Detection of perforators for DIEP flap breast reconstruction using deep learning on thermal images.

by Gunther Steenackers*, Warre Clarys*, Jan Verstockt*, Edgar Cardenas*, Filip Thiessen**

* *University of Antwerp, InViLab research group, Groenenborgerlaan 171, B-2020 Antwerp, Belgium;*
** *Gynecological Oncology Unit, Department of Obstetrics and Gynecology, Multidisciplinary Breast Clinic, Antwerp University Hospital, University of Antwerp, Antwerp, Belgium.*
Correspondence: gunther.steenackers@uantwerpen.be

Abstract

As the global cancer burden is expected to rise by 47% from 2020 to 2040, cost-effective diagnostic and therapeutic strategies are imperative to mitigate the increasing incidence and prevalence of breast cancer. As part of improving the success rate of breast reconstruction surgery, this research investigates the implementation of artificial intelligence for optimal perforator selection to assist the surgeon. This paper describes a complete overview of the methodology, including classification, detection-based models, and segmentation models. In the case of perforator mapping, segmentation models tend to be more appropriate as the loss functions indicate more steady convergence over time.

1. Introduction

Breast reconstruction is a critical component of the treatment process for breast cancer patients. In autologous flap reconstructions, a separation can be made between free flaps and pedicled flaps. This study focuses specifically on free flap reconstruction techniques. The emphasis will be on the DIEP flap method, as it is the predominant reconstructive technique used for 32% of autologous breast reconstructions [4], [5]. In general, the reconstruction procedure involves the removal of skin and fat from the abdomen, which is then transplanted to the breast area. A perforator is defined as a vessel that perforates an envelope of the target tissue to be transferred [6]. To reduce operative time, minimize complications and ensure an overall better result, it is crucial to select the most suitable perforator [5]. The initial step of the selection procedure is locating all the perforators in the abdomen. To identify the optimal tissue in the abdominal area, it is essential to locate the perforators. These perforators supply the skin and subcutaneous fat with blood, making the surrounding area the most suitable tissue if the blood flow is sufficient. Since patients have multiple perforators, selecting the appropriate one is a challenging step during the surgical procedure. Using infrared imaging, it is possible to determine and map the location of perforators, which is often done preoperatively. To confirm the locations and optimize tissue selection, a similar technique can be used intraoperatively [1].

2. Deep learning

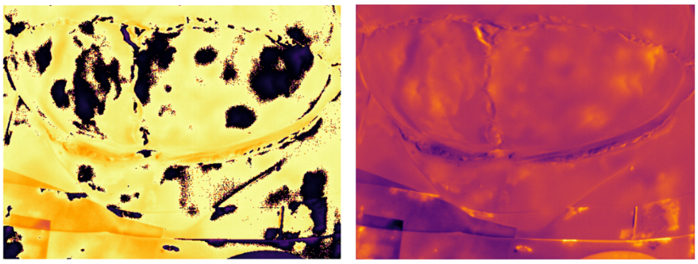
Selecting the ideal perforator for DIEP flap reconstruction, as previously mentioned, is of great importance within the realm of breast reconstruction procedures. This task demands significant time and precision on the surgeon's part, one has to be fulfilled with great knowledge of the subject. Using deep learning models, the surgeon's decision-making could be expedited by the model's provided input [3]. This is especially vital during surgery, due to the patient's heightened vulnerability to infections in the intraoperative setting. The ultimate goal is to transform the role of an AI model used as a helpful, practical tool for the surgeon into an accurate autonomous tool that can decide on the identification of the most ideal perforator within a split second, which the surgeon can quickly validate. Using an infrared camera, thermal images are made of the abdomen. These images can then be reviewed by the deep learning model to identify the most ideal perforator.

3. Methodology

In essence, the adopted methodology is as follows:
1. Preprocessing: At the beginning of the research unprocessed, thermographic HDF5 data files were provided and had to undergo a series of operations one by one.
2. Augmentation: Due to limited accessibility to data, the entire dataset consisted of 62 measurements. To obtain sufficient data for the development of an AI model, the dataset must be expanded.
3. CNN: As mentioned earlier in the literature study, a convolutional neural network will be used for the deep learning model. Two different approaches will be tested to determine the best structure that can be used for perforator detection.

3.1 Preprocessing data

The data used in this project, gathered using a thermal camera, was captured by Dr. Thiessen. Thermal images were taken during pre-, intra- and postoperative stages, meaning all the data can be classified in one of these categories. This was done using Python and Pandas in combination with Excel. Python was used to display every H5 file and its frames, and afterword's Pandas was used to save certain information (pre-, intra-, or postoperative; first- and last useful frame, coordinates of umbilicus …) into an Excel spreadsheet. Determining the first and last useful frame is an essential part of the entire process as these two frames will be used to build a representative ground truth. In this case, the first usable frame is the frame immediately following the cold challenge and is called the 'first useful frame'. The 'last useful frame' is the frame after reheating of the abdomen. To obtain a ground truth image, the first useful frame was subtracted from the last useful frame. To correctly subtract the frames, the temperature scales must be normalized. The temperature scale of the first and last useful frame should be similar to obtain a feasible result. The subtracted images before and after normalization are displayed in Figure 1.



(a) (b)

Fig. 1. *(a) Subtracted image without temperature normalization; (b) Subtracted image with temperature normalization.*

3.2 Data Augmentation

Expanding a dataset in general can be done in three ways: collecting additional data, data augmentation or generating synthetic data. Freely available and qualitative data is limited. As a result, the initial dataset only contained 62 measurements. However, this dataset included measurements from each phase of the DIEP flap reconstruction procedure. As the main objective of the AI model is supporting the surgeon in the decision-making for the optimal perforator, measurements during the postoperative and certain measurements during the intraoperative stage are not usable. (Unusable intraoperative measurements include measurements taken after removal of the flap.) After filtering the dataset, a usable dataset of 31 measurements remains. To make the most of the available data, intraoperative imaging is also used for the deep learning model due to the similarities in the pre-and intraoperative images. It has to be taken into account that these models will be less accurate for preoperative operation than models where only preoperative images are used. The effect of this has to be carefully monitored and the method has to be adapted to these findings. To execute data augmentation on the dataset, the library 'Albumentations' is imported into Python. The selected augmentation techniques presented in Table 1 were used to transform the dataset from 31 to 496 useful images. When applying augmentation techniques, two components are crucial to check:

- Ensuring that the size of the augmented images remains equal to the original image (640x480 pixels).
- Ensuring that the augmented data represents the underlying distribution of original data.

HoizontalFlip	RandomSizedCrop	Spatter
VerticalFlip	Rotate	GaussNoise
Flip	RandomRotate90	RandomScale
Crop	Sharpen	Resize
RandomCrop	RandomGamma	Transpose

Table 1. *Representation of adopted augmentation techniques.*

3.3 Convolutional neural network (CNN)

The approach is based on U-net architecture, which is a frequently used architecture in the biomedical scene. The most important parameters are:

- Learning rate = 0.0001. Determines the extent of adjustment to the weights, of the neurons in the network, based on the loss during the gradient descent process. The learning rate is a parameter that can be used for tuning the model.
- Optimizer = Adam. An iterative optimization algorithm that has been proven to work well across a wide range of deep learning architectures, including CNN. The algorithm combines the effect of gradient descent of 'Momentum' with gradient descent of 'RMSprop' to minimize the loss during the training of neural networks [2], [3].

- Amount of epochs. The number of epochs determines the amount of training the AI model will undergo. Large datasets in combination with a high number of epochs (e.g. 500) will result in long simulation times, and conversely, smaller datasets with fewer epochs may lead to shorter training times. In this research, the number of epochs was often limited to 15 due to a lack of computational power.
- Loss function. This function is used to evaluate the performance of the deep learning model and indicates whether or not the deep learning model is converging. By plotting both training and validation loss, the AI model can be validated.
- Data classification and distribution. The entire dataset is divided into three separate datasets: training, validating, and testing. A commonly used split size is 70 – 30, which is also adopted in this research. This means 70% of the dataset is assigned to training and the other 30% is equally distributed in validation and testing. The choice of split size depends on the amount of data. In case of more data, the split size can be taken on 90 -10.

The model uses U-net architecture in the deep learning environment PyTorch. The U-net architecture is widely used in medical applications for pattern and object detection. The name of this technique refers to the shape of the architecture. The architecture consists of two halves: the encoder which represents the left half of Figure 2 and the decoder which represents the right half of Figure 2.

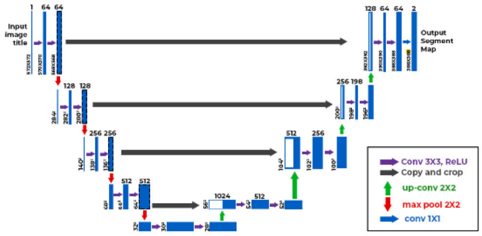


Fig. 2. *U-net architecture [26].*

The encoder extracts features from the captured input images and uses two types of layers: convolutional layers and max-pooling layers. A series of convolutional layers will detect certain patterns and features in the images. Afterwards, this information passes through the max-pooling layer, where it is down-sampled. This will reduce the information in the architecture and will significantly reduce the necessary computation power. The decoder will up-sample the features detected in the encoder, ultimately creating the final mask and output image. Here, the so-called “skip connections” are used, depicted as the black arrows in Figure 2. These skip connections help retrieve the lost information in the max-pooling layers of the encoder, resulting in a combination of low- and high-quality patterns and features used to build the final output [4], [5].

4. Classification and detection

For this approach, the subtracted image is also used as the input. The output is a mask containing three layers and a background, where every layer is represented by a number. Thus, the mask will be a matrix with an equal shape as the image and every cell of this matrix contains a number between zero and three. This matrix represents a PNG image, with the zero values being transparent in the final image. The background is represented by zeros, the general abdominal area is represented by ones, the umbilicus is represented by twos and the perforators are represented by threes. As the ground truth needs to be created for all the training images, a Python program was developed using thresholds based on RGB domains. Every pixel is evaluated and given a number from zero to three based on the color of that cell, with ones getting priority over twos and threes, and twos getting priority over threes. The number zero will only be appointed if the RBG value of the cell falls out of all the RGB domains.

5. Results and validation

The model used is a variation of the model presented in [4] and was partially enhanced and changed to fit the goals of this research more closely. For the model, a couple of parameters were tested to improve the results and to map their influence on the results. The main two parameters tested are the neural network's complexity and the amount of epochs used. Expanding the research could involve mapping other influences such as the weights distributed to the network and the amount and quality of the training data. The best results were obtained using a semi-complex neural network and at least 15 epochs, achieving better results as the number of epochs increased. After 15 epochs, the model returns an accuracy of around 92,2% or a loss of 7,8%, and the training process took about 10 minutes. As mentioned earlier, the model becomes more accurate with a higher number of epochs. This led to an increase in computational power as well as accuracy. Figure 1 illustrates an accuracy of 95,5% or a loss of 4,5% after 50 epochs.

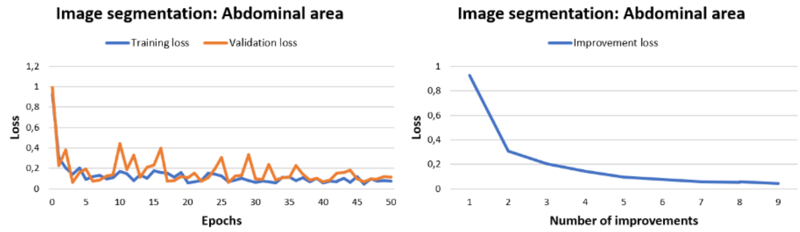


Fig. 3. *General loss of model (left) after 50 epochs; Improvement loss of model (right) after 50 epochs.*

To visualize and validate the actual accuracy of the model, three images can be compared. The first image is the subtracted input image, the second image contains the ground truth mask, and the last image displays the mask generated as output by the model. Figure 4 presents the visual validation for 28 images. Upon comparing the input and output masks, it is evident that the model is reasonably accurate but could benefit from further improvement through the utilization of additional epochs. This observation aligns with the validation loss in Figure 3.

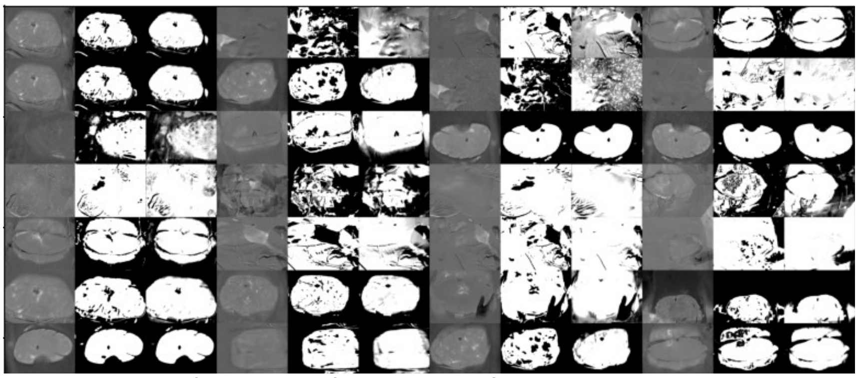


Fig. 4. *Visual validation of the model after 50 epochs for the general abdominal area. With the first, fourth, seventh and tenth columns showing the input image.*

6. Conclusion

After evaluating two CNN model architectures for perforator detection, it became evident in this case that image segmentation models employing U-net architectures achieved the most accurate results for identifying perforators. Although both tested image segmentation models performed relatively well, the U-net architecture model (Retna), with specific changes, is currently favored for two key reasons. Firstly, it allows the possibility to perform image segmentation based on multi-color masks, in contrast to the binary masks used in the blood vessels segmentation model. Secondly, it is more efficient in terms of time and computer power. It can also be concluded that using pre- and intraoperative training data is an efficient way of addressing the challenge of small datasets. Due to their large similarities, good results are obtained for perforator detection in both pre-and intraoperative images. However, to enhance the models' performance, accuracy, and robustness, acquiring a larger, more representative dataset is imperative and can be obtained by conducting more measurements using the DIRT technique.

Acknowledgments

This research was funded by the Fonds Wetenschappelijk Onderzoek (FWO) via support for the FWO TBM research project, "The use of dynamic infrared thermography for perforator mapping and quality improvement in autologous breast reconstructions." (FWO TBM FN7023). The authors also acknowledge the work of UAntwerp master students Iven Suykens and Laurens Schouwaerts, which resulted in the paper.

REFERENCES

[1] F. Thiessen et al., "Dynamic infrared thermography (DIRT) in Deep Inferior Epigastric Perforator (DIEP) flap breast reconstruction: standardization of the measurement set-up.," *Gland Surg*, 2019, doi: 10.21037/gs.2019.12.09.

[2] "Gentle Introduction to the Adam Optimization Algorithm for Deep Learning - MachineLearningMastery.com." Accessed: Dec. 19, 2023. [Online]. Available: https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/

[3] "What is Adam Optimizer?" Accessed: Dec. 19, 2023. [Online]. Available: https://www.analyticsvidhya.com/blog/2023/09/what-is-adam-optimizer/

[4] X. X. Yin, L. Sun, Y. Fu, R. Lu, and Y. Zhang, "U-Net-Based Medical Image Segmentation," *J Healthc Eng*, vol. 2022, 2022, doi: 10.1155/2022/4189781.