms E_{Seg}, E_{Scale} and E_{Pr} are obtained from separate experiments (Section 3.6), the first and the third by measurement and the second by simulation, not on a single image carried through the chain, and combined into the standard uncertainty of the fabricated-patch area (Equation (14)). ε_A: relative area error; H: Hausdorff distance; ASD: average symmetric surface distance; A_u, A_o: under- and over-coverage; FDM: fused deposition modelling. Box colours group the elements by stage: blue, acquisition; green, segmentation; grey, patch geometry and combination of the terms; orange, fabrication.

Because under-coverage is the unsafe error, the offset is sized from its empirical distribution and not fixed arbitrarily. For each image i, the margin μ_i is the smallest uniform normal offset of the automatic patch footprint P_i (the rasterized contour of Algorithm 1 at $\delta = 0$) that fully covers the reference mask G_i, that is the directed distance of Equation (7) from the reference to the footprint,

$$\mu_i = \vec{h}(G_i, P_i). \tag{15}$$

The clinical images are uncalibrated and differ in resolution, so a margin in pixels has no common meaning across them; μ_i is therefore expressed relative to the equivalent diameter of the reference, $D_i = 2\sqrt{|G_i|/\pi}$, and the offset that achieves full coverage in a fraction τ of the cases is the empirical τ-quantile of the relative margins,

$$\delta^*(\tau) = Q_\tau(\{\mu_i/D_i\}), \tag{16}$$

evaluated per back-end, with $\tau = 0.95$ as a conventional coverage level; other levels follow from the same quantile. The relative offset becomes a length once the size of the wound is known: for a wound of equivalent diameter D the offset of Algorithm 1 is $\delta = \delta^*(\tau)\, D$. Three checks accompany the rule. The quantile is estimated and tested on different images by five-fold cross-validation repeated 2000 times, δ^* being taken from four folds and the fraction of fully covered wounds counted on the fifth. The radial offset of Algorithm 1 is compared with the normal offset of Equation (15) by computing, for each image, the smallest radial offset that covers the reference. Finally, the cost of the margin is reported as the over-coverage A_o/A_G that it produces. The automatic masks used for this analysis were regenerated with the same code as Experiment A; they reproduce its

per-image scores (identically for GrabCut and watershed, and with a Dice coefficient within 0.003 for SAM).

3.6. Experimental Design

The pipeline was evaluated in three experiments and a simulation of the scale calibration, each isolating one stage of the budget.

Experiment A (clinical segmentation and propagation). A set of 149 clinical DFU photographs was assembled from two public sources, the Lower Limb and Feet Wound Image Dataset [41] and a public DFU image dataset [42], retaining superficial, single and clearly delineable wounds acquired in an approximately fronto-parallel view (median image size 331×331 px; median equivalent diameter of the reference 94 px). Each image was manually delineated by a single operator to give the reference, then segmented by the three back-ends seeded from that reference; the mask-level metrics and the propagated patch-level area error were computed (RQ1, RQ2), with the perilesional offset set to $\delta = 0$ so that the offset, derived in Section 3.5 and applied in RQ3, does not enter the segmentation-error characterization. As a check of the dependence on the seed, the reference-free mode of Section 3.3 was also run on the 149 images with no operator input, and its masks were scored against the same reference; these results are reported separately (Section 4.1) and do not enter the error budget.

Experiment B (absolute geometric accuracy). Six planar shapes of known CAD geometry, each printed in three replicates (18 parts), were photographed under the same calibrated protocol, with the scale set through the two-point tool of Section 3.1, and the reconstructed in-plane geometry was regressed against the nominal CAD size. This measures the photograph-to-geometry error (segmentation plus calibration) on objects of known nominal geometry; because the printed parts are compared with the nominal CAD, the comparison also contains the printing deviation quantified in Experiment C.

Experiment C (printing accuracy). The same 18 parts were measured against their source STL files to isolate the printing term, separating the in-plane area error from the out-of-plane thickness, which is a user-set parameter and not a reconstructed quantity.

Scale-calibration simulation. The scale term was simulated, not measured. The two endpoints of a reference of length L (in pixels) were perturbed with isotropic Gaussian noise of SD σ per coordinate, the perturbed length L' was computed and the relative area error $\varepsilon_{\text{Scale}} = (L/L')^2 - 1$ was recorded, for 20,000 draws. Because the clinical images contain no reference, L was set to 117 px, the median of the longer bounding-box side of the reference contours, which represents a reference about as long as the wound itself in these images. The repeatability of point placement was not measured in this study and $\sigma = 3$ px was assumed; the dependence on σ was explored between 0.5 and 5 px (Section 4.3). For $\sigma \ll L$ the simulation is summarized by the closed form

$$E_{\text{Scale}} \approx 2\sqrt{2}\,\frac{\sigma}{L}, \tag{17}$$

so that the term can be recomputed for any measured placement noise and reference length: it depends only on the ratio σ/L, hence on the resolution of the image and on the length of the reference in the image, and not on a fixed physical size.

Phantoms were fabricated by FDM in polylactic acid (PLA) on a Creality K1C printer (Shenzhen Creality 3D Technology Co., Ltd., Shenzhen, China) at a 0.2 mm layer height; planar reference dimensions were measured with a digital caliper (CDJB15; Borletti, LTF S.p.A., Antegnate, Italy). All clinical references were delineated by a single operator, a limitation discussed in Section 5.5.

3.7. Statistical Analysis

The statistical analysis addressed four objectives: comparing the per-image mask-level and patch-level metrics across the three back-ends, assessing the association between mask overlap and patch area error, evaluating the absolute geometric accuracy of the phantom reconstructions, and checking the offset rule on images not used to derive it. The procedures used for each objective are described below.

Normality was assessed with the Shapiro–Wilk test; the per-image scores departed from normality for every metric and back-end ($p < 0.001$), so the distributions are summarized by the median and the interquartile range (IQR), with the mean and the SD reported alongside for comparison with the literature, and non-parametric tests are used throughout.

No case is excluded from the primary analysis. All valid segmentations are retained ($n = 149$, 149 and 146 for SAM, GrabCut and watershed, the last returning no mask on three images), and the paired comparisons use the 146 images on which the three back-ends returned a mask, so that they are compared on the same images. Differences across the back-ends are evaluated with the Friedman test, with Kendall's W as effect size [43], followed by post hoc Wilcoxon signed-rank tests for the three pairwise comparisons, with Bonferroni-adjusted p-values and the effect size $r = |Z|/\sqrt{n}$, where Z is the standardized test statistic and n the number of pairs [44]; values of r of about 0.1, 0.3 and 0.5 are conventionally read as small, medium and large. The Wilcoxon signed-rank test also compares under- and over-coverage within each back-end.

In the first version of this work the per-image outliers of the patch area error were removed before the analysis, separately for each back-end, by Tukey's rule (values beyond 1.5 IQR from the quartiles). That rule discards upper-tail cases of the very outcome under comparison, and a different number of cases for each back-end (5, 14 and 10), so it is no longer applied in the primary analysis. The analysis after exclusion ($n = 144$, 135 and 136; 125 images common to the three back-ends) is retained as a sensitivity analysis in Appendix A, and conclusions are drawn only from the findings that hold in both.

The relationship between mask overlap and the patch area error (RQ2) is summarized by the Spearman rank correlation, pooled and per back-end; each image contributes one observation per back-end, so the pooled coefficient is descriptive. Absolute geometric accuracy on the phantoms is reported by linear regression of reconstructed size against nominal size. The perilesional offset is assessed by repeated five-fold cross-validation (Section 3.5). The significance level is $\alpha = 0.05$, and the analyses use Python (v3.11.9; NumPy v1.26.4, SciPy v1.13.1 and pandas v2.2.1) and R (v4.3). The reference is delineated by a single operator, so no inter-operator agreement is reported (Section 5.5).

3.8. Software Architecture

IMTOP is a desktop application built around a Python back-end and an HTML and JavaScript front-end embedded in a Qt WebEngine view (PyQt6 v6.11.0), the two sides communicating through a Qt WebChannel bridge (Figure 4). The front-end provides an interactive image canvas, a two-point scale tool, a segmenter selector, a parameter panel, a per-metric overlay list and a three-dimensional patch viewer; the patch mesh is built and rendered on the client with Three.js (release r128), and its STL representation is generated from the displayed mesh and passed to the back-end for writing. The back-end performs image input and output, scale computation, segmentation with three interchangeable back-ends together with mask post-processing, and metric computation, and it writes all output files (reference mask, STL and metric records in JSON). The pipeline of Figure 1 is governed by a state machine whose stages are unlocked by specific back-end events, the segmentation event unlocking both the metrics and the patch stages; backward navigation is always available and loading a new image resets the pipeline. The mapping from the conceptual

pipeline to the implemented interface is shown in Figure 5. Communication between the two sides is asynchronous, through named slots and signals (Figure 4). The segmentation model is loaded once, on first use, on a worker thread so that the interface never blocks; only the per-image encoding pass is repeated for a new image. At the metrics stage the operator sees, for the selected back-end, the overlap, boundary and coverage metrics overlaid on the image before accepting the segmentation, and at the patch stage the thickness, offset and edge parameters are set interactively; this is the decision loop that the error budget of Section 3.5 is meant to inform. The application will be released to support reproducibility (data availability statement).

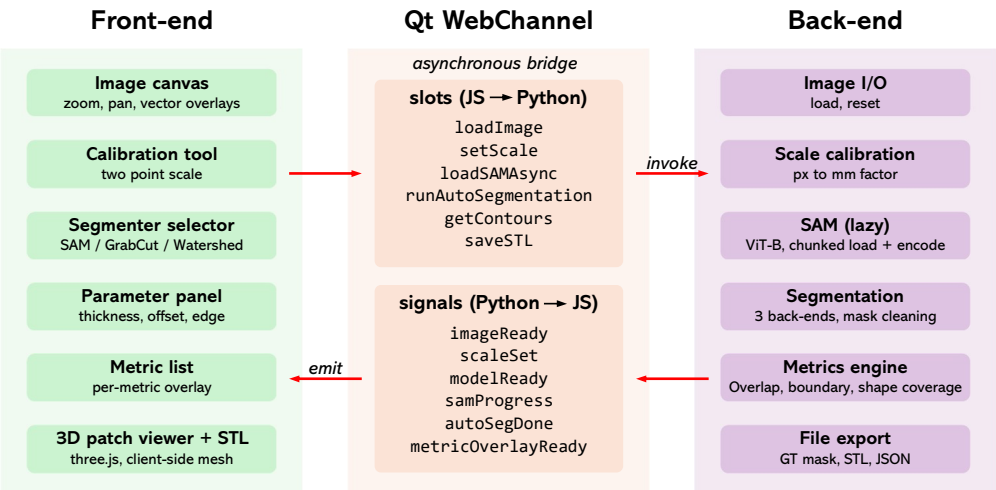

Figure 4. Software architecture: an HTML/JavaScript front-end (**left**) and a Python back-end (**right**) over a Qt WebChannel bridge (**centre**), with the implemented slots (calls from the front-end to the back-end) and signals (events emitted by the back-end) on the bus.

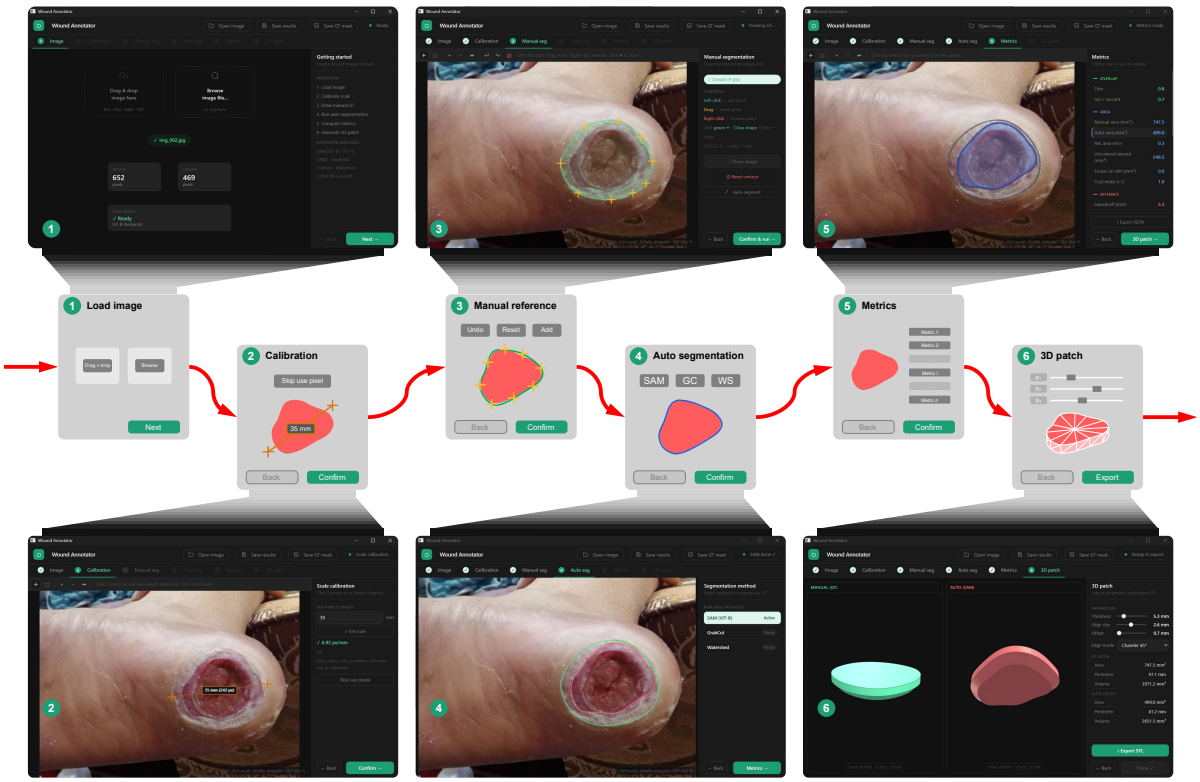

Figure 5. From pipeline concept to implemented interface, shown at the full width of the page for legibility. The central row sketches each stage as an interface mock-up, and the screenshots above

and below show the corresponding view in the released application (whose window is titled Wound Annotator); the numbers 1 to 6 correspond to the six stages of Figure 1: (**1**) image loading; (**2**) two-point scale calibration on a reference of known length; (**3**) manual reference contour drawn by the operator (control points and spline); (**4**) reference-seeded segmentation with the selected back-end; (**5**) metrics panel with the per-metric overlay; (**6**) three-dimensional view of the reference and automatic patches with the thickness, offset and edge parameters.

4. Results

Findings are reported in the order of the three research questions; their interpretation is deferred to Section 5. Unless stated otherwise the results refer to all valid cases, without exclusion of outliers; the corresponding results after exclusion are in Appendix A. The per-image overlap and patch-area scores of Section 4.1 are the inputs to the error budget (RQ1, Section 4.3) and to the overlap-versus-error analysis (RQ2, Section 4.4).

4.1. Agreement of the Reference-Seeded Back-Ends with the Reference

Watershed failed to return a valid mask on three images (2%); SAM and GrabCut returned a mask on all 149. Table 2 summarizes the per-image scores on all valid cases and Figure 6 shows their distributions. Because the three back-ends are seeded from the reference, these values describe boundary refinement under a common initialization: they are not the accuracy of automatic segmentation and are not comparable with the unseeded values reported in the literature.

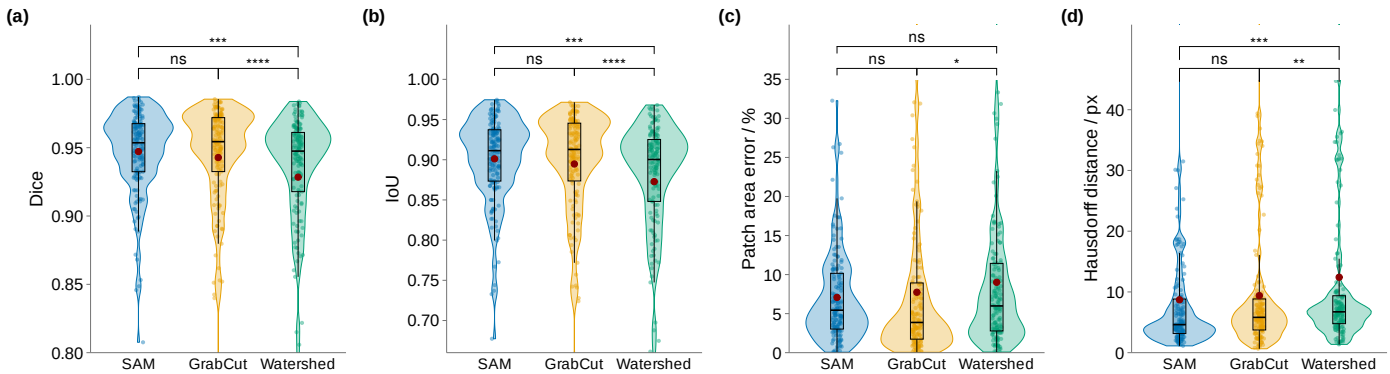

Figure 6. Per-image agreement with the reference for the three reference-seeded back-ends, all valid cases ($n = 149$, 149 and 146 for SAM, GrabCut and watershed): (**a**) Dice coefficient; (**b**) IoU; (**c**) patch area error; (**d**) Hausdorff distance, in pixels because the images are uncalibrated. Violins show the distribution, boxes the median and the interquartile range, red dots the mean. Brackets give the Bonferroni-adjusted paired Wilcoxon results on the 146 images common to the three back-ends (ns: not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$; **** $p < 0.0001$). The vertical axes are truncated for readability: 8, 7, 9 and 10 of the 444 points lie outside the displayed range in panels (**a**–**d**) and are included in all the statistics.

The distributions were skewed, so that medians and means differ, most clearly for the patch area error (medians of 5.4%, 3.9% and 6.0% against means of 7.1%, 7.7% and 9.0% for SAM, GrabCut and watershed) and for the Hausdorff distance. The Friedman test was significant for the four metrics ($p < 0.001$, $n = 146$), with small effect sizes (Kendall's W between 0.06 and 0.09; Table 3). In the post hoc comparisons SAM and GrabCut did not differ in Dice, IoU or Hausdorff distance, and both agreed with the reference better than watershed, with a higher Dice and IoU and a smaller Hausdorff distance (adjusted $p \leq 0.002$, r between 0.28 and 0.41). For the patch area error only the difference between GrabCut and watershed was significant (adjusted $p = 0.011$, $r = 0.24$).

Table 3. Comparisons between the back-ends. For each metric: Friedman omnibus test with Kendall's W as effect size, and post hoc paired Wilcoxon signed-rank tests with the standardized statistic Z, the Bonferroni-adjusted p-value (p_{adj}) and the effect size $r = |Z|/\sqrt{n}$. Primary analysis: all valid cases, on the $n = 146$ images on which the three back-ends returned a mask. Sensitivity analysis: after the Tukey exclusion, on the $n = 125$ images retained for the three back-ends. A positive Z means that the first back-end of the pair has the larger value. For the Friedman test the p-value is not adjusted.

Comparison	All Valid Cases ($n = 146$)			After Exclusion ($n = 125$)	
	Statistic	p_{adj}	Effect Size	p_{adj}	Effect Size
Dice					
Friedman omnibus test	$\chi^2(2) = 25.04$	<0.001	$W = 0.09$	<0.001	$W = 0.10$
SAM vs. GrabCut	$Z = -0.08$	1.000	$r = 0.01$	0.110	$r = 0.19$
GrabCut vs. watershed	$Z = +4.93$	<0.001	$r = 0.41$	<0.001	$r = 0.53$
SAM vs. watershed	$Z = +3.83$	<0.001	$r = 0.32$	0.057	$r = 0.21$
IoU					
Friedman omnibus test	$\chi^2(2) = 25.04$	<0.001	$W = 0.09$	<0.001	$W = 0.10$
SAM vs. GrabCut	$Z = -0.12$	1.000	$r = 0.01$	0.096	$r = 0.19$
GrabCut vs. watershed	$Z = +5.00$	<0.001	$r = 0.41$	<0.001	$r = 0.54$
SAM vs. watershed	$Z = +3.84$	<0.001	$r = 0.32$	0.054	$r = 0.21$
Patch area error					
Friedman omnibus test	$\chi^2(2) = 18.25$	<0.001	$W = 0.06$	<0.001	$W = 0.11$
SAM vs. GrabCut	$Z = +1.99$	0.141	$r = 0.16$	<0.001	$r = 0.35$
GrabCut vs. watershed	$Z = -2.92$	0.011	$r = 0.24$	<0.001	$r = 0.36$
SAM vs. watershed	$Z = -1.31$	0.567	$r = 0.11$	1.000	$r = 0.04$
Hausdorff distance					
Friedman omnibus test	$\chi^2(2) = 20.23$	<0.001	$W = 0.07$	<0.001	$W = 0.06$
SAM vs. GrabCut	$Z = -1.06$	0.866	$r = 0.09$	1.000	$r = 0.05$
GrabCut vs. watershed	$Z = -3.42$	0.002	$r = 0.28$	0.002	$r = 0.31$
SAM vs. watershed	$Z = -3.69$	<0.001	$r = 0.31$	0.058	$r = 0.21$

The sensitivity analysis after the Tukey exclusion (Table 3 and Appendix A) left the medians almost unchanged (5.2%, 3.4% and 5.7% for the patch area error) and lowered the means and SDs, and it changed the results for two pairs of back-ends (four of the twelve pairwise tests): the differences between SAM and watershed in Dice, IoU and Hausdorff distance were no longer significant (adjusted p between 0.054 and 0.058), whereas the difference between SAM and GrabCut in patch area error became significant (adjusted $p < 0.001$). The findings common to the two analyses are that SAM and GrabCut are indistinguishable in overlap and that GrabCut agrees with the reference better than watershed in overlap, area and boundary; the position of SAM relative to the other two depends on how the upper tail is handled.

Run without any operator input, the reference-free mode of the tool (an Otsu-derived prompt followed by SAM) returned a mask on 106 of the 149 images and found no candidate object on the other 43. Against the reference, its Dice coefficient was at least 0.90 on 12 images (8.1% of the 149, against 92.6% for the seeded SAM), at least 0.50 on 20 (13.4%), and below 0.10, or no mask was returned, on 123 (82.6%). On the 12 images with a Dice of at least 0.90, the median Dice was 0.953 and the median patch area error 5.8%, close to the seeded values. This mode therefore does not reliably localize the wound on clinical photographs, and the values of Table 2 describe the refinement of a boundary that the operator has already localized.

4.2. Absolute Geometric and Printing Accuracy

On the 18 phantoms (six planar CAD shapes printed in three replicates, Figure 7) the reconstructed in-plane size regressed against the nominal CAD with slope 0.994 and $R^2 = 0.9995$ (Table 4): mean dimensional error 0.80%, area error 0.32%, Dice 0.997, with a worst-case minor-axis underestimate of 2.51% on the crescent. Measured against the source STL, the parts had a mean in-plane area error of 0.69% and a dimensional error of $0.41 \pm 0.11\%$; the out-of-plane thickness deviated by +7.4% and is reported apart from the in-plane budget (its consequence for the dispensed volume is discussed in Section 5.2).

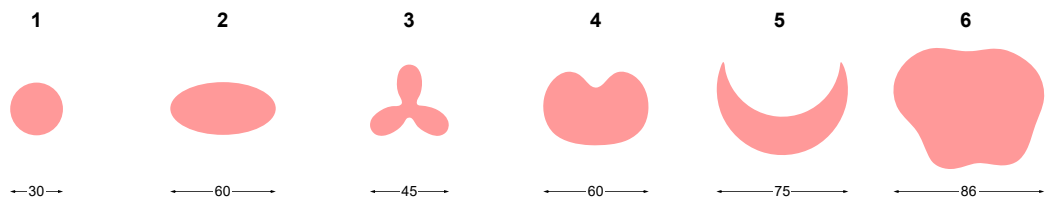

Figure 7. Nominal CAD geometries of the six planar phantom shapes of Experiments B and C, numbered 1 to 6 (disc, ellipse, trefoil, bean, crescent and freeform), with the nominal in-plane size in millimetres; each shape was printed three times in PLA by FDM (18 parts).

Table 4. Absolute geometric accuracy on the 18 phantoms (Experiment B: geometry reconstructed from the calibrated photograph against the nominal CAD model) and printing accuracy against the STL (Experiment C: printed parts measured against their STL models). The phantoms are planar PLA parts printed by FDM (six shapes, three replicates each).

Quantity	Value
Regression slope (reconstructed vs. nominal)	0.994
R^2	0.9995
Mean dimensional error (Experiment B)	0.80%
Mean area error (Experiment B)	0.32%
Mean Dice (Experiment B)	0.997
Worst-case minor-axis underestimate (crescent)	2.51%
Printing in-plane area error (Experiment C)	0.69%
Printing dimensional error (Experiment C)	$0.41 \pm 0.11\%$

Two remarks bound the scope of these values. First, they were obtained on rigid, planar PLA parts printed by FDM and photographed in a fronto-parallel view, not on the dressing material and not on a curved surface, so they hold for the tested fabrication set-up only. Second, Experiment B is the closest available approximation to an end-to-end test on a known object: each phantom was printed, photographed, calibrated with the two-point tool, segmented and reconstructed, and the reconstructed geometry was compared with the nominal CAD model, so that the 0.32% area deviation is the joint result of the printing, calibration, segmentation and contour-fitting errors of the same object, which may partly compensate one another (it is smaller than the printing deviation measured in Experiment C). The loop was not closed by printing the reconstructed contour and measuring the new part; that step would add one further printing deviation of the size measured in Experiment C.

4.3. Error Budget (RQ1)

Table 5 reports the three terms and their combination on all valid cases. The segmentation term had a bias of −5.7%, +1.8% and −5.5% for SAM, GrabCut and watershed and a standard uncertainty of 7.1%, 13.0% and 12.8%; the typical single-case magnitude (median of $|\varepsilon_{\mathrm{Seg}}|$) was 5.4%, 3.9% and 6.0%. The simulated scale term had a negligible bias

(+0.3%) and a standard uncertainty of 7.3% at $\sigma = 3$ px and $L = 117$ px (median magnitude 4.9%; Figure 8b), in agreement with the closed form of Equation (17), $2\sqrt{2} \times 3/117 = 7.3\%$. A longer reference reduces it in inverse proportion to its length: at the same σ it becomes 3.4%, 1.7% and 0.8% for $L = 250$, 500 and 1000 px. The scale and printing terms are the same for the three back-ends, because the scale factor is set by the operator's two points and applied identically to every mask, and printing acts on the final STL.

Table 5. Error budget of the total area error of the patch, $\varepsilon_{\text{patch}}$, all valid cases. For each term: bias (mean signed error), standard uncertainty (SD of the signed error) and median absolute error. The segmentation term is measured on the clinical images against the single-operator reference. The scale term is simulated (20,000 draws of the two-point calibration model, with an assumed placement noise $\sigma = 3$ px per coordinate on a reference of $L = 117$ px) and the printing term is modelled as a normal distribution centred on the mean in-plane area error measured on PLA phantoms printed by FDM, with an assumed SD of 0.10%; both are common to the three back-ends. The first-order rows are computed from the unrounded values of the terms; the variance shares are E_k^2/E_{patch}^2. The combined values are propagated estimates and not an end-to-end measurement. Negative values indicate a patch smaller than its reference.

Quantity	SAM	GrabCut	Watershed
Segmentation term (measured, Experiment A)			
Cases, n	149	149	146
Bias B_{Seg} (%)	−5.7	+1.8	−5.5
Standard uncertainty E_{Seg} (%)	7.1	13.0	12.8
Median $\lvert\varepsilon_{\text{Seg}}\rvert$ (%)	5.4	3.9	6.0
Scale term (simulated, $\sigma = 3$ px, $L = 117$ px)			
Bias B_{Scale} (%)		+0.3	
Standard uncertainty E_{Scale} (%)		7.3	
Median $\lvert\varepsilon_{\text{Scale}}\rvert$ (%)		4.9	
Printing term (PLA phantoms, Experiment C)			
Bias B_{Pr} (%)		+0.69	
Standard uncertainty E_{Pr} (%, assumed)		0.10	
First-order combination (Equations (13) and (14))			
Bias B_{patch} (%)	−4.7	+2.8	−4.6
Combined standard uncertainty E_{patch} (%)	10.2	14.9	14.7
Variance share of segmentation (%)	49	76	75
Variance share of scale (%)	51	24	25
Variance share of printing (%)		<0.1	
Monte Carlo propagation (Equation (12), 20,000 draws)			
Mean of $\varepsilon_{\text{patch}}$ (%)	−4.8	+2.8	−4.7
Standard deviation of $\varepsilon_{\text{patch}}$ (%)	10.0	15.0	14.7
Median $\lvert\varepsilon_{\text{patch}}\rvert$ (%)	7.4	7.3	8.0
95th percentile of $\lvert\varepsilon_{\text{patch}}\rvert$ (%)	21.8	32.4	30.6
Draws with $\lvert\varepsilon_{\text{patch}}\rvert > 5\%$ (%)	65	64	67
Draws with $\lvert\varepsilon_{\text{patch}}\rvert > 10\%$ (%)	36	37	40

To first order the combined standard uncertainty was 10.2%, 14.9% and 14.7%, with a bias of −4.7%, +2.8% and −4.6%. The Monte Carlo propagation gave the same values within 0.3 percentage points (SD 10.0%, 15.0% and 14.7%; mean −4.8%, +2.8% and −4.7%), so the linearization holds for the first two moments. The distribution of $\lvert\varepsilon_{\text{patch}}\rvert$ was right-skewed (Figure 9a), with a median of 7.4%, 7.3% and 8.0% and a 95th percentile of 21.8%, 32.4% and 30.6%; the absolute error exceeded 5% in about two-thirds of the draws and 10% in 36% to 40% of them (Figure 9b).

In the variance decomposition (Figure 8a) printing was negligible (<0.1%). Segmentation and scale contributed almost equally for SAM (49% and 51%), and segmentation was the larger contributor for GrabCut and watershed (76% and 75%). The order of the two terms depends on the assumed ratio σ/L (Figure 8b) and on the handling of the upper tail: after the Tukey exclusion the scale term was the larger contributor for SAM and GrabCut (62% and 53%) and the two terms were equal for watershed (Appendix A). What holds in both analyses, at the assumed σ/L, is that printing is negligible and that segmentation and scale are of the same order.

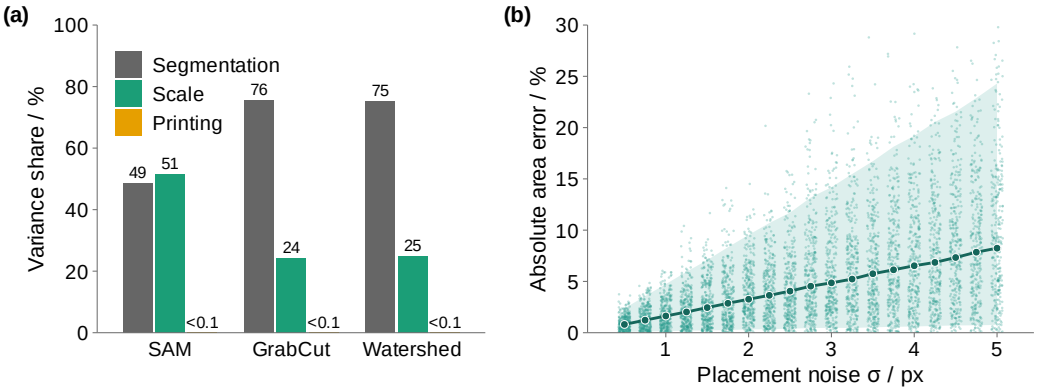

Figure 8. (**a**) Share of the variance of the total area error $\varepsilon_{\text{patch}}$ contributed by the segmentation, scale and printing terms, E_k^2/E_{patch}^2 (Equation (14)), per back-end (all valid cases; $n = 149$, 149 and 146 for SAM, GrabCut and watershed). (**b**) Simulated scale term: absolute relative area error against the placement noise σ of the two calibration points, for a reference of $L = 117$ px (line: median and band: 5th to 95th percentile of 20,000 draws; dots: 300 of the draws for each value of σ, those above 30% not being displayed). The budget of Table 5 uses $\sigma = 3$ px.

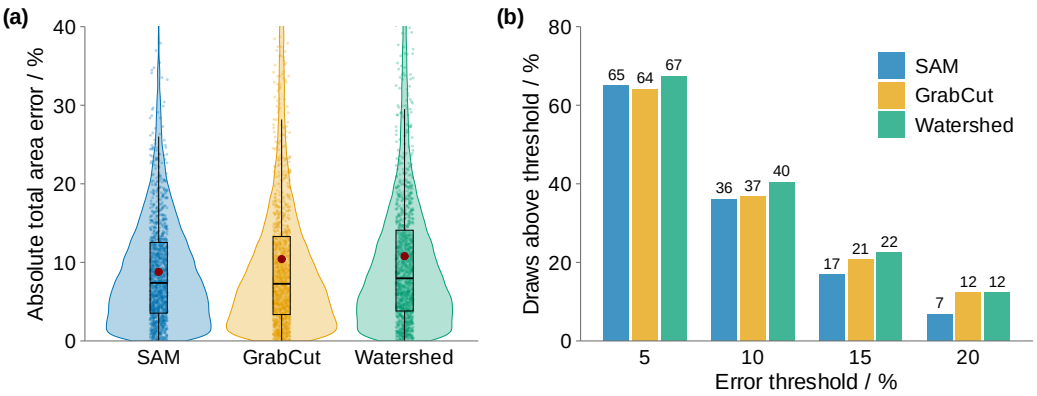

Figure 9. Monte Carlo distribution of the absolute total area error $|\varepsilon_{\text{patch}}|$ (all valid cases; 20,000 draws per back-end): (**a**) distribution per back-end (violin and box computed on all draws; box: median and interquartile range; red dot: mean; dots: 1200 of the draws; 1.7% of the draws lie above the displayed 40%); (**b**) percentage of draws exceeding thresholds of 5%, 10%, 15% and 20%. Each draw combines a resampled clinical segmentation error with a simulated scale error and a printing error (Equation (12)); the figure therefore shows a propagated estimate, not measured clinical cases.

These combined values must be read with caution. They are estimates propagated from separate experiments, with a segmentation term measured against a single-operator reference, a simulated scale term and a printing term measured on PLA phantoms; they are not a measured end-to-end error of a clinical case, and they do not include the terms omitted from the budget (Section 3.5).

4.4. Overlap as a Predictor of Patch Area Error (RQ2)

This analysis pools all valid cases ($n = 444$), like the rest of the primary analysis, so that the correlation is not conditioned on the outcome variable and the high-error cases are retained. Mask overlap was correlated with, but did not determine, the patch area error. The Dice coefficient and the patch area error were negatively and monotonically associated (Spearman $\rho = -0.68$, $p < 0.001$), and the association held within each back-end ($\rho = -0.62$, -0.65 and -0.75 for SAM, GrabCut and watershed; Figure 10). At a fixed overlap the patch error still varied over a wide range: among the 235 cases with Dice ≥ 0.95 the patch area error spanned 0.1% to 8.9% (median 3.3%, IQR 1.7% to 5.2%). For a given Dice value q the relative area disagreement is bounded above by $2(1-q)/q$, about 10.5% at $q = 0.95$, so the observed spread covers most of the feasible range: Dice constrains the magnitude of the disagreement but not its sign. Consistently, the comparisons between back-ends gave different verdicts for overlap and for patch area error (Table 3). SAM exceeded watershed in Dice (adjusted $p < 0.001$, $r = 0.32$) without differing from it in patch area error (adjusted $p = 0.567$); SAM and GrabCut, indistinguishable in Dice (adjusted $p = 1.000$), had median patch area errors of 5.4% and 3.9%, a difference that was not significant on all valid cases (adjusted $p = 0.141$) and was significant after the Tukey exclusion (adjusted $p < 0.001$). In both analyses, therefore, one pair of back-ends was ranked differently by overlap and by patch area error, although not the same pair.

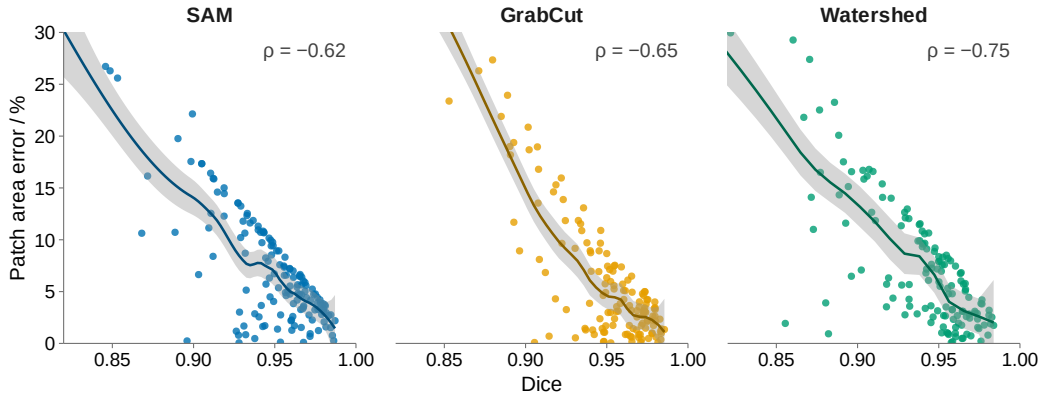

Figure 10. Patch area error against the Dice coefficient for each reference-seeded back-end (all valid cases; $n = 149$, 149 and 146 for SAM, GrabCut and watershed; $n = 444$ pooled). Each point is one image; the curve is a locally weighted (LOESS) trend with its 95% confidence band; ρ is the Spearman rank correlation. Sixteen points with Dice below 0.82 or patch area error above 30% fall outside the displayed window and are included in the statistics.

4.5. Coverage and Perilesional Offset (RQ3)

At zero offset the areal coverage c had a median of 93.4%, 96.3% and 92.7% (mean 92.1%, 95.0% and 90.5%) for SAM, GrabCut and watershed, and few cases were already fully covered (0.7%, 6.0% and 0.7% of the cases above 99.5% coverage), so an outward margin is needed in almost every case.

The disagreement was clinically asymmetric for two of the three back-ends (Figure 11). The median under-coverage A_u/A_G was 6.6%, 3.7% and 7.3% of the reference area and the median over-coverage A_o/A_G was 0.7%, 3.4% and 1.8%; under-coverage exceeded over-coverage for SAM and watershed (paired Wilcoxon, $n = 149$ and 146, $p < 0.001$, $r = 0.70$ and 0.57) but not for GrabCut ($n = 149$, $p = 0.836$, $r = 0.02$). Accordingly, the masks of SAM and watershed were smaller than the reference by a median of 4.8% in area, a systematic component of the segmentation term (the biases of Table 5), whereas GrabCut was centred on the reference (median -0.6%). The same pattern was found after the Tukey exclusion (Appendix A).

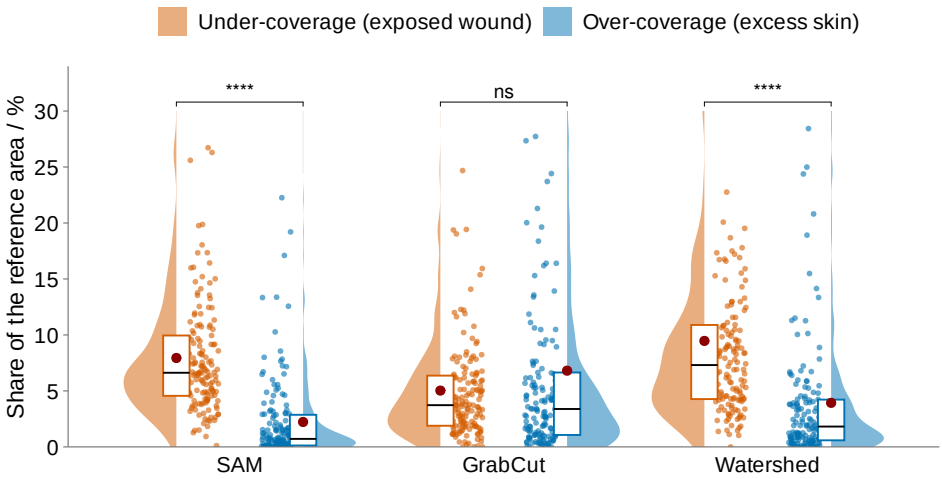

Figure 11. Under-coverage (wound left exposed, A_u / A_G) and over-coverage (excess on healthy skin, A_o / A_G) at zero offset, per reference-seeded back-end, as percentages of the reference area (all valid cases; $n = 149$, 149 and 146 for SAM, GrabCut and watershed). Half-violins show the distributions, dots the individual images, boxes the median and the interquartile range, red dots the mean. Brackets give the paired Wilcoxon test of under- against over-coverage (ns: not significant; **** $p < 0.0001$). Nine under-coverage and nine over-coverage values above 30% are not displayed and are included in the statistics.

Table 6 and Figure 12 report the perilesional offset. The margin needed for full coverage had a median of 7.2%, 5.9% and 7.0% of the equivalent diameter, and the offset that covers the wound in 95% of the cases, $\delta^*(0.95)$, was 16.4%, 14.3% and 20.6% of the equivalent diameter for SAM, GrabCut and watershed. In cross-validation, with the offset estimated and tested on different images, the wound was fully covered in 94.3%, 94.2% and 94.2% of the held-out cases (2.5th to 97.5th percentile over the repetitions: about 92.5% to 95.3%), against the nominal 95%, so within this dataset the rule transfers with a loss below one percentage point.

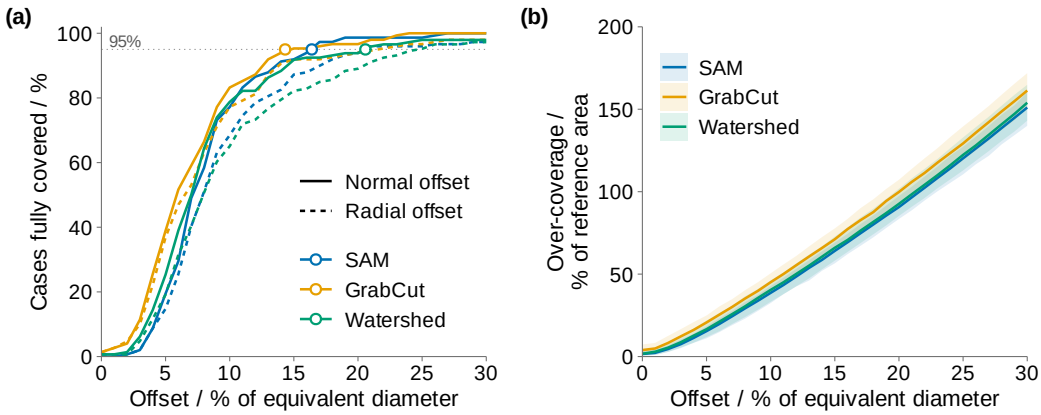

Figure 12. Perilesional offset (all valid cases; $n = 149$, 149 and 146 for SAM, GrabCut and watershed). (**a**) Percentage of cases in which the patch enlarged by a given offset covers the reference entirely, for the normal offset of Equation (15) (solid lines) and for the radial offset of Algorithm 1 (dashed lines); the offset is expressed as a percentage of the equivalent diameter of the reference; circles mark $\delta^*(0.95)$ and the dotted line the 95% level. (**b**) Cost of the margin: median over-coverage produced by a normal offset, as a percentage of the reference area (band: interquartile range).

Table 6. Perilesional offset sized on the under-coverage, per reference-seeded back-end (all valid cases). Normal offset: μ/D, margin needed to cover the reference fully (Equation (15)) relative to the equivalent diameter D of the reference, as median [IQR]; δ^*, its 95th percentile $\delta^*(0.95)$ (Equation (16)); percentage of cases in which the patch enlarged by δ^* covers the reference entirely, in-sample and in five-fold cross-validation (mean [2.5th–97.5th percentile] over 2000 repetitions). Radial offset: the same percentage when δ^* is applied radially from the centroid, as in Algorithm 1, and the radial offset δ^*_{rad} needed to reach 95%. Over-coverage: median excess area on healthy skin at δ^*, as a percentage of the reference area. δ (mm): δ^* converted into millimetres for assumed equivalent diameters of 10/20/40 mm.

Back-End	n	Normal Offset				Radial Offset		Over-Coverage (%)	δ (mm)
		μ/D (%)	δ^* (%)	In-Sample (%)	Cross-Validated (%)	Coverage (%)	δ^*_{rad} (%)		
SAM	149	7.2 [5.5–9.2]	16.4	94.6	94.3 [92.6–95.3]	88.6	20.6	72	1.6/3.3/6.6
GrabCut	149	5.9 [4.0–8.8]	14.3	94.6	94.2 [92.6–95.3]	91.9	21.4	68	1.4/2.9/5.7
Watershed	146	7.0 [4.9–9.2]	20.6	94.5	94.2 [92.5–95.2]	89.7	24.5	96	2.1/4.1/8.2

The offset is dimensionless because the images are uncalibrated; its value in millimetres follows from the size of the wound. As an illustration, for wounds of 10, 20 and 40 mm equivalent diameter the SAM offset corresponds to 1.6, 3.3 and 6.6 mm (Table 6); these are conversions under an assumed wound size, not measurements on calibrated clinical images. The relative margin decreased with the size of the wound in the image (Spearman $\rho = -0.48$, -0.41 and -0.46 between μ_i/D_i and D_i), so a single relative offset is governed by the wounds that are small in the image and is conservative for the large ones.

The margin increases the over-coverage. At $\delta^*(0.95)$ the patch extended over healthy skin by a median of 72%, 68% and 96% of the reference area (Figure 12b), because the offset is sized on the largest inward deviation along the contour and not on the average one.

Finally, the offset of Algorithm 1 is radial, whereas Equation (15) defines a normal offset. Applied radially, the same $\delta^*(0.95)$ fully covered 88.6%, 91.9% and 89.7% of the cases instead of 94.5% to 94.6%, and reaching 95% required a radial offset of 20.6%, 21.4% and 24.5% of the equivalent diameter (Figure 12a). The radial construction therefore provides less than the intended margin on elongated and concave contours, and a normal offset is needed for the rule to hold.

5. Discussion

IMTOP turns a single calibrated RGB photograph into a watertight, printable patient-specific patch geometry without depth sensing. Its accuracy was evaluated on fabricated planar phantoms of known geometry. For the clinical images, for which no physical ground-truth wound geometry was available, accuracy was assessed through the total area error propagated along the processing chain. The discussion is organized on three levels: the technical performance of the image-to-geometry chain (Section 5.1), the accuracy of fabrication (Section 5.2) and the clinical applicability of the result (Section 5.3); a comparison with the existing routes and the limitations of the study follow.

5.1. Technical Performance

RQ1 asked how the fabricated-patch error distributes across the segmentation, scale and printing stages, in magnitude and in variance. On the evaluated data printing was negligible, with less than 0.1% of the variance, and the budget was governed by segmentation and scale, which were of the same order: a standard uncertainty of 7.1% to 13.0% for the measured segmentation term against 7.3% for the simulated scale term. With all cases retained, the two terms contributed almost equally for SAM and segmentation was the larger contributor for GrabCut and watershed. This order depends on the handling of the

upper tail and on the calibration conditions: after the exclusion of the upper-tail cases the scale term is the larger one for SAM and GrabCut, and the order varies with the assumed placement noise and reference length through σ/L. With a reference of 500 px instead of 117 px, or with a placement noise of 1 px, the scale term would fall below 2.5% and segmentation would dominate for every back-end. The reference in the photograph should therefore be as long as practical, an indication that follows directly from Equation (17). The segmentation term also contains a bias, about −6% for SAM and watershed, which Equation (13) keeps apart from the dispersion: a systematic under-segmentation of this kind can be corrected, by the perilesional offset or by a calibration of the back-end, whereas the dispersion cannot.

That the joint photograph-to-geometry error is small on phantoms with sharp, high-contrast borders (0.32%) while the segmentation term is large on clinical images is consistent with an origin of the latter in the ambiguity of the wound boundary and not in the segmentation algorithms or the scale recovery, although the two experiments differ in reference (nominal CAD against a single operator) and in imaging conditions, and although a single-operator reference cannot separate border ambiguity from the systematic tendency of one tracer (Section 5.5). Stage-wise error propagation of this kind follows the GUM [15] and the dimensional-tolerance studies of FDM by Lieneke et al. [16] and Abas et al. [39]; to our knowledge it has not been applied to image-to-print wound dressings.

RQ2 asked whether mask overlap suffices to predict the fabricated-patch error. It does not, and part of the reason is analytic: for a given Dice the magnitude of the area disagreement is bounded but its sign is not, so two masks with equal overlap can leave the wound exposed or extend onto healthy skin by the same amount. Empirically, Dice was correlated with but did not determine the patch area error (Spearman $\rho = -0.68$), the error among the cases with Dice ≥ 0.95 spanned most of the feasible range, and in both the primary and the sensitivity analysis one pair of back-ends was ranked differently by overlap and by patch area error. Region overlap is therefore necessary but insufficient: the sign of the disagreement, captured by the coverage split, and its localization, captured by the boundary distance, are needed to characterize the fabricated object, since a scalar area error is itself blind to a displaced mask of equal area. This qualifies the common practice of ranking wound and DFU segmentation by overlap alone, from Goyal et al. [11] and Wang et al. [10] to the DFU segmentation challenge [32] and self-configuring nnU-Net [31], and is consistent with the insensitivity of Dice to boundary placement noted by Huttenlocher et al. [38].

All of the above refers to reference-seeded segmentation. The three back-ends refine a contour drawn by the operator and are scored against that same contour, so their agreement is an upper bound for an unseeded workflow: the two classical methods are confined to a band around the reference by construction, and SAM is limited by a bounding box taken from it. Their close agreement in overlap, with SAM [34] and GrabCut [27] differing little, is thus partly a consequence of the shared seed. The reference-free check quantifies the difference: without a seed, the reference-free mode of the tool, designed for high-contrast scenes, reached a Dice of at least 0.90 on 8% of the images (Section 4.1). On clinical photographs the operator seed is therefore part of the workflow, and whether the choice of segmenter matters less than the wound-border ambiguity common to all of them remains to be assessed by a comparison of unseeded methods.

5.2. Fabrication Accuracy

For the tested set-up, rigid and planar PLA parts printed by FDM, the in-plane printing error (0.69% in area, 0.41% in linear dimension) was an order of magnitude below the segmentation and scale terms. The conclusion that printing is negligible is specific to

this process: an extruded hydrogel dressing is expected to deviate more, and the printing term will be measured again on the final ink.

The printed thickness deviated by +7.4%. Thickness is a user-set parameter and not a reconstructed quantity, which is why it is excluded from the in-plane budget. For a drug-loaded hydrogel, however, the dispensed volume is the product of area and thickness, so the relative volume error is, to first order, the sum of the area and thickness errors: a deviation of this size would outweigh the in-plane printing error and is functionally relevant to local dosing. A volumetric budget, with the thickness measured on the dressing material, is required before any statement on dosing.

Experiment B shows that, on a known planar object and with the calibration points placed on the real photograph, the photograph-to-geometry chain reproduced the nominal geometry within 0.32% in area. A closed end-to-end test, in which the reconstructed contour is printed and measured again, is left to future work on phantoms; on a real wound such a test is not possible, because no physical ground truth is available.

5.3. Clinical Applicability

RQ3 asked how the residual segmentation uncertainty should set the perilesional margin. Because the disagreement is clinically asymmetric, under-coverage of the wound being the unsafe error and exceeding over-coverage for SAM and watershed, the offset is sized from the empirical distribution of the margins needed for full coverage, as their per-back-end 95th percentile. The resulting offsets, 14% to 21% of the equivalent diameter (about 3 to 4 mm for a wound of 20 mm), covered the wound in about 94% of the held-out cases in cross-validation, so within this dataset the rule is not an artefact of in-sample fitting. Its clinical use rests on three conditions (Section 5.5): the offset is relative to the outline of a single operator, its value in millimetres is a conversion under an assumed wound size, and the rule was validated within one dataset; a prospective test on calibrated images from a different centre is the next step. In decision-support terms, it converts an uncertainty estimate into a design parameter under an explicit, asymmetric cost, while the error budget attributes the residual error to the stage responsible for it; the rule is complementary to the assessment and monitoring step that precedes fabrication [24].

No established clinical threshold for acceptable patch-area error exists, whereas the literature reports tolerance ranges for tracking accuracy in augmented surgery, which allow a measured error to be judged directly [45]. The estimated error, a median of 7% to 8% and a 95th percentile of 22% to 32% in area, is therefore reported without a judgement of clinical adequacy, and the exceedance fractions of Figure 9b are given so that a tolerance of choice can be applied. Taken together, the results support a proof of concept of the geometric chain; clinical accuracy remains to be demonstrated.

5.4. Comparison with Existing Routes

Set against the existing routes, IMTOP is a deliberate trade-off, not a strict improvement. Image-based planimetry recovers wound geometry from a single photograph but stops at a measurement: Foltynski reports area [14], Yap et al. standardize smartphone acquisition [19], and Kim et al. predict healing [21], none producing a fabricable object, whereas IMTOP carries the same depth-free image through to a printable patch. At the other end, Chae et al. [9] also reach a printed DFU patch, but from a reconstructed RGB-D surface: their depth capture recovers the true wound-bed relief that our planar assumption gives up, while our single-photograph input removes their depth sensor. Filko et al. [22,23] likewise reconstruct and measure the wound from a robot-mounted RGB-D rig, validating on the reconstructed surface and not, as here, on the fabricated part. Cavazzana et al. [24], from the same group, segment and analyse the ulcer in three dimensions for objective

assessment; IMTOP is complementary to this, the fabrication step that can follow such an assessment, not an alternative to it.

5.5. Limitations and Future Work

Several limitations bound these results.

Single-operator reference. The reference was delineated by one operator with a geometric and not a clinical background, and no delineation by a second annotator or by a clinician could be obtained for this study. This may influence the results in three ways. First, the reference is not the true wound boundary: the segmentation term measures the agreement with one outline, and the error with respect to the true boundary, which also contains the operator-versus-truth term, is larger than budgeted. Second, a single tracer may place the margin with a systematic shift, for example on a gradual transition, and such a shift appears here as a bias: the under-segmentation of 5% to 6% found for SAM and watershed, and with it the asymmetry of the coverage and the size of the offset, may partly reflect the tracer and not the back-ends. Third, because the segmentation term cannot separate the intrinsic ambiguity of the wound border from the systematic tendency of one tracer, the attribution of the segmentation term to border ambiguity is provisional until multi-annotator, clinician-validated references are available.

Reference-seeded segmentation. The back-ends are seeded from the reference and scored against it, so the reported agreement is an upper bound and is not the accuracy of automatic segmentation. Run without a seed, the reference-free mode of the tool reached a Dice of at least 0.90 on 8% of the clinical images (Section 4.1); it was designed for phantoms, and no unseeded method designed for wounds was evaluated.

Uncalibrated clinical images. The clinical images contain no physical reference. The clinical distances are therefore in pixels, the offset is dimensionless and its value in millimetres is a conversion; neither the scale term nor the offset was measured on calibrated clinical data.

Assumed calibration noise. The scale term is simulated with an assumed placement noise of 3 px on a reference of 117 px. The repeatability of point placement by an operator was not measured, so the term is a scenario and not a measurement; Equation (17) and Figure 8b allow it to be rescaled to any measured value. The only measured evidence is indirect: the 0.80% dimensional error of Experiment B was obtained with calibration points placed on the phantom photographs and includes the realized placement error of that setting.

Partial budget. Besides the operator-versus-true-boundary term, the budget omits any departure from planarity. The pipeline assumes an approximately fronto-parallel view and the phantoms are planar, so the 0.32% phantom area error is a planar best case and does not bound the error on a curved foot. As an order of magnitude, under a weak-perspective approximation a wound plane tilted by an angle θ with respect to the image plane, with the reference not foreshortened, is reduced in area by the factor $\cos\theta$, that is by 1.5% at 10° and by 6.0% at 20°. The three terms are also treated as uncorrelated, which holds by construction here and may not hold in practice.

Fabrication. Printing was measured on PLA phantoms made by FDM and not on the dressing material, so the conclusion that printing is negligible is specific to that set-up, and the +7.4% thickness deviation is relevant to the dispensed volume (Section 5.2).

No end-to-end validation. The combined error is propagated from separate experiments. No object was carried through acquisition, calibration, segmentation, STL generation, printing and final measurement, and no clinical case was validated end to end.

Outlier handling. After the per-back-end Tukey exclusion used in the first version of this work, the means, the comparisons for two pairs of back-ends and the variance ranking differ from the primary results (Appendix A); only the findings common to the two analyses are retained as conclusions.

Offset. The sizing rule was cross-validated within one dataset only. The offset implemented in Algorithm 1 is radial and falls short of the normal margin on elongated or concave wounds (Section 4.5). Full coverage at the 95% level has a large cost in over-coverage.

Future work will (i) add multi-annotator, clinician-validated references; (ii) measure the repeatability of calibration-point placement to ground σ, with accuracy and precision estimated under controlled influence factors as in our metrological assessment of augmented-reality tooltip tracking [46]; (iii) acquire calibrated clinical images, fix the offset in millimetres, replace the radial offset with a normal one and validate the rule prospectively; (iv) bound the planarity error on curved or obliquely viewed targets; (v) run a closed end-to-end test on phantoms, in which the reconstructed contour is printed and measured, and develop and evaluate an unseeded prompt suited to clinical images; (vi) incorporate depth where available without sacrificing the frugal default; and (vii) integrate the patch geometry with the printable hydrogel ink under development within the same project, with the printing and thickness terms measured on that material, toward a dressing tested on the wound.

6. Conclusions

A single RGB photograph with a physical scale reference can be turned, without depth sensing, into a watertight printable patch geometry for superficial DFUs, and the area error of the patch can be budgeted stage by stage. On PLA phantoms, reconstruction and FDM printing were accurate to within 1% in area. On 149 public clinical images, with back-ends seeded from a single-operator reference and a simulated scale term, the median absolute total area error estimated by propagation was 7% to 8%. Mask overlap did not determine the patch error, and a normal offset of 14% to 21% of the wound equivalent diameter covered the wound in about 94% of cross-validated cases. These results support the proof of concept; clinician-validated references, calibrated images and tests on the dressing material are the next steps toward clinical use.

Author Contributions: Conceptualization, F.S. and R.L.; methodology, F.S. and L.U.; software, F.S. and G.A.; validation, F.S. and S.M.; formal analysis, F.S. and J.S.; investigation, F.S., L.U., S.M. and E.V.; resources, S.M. and E.V.; data curation, F.S. and G.A.; writing—original draft preparation, F.S. and L.U.; writing—review and editing, J.S., G.A., R.L., S.M. and E.V.; visualization, F.S.; supervision, S.M. and E.V.; project administration, R.L., J.S. and L.U.; funding acquisition, R.L., J.S. and L.U. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Politecnico di Torino and Fondazione CRT through the Call for Proposals for Collaborative Research, grant number 10609 (PERSEUS project: Patient-specific engineering of smart responsive wound dressings for enhanced diabetic foot ulcer management through AI- and XR-technologies).

Institutional Review Board Statement: Not applicable. The study did not involve new human participants: the clinical images were drawn from publicly available, anonymized datasets released by their creators for research use, and no new patient data were collected.

Informed Consent Statement: Not applicable.

Data Availability Statement: The IMTOP application, the analysis scripts, the phantom CAD models, the STL files and the dimensional measurements will be made publicly available on GitHub (https://github.com/federicosalerno-phd/IMTOP (accessed on 9 October 2026)) and archived on Zenodo with a permanent DOI upon publication. The clinical images are drawn from two public datasets, the Lower Limb and Feet Wound Image Dataset [41] (DOI 10.17632/hsj38fwnvr.3) and a public DFU image dataset [42], and are not redistributed here. The SAM ViT-B checkpoint (`sam_vit_b_01ec64.pth`) is distributed by its original authors [34] and is not redistributed here.

Acknowledgments: This work was financed by PERSEUS project (Patient-specific engineering of smart responsive wound dressings for enhanced diabetic foot ulcer management through AI- and XR-technologies) with the support of Politecnico di Torino and Fondazione CRT.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

3D	Three-dimensional
ASD	Average symmetric surface distance
CAD	Computer-aided design
CNN	Convolutional neural network
CPU	Central processing unit
DFU	Diabetic foot ulcer
FDM	Fused deposition modelling
GT	Ground truth (the manual reference, in figures)
GUM	Guide to the Expression of Uncertainty in Measurement
HTML	Hypertext Markup Language
IoU	Intersection over union
IQR	Interquartile range
JSON	JavaScript Object Notation
LOESS	Locally estimated scatterplot smoothing
PLA	Polylactic acid
RGB	Red, green, blue (colour image)
RGB-D	Colour (red, green, blue) and depth
RQ	Research question
SAM	Segment Anything Model
SD	Standard deviation
STL	Stereolithography (mesh file format)
ViT-B	Vision Transformer, base variant
WS	Watershed (in figures)

Appendix A. Sensitivity Analysis After the Exclusion of Outliers

In the first version of this work the per-image outliers of the patch area error were removed before the analysis by Tukey's rule (values beyond 1.5 IQR from the quartiles), separately for each back-end. Only upper-tail cases were affected: 5, 14 and 10 cases for SAM, GrabCut and watershed, above fences of 20.9%, 19.7% and 24.4%. Tables A1 and A2 repeat the descriptive statistics and the error budget on the resulting sets, and the right-hand columns of Table 3 repeat the tests on the 125 images retained for the three back-ends.

Compared with the primary analysis, the medians changed by at most 0.5 percentage points for the patch area error and by at most 0.4 px for the Hausdorff distance, whereas the means of the patch area error decreased by 10% to 35% and its SDs by 22% to 55%. In the budget the standard uncertainty of the segmentation term fell to 5.8%, 6.8% and 7.3%, below the simulated scale term for SAM and GrabCut, so that the scale term became the larger variance contributor for these two back-ends (62% and 53%) and the two terms contributed equally for watershed; the segmentation bias of GrabCut changed sign (−0.7% against +1.8%). The median absolute total area error was 7.1%, 6.3% and 7.3%, against 7.4%, 7.3% and 8.0% in the primary analysis. Under-coverage still exceeded over-coverage for SAM and watershed ($n = 144$ and 136, $p < 0.001$, $r = 0.71$ and 0.60) and not for GrabCut ($n = 135$, $p = 0.071$, $r = 0.16$).

Table A1. Agreement of the three reference-seeded back-ends with the single-operator reference after the per-back-end Tukey exclusion on the patch area error (sensitivity analysis; same quantities as Table 2). *n*: number of cases retained. IQR: interquartile range; SD: standard deviation.

Back-End	*n*	Statistic	Dice	IoU	Patch Area Error (%)	Hausdorff (px)
SAM	144	Median [IQR]	0.956 [0.936–0.968]	0.915 [0.880–0.937]	5.2 [3.0–9.5]	4.6 [3.2–8.4]
		Mean ± SD	0.950 ± 0.025	0.907 ± 0.044	6.4 ± 4.5	8.2 ± 10.4
GrabCut	135	Median [IQR]	0.957 [0.941–0.973]	0.917 [0.888–0.947]	3.4 [1.4–7.0]	5.4 [3.6–7.7]
		Mean ± SD	0.953 ± 0.024	0.912 ± 0.042	5.0 ± 4.7	7.7 ± 8.1
Watershed	136	Median [IQR]	0.950 [0.929–0.963]	0.905 [0.868–0.929]	5.7 [2.7–9.7]	6.6 [4.5–9.0]
		Mean ± SD	0.942 ± 0.029	0.892 ± 0.051	6.8 ± 5.3	9.1 ± 8.6

Table A2. Error budget of the total area error of the patch after the per-back-end Tukey exclusion (sensitivity analysis; same quantities and assumptions as Table 5).

Quantity	SAM	GrabCut	Watershed
Segmentation term (measured, Experiment A)			
Cases, n	144	135	136
Bias B_{Seg} (%)	−5.3	−0.7	−4.6
Standard uncertainty E_{Seg} (%)	5.8	6.8	7.3
Median $\lvert\varepsilon_{\mathrm{Seg}}\rvert$ (%)	5.2	3.4	5.7
Scale term (simulated, $\sigma = 3$ px, $L = 117$ px)			
Bias B_{Scale} (%)		+0.3	
Standard uncertainty E_{Scale} (%)		7.3	
Median $\lvert\varepsilon_{\mathrm{Scale}}\rvert$ (%)		4.9	
Printing term (PLA phantoms, Experiment C)			
Bias B_{Pr} (%)		+0.69	
Standard uncertainty E_{Pr} (%, assumed)		0.10	
First-order combination (Equations (13) and (14))			
Bias B_{patch} (%)	−4.3	+0.3	−3.6
Combined standard uncertainty E_{patch} (%)	9.3	10.0	10.3
Variance share of segmentation (%)	38	47	50
Variance share of scale (%)	62	53	50
Variance share of printing (%)		<0.1	
Monte Carlo propagation (Equation (12), 20,000 draws)			
Mean of $\varepsilon_{\mathrm{patch}}$ (%)	−4.3	+0.2	−3.7
Standard deviation of $\varepsilon_{\mathrm{patch}}$ (%)	9.1	9.9	10.2
Median $\lvert\varepsilon_{\mathrm{patch}}\rvert$ (%)	7.1	6.3	7.3
95th percentile of $\lvert\varepsilon_{\mathrm{patch}}\rvert$ (%)	19.3	19.9	20.9
Draws with $\lvert\varepsilon_{\mathrm{patch}}\rvert > 5\%$ (%)	64	59	64
Draws with $\lvert\varepsilon_{\mathrm{patch}}\rvert > 10\%$ (%)	34	30	36

References

1. Armstrong, D.G.; Tan, T.W.; Boulton, A.J.M.; Bus, S.A. Diabetic foot ulcers: A review. *JAMA* **2023**, *330*, 62–75. https://doi.org/10.1001/jama.2023.10578.
2. McDermott, K.; Fang, M.; Boulton, A.J.M.; Selvin, E.; Hicks, C.W. Etiology, epidemiology, and disparities in the burden of diabetic foot ulcers. *Diabetes Care* **2023**, *46*, 209–221. https://doi.org/10.2337/dci22-0043.
3. Chen, L.; Sun, S.; Gao, Y.; Ran, X. Global mortality of diabetic foot ulcer: A systematic review and meta-analysis of observational studies. *Diabetes Obes. Metab.* **2023**, *25*, 36–45. https://doi.org/10.1111/dom.14840.
4. Muwaffak, Z.; Goyanes, A.; Clark, V.; Basit, A.W.; Hilton, S.T.; Gaisford, S. Patient-specific 3D scanned and 3D printed antimicrobial polycaprolactone wound dressings. *Int. J. Pharm.* **2017**, *527*, 161–170. https://doi.org/10.1016/j.ijpharm.2017.04.077.

5. Winter, G.D. Formation of the scab and the rate of epithelization of superficial wounds in the skin of the young domestic pig. *Nature* **1962**, *193*, 293–294. https://doi.org/10.1038/193293a0.
6. Laurano, R.; Boffito, M.; Ciardelli, G.; Chiono, V. Wound dressing products: A translational investigation from the bench to the market. *Eng. Regen.* **2022**, *3*, 182–200. https://doi.org/10.1016/j.engreg.2022.04.002.
7. Tsegay, F.; Hisham, M.; Elsherif, M.; Schiffer, A.; Butt, H. 3D printing of pH indicator auxetic hydrogel skin wound dressing. *Molecules* **2023**, *28*, 1339. https://doi.org/10.3390/molecules28031339.
8. Murphy, S.V.; Atala, A. 3D bioprinting of tissues and organs. *Nat. Biotechnol.* **2014**, *32*, 773–785. https://doi.org/10.1038/nbt.2958.
9. Chae, H.J.; Lee, S.; Son, H.; Han, S.; Lim, T. Generating 3D bio-printable patches using wound segmentation and reconstruction to treat diabetic foot ulcers. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA, 18–24 June 2022; pp. 2539–2549.
10. Wang, C.; Anisuzzaman, D.M.; Williamson, V.; Dhar, M.K.; Rostami, B.; Niezgoda, J.; Gopalakrishnan, S.; Yu, Z. Fully automatic wound segmentation with deep convolutional neural networks. *Sci. Rep.* **2020**, *10*, 21897. https://doi.org/10.1038/s41598-020-78799-w.
11. Goyal, M.; Yap, M.H.; Reeves, N.D.; Rajbhandari, S.; Spragg, J. Fully convolutional networks for diabetic foot ulcer segmentation. In Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; pp. 618–623. https://doi.org/10.1109/SMC.2017.8122675.
12. Salerno, F.; Moos, S.; Ulrich, L.; Novaresio, A.; Vezzetti, E. A Methodology for the Dimensional and Mechanical Analysis of Surgical Guides. In *Proceedings of the International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering (ADM 2023), Florence, Italy, 6–8 September 2023*; Lecture Notes in Mechanical Engineering; Springer: Cham, Switzerland, 2024; pp. 184–193. https://doi.org/10.1007/978-3-031-52075-4_22.
13. Armstrong, D.G.; Lavery, L.A.; Harkless, L.B. Validation of a diabetic wound classification system: The contribution of depth, infection, and ischemia to risk of amputation. *Diabetes Care* **1998**, *21*, 855–859. https://doi.org/10.2337/diacare.21.5.855.
14. Foltynski, P. Ways to increase precision and accuracy of wound area measurement using smart devices: Advanced app Planimator. *PLoS ONE* **2018**, *13*, e0192485. https://doi.org/10.1371/journal.pone.0192485.
15. JCGM. *Evaluation of Measurement Data—Guide to the Expression of Uncertainty in Measurement (GUM)*; Technical Report JCGM 100:2008; Joint Committee for Guides in Metrology (JCGM), BIPM: Sèvres, France, 2008.
16. Lieneke, T.; Denzer, V.; Adam, G.A.O.; Zimmer, D. Dimensional tolerances for additive manufacturing: Experimental investigation for fused deposition modeling. *Procedia CIRP* **2016**, *43*, 286–291. https://doi.org/10.1016/j.procir.2016.02.361.
17. Hakimi, N.; Cheng, R.; Leng, L.; Sotoudehfar, M.; Ba, P.Q.; Bakhtyar, N.; Amini-Nik, S.; Jeschke, M.G.; Günther, A. Handheld skin printer: In situ formation of planar biomaterials and tissues. *Lab Chip* **2018**, *18*, 1440–1451. https://doi.org/10.1039/C7LC01236E.
18. Albanna, M.; Binder, K.W.; Murphy, S.V.; Kim, J.; Qasem, S.A.; Zhao, W.; Tan, J.; El-Amin, I.B.; Dice, D.D.; Marco, R.A.; et al. In situ bioprinting of autologous skin cells accelerates wound healing of extensive excisional full-thickness wounds. *Sci. Rep.* **2019**, *9*, 1856. https://doi.org/10.1038/s41598-018-38366-w.
19. Yap, M.H.; Chatwin, K.E.; Ng, C.C.; Abbott, C.A.; Bowling, F.L.; Rajbhandari, S.; Boulton, A.J.M.; Reeves, N.D. A new mobile application for standardizing diabetic foot images. *J. Diabetes Sci. Technol.* **2018**, *12*, 169–173. https://doi.org/10.1177/1932296817713761.
20. Ramachandram, D.; Ramirez-GarciaLuna, J.L.; Fraser, R.D.J.; Martínez-Jiménez, M.A.; Arriaga-Caballero, J.E.; Allport, J. Fully automated wound tissue segmentation using deep learning on mobile devices: Cohort study. *JMIR MHealth UHealth* **2022**, *10*, e36977. https://doi.org/10.2196/36977.
21. Kim, R.B.; Gryak, J.; Mishra, A.; Cui, C.; Soroushmehr, S.M.R.; Najarian, K.; Wrobel, J.S. Utilization of smartphone and tablet camera photographs to predict healing of diabetes-related foot ulcers. *Comput. Biol. Med.* **2020**, *126*, 104042. https://doi.org/10.1016/j.compbiomed.2020.104042.
22. Filko, D.; Marijanović, D.; Nyarko, E.K. Automatic robot-driven 3D reconstruction system for chronic wounds. *Sensors* **2021**, *21*, 8308. https://doi.org/10.3390/s21248308.
23. Filko, D.; Nyarko, E.K. 2D/3D wound segmentation and measurement based on a robot-driven reconstruction system. *Sensors* **2023**, *23*, 3298. https://doi.org/10.3390/s23063298.
24. Cavazzana, R.; Faccia, A.; Cavallaro, A.; Giuranno, M.; Becchi, S.; Innocente, C.; Marullo, G.; Ricci, E.; Secco, J.; Vezzetti, E.; et al. Enhancing clinical assessment of skin ulcers with automated and objective convolutional neural network-based segmentation and 3D analysis. *Appl. Sci.* **2025**, *15*, 833. https://doi.org/10.3390/app15020833.
25. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man. Cybern.* **1979**, *9*, 62–66. https://doi.org/10.1109/TSMC.1979.4310076.
26. Meyer, F. Topographic distance and watershed lines. *Signal Process.* **1994**, *38*, 113–125. https://doi.org/10.1016/0165-1684(94)90060-4.
27. Rother, C.; Kolmogorov, V.; Blake, A. "GrabCut": Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. https://doi.org/10.1145/1015706.1015720.

28. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 3431–3440. https://doi.org/10.1109/CVPR.2015.7298965.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 5–9 October 2015*; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2015; Volume 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28.
30. Mahbod, A.; Schaefer, G.; Ecker, R.; Ellinger, I. Automatic foot ulcer segmentation using an ensemble of convolutional neural networks. In Proceedings of the 2022 26th International Conference on Pattern Recognition (ICPR), Montreal, QC, Canada, 21–25 August 2022; pp. 4358–4364. https://doi.org/10.1109/ICPR56361.2022.9956253.
31. Isensee, F.; Jaeger, P.F.; Kohl, S.A.A.; Petersen, J.; Maier-Hein, K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **2021**, *18*, 203–211. https://doi.org/10.1038/s41592-020-01008-z.
32. Yap, M.H.; Cassidy, B.; Byra, M.; Liao, T.Y.; Yi, H.; Galdran, A.; Chen, Y.H.; Brüngel, R.; Koitka, S.; Friedrich, C.M.; et al. Diabetic foot ulcers segmentation challenge report: Benchmark and analysis. *Med. Image Anal.* **2024**, *94*, 103153. https://doi.org/10.1016/j.media.2024.103153.
33. Wang, C.; Mahbod, A.; Ellinger, I.; Galdran, A.; Gopalakrishnan, S.; Niezgoda, J.; Yu, Z. FUSeg: The foot ulcer segmentation challenge. *Information* **2024**, *15*, 140. https://doi.org/10.3390/info15030140.
34. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment Anything. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 2–6 October 2023; pp. 3992–4003. https://doi.org/10.1109/ICCV51070.2023.00371.
35. Ma, J.; He, Y.; Li, F.; Han, L.; You, C.; Wang, B. Segment anything in medical images. *Nat. Commun.* **2024**, *15*, 654. https://doi.org/10.1038/s41467-024-44824-z.
36. Ravi, N.; Gabeur, V.; Hu, Y.T.; Hu, R.; Ryali, C.; Ma, T.; Khedr, H.; Rädle, R.; Rolland, C.; Gustafson, L.; et al. SAM 2: Segment anything in images and videos. *arXiv* **2024**, arXiv:2408.00714. https://doi.org/10.48550/arXiv.2408.00714.
37. Dice, L.R. Measures of the amount of ecologic association between species. *Ecology* **1945**, *26*, 297–302. https://doi.org/10.2307/1932409.
38. Huttenlocher, D.P.; Klanderman, G.A.; Rucklidge, W.J. Comparing images using the Hausdorff distance. *IEEE Trans. Pattern Anal. Mach. Intell.* **1993**, *15*, 850–863. https://doi.org/10.1109/34.232073.
39. Abas, M.; Habib, T.; Noor, S.; Salah, B.; Zimon, D. Parametric investigation and optimization to study the effect of process parameters on the dimensional deviation of fused deposition modeling of 3D printed parts. *Polymers* **2022**, *14*, 3667. https://doi.org/10.3390/polym14173667.
40. JCGM. *Evaluation of Measurement Data—Supplement 1 to the "Guide to the Expression of Uncertainty in Measurement"—Propagation of Distributions Using a Monte Carlo Method*; Technical Report JCGM 101:2008; Joint Committee for Guides in Metrology (JCGM), BIPM: Sèvres, France, 2008.
41. Islam, M.M. *Lower Limb and Feet Wound Image Dataset for Medical Analysis*, V3; Mendeley Data: Oxford, UK, 2026. https://doi.org/10.17632/hsj38fwnvr.3.
42. laithjj. Diabetic Foot Ulcer (DFU) Dataset. Available online: https://www.kaggle.com/datasets/laithjj/diabetic-foot-ulcer-dfu (accessed on 24 August 2026).
43. Tomczak, M.; Tomczak, E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. *Trends Sport Sci.* **2014**, *21*, 19–25.
44. Fritz, C.O.; Morris, P.E.; Richler, J.J. Effect size estimates: Current use, calculations, and interpretation. *J. Exp. Psychol. Gen.* **2011**, *141*, 2–18. https://doi.org/10.1037/a0024338. Correction in *J. Exp. Psychol. Gen.* **2012**, *141*, 30. https://doi.org/10.1037/a0026092.
45. Salerno, F.; Contenti, A.; Ulrich, L.; Moos, S.; Fiandrotti, A.; Vezzetti, E. A Low-Cost Benchmarking Platform for Tracking Systems to Enable Augmented Surgery Tools in Clinical Practice. *SN Comput. Sci.* **2026**, *7*, 437. https://doi.org/10.1007/s42979-026-05071-9.
46. Salerno, F.; Ulrich, L.; Maculotti, G.; Moos, S.; Genta, G.; Vezzetti, E.; Galetto, M. A metrological approach for Augmented Reality tooltip tracking assessment. *Comput. Ind.* **2026**, *175*, 104430. https://doi.org/10.1016/j.compind.2025.104430.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.