



Aalborg Universitet

AALBORG UNIVERSITY
DENMARK

Preparing Infection Detection Technology for Hospital at Home after Lower Limb External Fixation

Annadatha, Sowmya; Hua, Qirui; Fridberg, Marie; Jensen, Tobias; Liu, Jianan; Kold, Søren Vedding; Rahbek, Ole; Shen, Ming

Published in:
Digital Health

DOI (link to publication from Publisher):
[10.1177/20552076221109502](https://doi.org/10.1177/20552076221109502)

Creative Commons License
CC BY-NC-ND 4.0

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

[Link to publication from Aalborg University](#)

Citation for published version (APA):

Annadatha, S., Hua, Q., Fridberg, M., Jensen, T., Liu, J., Kold, S. V., Rahbek, O., & Shen, M. (2022). Preparing Infection Detection Technology for Hospital at Home after Lower Limb External Fixation. *Digital Health*, 8, 1-20. <https://doi.org/10.1177/20552076221109502>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy

If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Preparing infection detection technology for hospital at home after lower limb external fixation

Digital Health
Volume 8: 1–20
© The Author(s) 2022
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/20552076221109502
journals.sagepub.com/home/dhj



Sowmya Annadatha¹ , Qirui Hua¹ , Marie Fridberg², Tobias Lindstrøm Jensen¹, Jianan Liu³, Søren Kold² , Ole Rahbek²  and Ming Shen¹

Abstract

Background: Patients with severe bone fractures and complex bone deformities are treated by orthopedic surgeons with external fixation for several months. During this long treatment period, there is a high risk of inflammation and infection at the superficial skin area (pin site). This can develop into a devastating, sometimes fatal, and always costly condition of deep bone infection.

Objective: For pin site infection surveillance, thermography technology could be the solution to build an objective and continuous home-based remote monitoring tool to avoid frequent nursing care and hospital visits. However, future studies of infection monitoring require a preliminary step to automate the process of locating and detecting the pin sites in thermal images reliably for temperature measurement, and this step is the aim of this study.

Methods: This study presents an automatic approach for identifying and annotating pin sites on visible images using bounding boxes and transferring them to the corresponding thermal images for temperature measurement. The pin site is detected by applying deep learning-based object detection architecture YOLOv5 with a novel loss evaluation and regression method, control distance intersection over union. Furthermore, we address detecting pin sites in a practical environment (home setting) accurately through transfer learning.

Results and conclusion: The proposed model offers the pin site detection in 1.8 ms with a high precision of 0.98 and enables temperature information extraction. Our work for automatic pin site annotation on thermography paves the way for future research on infection assessment on thermography.

Keywords

Object detection, machine learning, image analysis, infection, thermography, bone, pin site, orthopaedics

Submission date: 16 May 2022; Acceptance date: 6 June 2022

Introduction

An External Fixator (EF) in Figure 1 is a device orthopedic surgeons use to save, correct, and reconstruct lower limbs after fractures. It consists of an external frame connected to the bone inside the limb with pins and wires. The fixation of the bone is important to obtain a stable construct for bone healing. The pin site, as highlighted with the blue bounding box in Figure 2(b), is the skin region where the pins and wires pass through the skin. There is a high risk of infection at the pin site during treatment since it is an open

wound. The time period for treatment with an EF after a lower limb bone fracture is typically 3–12 months, and previous studies have reported that up to 50% of

¹Department of Electronic Systems, Aalborg University, Aalborg, Denmark

²Aalborg University Hospital, Aalborg, Denmark

³Vitalent Consulting, Gothenburg Sweden and Silo AI, Stockholm, Sweden

Corresponding author:

Ming Shen, Department of Electronic Systems, Aalborg University, Aalborg, Denmark

Ole Rahbek, Aalborg University Hospital, Aalborg, Denmark.

Email: mish@es.aau.dk



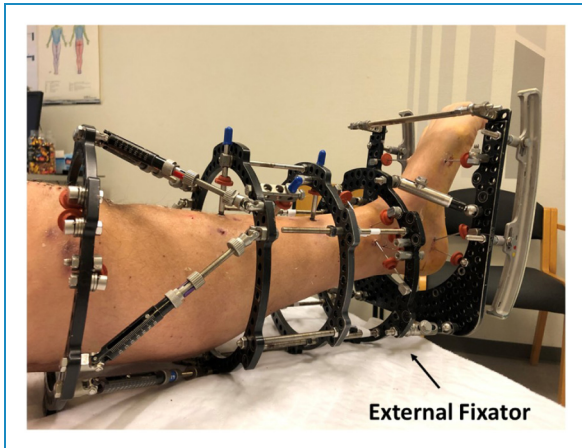


Figure 1. A patient with an external fixator (EF) after a lower limb bone fracture.

the patients suffer from mild to severe infection during that period.¹ Infection can become a serious situation for the patient; the construction can become less stable, resulting in slow or lack of bone healing, or the patient can become acute septic, and in the worst case, the

patient will need an amputation of the limb. During the treatment period with an EF, there is a need to monitor the patients frequently, and they have nursing care for cleaning and dressing the wounds to avoid infections one to several times weekly. Patients are examined at routine check-ups at the hospital and have X-rays every week or every few weeks throughout the whole treatment period as well. The post-surgical care regime and multiple outpatient clinic visits in the hospital are time and resource-consuming, both for patients and the healthcare system. Today, patients are discharged earlier from hospitals, shifting the postoperative surveillance to the home of the patient. The care is thus transferred from the highly specialized hospital to surveillance by municipality home nurses and general practitioners. During this shift, both information and knowledge of postoperative care are often lost, leaving the patient with uncertainty. A lack of agreement on the definition of surgical site infection by clinicians and little evidence of systematic measurement and monitoring of wounds after discharge from the hospital enhance this uncertainty. To summarize, there is a need to develop a point-of-care tool warning about early signs of infection to prevent complications of surgery.

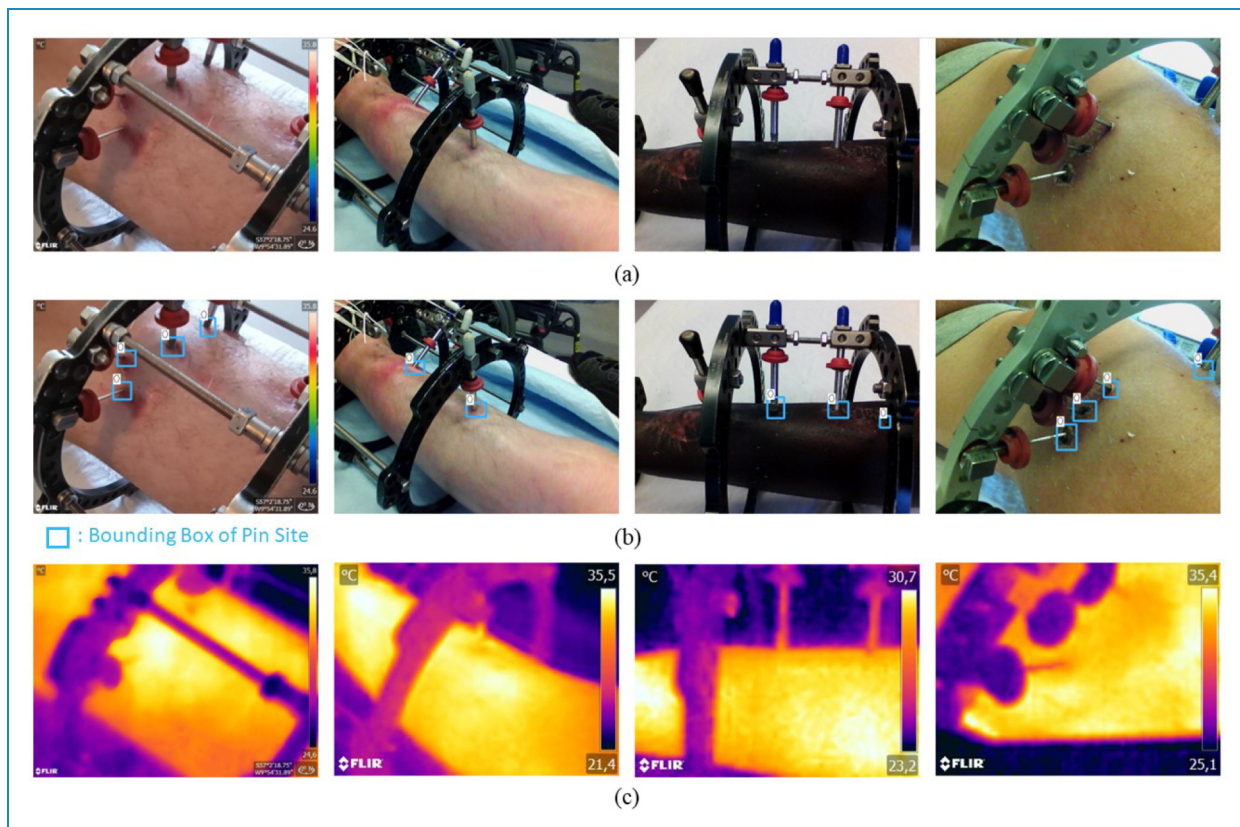


Figure 2. RGB visible images and infrared thermal images after external fixation. (a) RGB visual images, (b) images with bounding boxes annotating pin sites, as references for training, (c) thermal images.

It has been suggested that thermography is reliable in detecting signs of infection at pin sites.² The technology of those thermographic cameras is quickly evolving these days, and it is becoming possible to take good-quality infrared thermal images with a low-cost hand-held camera.³ These new infrared cameras even now come as an add-on to a mobile phone device, giving the possibility to monitor the infection signs remotely, by the patients from the patients' home setting and having the on-demand evaluation. Furthermore, analyzing the thermography by machine learning methods has shown powerful potential in the evaluation of health⁴ and diagnosis of diseases including breast cancer⁵ and Raynaud's phenomenon,⁶ but the training of model requires lots of data. To prepare the infection detection technology through temperature analysis, there is a need to annotate the essential pin site region on infrared thermal images. However, interfaces between skin and pins are not clearly visible to distinguish and mark based on their shapes and colours in most of the thermal images, for example, Figure 2 (c) while in RGB visible images like Figure 2 (b), they can be easily annotated in bounding boxes. For thermal images, pin sites that should have been monitored can be missed. To achieve higher accuracy of infection detection on thermography as well as the efficient collection of essential data, the idea is to first use the original visible images of RGB colours to detect the pin sites and transfer the selected regions to the corresponding infrared thermographic images. Considering a large amount of data of images, this repetitive task can take clinical manpower to complete and an automated and accurate annotation tool can effectively accelerate the detection with consistency.

In this paper, we propose a pin site detection model that automates the process of identifying and annotating regions of interest with bounding boxes on the corresponding RGB visible images and transfers the detected boxes to the targeted infrared thermal images. Therefore, this work contributes to preparing further research and potential applications in pin site infection detection using digital thermography for home-based care of patients after lower limb external fixation.

Literature review

Object detection

The implementation of an automatic pin site detection framework relies on object detection algorithms that can search and mark bounding boxes of the informative location from the given images. As a popular and important topic in computer vision, object detection has been widely studied and applied in many fields including face recognition,⁷ auto-driving⁸ and medical image analysis.⁹ Some traditional algorithms follow the exhaustive method with many sliding windows¹⁰ to find useful regions from the

image. But they consume considerable computational resources with poor accuracy and speed, and there is a doubt about their robustness to the diversity of images.¹¹ Furthermore, those works that utilize rule-based algorithms for desired area extraction like statistical pattern recognition algorithm¹² require rules or features defined manually.¹³ In the past decade, convolution neural networks (CNNs) have been introduced for object detection with possible superior performance in both accuracy and computation time, which greatly enhances the development of medical image assessment.¹¹

One-stage and two-stage object detectors

These CNN-based frameworks can be mainly divided into two groups, one-stage, and two-stage object detector. The former type, such as you only look once (YOLO),¹⁴ only contains one step of regression and classification that takes input images to learn the coordinates of bounding boxes and probabilities of classes. While the latter group, like R-CNN (regions with CNN features),¹⁵ fast R-CNN¹⁶ and faster R-CNN,¹⁷ achieves object detection by two steps, acquiring regions of interests in the first stage and then in the second stage making the decision from the selected regions. Compared to two-step frameworks that require separated stages to generate results, one-step methods require less consuming time and produce relatively lower but still acceptable accuracy.¹⁸ In the state of the art one-stage object detectors like YOLOv4¹⁹ and YOLOv5,²⁰ the accuracy is also improved with fast processing speed. Aiming to develop a lightweight application that can be deployed in mobile phones for patients, implemented by Pytorch, the one-stage framework YOLOv5 providing both processing speed and accuracy is chosen in this work.

Intersection over union in object detection

The evaluation of loss is significant to the bounding box regression in YOLO frameworks. To represent the similarity of the predicted box and ground-true box, the traditional intersection over union (IoU), as well as IoU loss function, is only determined by the overlapping area, which means different relative positions could result in the same IoU and degrades regression accuracy. Therefore, advanced methods have been suggested to comprehensively analyse the difference from various dimensions. For example, generalized IoU (GIoU)²¹ introduces a penalty term for non-overlapping cases. Distance-IoU (DIoU) and complete IoU (CIoU)²² take more aspects into account such as central distance and aspect ratio, but DIoU loss encounters difficulties when the center points of predicted boxes coincide while CIoU requires a relatively longer calculation time. In addition, one of the latest evaluation systems, control distance IoU (CDIoU),²³ can directly indicate the degree of difference between the predicted region and

ground-true box and shows better performance with reduced complexity. Considering having an optimized selection for the pin site data set, in this work, we incorporate DIoU, CIoU, and CDIoU methods into YOLOv4 and YOLOv5 and observe their performance.

Transfer learning

Generally, collecting sufficient medical image data is expensive and challenging, requiring professional equipment and experienced clinical staff, and therefore transfer learning approaches are widely utilized to construct medical imaging analysis models. There are lots of strategies for transfer learning, including instance weighting, feature transformation, model control, and parameter control.²⁴ Parameter control approaches are commonly utilized for deep learning-based models by sharing the network parameters of the source domain and target domain. An initial neural network can be trained by a large data set of the source domain and then finetuned only using a small set of data from the target domain to adapt. In this process, parameter sharing is achieved by freezing some layers of the pre-trained network on the source domain and training the last layers by instances from the target domain like deep adaptation networks (DAN),²⁵ or initializing the network parameters from the pre-trained network without freezing. These approaches have been applied in medical image diagnosis, including Alzheimer's disease²⁶ and knee osteoarthritis²⁷ and show good classification accuracy.

In our work, the photos taken by doctors in the hospital scenarios and by patients at home scenarios differ a lot in aspects including definition, pin site distance, and brightness, and hence the model trained by the large data set from the hospital has poor performance in the applications for patients at home. In addition, it is not possible to obtain sufficient patients' data due to the lack of infrared cameras available at their homes and logistic problem. As a result, we employ transfer learning to finetune the pre-trained model on the hospital data set so that the final model has better accuracy for patients at home.

Camera calibration and registration

The metal pins are not always clearly visible on thermal images, and therefore the region of interest on the skin, the pin site, is detected on the RGB visible images and then transferred to the corresponding thermal images for temperature extraction. However, the spatial difference due to the physical location causes a shifting error in the transferred bounding boxes. Therefore, calibration of the camera and registration of the images are mandatory before analyzing the annotated regions. The coordinate systems of cameras not at the same spatial position are different. So, the purpose of calibration is to derive a matrix for

converting coordinates from one system to another, based on which the images can be registered and aligned without any error in position and size.

With the rapid development in computer vision with the use of multiple cameras, lots of calibration and co-registration methods have been proposed. It is common and convenient to utilize a checkerboard or chessboard as the reference for calibration. By taking images of the checkerboard at a few orientations and annotating feature points, the parameters of the conversion matrix can be calculated.^{28,29} Functions provided by OpenCV can help to automate the process of locating points and estimating the parameters.³⁰

Methods

Data set

Visible and thermal image data set construction. In this work, the data sets contain RGB visible images and infrared thermography of external fixation from patients who are undergoing treatments for bone fracture. Table 1 introduces and classifies the data sets with different objectives and properties including size, collected method, scenario and image resolution. Besides, the specifications of digital cameras used for this work to capture images and construct the data set are presented in Table 2. To train the model for the pin site detection on visible images, data set P1 is collected using FLIR C3 cameras by professional clinical nursing staff the Aalborg University Hospital. Some samples of images in P1 are illustrated in Figure 3. After data augmentation, data set P1 is divided into a training group, a validation group, and a test group in the ratio of 8:1:1. In addition, data sets D1 and D2 are captured by patients themselves at home using FLIR C3 camera. With D1 data set, we have examined the performance of the detection model at home scenario as a practical application case, which was unsatisfactory due to scenario differences including uncontrolled illumination conditions, various angles and distances. As a result, the same data set, D1 is used to finetune the initial pin site model through transfer learning. To further validate the performance for harsh cases, this new transfer learned model is tested by another data set D2, which is also collected from home setting by patients and without any guidance from the research nurse and contains some images taken at a shorter distance. The pin site detection model training data sets are mainly collected from patients who have light skin tones from type 1 to type 2 based on the Fitzpatrick scale, and only the P1 data set contains 147 images of type 6 dark skin tone. Some of the patients have a certain stage of inflammation and infectious sign on their skin. Furthermore, most types of external frame constructions are included, containing various shapes taken from multiple orientations, as well as different numbers of thick and thin pins.

Table 1. Data sets collected and used in this work.

data set	Camera	Size	Visible image resolution	Thermal resolution	Scenario	Purpose
Hospital data set (P1)	FLIR C3	1042	640 * 480	80 * 60	Hospital by clinical staff	Images in this data set are taken by trained research nurses under the controlled setup in the hospital. Only visible images of this data set are used to train and test the initial pin site detection model (M1), which only detects and annotates regions of interest on visible images.
Home data set 1 (D1)	FLIR C3	197	640 * 480	80 * 60	Home by the patient	This data set is collected by the patient themselves, at home under uncontrolled illumination setup but with some help of nursing staff. We first test the initial pin site detection model (M1) with visible images of this data set and then use this set of images for finetuning the model M1 to obtain the transfer learned pin site detection model (M2) for home scenarios.
Home data set 2 (D2)	FLIR C3	429	640 * 480	80 * 60	Home by the patient	This data set is collected by patients at home, under uncontrolled illumination conditions and without any help from the nursing staff, and it contains some images taken at a short distance. With this data set, we evaluate the performance of the transfer learned pin site detection model (M2) on visible images based on shorter distances and from home settings.
Thermography data set 1 (T1)	FLIR One Pro	25	1440 * 1080	160 * 120	Hospital by clinical staff	The visible images and thermal images of pin sites in this data set are collected by trained research nurse under a controlled setup, which is utilized to test the complete function of the final pin site detection model (M3) on extracting temperature values from the detected area in the thermographic image.
Thermography data set 2 (T2)	FLIR T540	15	1280 * 960	464 * 348	Hospital by clinical staff	The visible images and thermal images of pin sites in this data set are collected by trained research nurse under controlled setup, which is utilized to test the complete function of the final pin site detection model (M3) on extracting temperature values from the detected area in the thermographic image.

Two sets of data (T1 and T2 in Table 1) including both RGB visible images and their respective infrared thermal images are acquired to validate the desired operation of the model. The effectiveness of the object detection model trained by the images taken by the FLIR C3 camera as well as temperature analysis is validated by these two data sets.

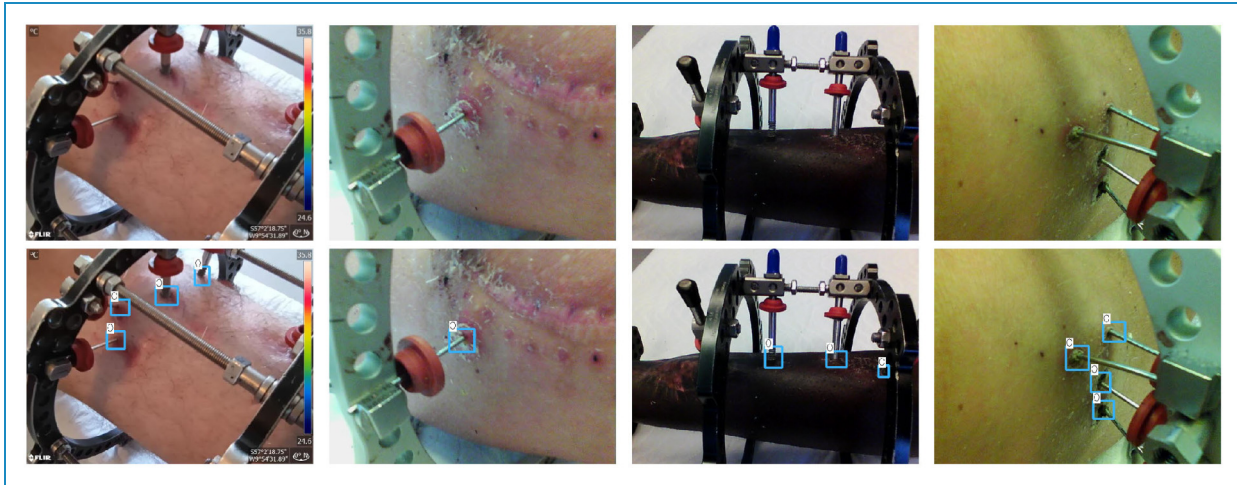
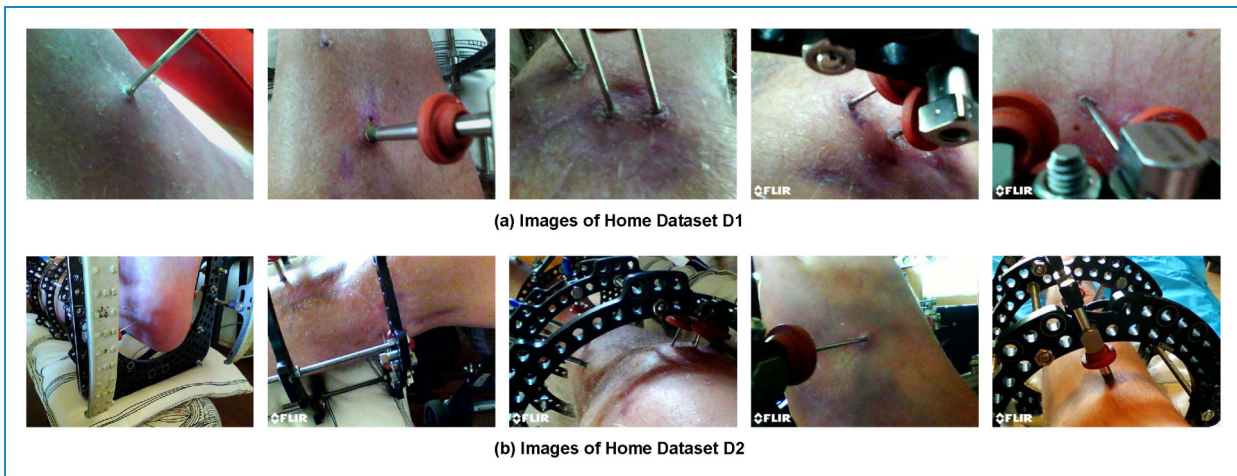
Data annotation. To construct a pin site detection model for infection assessment, the ground-true box plays an

important role as the region of interest for training. In our work, pins along with the skin interfaces on the images are annotated within bounding boxes as pin sites, which is completed manually through visual object tagging tool to guarantee precision. Some examples of annotated ground-true pin sites are illustrated in Figures 2 (b) and 3, which are marked with blue boxes representing regions of interest. It should be noted only the clear and visible pins are annotated in bounding boxes, while those pins that are hard to distinguish due to obstacles or

Table 2. Specifications of cameras used for data set construction.

FLIR type	Thermal resolution	Thermal sensitivity	Accuracy	Minimum focus distance
C3	80 * 60	100 mK	2°C or 2%	Thermal: 15 cm MSX: 1 m
T540	464 * 348	40 mK	2°C or 2%	Thermal: 15 cm MSX: 0.5 m
One Pro	160 * 120	70 mK	3°C or 5%	Thermal: 15 cm MSX: 0.3 m

MSX: Multi-spectral dynamic imaging

**Figure 3.** Hospital image data set of external fixation taken by FLIR C3 cameras, the second row contains manually annotated pin site boxes.**Figure 4.** Images in the home data set: (a) D1 data set and (b) D2 data set.

poor angles and distances are ignored as outliers. In addition, visible wounds are also included in the bounding boxes, allowing the measurement of clinical parameters for infection monitoring. Therefore, the center of the bounding box is defined by the interfacing point of the pin and skin as the pin site.

Data augmentation. There are three reasons for us to apply augmentation to our image data, which are listed below:

1. The size of the original base data set is too small with 1042 images, which is not sufficient to train a reliable pin site detection model with the desired accuracy.

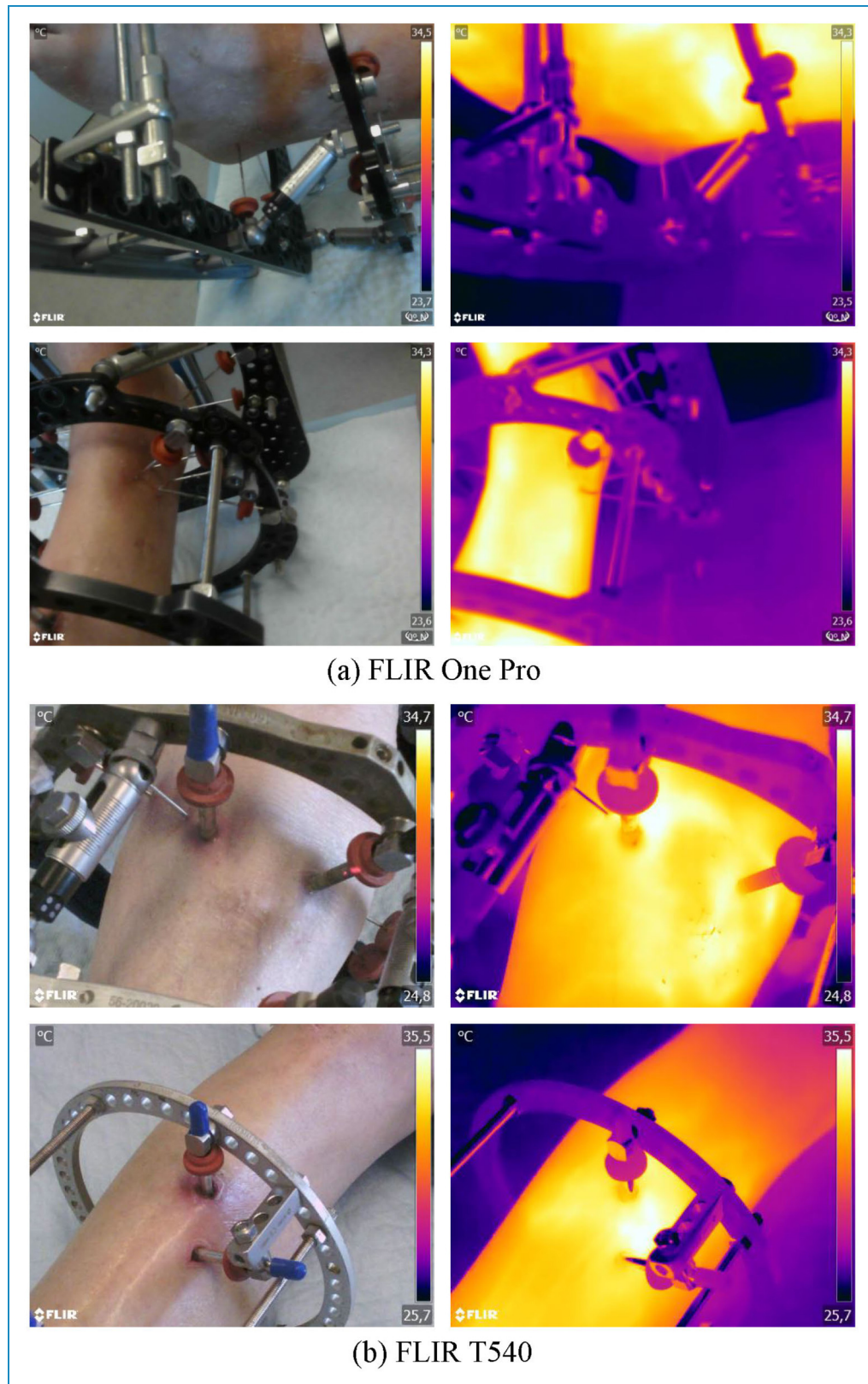


Figure 5. (a) Visual and thermal images taken by FLIR One Pro in T1 data set. (b) Visual and thermal images taken by FLIR T540 in T2 data set.

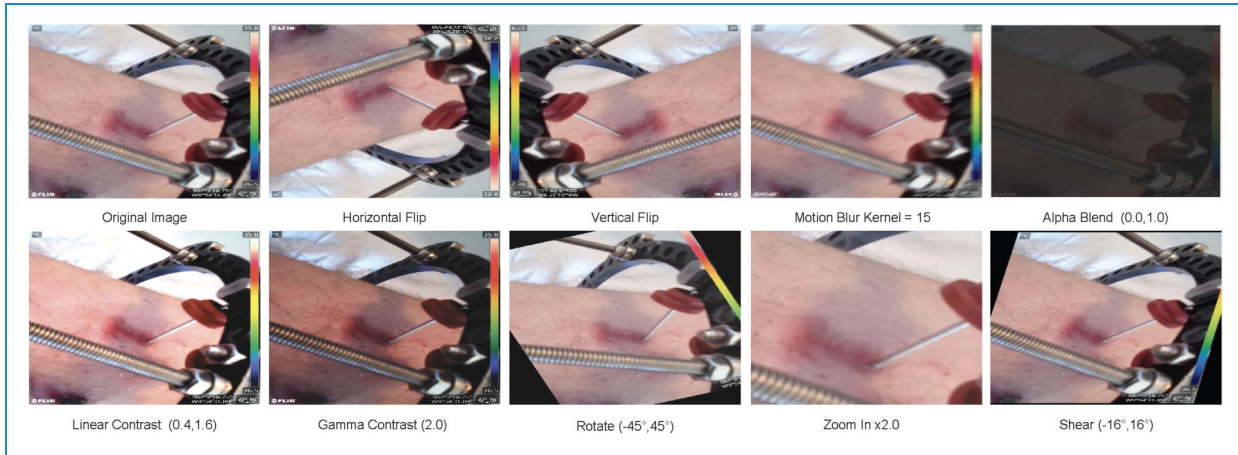


Figure 6. Augmentation techniques applied to the images, including flipping, motion blur, alpha blending, linear contrast, gamma contrast rotation, zooming in and shearing.

2. The collected data set is impossible to cover photos of all scenarios in real-world applications while the model needs to be trained with enough situations to achieve generalization. Therefore, it is essential to transform the original images into diverse conditions.
3. Augmenting the image data is adapted to address overfitting based on our experience.

Therefore, the data set from the hospital is scaled up and enriched with a wide range of scenarios for model training. In our work, the data augmentation is implemented by 8 techniques: horizontal flip, vertical flip, motion blur with the Gaussian kernel equal to 15, alpha blend with the overlay factor from 0 to 1, linear contrast adjustment via scales between 0.4 and 1.6, gamma contrast adjustment in the range 0 to 2, rotation between -45° and 45° , zooming in by two times and shearing between -16° and 16° . The exact transformations for a method with a given range are selected uniformly at random within its boundaries. For the alpha blend method, two branches of images are created with the pixel intensities multiplied by 100 and 0.2, respectively, which are then sampled by the overlay factor. Figure 6 presents an example of a raw photo augmented by these nine methods. Eventually, the size of the base image data set is increased to 10,409, and it is randomly divided into a training set, a validation set, and a test set, containing 8325, 1040, and 1044 images, respectively.

Thermography pin site detection model architecture

The objective of the pin site detection is to find and mark the regions of interest around the pin sites on the thermal images of external fixation, which prepares as a tool for future infection assessment. The processing flowchart of the complete model is illustrated in Figure 7, including data set construction, pin site detection model pre-training with the main hospital data set, finetuning through transfer

learning with the home data set D1, testing on visible images in the home data set D2, camera calibration, image registration, bounding boxes transfer, and maximum temperature extraction. As shown in the flow-chart, RGB visible cameras and infrared thermal cameras are calibrated, and thus thermal and visible images in the data set collected by those cameras are registered to remove the position and scale differences. Then, bounding boxes of pin sites are generated by the YOLOv5-based object detection model with CDIoU and transfer learning on visible images of external fixation. Because thermal images and visible images are well aligned after registration, those detected bounding boxes can be directly moved onto the respective thermal images. Based on each region of interest marked by the bounding box on temperature arrays of thermal images, we can obtain a set of maximum temperatures of their corresponding pin sites, which could be followed by future infection assessment research. For example, currently suggested by doctors, a threshold can be set for these maximum temperature values, above which a potential infection of the pin site is considered.

Pin site detection model on visible images. In our proposed model, the pin sites are detected from the corresponding visible images and marked with bounding boxes. After that, these bounding boxes are projected onto the respective infrared thermal images. To achieve it on visible images as the first step, we propose the pin site detection model based on the anchor-based object detection framework YOLOv5. Instead of using GIoU as its loss function in the basic YOLOv5 model, we have investigated other novel loss assessment methods such as CDIoU, CIoU and DIoU, and propose that CDIoU could provide excellent results in our targeted pin site data set. Furthermore, fully convolutional one-stage (FCOS)³¹ object detection framework also

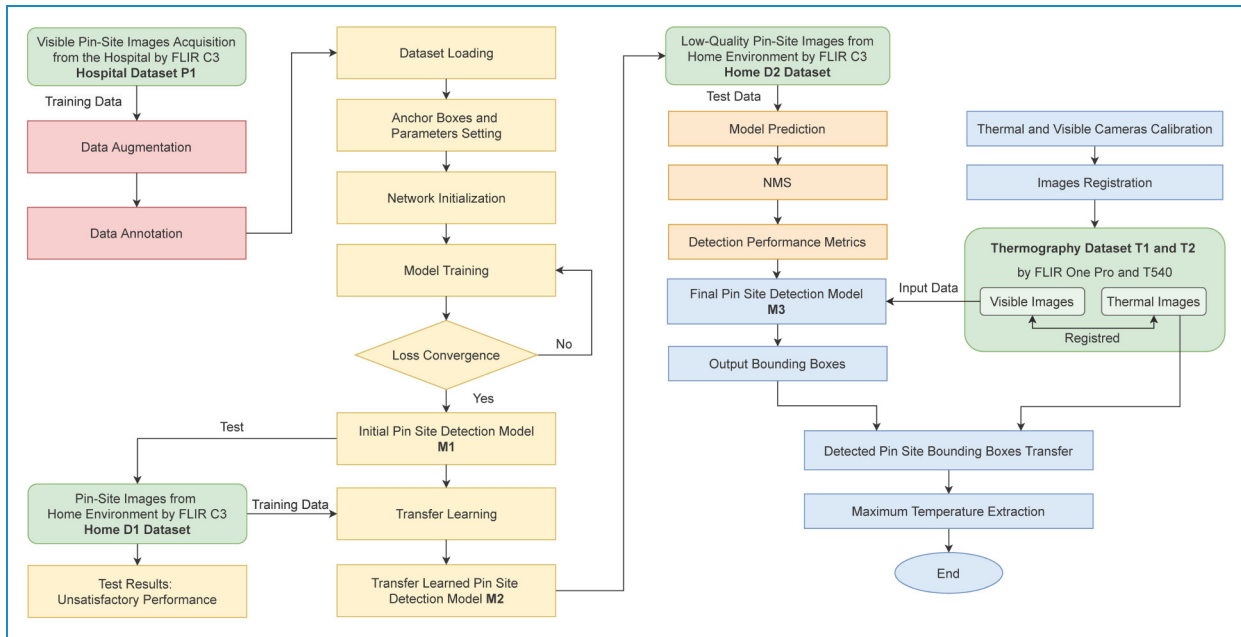


Figure 7. The flowchart of the construction of the pin site detection model.

has been explored to achieve an anchor-free structure for potential optimization. However, its results are not as expected, which are attached in the ‘Results and Discussion section for comparison. In addition, considering the challenges of applying the model to low-quality images and the lack of data from patients’ homes, transfer learning techniques are introduced to adapt the model. The initial proposed model is pre-trained over a large data set from our hospital with data augmentation. Then, the transfer learning approach is employed to finetune the model with another small data set collected from home scenarios by patients, so that the final model can suit practical applications for patients at home.

YOLOv5. The novel object detection model with the one-stage and anchor-based framework, YOLOv5, is utilized for detecting pin sites in this work. As shown in Figure 8, YOLOv5 consists of three essential components, backbone, neck, and head.²⁰ The first component, backbone, is a CNN used for features extraction and implemented by the cross stage partial network bottleneck with a focus layer. In the neck section, a path aggregation network (PANet) and a spatial pyramid pooling structure are applied to generate feature pyramids with improved information flow. As the final detection stage, the head leverages the features and anchor boxes to generate predicted outputs with scores for regression. In our cases, the classification of objects is not necessary as only the pins are required to detect.

As for activation functions, the middle and hidden layers employ the leaky rectified linear unit function while the final detection layer utilizes the sigmoid function.

Besides, as the default optimization function, stochastic gradient descent (SGD) is used for training.

Loss functions. In the ‘Detection section, the calculation of loss function is only based on the bounding box loss for regression. To represent the relative position relationship between the predicted boxes and ground-true boxes, IoU-based loss evaluation methods are popular and powerful. While different methods result in different results, the selection of a suitable approach for the target data set is important. For detecting pins in our work, CDIoU is proposed and incorporated into the YOLOv5 model. Considering the targeted pin site data set, the region of interest for doctors is the wound around a pin and hence the center of the bounding box is set at the position where the pin inserts into the skin. As the center is also a key feature in our image data set, we believe that for fast loss convergence, it is important to reduce the central point distance between the predicted bounding box and ground-true box without an increase in the inference time. Therefore, DIOU and CIOU featured by minimizing central distance directly are first considered, within which CIOU also takes the aspect ratio into account but consumes more time for calculation.²²

Furthermore, the latest method CDIoU evaluates both the central distance and shape similarity in the loss function through a minimum bounding rectangle of the predicted box and ground-true box. It is possible to reduce the center distance while considering the aspect ratio but without the performance degradation in the calculation time.²³

In conclusion, the loss analysis methods performed in this work include CIOU, DIOU, and CDIoU, and their

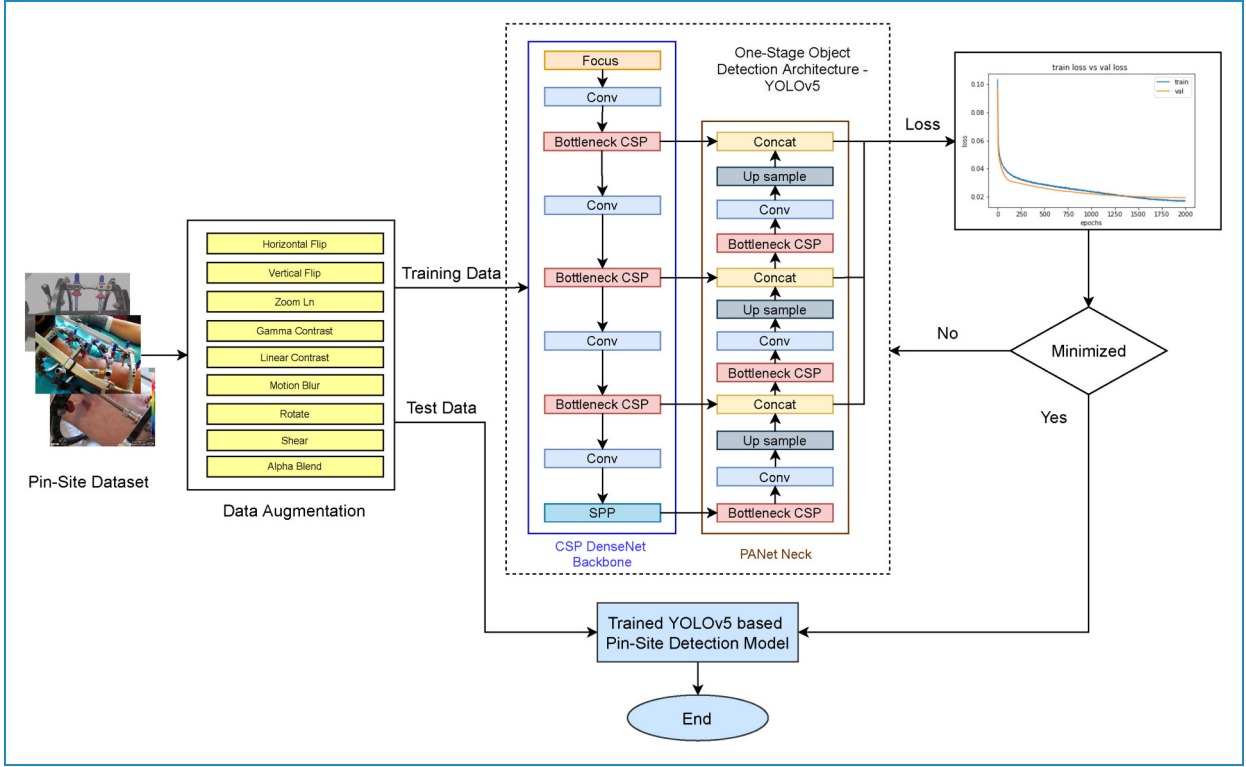


Figure 8. The architecture of the YOLOv5 object detection model.

detailed loss functions are described below. Assuming B represents the predicted bounding box and B^{gt} is the ground-true box, the basic IoU and its loss function \mathcal{L}_{IoU} are defined by:

$$IoU = \frac{|B \cap B^{gt}|}{|B \cup B^{gt}|}, \quad (1)$$

$$\mathcal{L}_{IoU} = 1 - IoU. \quad (2)$$

The central points \mathbf{b} and \mathbf{b}^{gt} of B and B^{gt} are used for the calculation of DIoU loss function as follows²²:

$$\mathcal{L}_{DIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} \quad (3)$$

where $\rho(\cdot)$ denotes the Euclidean distance and c represents the diagonal length of the minimum bounding box that encloses B and B^{gt} . Similarly, adding a factor of aspect ratio v and the trade-off parameter α , the CIoU loss is given by²²:

$$\mathcal{L}_{CIoU} = 1 - IoU + \frac{\rho^2(\mathbf{b}, \mathbf{b}^{gt})}{c^2} + \alpha v, \quad (4)$$

$$v = \frac{4}{\pi} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (5)$$

$$\alpha = \frac{v}{(1 - IoU) - v}, \quad (6)$$

where (w, h) and (w^{gt}, h^{gt}) are determined by the width and height of B and B^{gt} , respectively.

Besides, with less calculation than CIoU, as vectors AE , BF , CG and DH defined by Figure 9, the calculation of CDIoU is determined by²³:

$$diou = \frac{AE + BF + CG + DH}{4WY}, \quad (7)$$

$$\mathcal{L}_{CDIoU} = \mathcal{L}_{IoU} - diou \quad (8)$$

Transfer learning. The experience has been, that when the pin site detection model is trained on the images collected by a research team at the hospital, it obtains an excellent accuracy in the testing set. However, when this trained model is applied to images taken by the patients at their home setting (data set D1), we observe a considerable reduction in detection accuracy. It can be explained by the fact that the clinic nurses who take those images possess professional skills that those at-home patients do not have and that the characteristics of images vary greatly from patient to patient, including lighting conditions, appropriate distance to retain focus and optimal angles to make the pin sites most visible. Furthermore, home environments may not allow patients to take good images of the EF with pins. As the preliminary pin site detection model was trained from a data set where hospital

images are the majority, its actual performance for real patients at home is heavily suffered by the image quality problems from home scenarios. Specifically, the doctors would take the images of the external fixation from a remote distance to capture a full view of the pins, whereas most patients' images are taken at a relatively short distance. Meanwhile, it is not efficient to collect a large number of images from patients at home so that we can acquire a data set as large as the one from the hospital to retrain a model, limited by the imaging devices for patients due to logistic reasons.

As a result, to raise the detection accuracy on images taken by patients in practical applications, transfer learning is adopted in our work to deal with the low-quality images while the training data set is undersized. Firstly, the images from two environments are sorted into two groups accordingly. The main data set consists of a significantly larger number of images taken in the hospital by nurses as compared to the small-size data set D1 collected by patients in home environments. Then, the transfer learning steps

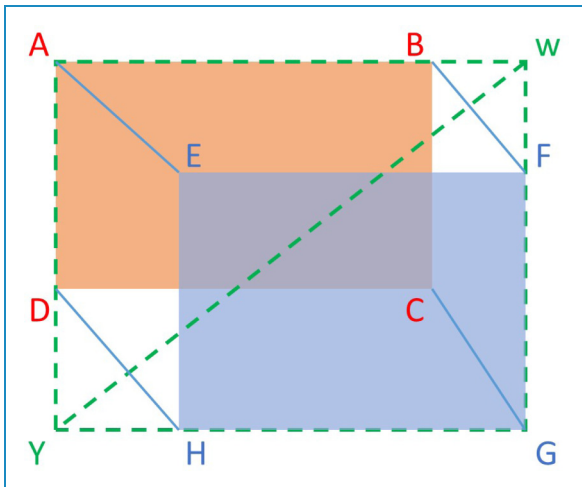


Figure 9. Calculation of control distance intersection over union (CDIoU) loss function. Rectangle $AWGY$ represents the smallest rectangle enclosing the region proposal box $ABCD$ and ground-true box $EFGH$.

via the data set D1 as demonstrated in Figure 7 are brought as a part of the overall model training procedure. The initial pin site detection model is pre-trained by the main hospital data set through the object detection framework as described in the previous section and then adjusted by the transfer learning techniques on the D1 data set. Transfer learning in our work is implemented by leveraging the pre-trained model as the initial setting and finetuning it with data set D1 as illustrated in Figure 10. We explore the performance of three approaches in Figure 10 to finetune the parameters of the pre-trained model during training the transfer learning model, including freezing 9 layers of the backbone (backbone freezing), freezing 23 layers of the backbone and neck (full freezing), and not freezing any layers (no freezing). The results of these three strategies could be compared and we propose no layer freezing method resulting in the best performance. Finally, the transfer learned pin site detection model is tested by data set D2, which is another set of low-quality images taken by patients, and its ability to detect challenging pin sites for real-world applications can be examined.

Thermal camera calibration and image registration. Sections mentioned previously present our YOLOv5 with CDIoU pin site detection model and transfer learning algorithm on visible images. As infrared and visible images are captured by two separated cameras from the same device, which indicates different spatial locations, they use two coordinate systems that lead to shifts between the projection of objects onto the thermal images and visible images. The purpose of calibration is to derive the calibration matrix between the thermal camera and visible camera coordinate systems, based on which the thermal and visible images can be registered and aligned without any position and size differences. After the displacement errors between two types of images are eliminated, the bounding boxes of pin sites generated by the detection model can be simply transferred to the thermal images without shifts. Therefore, each camera containing an image sensor and a thermal sensor needs to be calibrated individually. In this work, both the registered visible images and the converted temperature

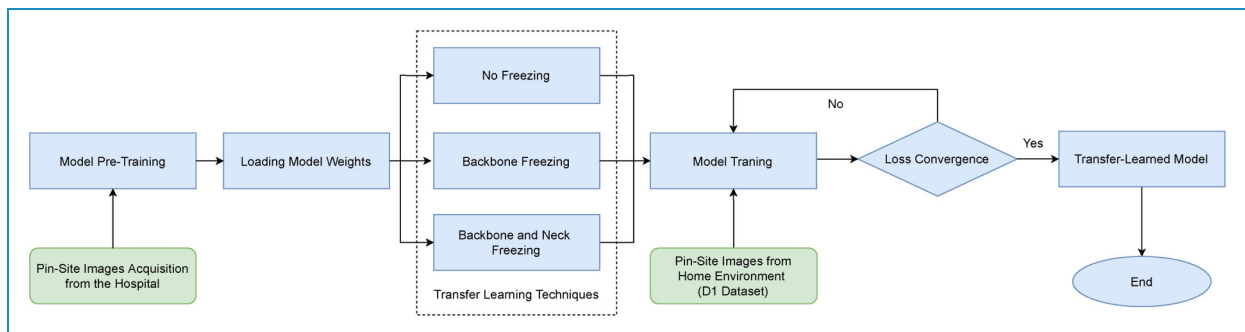


Figure 10. The processes of implementing transfer learning techniques.

values from the thermal sensor output are simply extracted from the FLIR Image Extractor tool <https://github.com/nationaldronesau/FlirImageExtractor>. This program for FLIR cameras provides functions for aligning thermal and visible images and for extracting the temperature values corresponding to visible pixels.

However, to further reduce the displacement error, we suggest two suitable calibration methods for thermographic images from the literature.

Geometric calibration of cameras has been extensively investigated in previous studies, for studies including computer vision. Zhang's calibration method²⁸ through a checkerboard is convenient and has been widely adopted. A three-dimensional (3D) point in the real world coordinate system can be denoted as $\mathbf{M}_w = [X_w, Y_w, Z_w, 1]^T$, with the augmented vector 1 added. Similarly, its corresponding two-dimensional (2D) pixels in visible image and thermal image can be represented by $\mathbf{m}_c = [u_c, v_c, 1]^T$ and $\mathbf{m}_t = [u_t, v_t, 1]^T$ in their pixel coordinate systems, separately. The relationship between the 3D and 2D points can be expressed by:

$$\begin{aligned} s_c \mathbf{m}_c &= s_c \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \mathbf{K}_c [\mathbf{R}_c | \mathbf{t}_c] \mathbf{M}_w \\ &= \mathbf{K}_c [\mathbf{r}_{c1} \quad \mathbf{r}_{c2} \quad \mathbf{r}_{c3} \quad \mathbf{t}_c] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned} \quad (9)$$

$$\begin{aligned} s_t \mathbf{m}_t &= s_t \begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} = \mathbf{K}_t [\mathbf{R}_t | \mathbf{t}_t] \mathbf{M}_w \\ &= \mathbf{K}_t [\mathbf{r}_{t1} \quad \mathbf{r}_{t2} \quad \mathbf{r}_{t3} \quad \mathbf{t}_t] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \end{aligned} \quad (10)$$

where s_c and s_t are scaling factors and \mathbf{K}_c and \mathbf{K}_t are intrinsic matrices of the thermal and visible cameras. $[\mathbf{R}_c | \mathbf{t}_c]$ and $[\mathbf{R}_t | \mathbf{t}_t]$ are extrinsic matrices of the thermal and visible cameras including the rotation parameters \mathbf{R}_c and \mathbf{R}_t as well as translate parameters \mathbf{t}_c and \mathbf{t}_t , and \mathbf{r}_{ci} and \mathbf{r}_{ti} represent the i^{th} column of the matrix \mathbf{R}_c and \mathbf{R}_t , respectively. Assume that the plane of the checkerboard for calibration is $Z_w = 0$ of the world coordinate system, so each calibration point on it has $Z_w = 0$, and hence the equations above can be simplified as:

$$\begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = s_c^{-1} \mathbf{K}_c [\mathbf{r}_{c1} \quad \mathbf{r}_{c2} \quad \mathbf{t}_c] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = \mathbf{H}_c \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (11)$$

$$\begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} = s_t^{-1} \mathbf{K}_t [\mathbf{r}_{t1} \quad \mathbf{r}_{t2} \quad \mathbf{t}_t] \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} = \mathbf{H}_t \begin{bmatrix} X_w \\ Y_w \\ 1 \end{bmatrix} \quad (12)$$

where the 3×3 homography matrices \mathbf{H}_c and \mathbf{H}_t define the relation between the point of the real world and the pixels of the images. By annotating at least 4 calibration points on the checkerboard with known 3D coordinates and detecting their projected pixels with 2D coordinates on the images, the calibration matrices \mathbf{H}_c and \mathbf{H}_t of the thermal and visible cameras can be calculated through OpenCV. The standard calibration method would have to derive the intrinsic matrix and extrinsic matrix, but for our case, we only require the homography \mathbf{H}_c and \mathbf{H}_t to obtain the transformation matrix between coordinate systems of thermal and visible cameras for registration. The visible images can be registered with the thermal images by the function:

$$\begin{bmatrix} u_t \\ v_t \\ 1 \end{bmatrix} = \mathbf{H}_t \mathbf{H}_c^{-1} \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = \mathbf{W} \begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} \quad (13)$$

where $\mathbf{W} = \mathbf{H}_t \mathbf{H}_c^{-1}$ is the transformation matrix for image registration.

However, the conventional method must rely on a checkerboard and its cross dots as the image and world points, which cannot be detected by thermal cameras and thus is not applicable directly. To address this problem, the calibration of thermal images includes a step to design patterns with clear feature points that can be recognized by infrared cameras before applying conventional calibration algorithms. As Method 1 shown in Figure 11, it leverages a foam metal plate³² made of cold and black metal aluminum disks inserted into a pink foam board at the ambient temperature. Thus, it consists of round dots that vary in temperature from that of the plate as thermographic patterns, and hence both the RGB visible camera and the infrared thermal camera can capture the recognizable features of the board. As a result, it is now possible to apply the conventional calibration technique introduced above²⁸ to extract the homography matrices \mathbf{H}_c and \mathbf{H}_t of the visible and thermal cameras, which is implemented by OpenCV. The transformation matrix \mathbf{W} , calculated based on homography matrices, converts every pixel on the visible images from the coordinate system of the visible camera to the coordinate system of the thermal camera as shown in the equation (13), so the visible images are registered with thermal images and the shifts are removed. The blank pixels of the image after registration as well as redundant parts of the image will be cut off to adjust the size of both types of images to be consistent.

An alternative³³ simplifies the calculation of the transformation matrix for calibration and also requires a foam metal board as a checkerboard. Illustrated in Figure 11 as Method 2, pixels are translated by calculating a root mean

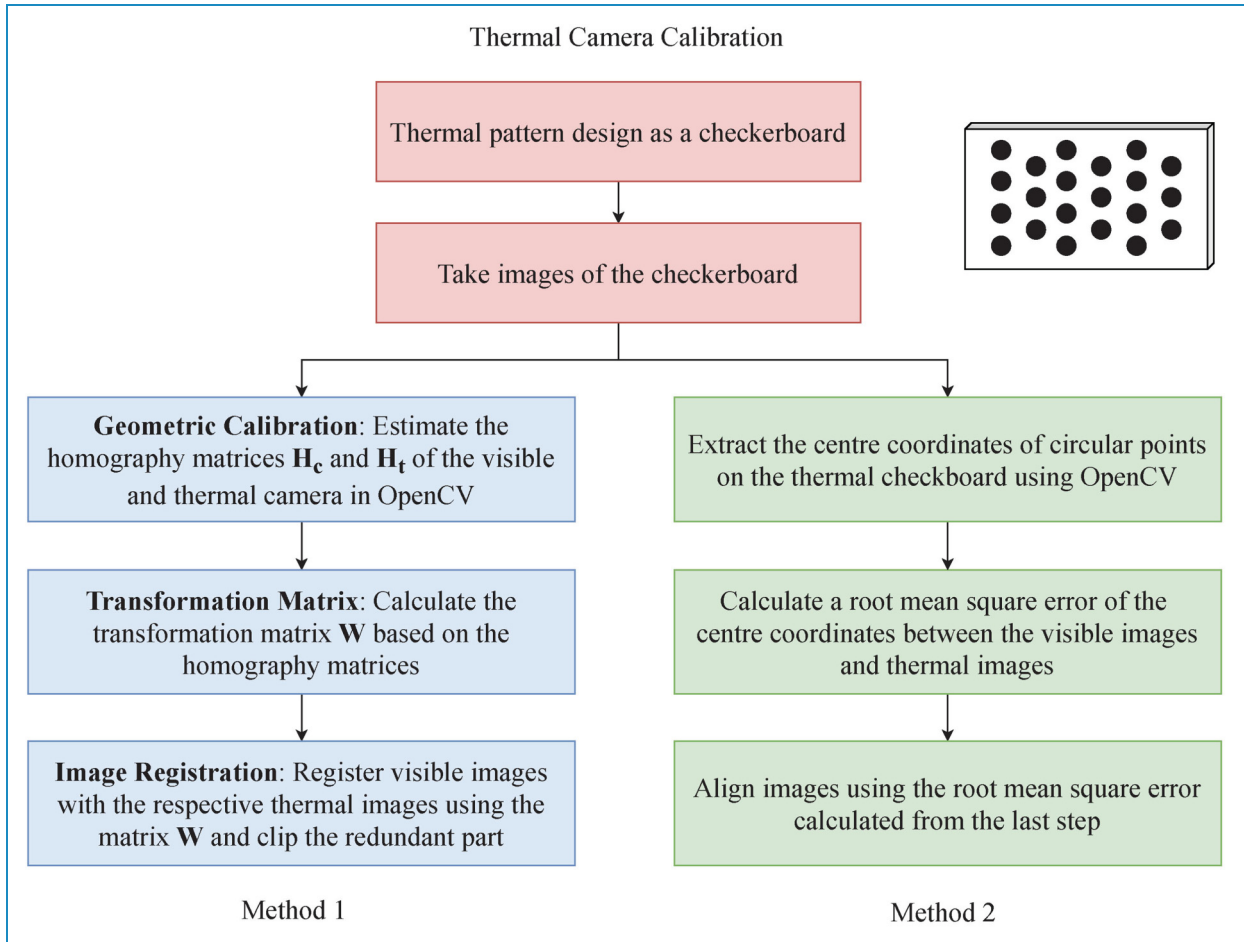


Figure 11. Two suggested methods for thermal camera calibration.

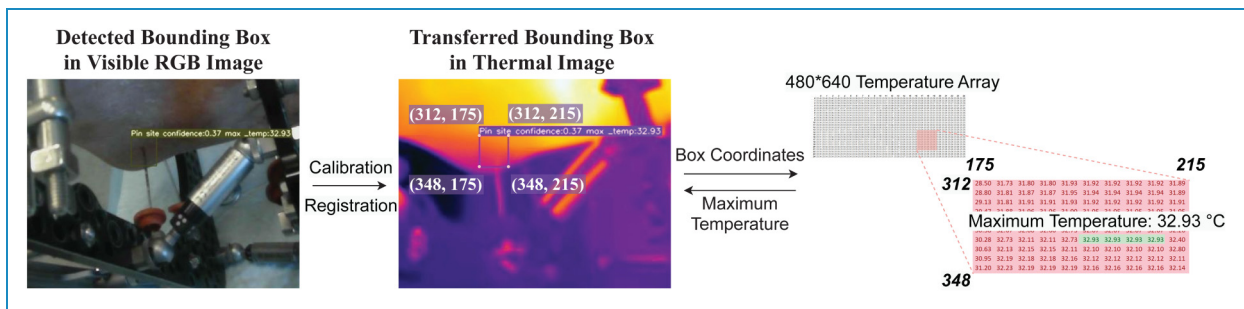


Figure 12. Maximum temperature extraction.

square error (RMSE) of the centroids of the metal aluminum disks between the RGB visible image and infrared thermal images. It is feasible as the thermal and visible cameras are on the same plane and vertical line as the FLIR cameras we used in this work.

Maximum temperature extraction. The maximum temperature around each pin site region might be significant for the infection assessment according to our pilot research. As a result, the last step of the model is to automate the

process of extracting the maximum temperature from the detected bounding box, which is shown in Figure 12. The pixel-level temperature values are directly readable from the FLIR images, as these follow the exchangeable image file format standard. After the bounding boxes are transferred to the calibrated thermal images, their coordinates can directly be obtained. As pin site regions on the temperature array are determined by the coordinates, the maximum temperature of each pin site can be calculated accordingly. As presented in Figure 12, the maximum temperature

within the detected bounding box of the pin site is extracted based on the coordinates of 4 corners and relevant information in the temperature array of the thermography.

Model configuration and evaluation methods

The pin site detection model training is set as the YOLOv5-small configuration and carried out according to the processes shown in Figure 7. As the default anchor box setting for the COCO data set is used for the pin site data set as the best possible recall value for pin site annotations reaches 99% of the COCO data set. SGD optimizer is set with momentum 0.937 and weight decay 0.0005. The initial learning rate lr is 0.01 and IoU training threshold is 0.2. In addition, the anchor threshold is set as 4.0 to decide which anchor the pin site is associated with. With the settings, the pin site detection model is pre-trained using the large data set from hospital environments and then fine-tuned by the small practical data set from home environments (D1 data set) using transfer learning techniques. Finally, the output of the model is filtered by non-maximum suppression (NMS) to generate the detection results.

To evaluate the performance of the model for pin site detection, the basic statistics are observed on the test data set like true positive (TP), false positive (FP), false negative (FN), representing true detections, missed detections and false detections, separately. In addition, inference time for the analysis of detection speed and advanced metrics including precision (P), recall (R), F_1 -score and average precision (AP) are employed, whose calculation functions are defined by:

$$\text{Precision : } P = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (14)$$

$$\text{Recall : } R = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (15)$$

$$F_1 = \frac{2 \times P \times R}{P + R}, \quad (16)$$

$$\text{AP} = \frac{1}{101} \sum_{R \in \{0,0.01,0.02,\dots,1\}} \max_{R' \geq R} P(R'). \quad (17)$$

The precision indicates the ratio of correct detection and recall defines the percentage of positives found by the model. Furthermore, AP is calculated by the area under the precision-recall rate curve, $P(R')$, which is defined by the precision at recall R' . Note that as there is only one object class in our target data set, the mean AP (mAP) is equal to AP.

Results and discussion

Basic model performance

To evaluate the basic performance in an ideal setting (hospital environment), we compare the proposed pin site detection model (YOLOv5 with CDIoU) with other conventional

models such as FCOS and YOLOv4, which are trained and tested with the hospital data set. Their performance metrics are shown and compared in Table 3.

YOLOv4 vs YOLOv5. It can be found that YOLOv5 object detectors lead to much faster operation and higher detection precision than the models based on YOLOv4. The detected boxes of pin sites generated by YOLOv5 models with CDIoU, DIoU and CIoU have a high precision from 0.97 to 0.98, outperforming the anchor-free FCOS model and YOLOv4 models by around 0.05. Similarly, mAP and F_1 score results also indicate that YOLOv5 models excluding the FCOS method still exhibit better detection accuracy, improving the mAP and F_1 by about 0.02 compared to YOLOv4 models. YOLOv5 has proven its convincing ability to identify and label all required pin sites from the images with excellent accuracy, and its improvements in accuracy over YOLOv4 do not come at the cost of complex computation. While YOLOv4 is implemented in Darknet, YOLOv5 is developed with PyTorch, which makes it light-weight and fast as a real-time pin site detection model. In our case, the calculation YOLOv5 models with DIoU, CIoU, and CDIoU require around 16.8 mega floating-point operations per second (MFLOPS) for pin site detection from one input image, which takes approximately 1.8 ms.

Besides, we have examined the anchor-free model using the YOLOv5 structure incorporated with FCOS. However, it performs worse than the traditional anchor-based frameworks YOLOv5 with CDIoU with obvious reduction in the recall rate and F_1 score and is therefore not recommended.

Table 3. Performance metrics of YOLOv5 and YOLOv4 pin site detection models with CDIoU, DIoU, CIoU or YOLOv5 + FCOS on the hospital data set.

Models	Hospital Data set			
	Precision	Recall	F_1	mAP [0.5]
YOLOv5 + CDIoU	0.976	0.949	0.96	0.961
YOLOv5 + FCOS	0.933	0.759	0.837	0.737
YOLOv5 + DIoU	0.97	0.94	0.96	0.957
YOLOv5 + CIoU	0.98	0.94	0.96	0.96
YOLOv4 + CDIoU	0.93	0.93	0.93	0.94
YOLOv4 + DIoU	0.92	0.96	0.93	0.938

mAP: mean average precision; CDIoU: control distance intersection over union; DIoU: distance-IoU; CIoU: complete-IoU; YOLO: you only look once; FCOS: fully convolutional one-stage.

Loss functions. The proposed combination of object detection framework and loss function, YOLOv5 with CDIoU, contributes to the leading metrics in the tests. More TP bounding boxes can be detected with the CDIoU loss, achieving an outstanding precision of 0.976 and nearly 0.1 higher recall rate than DIoU and CIoU. It is because the CDIoU loss can represent the difference in central distance and aspect ratio more comprehensively. Furthermore, in the NMS process, two or more predicted bounding boxes generated by DIoU and CIoU models with similar center points could have the same high IoU values in an image, which causes random selection of proposals and hence a decline in the recall rate. However, the model using CDIoU would result in different IoU values for NMS due to another loss calculation method considering both the centroid and aspect ratio. As a result, some proposals randomly discarded with DIoU and CIoU methods would be picked by using the CDIoU approach instead, so the recall rate is benefited. Such an improvement does not come at the cost of computational time, as the model using CDIoU stays a similar inference time to models with DIoU and CIoU at around 1.8 ms.

Improvements through transfer learning

With the satisfactory results in the test, the models based on the YOLOv5 framework (without transfer learning) are applied to more images from different scenarios in the D1 data set to verify their performance. However, as presented in Table 4, the models with good metrics in the previous test obtain relatively poor detection precision in applications for patients. The mAP drops to around 0.3 with an AP of 0.58, using YOLOv5 models for patients. It could be deduced that patients lacking specialized skills lower the quality of images, which creates a challenge for the models trained by the high-quality images.

To address this issue, we take advantage of transfer learning to improve the pin site detection accuracy for the home data set without the need for a large amount of training data. A transfer learning technique based on parameter sharing is employed, trained, and tested using the D1 data set. We have an experiment on three strategies of defining sharing parameters for our targeted data set, including freezing backbone, freezing neck and backbone, and freezing nothing. Obviously, the no-freezing method is chosen for its fastest convergence rate based on the loss curve and highest mAP value with the same number of epochs. Finally, the finetuned model using YOLOv5 and CDIoU and freezing no layers in transfer learning achieves an mAP of 0.67 with high precision as shown in Table 5, which is greatly improved compared to the results before introducing transfer learning in Table 4.

In addition, the models based on YOLOv5 after transfer learning are further evaluated by another Data set D2, which contains low-quality images intentionally collected

Table 4. Performance metrics of YOLOv5 pin site detection models with CDIoU, DIoU or CIoU trained by the hospital data set and tested by the data set D1 from patients at home.

Models	Home data set D1				
	Precision	Recall	F_1	mAP [0.5]	Inference time, ms
YOLOv5 + CDIoU	0.58	0.42	0.48	0.31	2.5
YOLOv5 + DIoU	0.55	0.41	0.47	0.29	2.4
YOLOv5 + CIoU	0.63	0.41	0.5	0.32	2.9

CDIoU: control distance intersection over union; DIoU: distance-IoU; CIoU: complete-IoU; YOLO: you only look once; FCOS: fully convolutional one-stage.

Table 5. Performance comparison of the YOLOv5 and CDIoU pin site detection model incorporated with TL methods using the home data set D1.

TL methods	Home Data set D1				
	Precision	Recall	F_1	mAP [0.5]	Inference time, ms
No freezing	1	0.66	0.79	0.67	6.8
Backbone freezing	0.76	0.66	0.71	0.61	9.8
Backbone and neck freezing	0.68	0.54	0.6	0.41	10.2

mAP: mean average precision; CDIoU: control distance intersection over union; TL: transfer learning; YOLO: you only look once.

from patients at home. The metrics of their performance are concluded in Table 6 and compared to the models with the same object detection architectures but without transfer learning. There are considerable improvements in precision and recall rate by 0.1–0.11 and 0.13–0.14, separately, through implementing the transfer learning technique. And generally, the introduction of transfer learning increases the mAP scores by at least 0.17. The CDIoU loss incorporated with transfer learning still brings supreme performance, including detection accuracy and speed rather than DIoU and CIoU. It earns 0.01 higher recall rate and 0.008 better mAp score than other models while the inference time is 0.1 ms faster. However, the final pin site detection precision of 0.7 and recall rate at 0.58 in a new data set D2 is not as good as the performance with the test data set. To gain the generalization ability over a wide range of data collected in various environments,

further development as well as a better design of model structure and parameters is necessary. The detected pin site results with red bounding boxes on images in the hospital data set and home data set are visualized in Figure 13 and Figure 14, respectively, which demonstrate the detection capability of the proposed YOLOv5 model with CDIoU and finetuned by transfer learning.

Results visualization

Pin site detection in thermography. Figure 15 illustrates the results of detected pin sites on thermography. The proposed pin site detection model based on YOLOv5 is firstly applied to the visible images in Figure 15(a). Then, the generated bounding boxes are transferred to the corresponding thermal images and illustrated as pin sites in Figure 15(b), thanks to the calibration of visible and thermal cameras

Table 6. Performance metrics of YOLOv5 pin site detection models with and without TL using the data set D1 and tested on the home data set D2 from patients.

Models	Home data set D2				Inference time, ms
	Precision	Recall	F_1	mAP [0.5]	
TL (D1) + YOLOv5 + CDIoU	0.7	0.58	0.64	0.518	1.9
TL (D1) + YOLOv5 + DIoU	0.69	0.57	0.63	0.51	2.0
TL (D1) + YOLOv5 + CIoU	0.7	0.57	0.63	0.5	2.2
YOLOv5 + CDIoU	0.6	0.44	0.51	0.34	1.8
YOLOv5 + DIoU	0.59	0.44	0.5	0.36	2.1
YOLOv5 + CIoU	0.59	0.44	0.5	0.34	2.1

mAP: mean average precision; CDIoU: control distance intersection over union; TL: transfer learning; YOLO: you only look once.

and registration of images in advance. Although it is not possible to directly apply object detection models to thermal images where pins of EFs are not clearly visible, our proposed method leverages object detection model on visible images and camera calibration methods to find out pin sites and project detected bounding boxes onto the thermal images. Primarily detecting pin sites on visible images could have higher accuracy and calibration minimizes the displacement errors between cameras to transfer bounding boxes, contributing to the precise annotation of pin sites on thermal images.

Maximum temperature extraction. Based on the regions of interest for all pin sites generated by the model, the maximum temperature can be easily measured for infection evaluation. Figures 16 and 17 show maximum temperature values for some pin sites in data sets T1 and T2, respectively. These results are obtained by applying the proposed pin site detection model to visual images and extracting from the aligned bounding boxes on their corresponding thermal images. Not limited to extracting maximum temperatures within the annotated area, more analysis methods are possible to be performed to the extracted data from the regions, such as deriving a function of temperature and area.³⁴ We believe further studies on infection could be efficiently carried out with the help of our proposed model.

Future work

The key task in our future work is to apply the pin site detection tool for studies focusing on infection detection. The temperatures at the pin site are associated with the early signs of wound inflammation and infection, and a thermography-based analysis might be an efficient tool to detect it.^{2,34} We will extend the size of our data set, and the developed pin site detection will be applied to automate the process of annotation to explore possible patterns of infection related to thermal imaging.

In addition, the results presented above demonstrate the general performance of the automatic pin site detection tool and indicate that it is ready for application in the research in infection detection for annotation of regions of interest.

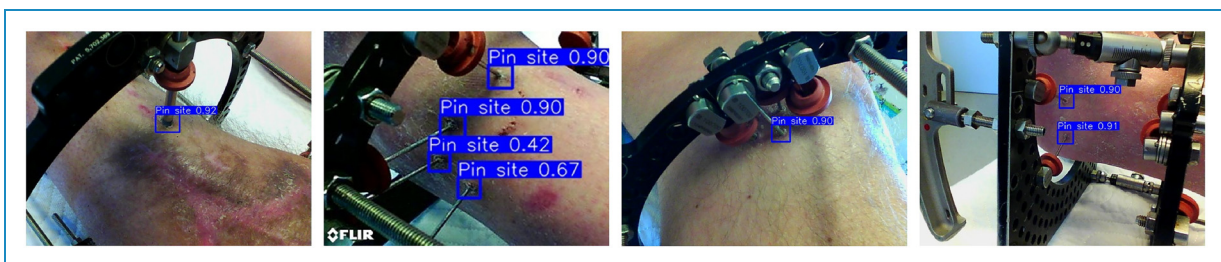


Figure 13. Pin site detection results in bounding boxes with confidence scores, generated by the proposed model using YOLOv5 with CDIoU on the hospital data set. YOLO: you only look once; CDIoU: control distance intersection over union.

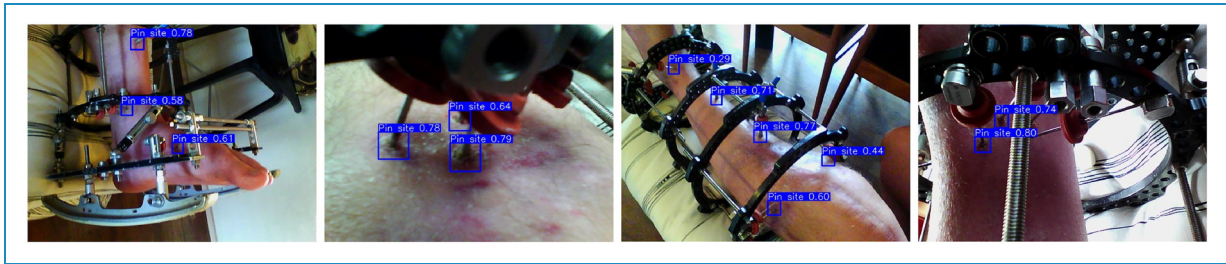


Figure 14. Pin site detection results in bounding boxes with confidence scores, generated by the model pre-trained using YOLOv5 with CDIoU and finetuned by transfer learning on the home data set. YOLO: you only look once; CDIoU: control distance intersection over union.

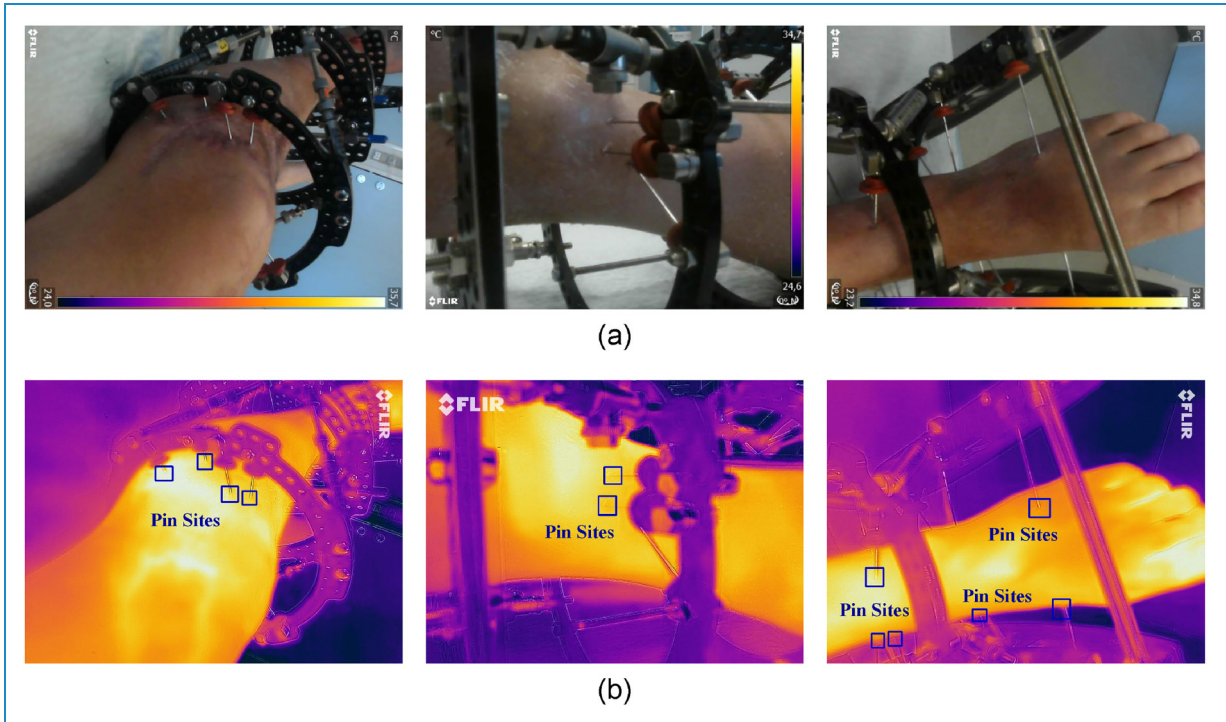


Figure 15. Applications of the pin site detection model on the thermal images: (a) original visible images, (b) infrared thermal images with pin site bounding boxes generated by the proposed model on the original visible images.

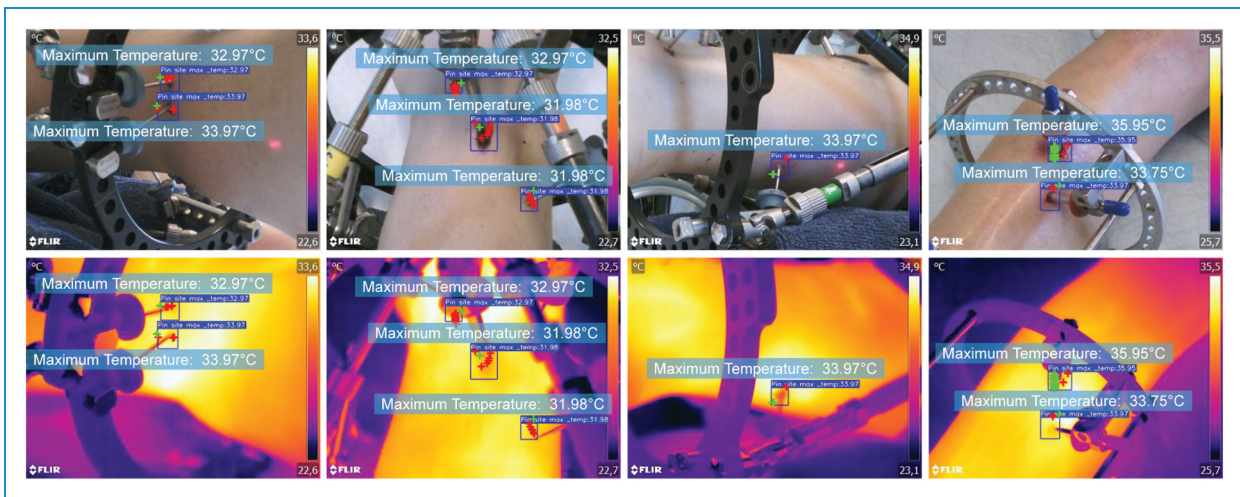


Figure 16. Measurement of maximum temperature in T1 data set.

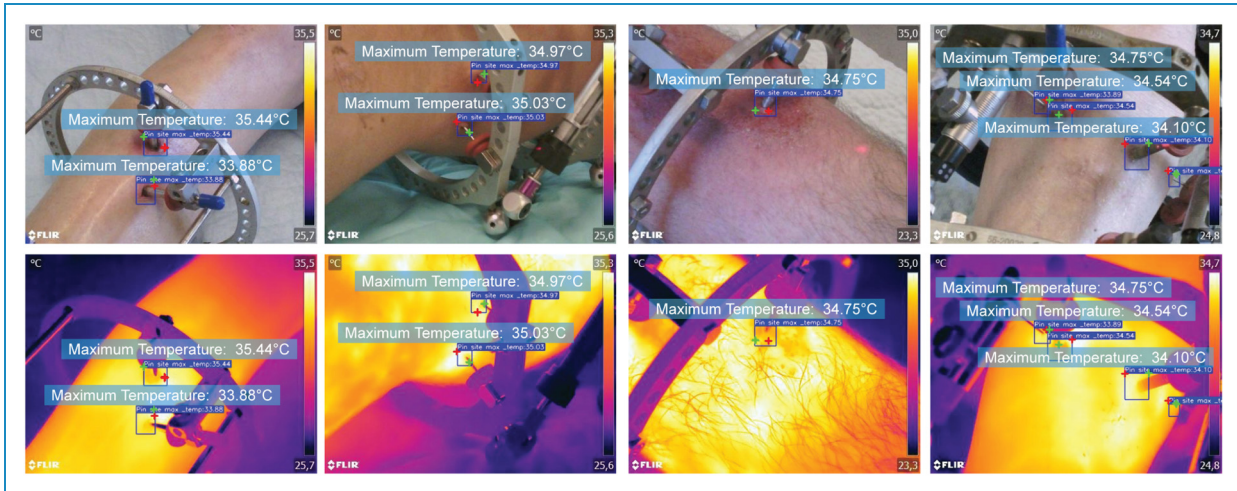


Figure 17. Measurement of maximum temperature in T2 data set.

However, we notice that there is some room for improvement in our future work. First of all, the data sets in this work are collected from three different cameras while we were exploring the choice of cameras with the consideration of factors such as resolution, temperature accuracy, operating distance, costs and operational difficulties, which may be considered as weaknesses to the study. On the other hand, the use of diverse types of cameras and the different conditions, such as obtained by hospital staff or the patient himself, could introduce a strength to the proposed model since its generalization has been further expanded under different conditions with variations in image specifications. In future infection studies, the appropriate type of thermal camera will be selected for collecting the extended data set after careful assessment of those factors. Besides, a wider diversity of data related to pin site detection studies, such as skin tones, infection level and fracture type, will be collected and recorded in the future. The current pin site detection model has been trained primarily on images of light skin tones, as it reflects the predominant skin tones of actual patients from Scandinavian countries in our clinic. However, diversity is warranted, and more skin tones are essential for future refinement of the model. As for camera calibration and image registration, we simplify this step and obtain the registered visible images and temperature values from the thermal images with the help of the FLIR Image Extractor tool due to the limitation of materials. To transfer the region of interest to the thermal image more precisely, one of the two recommended calibration methods described in the ‘Method section will be implemented.

Conclusion

In this paper, we present a pin site detection model preparing for exploring a new technique of infection detection

through thermography to monitor patients after lower limb external fixation. The challenge of annotating pin sites on thermal images has been overcome by applying the deep learning-based object detection model to the respective visible images and registering them with their corresponding thermal images. For locating pin sites on visible images, YOLOv5 object detection architecture is selected in this work, incorporated with a novel loss regression method by CDIoU and optimized by transfer learning to be adapted for patient applications in practical environments. Thanks to image registration, the detected bounding boxes can be transferred to the thermal images without losing accuracy and are ready for further temperature analysis. We have applied the proposed model to the data set, and the expected pin site regions and temperature values are successfully annotated and detected.

The development of the pin site detection model is significant to the future goal of studying the relationship between temperature and pin site infection in patients with EFs with the following advances.

1. Provide an automatic pin site detection tool to extract targeted temperature information from a large number of thermal images, improving the efficiency of necessary data collection.
2. Ensure the accuracy of marking pin sites by the conjoint utilization between visible images and thermal imaging through camera calibration and image registration.
3. Optimize the YOLOv5 object detection architecture for pin site detection in general scenarios, which is achieved by integrating the CDIoU loss function and cooperating with the transfer learning algorithm.
4. Pave the way for our further research to reveal the underlying patterns of infection detection through thermography.

Being able to extract targeted temperature information from a large number of images automatically, this time-consuming work is therefore simplified and necessary data for analysis would be available. In addition, we leverage the object detection model, YOLOv5, in this work for pin site detection. Instead of directly using the original model, the selection of loss functions is considered for potential optimization towards our aimed applications. Our experiment indicates CDIoU outperforms other conventional methods and is selected. It also verifies the effectiveness of the CDIoU loss function integrated into the YOLOv5 architecture in practical applications. Besides, the transfer learning algorithm is incorporated, and the motivation comes from our experiment results that the YOLOv5 pin site detection model trained in a hospital data set has a reduced accuracy when applied to a patient's home environment. Despite the limitation of the size of the home scenario data set, after being finetuned transfer learning, the model manages to increase precision and recall rate by 0.1 and 0.14, respectively, showing the adaptability in the home scenario. In general, compared to the original model, for the data set from home scenarios, the introduction of CDIoU loss function and transfer learning improves the accuracy performance and maintains inference time similarly. The experimental results in this paper have demonstrated the desired capability as well as the performance of the proposed pin site detection model that satisfies the requirements of detecting pin sites above.

By calibrating and registering the thermal camera and visible camera, the detected bounding boxes of pin sites on the visible images can be transferred to the thermal images for temperature analysis and infection detection. Once we could determine the assessment criteria of pin site infection on thermal images, the patients could use their mobile phones with the proposed pin site detection model for remote monitoring every day at home. In total, this paper exhibits the first work applying an object detection model on thermographic images for pin site detection, which paves the way for our future research on infection analysis with temperature information.

Acknowledgements: The authors would like to thank the orthopedic patients who provided us with image data in this research.

Contributorship: SA and QH researched literature and wrote the first draft of the manuscript. SA designed the model and analysed data. MF and SA were involved in the data collection. All authors reviewed and edited the manuscript and approved the final version of the manuscript.


Conflict of interests: The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.


Ethical approval: Ethical approval was not sought for the present study because the data we collected is not sensitive, has no ethical risks and is completely anonymous without personal information.


Funding: The authors received the following financial support for the research, authorship, and/or publication of this article: This work was supported by Innovation Fund Denmark (Project No.: 0161-00023).

Guarantor: Ming Shen.

ORCID iDs: Sowmya Annadatha  <https://orcid.org/0000-0002-0382-4019>

Qirui Hua  <https://orcid.org/0000-0002-3265-6683>

Søren Kold  <https://orcid.org/0000-0002-3387-1473>

Ole Rahbek  <https://orcid.org/0000-0002-5602-4533>

References

- Potgieter MS, Pretorius HS, Preez GD et al. Complications Associated with Hexapod Circular Fixation for Acute Fractures of the Tibia Diaphysis: A Retrospective Descriptive Study At a High Volume Trauma Centre. *Injury* 2020; 51(2): 516–521. DOI: 10.1016/j.injury.2019.11.012.
- Rahbek O, Husum HC, Fridberg M et al. Intrarater Reliability of Digital Thermography in Detecting Pin Site Infection: A Proof of Concept Study. *Strateg Trauma Limb Reconstr* 2021; 16(1): 1–7. DOI: 10.5005/jp-journals-10080-1522.
- Obinah MPB, Nielsen M and Hlmich LR. High-end Versus Low-end Thermal Imaging for Detection of Arterial Perforators. *Plast Reconstr Surg Glob Open* 2020; 8(10): 3175. DOI: 10.1097/GOX.0000000000003175.
- Cho Y, Bianchi-Berthouze N and Julier SJ. DeepBreath: Deep learning of breathing patterns for automatic stress recognition using low-cost thermal imaging in unconstrained settings. In *2017 Seventh international conference on affective computing and intelligent interaction (acii)*. pp. 456–463. doi:10.1109/ACII.2017.8273639.
- Roslidar R, Saddami K, Arnia F et al. A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (acii)*. pp. 77–81. doi:10.1109/CYBERNETICSCOM.2019.8875661.
- Filippini C, Cardone D, Perpetuini D et al. Convolutional Neural Networks for Differential Diagnosis of Raynaud's Phenomenon Based on Hands Thermal Patterns. *Appl Sci* 2021; 11(8): 36–14. DOI: 10.3390/app11083614.
- Taigman Y, Yang M, Ranzato M et al. DeepFace: Closing the gap to human-level performance in face verification. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1701–1708. doi:10.1109/CVPR.2014.220.
- Chen C, Seff A, Kornhauser A et al. Deepdriving: Learning affordance for direct perception in autonomous driving. In *2015 IEEE International Conference on Computer Vision (ICCV)*. pp. 2722–2730. doi:10.1109/ICCV.2015.312.
- Litjens G, Kooi T, Bejnordi BE et al. A Survey on Deep Learning in Medical Image Analysis. *Med Image Anal* 2017; 42: 60–88. DOI: 10.1016/j.media.2017.07.005.

10. Felzenszwalb P, McAllester D and Ramanan D. A discriminatively trained, multiscale, deformable part model. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 1–8. doi:10.1109/CVPR.2008.4587597.
11. Xiao Y, Tian Z, Yu J et al. A Review of Object Detection Based on Deep Learning. *Multimed Tools Appl* 2020; 79(33): 23729–23791. DOI: 10.1007/s11042-020-08976-6.
12. Filko D, Antoni D and Huljev D. Wita application for wound analysis and management. In *The 12th IEEE International Conference on e-Health Networking, Applications and Services*. pp. 68–73. doi:10.1109/HEALTH.2010.5556533.
13. Anisuzzaman DM, Wang C, Rostami B et al. Image-based Artificial Intelligence in Wound Assessment: A Systematic Review. arXiv:2009.07141, 2020.
14. Redmon J, Divvala S, Girshick R et al. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 779–788. doi:10.1109/CVPR.2016.91.
15. Girshick R, Donahue J, Darrell T et al. Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 580–587. doi:10.1109/CVPR.2014.81.
16. Girshick R. Fast R-CNN. In *2015 IEEE International Conference on Computer Vision*. pp. 1440–1448. doi:10.1109/ICCV.2015.169.
17. Ren S, He K, Girshick R et al. Faster R-CNN: Towards Real-time Object Detection with Region Proposal Networks. *IEEE Trans Pattern Anal Mach Intell* 2017; 39(6): 1137–1149. DOI: 10.1109/TPAMI.2016.2577031.
18. Zhao ZQ, Zheng P, Xu ST et al. Object Detection with Deep Learning: A Review. *IEEE Trans Neural Netw Learn Syst* 2019; 30(11): 3212–3232. DOI: 10.1109/TNNLS.2018.2876865.
19. Bochkovskiy A, Wang CY and Liao HYM. YOLOv4: Optimal speed and accuracy of object detection. 2004.10934, 2020.
20. Jocher G, Stoken A, Borovec J et al. ultralytics/yolov5: v5.0 - YOLOv5-P6 1280 models, AWS, Supervisely and YouTube integrations, 2021. doi:10.5281/zenodo.4679653.
21. Rezatofighi H, Tsoi N, Gwak J et al. Generalized intersection over union: A metric and a loss for bounding box regression. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 658–666. doi:10.1109/CVPR.2019.00075.
22. Zheng Z, Wang P, Liu W et al. Distance-IoU Loss: Faster and Better Learning for Bounding Box Regression. *Proceedings of the AAAI Conference on Artificial Intelligence* 2020; 34(07): 12993–13000. DOI: 10.1609/aaai.v34i07.6999.
23. Chen D and Miao D. Control distance IoU and control distance IoU loss function for better bounding box regression. arXiv:2103.11696, 2021.
24. Zhuang F, Qi Z, Duan K et al. A Comprehensive Survey on Transfer Learning. *Proceedings of the IEEE* 2021; 109(1): 43–76. DOI: 10.1109/JPROC.2020.3004555.
25. Long M, Cao Y, Wang J et al. Learning transferable features with deep adaptation networks. In *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. pp. 97–105.
26. Maqsood M, Nazir F, Khan U et al. Transfer Learning Assisted Classification and Detection of Alzheimer's Disease Stages Using 3D MRI Scans. *Sensors* 2019; 19(11): 2645. DOI: 10.3390/s19112645.
27. Byra M, Wu M, Zhang X et al. Knee Menisci Segmentation and Relaxometry of 3D Ultrashort Echo Time Cones MR Imaging Using Attention U-Net with Transfer Learning. *Magn Reson Med* 2020; 83(3): 1109–1122. DOI: 10.1002/mrm.27969.
28. Zhang Z. A Flexible New Technique for Camera Calibration. *IEEE Trans Pattern Anal Mach Intell* 2000; 22(11): 1330–1334. DOI: 10.1109/34.888718.
29. Choinowski A, Dahlke D, Ernst I et al. Automatic calibration and co-registration for a stereo camera system and a thermal imaging sensor using a chessboard. In *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences-ISPRS Archives* doi:10.1109/34.888718.
30. Cardone D, Spadolini E, Perpetuini D et al. Automated Warping Procedure for Facial Thermal Imaging Based on Features Identification in the Visible Domain. *Infrared Phys Technol* 2021; 112: 103595. DOI: 10.1016/j.infrared.2020.103595.
31. Tian Z, Shen C, Chen H et al. FCOS: Fully convolutional one-stage object detection. In *2019 IEEE/CVF International Conference on Computer Vision*. pp. 9626–9635. doi:10.1109/ICCV.2019.00972.
32. Zhang H. Multiple methods of geometric calibration of thermal camera and a method of extracting thermal calibration feature points, 2018. <https://henryzh47.github.io/Thermal-Camera-Calibration/>.
33. Yang MD, Su TC and Lin HY. Fusion of Infrared Thermal Image and Visible Image for 3D Thermal Model Reconstruction Using Smartphone Sensors. *Sensors* 2018; 18(7): 2003. DOI: 10.3390/s18072003.
34. Ramirez-GarciaLuna JL, Bartlett R, Arriaga-Caballero JE et al. Infrared Thermography in Wound Care, Surgery, and Sports Medicine: A Review. *Front Physiol* 2022; 13: 838528. DOI: 10.3389/fphys.2022.838528.