

Leveraging Deep Neural Networks for Automatic and Standardised Wound Image Acquisition

Ana Filipa Sampaio¹^a, Pedro Alves¹^b, Nuno Cardoso¹^c, Paulo Alves^{2,3}^d, Raquel Marques³^e,
Pedro Salgado⁴ and Maria João M. Vasconcelos¹^f

¹Fraunhofer Portugal AICOS, Porto, Portugal

²School of Nursing Department, Universidade Católica Portuguesa (UCP), Porto, Portugal

³Centre for Interdisciplinary Research in Health, Instituto de Ciências da Saúde, UCP, Porto, Portugal

⁴F3M Information Systems, S.A, Portugal

Keywords: Skin Wounds, Mobile Health, Object Detection, Deep Learning, Mobile Devices.

Abstract: Wound monitoring is a time-consuming and error-prone activity performed daily by healthcare professionals. Capturing wound images is crucial in the current clinical practice, though image inadequacy can undermine further assessments. To provide sufficient information for wound analysis, the images should also contain a minimal periwound area. This work proposes an automatic wound image acquisition methodology that exploits deep learning models to guarantee compliance with the mentioned adequacy requirements, using a marker as a metric reference. A RetinaNet model detects the wound and marker regions, further analysed by a post-processing module that validates if both structures are present and verifies that a periwound radius of 4 centimetres is included. This pipeline was integrated into a mobile application that processes the camera frames and automatically acquires the image once the adequacy requirements are met. The detection model achieved $mAP@.75IOU$ values of 0.39 and 0.95 for wound and marker detection, exhibiting a robust detection performance for varying acquisition conditions. Mobile tests demonstrated that the application is responsive, requiring 1.4 seconds on average to acquire an image. The robustness of this solution for real-time smartphone-based usage evidences its capability to standardise the acquisition of adequate wound images, providing a powerful tool for healthcare professionals.


1 INTRODUCTION


The prevalence of chronic wounds is high worldwide, being considered a major public health problem. Chronic wounds are those that do not progress through a normal and timely sequence of repair (Werdin et al., 2009) and can include pressure, venous, arterial, diabetic ulcers, and others. The study of (Martinengo et al., 2019) indicates a pooled prevalence of 2.21 per 1000 population for chronic wounds of mixed etiologies, and the prevalence for chronic leg ulcers was estimated at 1.51 per 1000 population. In the United States, chronic wounds affect the qual-


ity of life of nearly 2.5% of the total population, having a significant impact on healthcare (Sen, 2021). In Portugal, the prevalence is 0.84 per 1000 population, and patients with more than 80 years of age present a prevalence of 5.86 (Passadouro et al., 2016).


Wound healing is a complex process that involves several phases and proper monitoring of its evolution. The rise of the digital health sector can help healthcare professionals properly monitor wound evolution by storing clinical information in each monitorization, including wound details and images. According to the study of (Chen et al., 2020) current evidence suggests that telemedicine has similar efficacy and safety with conventional standard care of chronic wounds.


An essential step for proper wound documentation is wound image capturing (Mukherjee et al., 2017),(Anisuzzaman et al., 2022b); however, poor image quality can negatively influence later assessments. Many of the approaches reported in the literature for automatized wound monitoring address


^a <https://orcid.org/0000-0003-1770-4429>

^b <https://orcid.org/0000-0002-0372-4755>

^c <https://orcid.org/0000-0001-5736-7995>

^d <https://orcid.org/0000-0002-6348-3316>

^e <https://orcid.org/0000-0002-6701-3530>

^f <https://orcid.org/0000-0002-0634-7852>

the inadequate image quality using computer vision methodologies to remove acquisition artefacts and correct colour and perspective distortions in the captured images (Lu et al., 2017; Cui et al., 2019). Still, the application of pre-processing strategies may be insufficient for some tasks since it does not ensure that all the relevant structures needed for an adequate wound assessment (e.g. the periwound skin) are properly visible. Subsequently, efforts to enhance the image quality at the acquisition step have been conducted by either providing acquisition guidelines or using semi-automated acquisition processes.

Many of the attempts to automatise the acquisition of wound images resort to object detection models, namely convolutional neural networks (CNNs) to locate the wound bed in the captured image field. The work in (Goyal et al., 2018) first presented an algorithm to detect and locate diabetic foot ulcers in real-time. The authors trained a Faster R-CNN model with Inception-V2, obtaining a mean average precision (mAP) of 91.8%. In (Faria et al., 2020), a methodology for automated image acquisition of skin wounds via mobile devices is presented. The application of image focus validation combined with a detection deep neural network (DNN) for wound bed localisation (an SSDLite model with a MobileNetV2 backbone) achieved an mAP at a 0.50 intersection over union (IOU) threshold ($mAP@.50IOU$) of 86.46% and an inference speed of 119 milliseconds (ms) for smartphone usage, ensuring simultaneous image quality and adequacy in real-time. More recently, a similar wound localiser for mobile usage was presented in (Anisuzzaman et al., 2022a). Two detection DNNs, SSD with a VGG16 backbone and YOLOv3, were trained in a private dataset of 1010 images of venous, pressure and diabetic foot ulcers, achieving mAP values of 86.4% and 93.9%, respectively. In (Scabba et al., 2022), the authors trained a RetinaNet model with three different feature extractors, all pre-trained on the Common Objects in Context (COCO) dataset. The best performance was achieved with the MobileNet backbone, leading to a test $mAP@.50IOU$ of 0.762.

Despite relying mainly on private datasets for the models' development, most of these approaches also resorted to a public database, Medetec (Medetec, 2014), to expand the training set (Faria et al., 2020) and to assess the performance of their approach in other datasets (Anisuzzaman et al., 2022a). However, even being used in many of the aforementioned works, the Medetec database does not provide information regarding the wound region location in its images (its usage was enabled by private annotations).

The high detection performances reported by

these approaches highlight the potential of DNNs for the analysis of wound images. Yet, the majority of these works only use it to narrow the wound region for further processing by other algorithms, such as segmentation or classification models. The exception is the systems described in (Anisuzzaman et al., 2022a; Faria et al., 2020), which integrate the trained wound detectors in mobile wound acquisition applications. Still, their frameworks do not use detection information to analyse the adequacy of the captured images.

Thus, considering the importance of adequate wound images and the lack of solutions that incorporate this validation in the image acquisition step, this work presents an approach to standardise wound image acquisition by automating the process while guaranteeing wound detection and periwound skin. For this purpose, a new dataset of wound images with reference markers was constructed and used to develop a wound and marker detection model based on a RetinaNet architecture with a MobileNetV2 backbone. This model was used as the basis for an image adequacy validation module, responsible for ensuring that the acquired images are compliant with the requirements recommended by clinical experts: the presence of the wound and marker regions and inclusion of 4 centimetres (cm) of periwound skin in the captured image. A mobile application incorporating the proposed pipeline was implemented and tested on several images, enabling a real-time, automated and smooth process for image acquisition and assessment. Therefore, this work advances the previous state-of-the-art approaches by providing a detection model for simultaneous wound and reference marker localisation, taking advantage of its potentialities to propose a simple yet effective strategy for validating the adequacy of wound images, and integrating this pipeline in a user-friendly mobile solution that aims to facilitate monitoring procedures while ensuring the quality of the wound images acquired.

2 METHODOLOGY

The proposed approach to automate the wound image acquisition relies on the application of deep learning detection models for the correct localisation of the wound bed and the reference marker in the field of view, in association with a domain-adapted post-processing procedure to verify the presence of all the relevant structures - open wound, marker and periwound skin. This strategy ensures that all the information needed for a complete visual assessment of the wound status and the automated extraction of wound properties from the image are present in the

captured images, providing a useful tool for image acquisition both for manual clinical monitoring of wounds and for automated assessment solutions.

For the development of this methodology, a novel dataset - the Clinical Wound Support (CWS) dataset - was constructed, as described in section 2.1. The succeeding sections detail the hyperparameter tuning process used to find the most promising model (section 2.2.1), as well as the dataset preparation for the training of the detection models (section 2.2.2), the implementation of the adequacy validation module (section 2.3) and the integration of the whole acquisition pipeline in a mobile application (section 2.4).

2.1 Dataset Construction

The scarce datasets with reference markers available in the literature are based on complex and expensive markers and do not provide annotations regarding the marker regions (Yang et al., 2016). Therefore, a novel dataset comprising images of wounds and reference markers was constructed. The images were acquired by clinical professionals from nine Portuguese healthcare institutions using a smartphone application developed for this purpose using commercially available adhesive reference markers¹. These reference markers were selected due to their ready accessibility, ease of usage and standardised size (2×2 cm).

The dataset comprises monitorisations of different types of chronic wounds: diabetic foot ulcers, pressure ulcers and leg ulcers (venous, arterial and mixed), among others. Although that was not always possible, in most cases, the wound was followed up weekly over a 4-week period, resulting in a maximum of 5 monitorisations per wound to capture varying healing states of each type of wound. The resulting images were annotated in terms of wound and marker boundaries by three specialists in the field of wound care, each responsible for a distinct subset of wound images. The dataset distribution is presented in table 1, and annotated examples are depicted in fig. 1.

Table 1: Composition of the acquired CWS dataset, in terms of wounds, monitorisations and captured images.

Wound type	Pressure ulcer	Venous leg ulcer	Arterial leg ulcer	Leg ulcer of unknown etiology	Diabetic foot ulcer	Other ulcer types	Total
No. wounds	29	10	2	4	5	1	51
No. monitor.	61	36	7	18	17	6	145
No. images	65	52	8	20	17	6	168

2.2 Model Development

The wound and marker localisation module was developed using a RetinaNet detection architecture with

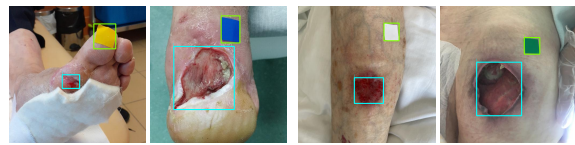


Figure 1: Illustrative examples of skin wound images and respective ground truth localisation (open wound in cyan and reference marker in green).

a MobileNetV2 backbone network. MobileNetV2 (Sandler et al., 2018) belongs to a family of CNNs optimised for mobile deployment through the application of lightweight filtering operations and an inverted residual structure, being able to provide competitive performances with a reduced number of parameters and a lower computational complexity. RetinaNet is a one-stage detector that combines off-the-shelf CNNs with a feature pyramid network to enable the detection of objects at multiple scales and uses a focal loss function to account for possible data imbalances during training (Lin et al., 2018). The TensorFlow Python implementation of the model was used, resorting to the Object Detection API for model training and optimisation. This model was selected considering the need for deployment in a mobile scenario, as it offers a good trade-off between detection performance and inference time while being easily exported to a smartphone-compatible format (TensorFlow Lite²). This was assessed based on the benchmarks in (Huang et al., 2017; Ten, 2021).

2.2.1 Model Optimisation

The detection model configuration was optimised using a random search hyperparameter tuning process, applied instead of grid search optimisation due to time constraints. Six different batch sizes and learning rate combinations were analysed. These two hyperparameters were selected for tuning due to their high impact on the training results. The corresponding values were randomly sampled from adequate ranges based on the recommendations of (Bengio, 2012). Batch sizes of 8 and 16 were considered, as these provide sufficient training instances in each learning iteration while complying with the dynamic memory constraints of the hardware used. The learning rate values were sampled from a logarithmically scaled range between 10^{-6} and 0.01; a constant learning rate was considered for all the iterations of each experiment. Also, in view of the impact of the input image size on the computational demands of the detection models and the other processing steps, two image sizes were assessed: 320×320 and 512×512 pixels. These hyperparameter combinations were tested for both input

¹<https://www.apli.com/en/product/18326>

²<https://www.tensorflow.org/lite>

Table 2: Hyperparameter values tested for the six experiments conducted to optimise the detection model.

Experiment	Batch size	Learning rate	Image size
1	8	7.56463E-03	320 / 512
2	8	1.12332E-05	320 / 512
3	16	5.59081E-04	320 / 512
4	8	6.57933E-05	320 / 512
5	8	1.14976E-04	320 / 512
6	16	7.56463E-03	320 / 512

sizes, resulting in the experiments shown in table 2.

A 3-fold cross-validation was followed for the hyperparameter tuning experiments, with details regarding the data preparation being provided in the next section. The evaluation of the tuning results was conducted on the validation subsets of each fold, considering performance metrics appropriate for object detection tasks - mAP and average recall (AR), computed for *IOU* thresholds of 0.50 and 0.75 in the case of mAP, and for a maximum number of 10 detections for AR. A higher focus was attributed to *mAP@.75IOU* because the higher *IOU* threshold imposed leads to a more representative metric of the model's ability to provide accurate regions of interest; *mAP@.50IOU* was evaluated for the state-of-the-art comparison (Goyal et al., 2018; Faria et al., 2020; Anisuzzaman et al., 2022a; Scebbba et al., 2022).

Each model was optimised by applying the Adam optimisation algorithm, using the Huber regression and focal cross-entropy functions for the localization and classification losses, respectively. Although each model was trained for 5000 epochs, during this process, its validation performance was periodically monitored, and the network parameters that yielded the optimal validation *mAP@.75IOU* were considered as final for the respective experiment to prevent overfitting. The network was initialised using the parameters resulting from pre-training in the public COCO dataset (Lin et al., 2014). All the non-mentioned training conditions and hyperparameters were set to default values provided with the model implementation, as specified in (Con, 2022).

2.2.2 Data Preparation

To develop the model, 75% of the dataset was used for training, with the remaining 25% being separated as test data for the final model assessment. For the hyperparameter tuning experiments, the training data was split into three distinct folds of train and validation sets. The division of the data instances in the several subsets was performed in a stratified manner, at the level of the wounds and not the images, to ensure

that all images of each wound were kept in the same subset and that the training and test subsets comprise representative examples of each wound type, body location and skin phototype. This process resulted in a distribution of 98 training images from 29 different wounds and 70 test images from 22 wounds. The images, already square as a result of their acquisition process, were resized according to the input size needed for the trained network. Geometric image transformations, in particular 90°, 180° and 270° rotations, and vertical and horizontal flips, were applied dynamically during training to augment the instances.

2.2.3 Final Model Training and Mobile Deployment

The final model configuration was identified as the one that resulted in the highest performance metrics. After the determination of the best experience from hyperparameter tuning, the detection model was re-trained using the whole training data considering the corresponding hyperparameter values, for a fixed number of epochs. The number of training epochs was established as the average number of epochs needed for the convergence of the models trained for each cross-validation fold of the selected experiment. The trained model parameters were then quantised and optimised for mobile deployment using the Tensorflow Lite framework. The best model of each input size was deployed in a mobile device and compared to assess its influence on the model's inference speed and its impact on the mobile acquisition workflow.

2.3 Image Adequacy Validation

The validation of the image adequacy was performed based on the outputs of the detection model by assessing if both the open wound and reference marker were detected. In addition to the open wound, the surrounding area, known as the periwound area, also contains important information regarding the state of the wound. Therefore, there is a need to ensure that the captured image also contains this area. Following experts' recommendation, 4 cm in all directions of the open wound border were considered.

Given the marker size of 2×2 cm, the largest side of the detected marker's bounding box was used as a reference, ensuring that, even if the marker is not perpendicular to the camera, the side closer to 2 cm will be used. Then, the verification is performed by ensuring that there is a distance of at least 2 times the reference is present around the four sides of the wound's bounding box.

2.4 Mobile Application

The methodology presented in the previous subsections was deployed as an Android application running on a smartphone with the aim of supporting healthcare professionals in wound monitoring procedures.

This application allows the automatic acquisition of skin wound images in an easy and intuitive manner while providing real-time feedback about the wound bed and marker's localisation and guaranteeing the proper periwound area presence (see fig. 2).

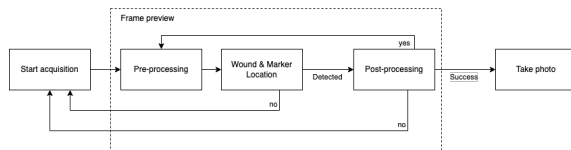


Figure 2: Automatic wound image acquisition diagram.

The user is first guided to centre the skin wound in the camera feed through a central target while the wound and marker localisation module is running (fig. 3 left). In the detection module, the raw images are cropped to a square at its centre with a height equal to its total width and resized to the model's input size (pre-processing, see fig. 2). Once the detection is complete, the user receives the indication through the left icon of the screen, and the adequacy validation module (post-processing block in fig. 2) starts running (fig. 3 right). The detection and adequacy validation modules keep running frame by frame until either one frame fails the detection (therefore restarting the process) or the acquisition is successfully completed. An image is automatically acquired when 20 consecutive frames or all the frames acquired in the span of 1250ms detect a wound and a marker with a score of at least 50% each. In case the developed modules fail, the user is always capable of manually acquiring an image by clicking on the bottom right button on the screen. After the acquisition, to make sure the user did not move after the last frame was processed, the acquired image is also verified with the detection and adequacy validation modules. In case of failure, the user has the option to retry the acquisition or to still continue with the previously acquired image.

3 RESULTS AND DISCUSSION

This section presents the results of the model optimisation experiments, as well as an in-depth analysis of the best-performing model in terms of detection performance and mobile inference metrics.

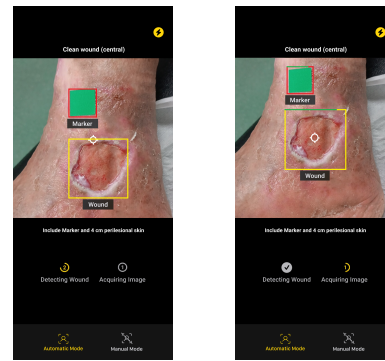


Figure 3: Application screenshots of real-time wound localization (left) and image adequacy validation (right).

3.1 Hyperparameter Tuning

To identify the most adequate hyperparameter combination, the performance achieved in the different experiments conducted for each image size was compared in terms of $mAP@.75IOU$ and $AR@10$, for each detected class (open wound or reference marker) and total. The average cross-validation $mAP@.75IOU$ of the hyperparameter tuning experiments is presented in fig. 4. Although the $AR@10$ results were also taken into account in the comparative analysis of the six experiments, those values exhibited a similar tendency as the ones of $mAP@.75IOU$ and are not graphically represented. Figure 4 shows that all experiments resulted in a similar performance for the reference marker detection, so the best hyperparameter combination was selected based mostly on the wound detection ability provided. Thus, through the analysis of the values reported for the wound class, experiment 5 yielded the highest $mAP@.75IOU$ and $AR@10$ metrics for both image sizes, so the associated hyperparameters were considered for the final detection model implementation.

For this training setup, the model trained on the 320×320 images obtained cross-validation $mAP@.75IOU$ and $AR@10$ values of 0.46699 ± 0.01170 and 0.48186 ± 0.03115 for the wound class, and 0.99010 ± 0.01400 and 0.83523 ± 0.00923 regarding the detection of the marker. Similarly, the detection performance of the 512×512 model resulted in $mAP@.75IOU$ and $AR@10$ metrics of 0.46107 ± 0.04047 and 0.51709 ± 0.06306 for the open wounds, and 0.99928 ± 0.00102 and 0.86706 ± 0.00871 for the marker class. Concerning the cross-validation $mAP@.50IOU$ for the open wound class, the 320×320 network achieved an average of 0.66188 ± 0.05608 , while the 512×512 model obtained 0.78117 ± 0.11202 ; the $mAP@.50IOU$ values for the marker class are approximately equal to the

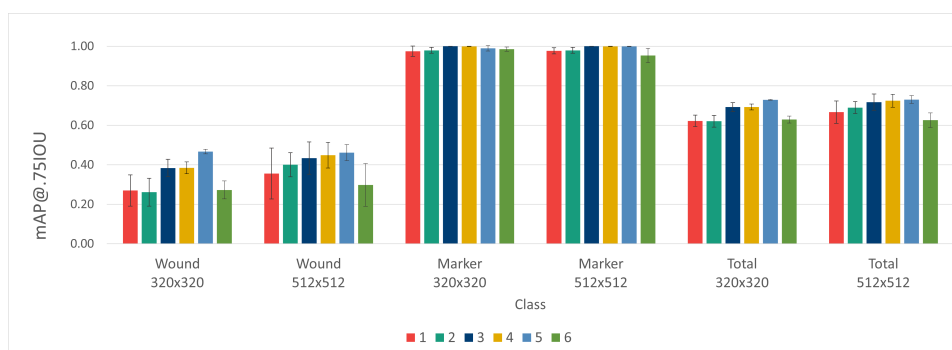


Figure 4: Average class-wise 3-fold cross-validation $mAP@.75IOU$ results obtained with the different hyperparameter settings tested in each experiment for the two image sizes analysed. The represented values were obtained as the average validation results for the three cross-validation splits. The total mAP was determined by averaging the results of the two classes (wound and marker). The error bars correspond to the standard deviation over all splits.

$mAP@.75IOU$ values reported. Both models exhibit superior performance for marker detection in relation to the open wound class, possibly due to the simplicity of the marker shape and the lower variability of the marker’s regions in the dataset. Moreover, despite providing good performance for the identification of wounds in the images (materialised in the high $mAP@.50IOU$ values achieved), the models demonstrated an impaired ability to accurately distinguish the region of interest corresponding to the whole open wound, as indicated by the lower mAP obtained when imposing a stricter IOU threshold of 0.75.

3.2 Selection of the Best Model

Table 3 shows the test results of the final RetinaNet detection model for each class after being re-trained on the whole training set with the optimal hyperparameter set-up (batch size of 8 and learning rate of 1.14976×10^{-4}). Comparing the test metrics with the cross-validation results presented in the previous section, we can observe that there was a general decline in the models’ performance, both in terms of detection accuracy and sensitivity (reflected by the $mAP@.50IOU$ and $AR@10$ metrics, correspondingly), even though this deterioration was less severe in the case of the marker class. This difference may be due to the occurrence of overfitting during model training or to the variability of properties in the training and test set images. Moreover, the discrepancy between the cross-validation and test performances was more expressive for the 320×320 model. Therefore, despite the similar cross-validation results demonstrated by the models trained with the two image sizes, in the test set the 512×512 demonstrated considerably higher detection robustness.

To assess the applicability of each detection model for real-time usage, both models were deployed on

Table 3: Class-wise detection performance of the best models, re-trained in the overall training set and evaluated using the test data from the CWS dataset.

Evaluated metric/ Image size		320 × 320	512 × 512
$mAP@.75IOU$	Wound	0.23712	0.39141
	Marker	0.92079	0.95050
$AR@10$	Wound	0.36563	0.47031
	Marker	0.79412	0.83971

Table 4: Processing times (in milliseconds) of the different steps of the image acquisition module using the best-performing models associated with each input image size, assessed in mid and high-range (MR, HR) smartphones.

	320 × 320		512 × 512	
	MR	HR	MR	HR
Total time per frame	132.00	103.43	393.52	274.79
Pre-processing	1.21	1.28	1.39	1.47
Detection inference	130.26	101.53	391.41	272.76
Adequacy validation	0.00	0.00	0.06	0.02
Total acquisition time	1313.98	1296.13	1400.63	1394.67
Avg. no. frames per acq.	12	12	5	6
Model size	10.9 MB		11.2 MB	

two mobile devices (mid and high range), namely Xiaomi Poco X3 NFC and Samsung S21. Since the acquisition of each image involves the analysis of several camera frames by the application, the time required by each processing step - the pre-processing to get compatible image shape and dimensions, the detection model’s inference and the validation of the image structures’ adequacy (open wound, reference marker and periwound area) - was examined for all the frames analysed, along with the total acquisition time and the average number of frames processed for each image acquired. The results of the mobile assessment are presented in table 4. This evaluation was carried out by using the developed image acquisition application to capture images of the test set images.

From the times presented for each step, it is possible to verify that the inference time of the detection

model is the most time-consuming step of the acquisition process, being the major contributor to the time taken to analyse each frame and, subsequently, for the whole image acquisition flow. Despite the higher inference times required by the 512×512 model (in relation not only to the 320×320 model but also to the approach described in (Faria et al., 2020)), it is still able to provide a responsive and smooth acquisition process in both high and mid-range devices, with a sufficient amount of evaluated frames in the stipulated acquisition time. This, in combination with its superior detection performance in the test set, motivated its selection as the final model to be deployed in the final acquisition application.

3.3 Final Model Assessment

To allow the comparison with the detection approaches described in the state of the art and mentioned in section 1, the $mAP@.50IOU$ metric computed for the wound class was also analysed. Comparing the cross-validation $mAP@.50IOU$ of the developed model (0.78117 ± 0.11202) with the corresponding performances reported in (Goyal et al., 2018; Faria et al., 2020) (0.918 and 0.865), it is possible to verify that, despite the lower metric value achieved, it is associated with a similar magnitude order. In contrast, the obtained test metric (0.66829) exhibits a greater deviation in relation to the test performances of (Anisuzzaman et al., 2022a; Scebbba et al., 2022) (0.939 and 0.762), which might be due to the presence of borderline wound cases in the test set, further discussed below. Still, the direct comparison of these metrics does not provide a fair analysis of the different models' performance since they were evaluated in different datasets, whose information is not available. Moreover, the models proposed in other works were solely focused on the detection of the open wound area, while this work provides a model that is able to detect directly both the open wound and reference marker. Hence, future experiments should be conducted to enable a more even comparison, by training or at least evaluating the models in the same dataset.

In addition, to validate the detection ability of the network selected as the final model, its prediction results in the test images were analysed in detail. Figure 5 comprises instances for which the detection of the model was not as satisfactory, whereas fig. 6 presents examples of correct detections.

The inspection of the test predictions shows that the model's detection ability is flawed under very specific acquisition conditions, providing insights regarding the possible causes of the inferior test performance displayed. Particularly, it demonstrated diffi-

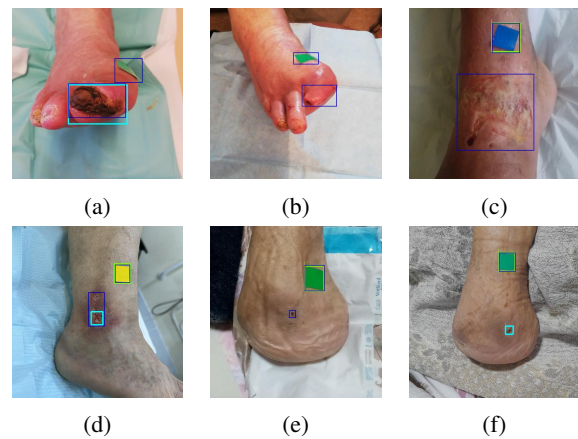


Figure 5: Examples of test images that highlight the performance limitations of the 512×512 model. These include images acquired in a skewed perspective (a and b), superficial wounds (c and d) and wounds in the open/healed transition (e and f). The ground truth annotations are marked in dark blue, while light blue and green boxes outline the wound and marker regions detected by the network.

culty in correctly detecting the marker and/or wound region when the image is captured from a skewed perspective, a common occurrence in diabetic foot ulcer cases, as showed in figs. 5a and 5b. In some cases of more superficial leg ulcers, especially with heterogeneous lighting conditions and the presence of reflections, the model is not capable of detecting the wound region (fig. 5c), or provides only a partial detection that does not include the whole region of interest (fig. 5d), which might affect the verification of the periwound area in the photo and be an obstacle to the usage of this information to further extract wound properties in automated systems. Notwithstanding, these limitations can be overcome in the acquisition moment by using adequate lighting and adjusting the smartphone's positioning in relation to the wound. The model also misdetects wounds that were recently healed or close to healing, by incorrectly detecting the scab region in fig. 5f as an open wound and by missing the detection of the small wound region in fig. 5e. It is worth highlighting that these cases are ambiguous even for trained clinical experts, so this limitation is not critical for a practical scenario.

In spite of failing in some of the described circumstances, the model presented a satisfactory detection performance in a varied set of conditions, being able to detect both the wound and marker in images acquired from a skewed perspective (fig. 6c) and inconsistent lighting (such as the shadows present in fig. 6d). It demonstrated a consistent detection ability for the multiple wound types and body locations of the dataset - observable in the detections for hip pressure ulcers (figs. 6a, 6d and 6e), leg ulcers (fig. 6b)

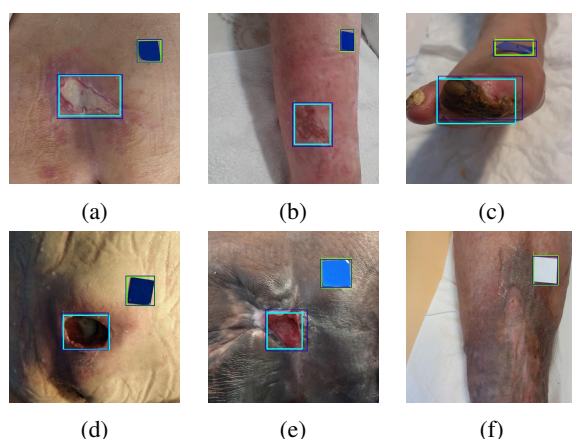


Figure 6: Examples of test images with correct predictions generated by the 512×512 model in varied settings. These comprise images from different wound types and locations (hip pressure ulcers in a, d and e, leg ulcer in b and diabetic foot ulcer in c), and images with healed wounds (f). The ground truth annotations are marked in dark blue, while light blue and green boxes outline the wound and marker regions detected by the network.

and diabetic foot ulcers (fig. 6c) -, while showing a reduced number of false positives, as it is possible to observe by the lack of wound area detections in images without open wounds (fig. 6f). Remarkably, notwithstanding the low representation of higher phototypes in the dataset, it also maintained a good detection ability for varying skin phototypes (displayed by the heterogeneous types of skin in fig. 6), evidencing its robustness and suitability for usage in clinical scenarios.

Thus, regardless of the apparently superior performances reported in the literature, the developed model was able to locate the wound and reference marker regions consistently for different image acquisition conditions. Its combination with the image validation module and the integration of this pipeline in a mobile-based solution make this the first methodology (to the authors' knowledge) to exploit the potential of detection architectures to ensure wound image adequacy in the acquisition process, providing a real-time acquisition system that captures images with sufficient information for the analysis of wound healing by human experts and for the development of automated wound characterisation algorithms.

4 CONCLUSIONS

This work presented a pipeline for the automation of wound image acquisition and adequacy validation, which leverages the potential of deep detection neural networks to guarantee that relevant skin regions are

captured while providing a metric reference, both elements of great utility to support further wound analysis tasks.

A deep neural network - RetinaNet with a MobileNetV2 backbone - was used to detect the open wound region and a reference marker, to be used as a metric reference. This model was developed using a new dataset of wounds and markers of diverse types and characteristics. Two models associated with different image sizes of 320×320 and 512×512 were tuned and deployed in mobile devices. Their comparison showed that the 512×512 model offers a superior detection performance while still providing inference times suitable for real-time image acquisition, being selected as the final model. Although this model yielded detection metrics slightly inferior to those reported in the literature for similar models (evaluated in other datasets), it demonstrated a satisfactory ability to detect wounds and markers in a variety of acquisition conditions, evidencing its robustness.

The developed detection model was used as the basis for an image adequacy validation module, responsible for ensuring the presence of the reference marker, wound area and recommended 4 cm radius of periwound skin (validated using the marker's metric reference) in the captured picture. A mobile application incorporating both modules was implemented and tested for the acquisition of several wound images. This assessment showed that the proposed approach contributed to the acquisition of adequate images through an easily usable interface that streamlines the acquisition process. Nonetheless, the periwound radius verification could be improved by getting a more rigorous marker segmentation and using the resulting sides or diagonal as a metric reference instead of its bounding box, as well as resorting to perspective correction methods to account for possible image distortions.

In spite of the identified refinement possibilities, the integration of the developed detection network with the image adequacy validation module provides a powerful tool to assist wound image acquisition, enabling its standardisation. Consequently, it facilitates the acquisition of more adequate images, with the representation of all the structures relevant for wound healing assessment, offering an invaluable asset for both the visual inspection by clinical experts and the automated wound properties extraction in more sophisticated systems. Although its accuracy in the wound region identification should be improved for the latter application scenario (through the expansion of the dataset with private or public images or the exploration of new detection architectures), the robustness demonstrated in varied acquisition settings, com-

bined with its user-friendly, responsive mobile interface, proved it is suitable for usage by clinical professionals in real-world scenarios.

ACKNOWLEDGEMENTS

This work was done under the scope of ClinicalWoundSupport - "Wound Analysis to Support Clinical Decision" project (POCI-01-0247-FEDER-048922), according to Portugal 2020 is co-funded by the European Regional Development Fund from European Union, framed in the COMPETE 2020. The authors would like to thank Paula Teixeira from Unidade Local de Saúde de Matosinhos for the collaboration in the data annotation process.

REFERENCES

- (2021). TensorFlow 2 Detection Model Zoo. https://github.com/tensorflow/models/blob/master/research/object_detection/g3doc/tf2_detection_zoo.md.
- (2022). Configuration settings for the models of the tensorflow object detection API. https://github.com/tensorflow/models/tree/master/research/object_detection/configs/tf2.
- Anisuzzaman, D., Patel, Y., Niezgodá, J. A., Gopalakrishnan, S., and Yu, Z. (2022a). A mobile app for wound localization using deep learning. *IEEE Access*, 10:61398–61409.
- Anisuzzaman, D., Wang, C., Rostami, B., Gopalakrishnan, S., Niezgodá, J., and Yu, Z. (2022b). Image-based artificial intelligence in wound assessment: A systematic review. *Advances in Wound Care*, 11(12):687–709.
- Bengio, Y. (2012). Practical recommendations for gradient-based training of deep architectures. In *Neural networks: Tricks of the trade*, pages 437–478. Springer.
- Chen, L., Cheng, L., Gao, W., Chen, D., Wang, C., Ran, X., et al. (2020). Telemedicine in chronic wound management: systematic review and meta-analysis. *JMIR mHealth and uHealth*, 8(6):e15574.
- Cui, C., Thurnhofer-Hemsi, K., Soroushmehr, R., Mishra, A., Gryak, J., Domínguez, E., Najarian, K., and López-Rubio, E. (2019). Diabetic wound segmentation using convolutional neural networks. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1002–1005. IEEE.
- Faria, J., Almeida, J., Vasconcelos, M. J. M., and Rosado, L. (2020). Automated mobile image acquisition of skin wounds using real-time deep neural networks. In *Medical Image Understanding and Analysis: 23rd Conference, MIUA 2019, Liverpool, UK, July 24–26, 2019, Proceedings 23*, pages 61–73. Springer.
- Goyal, M., Reeves, N. D., Rajbhandari, S., and Yap, M. H. (2018). Robust methods for real-time diabetic foot ulcer detection and localization on mobile devices. *IEEE journal of biomedical and health informatics*, 23(4):1730–1741.
- Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S., et al. (2017). Speed/accuracy trade-offs for modern convolutional object detectors. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7310–7311.
- Lin, T., Goyal, P., Girshick, R., He, K., and Dollar, P. (2018). Focal loss for dense object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 42(2):318–327.
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco: Common objects in context. In *European conference on computer vision*, pages 740–755.
- Lu, H., Li, B., Zhu, J., Li, Y., Li, Y., Xu, X., He, L., Li, X., Li, J., and Serikawa, S. (2017). Wound intensity correction and segmentation with convolutional neural networks. *Concurrency and computation: practice and experience*, 29(6):e3927.
- Martinengo, L., Olsson, M., Bajpai, R., Soljak, M., Upton, Z., Schmidtchen, A., Car, J., and Järbrink, K. (2019). Prevalence of chronic wounds in the general population: systematic review and meta-analysis of observational studies. *Annals of epidemiology*, 29:8–15.
- Medetec (2014). Wound database. data retrieved from <http://www.medetec.co.uk/files/medetec-image-databases.html>.
- Mukherjee, R., Tewary, S., and Routray, A. (2017). Diagnostic and prognostic utility of non-invasive multimodal imaging in chronic wound monitoring: a systematic review. *Journal of medical systems*, 41:1–17.
- Passadouro, R., Sousa, A., Santos, C., Costa, H., and Craveiro, I. (2016). Characteristics and prevalence of chronic wounds in primary health care. *Journal of the Portuguese Society of Dermatology and Venereology*, 74(1):45–51.
- Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4510–4520.
- Scebba, G., Zhang, J., Catanzaro, S., Mihai, C., Distler, O., Berli, M., and Karlen, W. (2022). Detect-and-segment: A deep learning approach to automate wound image segmentation. *Informatics in Medicine Unlocked*, 29:100884.
- Sen, C. K. (2021). Human wound and its burden: updated 2020 compendium of estimates. *advances in wound care*, 10(5):281–292.
- Werdin, F., Tennenhaus, M., Schaller, H.-E., and Rennekampff, H.-O. (2009). Evidence-based management strategies for treatment of chronic wounds. *Eplasty*, 9.
- Yang, S., Park, J., Lee, H., Kim, S., Lee, B.-U., Chung, K.-Y., and Oh, B. (2016). Sequential change of wound calculated by image analysis using a color patch method during a secondary intention healing. *PLoS one*, 11(9):e0163092.