Health and Technology (2020) 10:891–903
https://doi.org/10.1007/s12553-020-00423-8

ORIGINAL PAPER

A novel approach to keypoint detection for the aesthetic evaluation of breast cancer surgery outcomes

Tiago Gonçalves[1] · Wilson Silva[2] · Maria J. Cardoso[3] · Jaime S. Cardoso[2]

Received: 27 January 2020 / Accepted: 24 March 2020 / Published online: 20 April 2020

Abstract

The implementation of routine breast cancer screening and better treatment strategies made possible to offer to the majority of women the option of breast conservation instead of a mastectomy. The most important aim of breast cancer conservative treatment (BCCT) is to try to optimize aesthetic outcome and implicitly, quality of life (QoL) without jeopardizing local cancer control and overall survival. As a consequence of the impact aesthetic outcome has on QoL, there has been an effort to try to define an optimal tool capable of performing this type of evaluation. Starting from the classical subjective aesthetic evaluation of BCCT (either by the patient herself or by a group of clinicians through questionnaires) to an objective aesthetic evaluation (where machine learning and computer vision methods are employed), leads to less variability and increasing reproducibility of results. Currently, there are some offline software applications available such as BAT© and BCCT.core, which perform the assessment based on asymmetry measurements that are computed based on semi-automatically annotated keypoints. In the literature, one can find algorithms that attempt to do the completely automatic keypoint annotation with reasonable success. However, these algorithms are very time-consuming. As the course of research goes more and more into the development of web software applications, these time-consuming tasks are not desirable. In this work, we propose a novel approach to the keypoints detection task treating the problem in part as image segmentation. This novel approach can improve both execution-time and results.

Keywords Keypoint detection · Aesthetic evaluation · Breast cancer · Deep neural networks

1 Introduction

Breast cancer ranks as the most frequent cancer among women [17, 33]. Despite being a highly mutable and rapidly evolving disease, it is estimated that most breast cancers are curable if properly detected and treated [25]. Thanks to the widespread use of breast cancer screening and better treatments, survival rises to 90% at 10 years in the majority of early detected cases [38]. Under this paradigm, it is possible to surgically treat most cancers with conservative approaches, i.e., lumpectomy, which consists on the removal of the cancerous tissue with a rim of healthy tissue (free margin) [34]. This type of treatment is commonly known as breast cancer conservative treatment (BCCT), it mandates adjuvant radiotherapy to the breast and has similar survival rates to the more radical mastectomy-based approaches, which consist of the removal of the entire breast, even if nowadays, the use of breast reconstruction has become more frequent [34]. In both cases, mastectomy with breast reconstruction, and BCCT, it is possible, to obtain good cosmetic results, and consequently improve patient's quality of life.

The assessment of cosmetic results has become crucial to those performing breast cancer treatment since it is also a means of evaluating the quality of the treatment and valuable input to the improvement of current techniques. Moreover, it has become important to define an objective evaluation standard for the cosmetic outcome of breast surgical procedures, as a means to measure the performance of new surgical and also radiotherapy techniques [10].

✉ Tiago Gonçalves
tiago.f.goncalves@inesctec.pt

[1] INESC TEC, Porto, Portugal

[2] INESC TEC and Faculdade de Engenharia, Universidade do Porto, Porto, Portugal

[3] Champalimaud Foundation and Nova Medical School, Lisboa, Portugal

Currently, to perform the aesthetic assessment, a vast majority of extracted features are related to asymmetry measurements [6]. To facilitate the extraction of such features, it is fundamental to mark breast keypoints. Some semi-automatic methods are already available (see Section 2), but they need the user to input information and are not considered efficient (i.e., their execution time is high and they are computationally complex).

The advent of machine learning and deep learning opened the possibility to design novel algorithms based on deep neural networks (DNN) which may be fully end-to-end (i.e., receive an image and output the aesthetic assessment score). Currently, the state-of-the-art algorithm for keypoint detection is a hybrid method based on a DNN and on traditional computer vision methods, which makes it computationally heavy and slow (see Section 3). Several conventional image analysis methods are fast to train (or require no train at all) but slow to apply to test images. Deep learning methods are slow to train but fast to apply. As such, while benefiting in terms of accuracy, the hybrid method keeps the worst of both in terms of efficiency, being very slow to train and to test.

The main objective of this work is the development of a novel breast keypoint detection algorithm that addresses the efficiency problem while maintaining or improving the accuracy (see Section 4). A study of algorithm performance based on execution time has also been conducted, since, at the long term, the intention is to deploy such algorithms into a web-based application that could be accessed by a diversity of devices and operative systems.

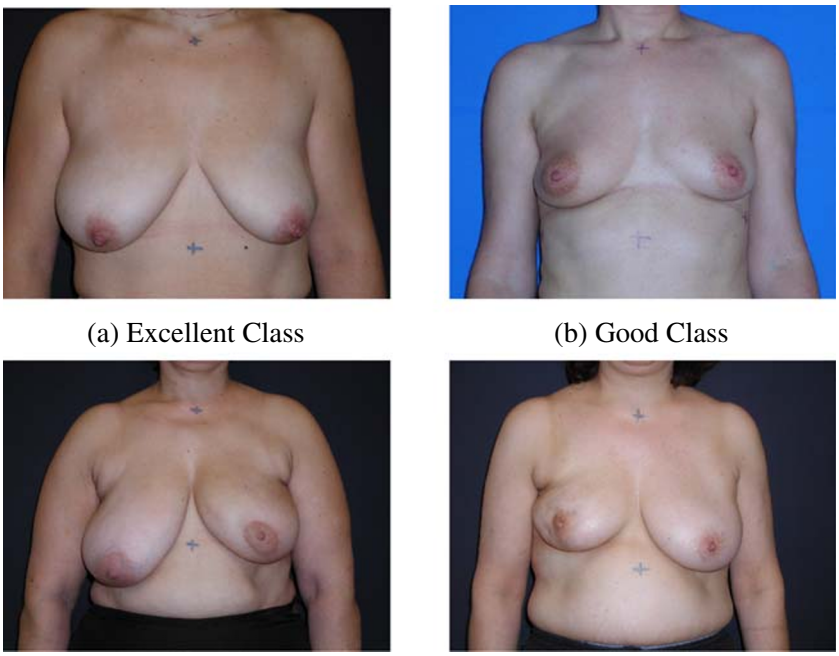
2 Aesthetic classification of BCCT outcomes

2.1 Assessment

Usually, the evaluation of the aesthetic outcome of breast cancer surgery is performed by an observer who analyses several features (colour, shape, geometry, irregularity and roughness) of the treated breast in comparison with the untreated one (in case of a unilateral treatment). The main assumption is that better cosmetic results are a consequence of more similar breasts. This approach has the advantage of being able to adapt to the appearance of new emerging oncoplastic techniques, in which, both breasts are subjected to surgery, leading to a more challenging comparison [25]. The assessment is done according to the Harvard scale, introduced by Jay Harris [20], which classifies the overall cosmetic result into four classes, according to the degree of differences between the treated breast and the untreated one:

1. **Excellent:** if the treated breast is nearly identical to the untreated one - see Fig. 1(a).
2. **Good:** if the treated breast has some differences when compared with the untreated one - see Fig. 1(b).
3. **Fair:** if the treated breast is different from the untreated one, but not seriously distorted see - Fig. 1(c).
4. **Poor:** if the treated breast is seriously distorted - see Fig. 1(d).

To perform the classification of a photograph into the previously presented four classes (excellent, good, fair and poor), there are several valid approaches. Nonetheless, one



(a) Excellent Class (b) Good Class

(c) Fair Class (d) Poor Class

Fig. 1 Harvard scale examples

can group them into two different clusters: subjective and objective methods.

2.1.1 Subjective methods

Generally, subjective methods include patients' self-evaluation and the evaluation by a single observer or through a panel of observers. Regarding patients' self-evaluation through PROMs (Patient Reported Outcome Measures), it can be argued that it is the one that best translates the psychosocial adaptation of patients to the result, being, thus, a simple way to assess the cosmetic outcome; however with low reproducibility, due to the influence of several factors such as age, socio-economic status or the fear to criticize the treatment itself or the responsible caregiver, which will impact on how patients' see themselves after the treatment and, consequently, impact the final evaluation [1]. Also, the assessment by a panel of observers can prove itself to be a long and costly process [9].

2.1.2 Objective methods

To overcome the lack of objectivity and reproducibility related to subjective methods, objective methods were introduced. Historically, it was Richard Pezner the one who introduced the first objective measure of the breast asymmetry evaluation: breast retraction assessment (BRA) [28]. This descriptor shows the amount of retraction of the treated breast by comparing it to the untreated breast. Fig. 2 and Eq. 1 show how to get BRA points and how to compute BRA value, respectively.

$$\text{BRA} = \sqrt{(x_R - x_L)^2 + (y_R - y_L)^2} = \|R - L\|_2 \quad (1)$$

Pezner et al. were then followed by several other authors who also introduced important objective descriptors [24, 35, 36]. Also, this introduction of objective measures created the conditions for the development of new software and the design of novel computer vision and machine learning algorithms for the automatic aesthetic classification of BCCT outcomes. This will be presented in the next Section, with the description of the two computer-aided systems that are present in the literature: BAT©, and BCCT.core.

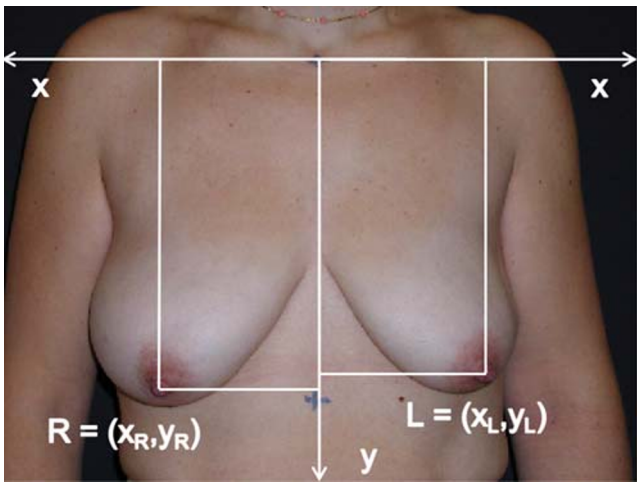


Fig. 2 Illustrative example of BRA [27]

2.2 Computer-aided aesthetic classification of BCCT outcomes

2.2.1 Breast analysing tool - BAT©

Proposed by Fitzal et al. [16], BAT© considers the Breast Symmetry Index (BSI), to assess the cosmetic outcome of BCCT. This measure takes into account the size differences between right and left breasts, producing reproducible results. This system has proven to be capable of differentiating between good and fair cosmetic outcomes and has a high correlation with subjective votes from experts. On the other hand, it is not able to differentiate between excellent and good, or fair and poor cosmetic outcomes and has a low correlation with the patients' self-evaluation. Fig. 3 shows the interface of BAT©.

2.2.2 BCCT.core

Proposed by Cardoso and Cardoso [6], this tool was the first one to perform automatic feature extraction from digital photographs as a way of capturing some relevant factors for the overall score. To properly do this, they focused on asymmetry, colour differences and scar visibility features. Also, to compute asymmetry features, there is the need to mark specific keypoints in the image (sternal notch, the scale mark, the nipples and left and right breasts contours, adjusted with an active contour based on splines with control points). Although it was possible to manually mark these keypoints, the main goal has always been to

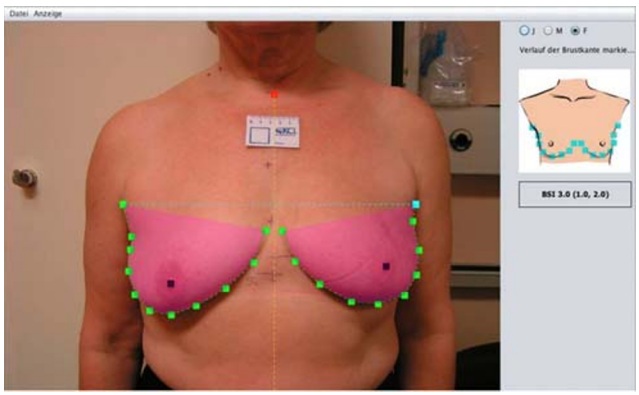


Fig. 3 BAT© software interface, from [21]

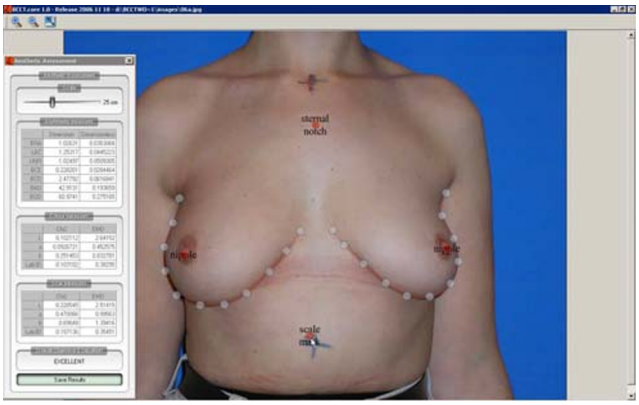


Fig. 4 BCCT.core software interface, from [6]

do this labelling in a fully automated way, to overcome the variability in the annotation process. Fig. 4 shows the interface of BCCT.core software interface. A brief description of the algorithms proposed in the literature to automate the keypoint detection process is presented next.

Breast Contour Cardoso and Cardoso [5] were the first ones to present an algorithm capable of automatically detecting breast contours in digital photographs. To do this, they compute the gradient of the image and model it as a weighted graph based on pixel gradient, value and position. Assuming that the two endpoints of the breast contour are known, the problem is focused on finding the shortest path between both endpoints that goes through the breast contour. Later on, to help on this task, Sousa et al. introduced the use of shape priors to facilitate the process [32].

Endpoints Following the work on breast contour detection, Cardoso et al. proposed a method for the automatic detection of the endpoints [11]. This method assumes that

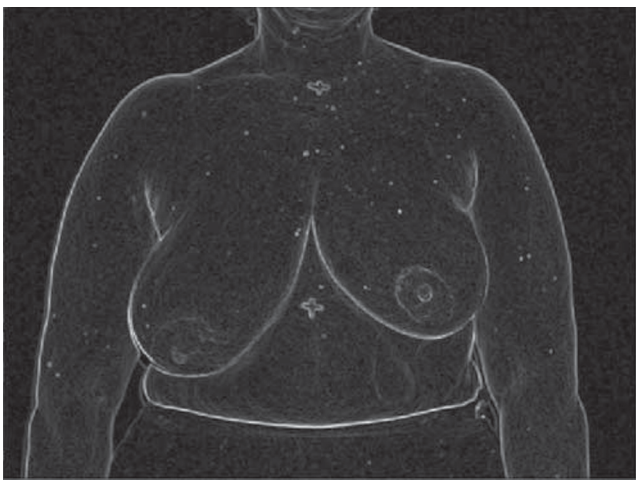


Fig. 5 Image gradient, from [11]

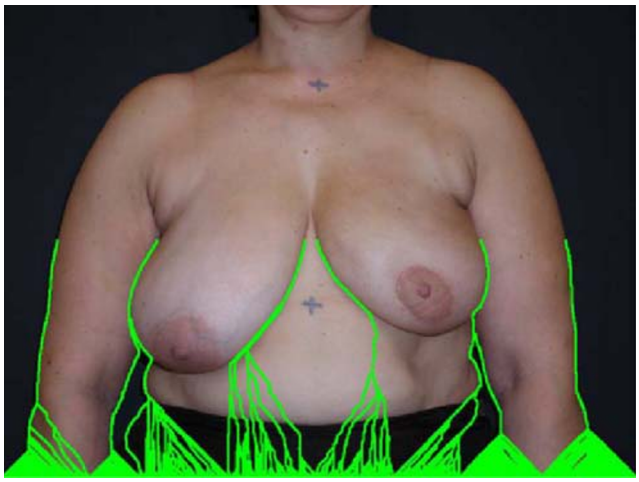


Fig. 6 Shortest paths from the bottom to the middle

the photo only contains the torso of the patient, which means that the external endpoint of the breast contour can be assumed to be at the point of the body where the arm contour intersects the trunk contour. However, in most of the photographs, patients are in the arms-down position, so the arm's contour is almost indistinguishable from the trunk's contour. This means that the external endpoint of the breast can be defined as the highest point of the breast contour. Figures 5, 6, 7, 8, 9, 10, 11 to 12 present a graphical sequence of the algorithm proposed by Cardoso et al. in [11].

Nipples Breast surface is generally characterized as a featureless shape. So, the nipple should be the most prominent feature on it. Taking this into account, Cardoso et al. proposed a method [7] that uses a Harris corner descriptor to detect possible nipple locations and then applies a closed contour method to find areola contours around those

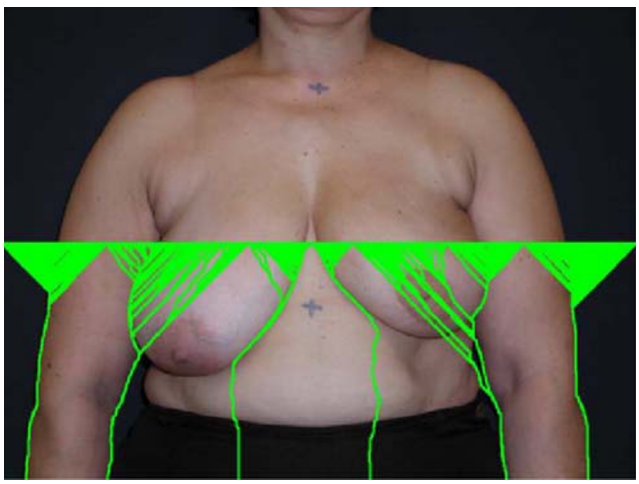


Fig. 7 Shortest paths from the middle to the bottom

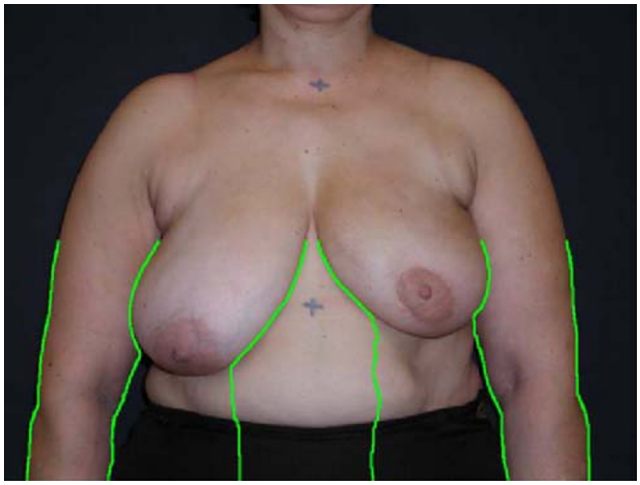


Fig. 8 Strong paths between middle and bottom

Fig. 10 Strong paths between top and bottom

points (see Figs. 13 and 14). High-level features (Harris corner quality factor, the average magnitude of the directional derivative of the contour, shape factor of the contour, equivalent diameter of the contour) are extracted and the best pair candidate/closed contour is selected by a support vector machine (SVM) classifier trained on the extracted features.

2.3 Conclusions

Both BAT© and BCCT.core established the fundamental use-cases for the development of computer-aided systems for the aesthetic assessment of BCCT. However, these applications still require information given by the user and none of them has been considered the gold standard for this task. Therefore, there is a need to study the applicability of different algorithms on this task (e.g., deep learning), especially if they could improve performance and be easily integrated with other platforms [8].

3 State-of-the-art keypoint detection algorithms for the aesthetic evaluation of BCCT

3.1 A deep keypoint detection algorithm

Recently, Silva et al. proposed a novel method [30] that uses a deep neural network (DNN) for the keypoint detection task, opening the possibility to follow an integrated learning approach. Following the ideas of Cao et al. [4] and Belagiannis et al. [2], Silva et al. proposed an architecture that first learns how to regress *heatmaps* (which are obtained by applying a Gaussian kernel to the keypoints) and, after iterative tuning of this heatmap regression, it can predict keypoint localization (see Fig. 15). To perform the heatmap regression, the U-Net model [29] is used. Then, to do the keypoint regression, the original images are multiplied by the refined heatmaps (to improve the initial

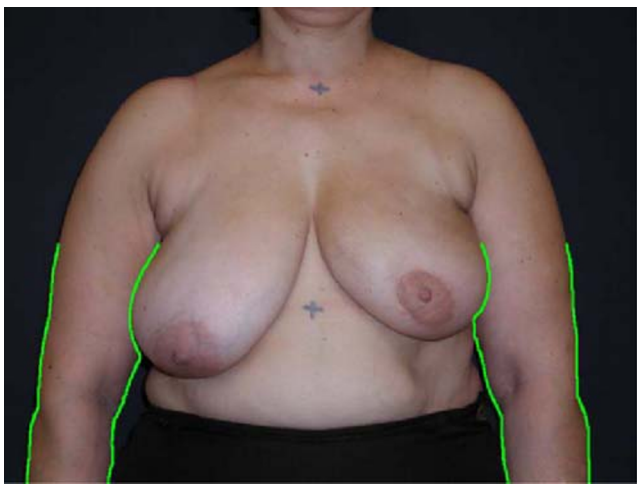


Fig. 9 Selected shortest paths

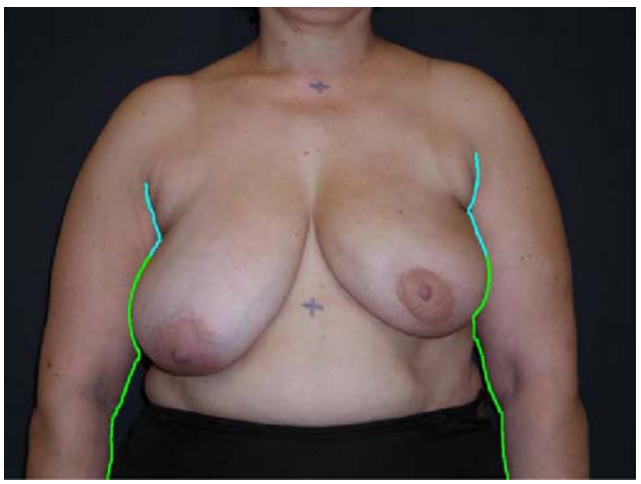


Fig. 11 The endpoints are the highest points of the shortest path

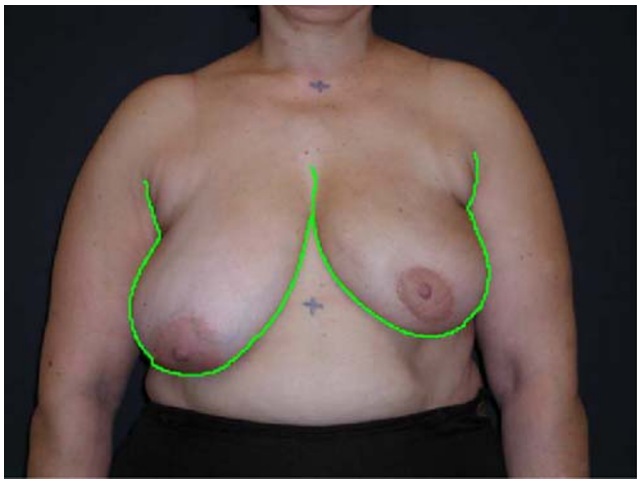


Fig. 12 Automatic breast contour detection with the shortest path algorithm

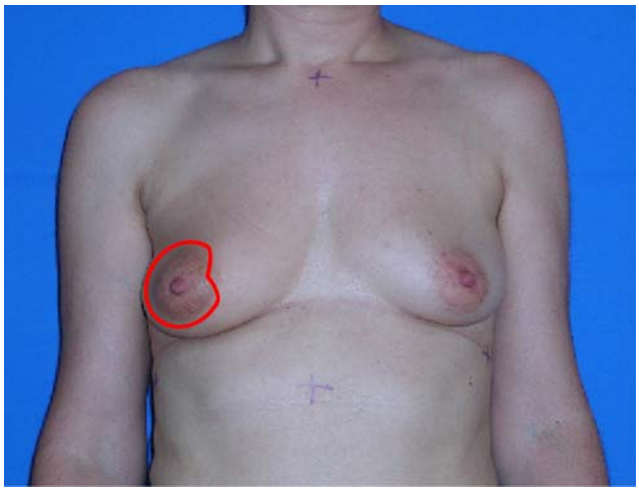


Fig. 14 Harris corner detection. [7]

fuzzy localization of keypoints) and are fed to a keypoint regression module composed of three blocks: VGG16 [31], four convolutional layers and three dense layers. The entire model is trained, using the iterative refinement of the regression of heatmaps as a regularization term of the loss function.

3.2 A hybrid keypoint detection algorithm

Silva et al. in their original work also proposed a hybrid approach to the detection of keypoints (see Fig. 16). With this hybrid approach, the endpoints and the nipples are obtained with the use of the deep learning algorithm, whereas the contour is detected with the conventional shortest path algorithm presented in Section 2. The main difference here is that the endpoints given as input to the shortest path algorithm are obtained with the Deep Keypoint Detection Algorithm, explained in the previous Subsection, instead of being specified by the user. This hybrid algorithm led to an improvement in the results.

3.3 Conclusions

The work of Silva et al. presented important contributions towards the development of deep learning algorithms capable of detecting keypoints in photographs of women's torso, after being subjected to BCCT. On the other hand, the Hybrid Keypoint Detection Algorithm performed better on the breast contour detection. However, it is not very efficient because it models images as graphs based on gradients and uses the shortest path algorithm to find the breast contour. This procedure is very time-consuming when compared with the inference process of a deep learning model. Also, if one intends to integrate such algorithms into a web

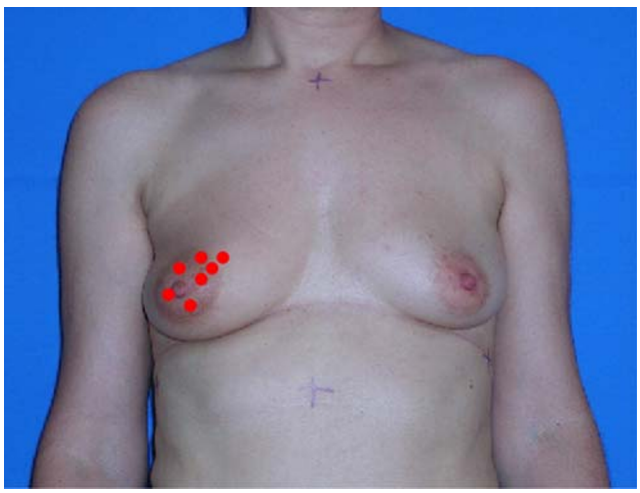


Fig. 13 Nipple candidates detection [7]

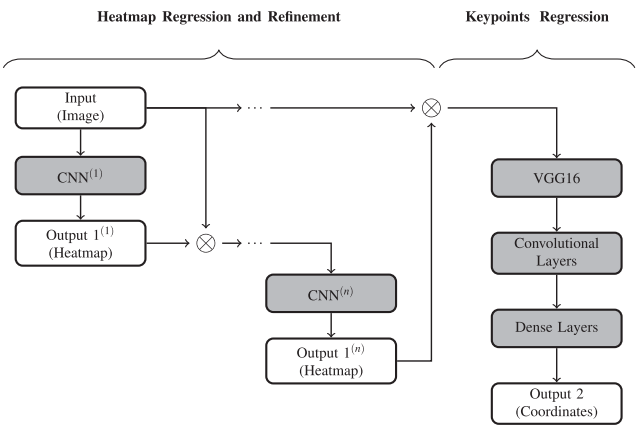


Fig. 15 Deep Keypoint Detection Algorithm, from [30]

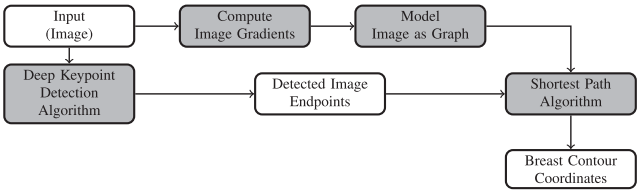


Fig. 16 Hybrid Keypoint Detection Algorithm, from [30]

application, it is of utmost importance that performance measurements (e.g., loading time, execution time) are taken into account when testing and designing novel methods.

4 A new approach to keypoint detection for the aesthetic evaluation of BCCT

To overcome this dependence on the shortest path algorithm (which guarantees the best results), we investigated other deep learning approaches to detect the breast contour, in particular, through image segmentation.

4.1 Medical image segmentation overview

Image segmentation is the task of dividing an image into regions of interest (ROI), guaranteeing that every pixel belonging to that ROI is similar in terms of characteristics [19]. In fact, in medical applications, image segmentation is one of the most important steps in an application pipeline.

4.2 Deep image segmentation

Before the advent of deep learning and convolution neural networks (CNN), traditional image segmentation would rely on domain knowledge and on feature engineering techniques to produce final results. Some well-known techniques under this scope are thresholding, edge-based, region-based, deformable models or graph cuts. Currently, thanks to the high performance of CNNs in image-related tasks (e.g., image classification) [31], the state-of-the-art segmentation models are almost all based on deep learning. The theory behind CNNs for image segmentation was proposed by Long et al., who developed and implemented a fully convolutional network (FCN), trained end-to-end and/or pixel-to-pixel on semantic segmentation [22]. This architecture was established as the baseline for the development of novel and improved ones, such as U-Net, U-Net++, Global Convolutional Network (GCN) or DeepLabV3+.

4.2.1 U-net

U-Net was proposed by Ronneberger et al. in 2015 for biomedical image segmentation tasks [29]. It was built on top of the FCN, with added successive layers with pooling operators replaced by upsampling operators to the baseline network to increase the resolution of the output. Moreover, U-Net does not have fully connected layers and its segmentation map only contains the pixels for which the full context of the image is available in the input image.

4.2.2 U-net++

Published in 2018 by Zhou et al., U-Net++ is the result of several modifications of the basic U-Net architecture regarding the skip connections, which became nested and dense. The intuition behind this approach is related to the argument that the network would deal with an easier learning task when the feature maps from the decoder and encoder modules are semantically similar; in U-Net this does not happen, since its skip connections are plain [40].

4.2.3 GCN

Focusing on the two main problems of semantic segmentation (classification and localization), Peng et al. designed a novel FCN-based architecture, GCN, which was capable of addressing both problems: the localization task was addressed by the fact that the network itself was fully convolutional, while the classification task was handled by using large kernel size in the network architecture to enable densely connections between feature maps and per-pixel classifiers, leading, thus, to an enhancement in the ability to handle different transformations [26]. They also introduced a *boundary refinement module* that aims to improve localization performance near object boundaries.

4.2.4 DeepLabV3+

This network is the result of successive iterations on the basic DeepLab architecture [12]. The entire work is based on the use of of *atrous convolution* or dilated convolution, which allows the computation of features at different resolutions, without the need to learn extra parameters. This prevents the use of consecutive pooling operations (typically employed in CNNs) which make the network invariant to local transformations (e.g., translations), an undesired property in semantic segmentation tasks, where spatial information is required. Besides, DeepLabV3+ introduced an encoder-decoder module to refine segmentation results and uses the Xception [13] model as the backbone for the segmentation task.

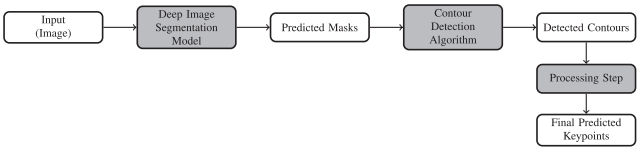


Fig. 17 Proposed approach for an Image Segmentation-Based Keypoint Detection Algorithm

4.3 Segmentation for breast contour detection

The intuition behind this approach is that it is easier to detect breast contours if one is capable to detect breasts first [18]. This can be seen as a problem of semantic segmentation, where both breasts are considered the foreground and the rest is considered the background of the image. The main hypothesis is that if it is possible to perform the segmentation of both breasts with high precision, one could proceed to an algorithm of contour detection and then accurately extract the keypoints related to the breast contours. With segmentation, the goal is to learn a single solution (i.e., one image corresponds to one mask). This is important, because, if the DNN is capable of predicting the correct mask, the set of points of the detected contour will contain a subset of points that belong to the real breast contour. On the other hand, with keypoint regression, there is a higher degree of variability, where the DNN can predict points that belong to the real breast contour and points that do not, negatively influencing the algorithm performance. Furthermore, when compared with the hybrid keypoint detection algorithm, explained in Section 3, it would be expected that this approach would bring improvements in terms of results and performance, i.e., it would be faster than the hybrid keypoint detection algorithm. Fig. 17 shows the proposed pipeline for a keypoint detection algorithm based deep image segmentation. A deep image segmentation model, trained on our dataset, is used to generate the masks and a contour detection algorithm finds all the keypoints that correspond to the contours present in image masks. After a post-processing step, the breast contour keypoints are then extracted.

5 Experimental settings

5.1 Dataset

The available dataset to perform experiments is the aggregation of three different datasets: PORTO, with 120 images; TSIO, with 30 images; VIENNA, with 71 images. For each image, there are 37 ground-truth keypoints (4 endpoints, 30 points along the breast contours, 2 nipples and the sternal notch) resulting in a total of 74 coordinates. It is important to refer that this dataset has some internal variability: images from PORTO and TSIO datasets were obtained in equivalent conditions, i.e., all those images have a distinct and consistent background and do not have many shadows nor artificial variations in colour, while images from the VIENNA dataset have a noisy and variable background and were acquired under poor lighting conditions. Example images from these datasets are illustrated in Figs. 18, 19 and 20.

5.2 5-fold cross-validation train/test split

All experiences were done taking into account 5-fold cross-validation split into train and test sets. The train and test indices for each fold were generated using scikit-learn [3] and were saved for further use to guarantee that the same data was being used in the same fold in different experiments.

5.3 Silva et al. Deep Keypoint Detection Algorithm

The original Keras [14] implementation by Silva et al. was used [30]. This training scheme also requires the ground-truth heatmaps which were generated considering a Gaussian kernel centred at each keypoint with a pre-defined standard deviation. The model was trained during 300 epochs with Adadelta [39] as the optimizer. Regarding data augmentation, translations, rotations and flips were applied to all images, heatmaps and keypoints in an online setting during training.

Fig. 18 Example images from PORTO Dataset

Fig. 19 Example images from TSIO Dataset

Fig. 20 Example images from VIENNA Dataset

5.4 Silva et al. hybrid keypoint detection algorithm

The Python implementation of the shortest path algorithm by Silva et al. was used [30]. In this case, the predicted endpoints from the model of the previous Subsection were given as input to the shortest path algorithm.

5.5 Segmentation-based keypoint detection algorithm

This new proposed approach combines image segmentation and the model proposed by Silva et al. in a pipeline. First of all, there was the need for training a deep image segmentation model that could achieve good results in breast segmentation. To train such model, it was necessary to generate breast masks, which were obtained with the support of the ground-truth keypoints and images: every pixel inside the keypoints' area was given the 255 value; to the rest of the pixels was given the 0 value. Then, images were normalized to have pixel values between 0 and 1.Taking into account previous results with other models [18], for this experiment, it was decided to use the U-Net++ architecture as the deep segmentation model. The official Keras implementation by Zhou et al. [1] was trained and fine-tuned (the model has an encoder which is initialized with the ImageNet [15] weights) during 300 epochs with Adadelta [39] as the optimizer; during training, binary cross-entropy was selected as the loss function. Regarding data augmentation, image and mask translations, rotations and flips were employed in an online fashion during training. The trained U-Net++ model was then used to generate segmentation masks. From these masks, contours were extracted using the marching squares algorithm (a special case of the marching cubes algorithm [23]), implemented in scikit-image [37]. The intuition behind this contour detection step is that the detected contours will contain the breast contour keypoints, assuming that breast segmentation masks were well generated. As such, this first step outputs a variable number of contour keypoints, some of which are not desired, because they do not belong to what is considered the breast contour. As a post-processing step, the Silva et al. Deep Keypoint Detection Algorithm's predicted endpoints are projected onto the mask contours through the minimization of the Euclidean Distance between the mask contour keypoint and the predicted keypoint. At the end of this processing step, there are new 34 breast contour keypoints plus the nipples and sternal notch which were predicted by the Silva et al. Deep Keypoint Detection Algorithm (see Fig. 21). Fig. 22 shows a chronological scheme of this algorithm.

5.6 Study of algorithm performance

One of the main goals of this work was the study of the performance of these algorithms (see Sections 5.3, 5.4 and 5.5) to assess which one would fit better into a web-version of BCCT.core, capable of real-time interaction with deep learning models. To perform this study, the execution time of each algorithm on CPU (Intel® Core™ i7-2600 CPU @ 3.40GHz × 8) was measured on the test set of each cross-validation fold.

6 Results and discussion

Table 1 presents the average error distance (measured in pixels) and the average execution time (measured in seconds) of each model inference on the test set. It can be seen that the Segmentation-Based Keypoint Detection Algorithm surpasses both Deep and Hybrid Keypoint Detection Algorithms from Silva et al. in the endpoints and breast contours detection tasks, which were, to our

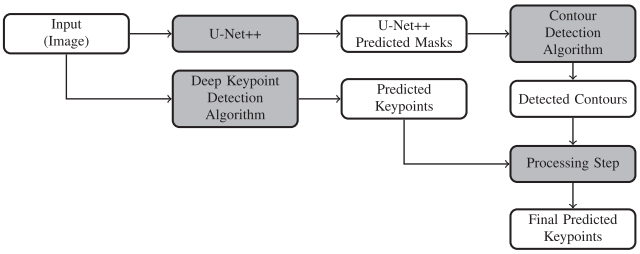


Fig. 21 A Novel Segmentation-Based Keypoint Detection Algorithm

[1] See: https://github.com/MrGiovanni/UNetPlusPlus

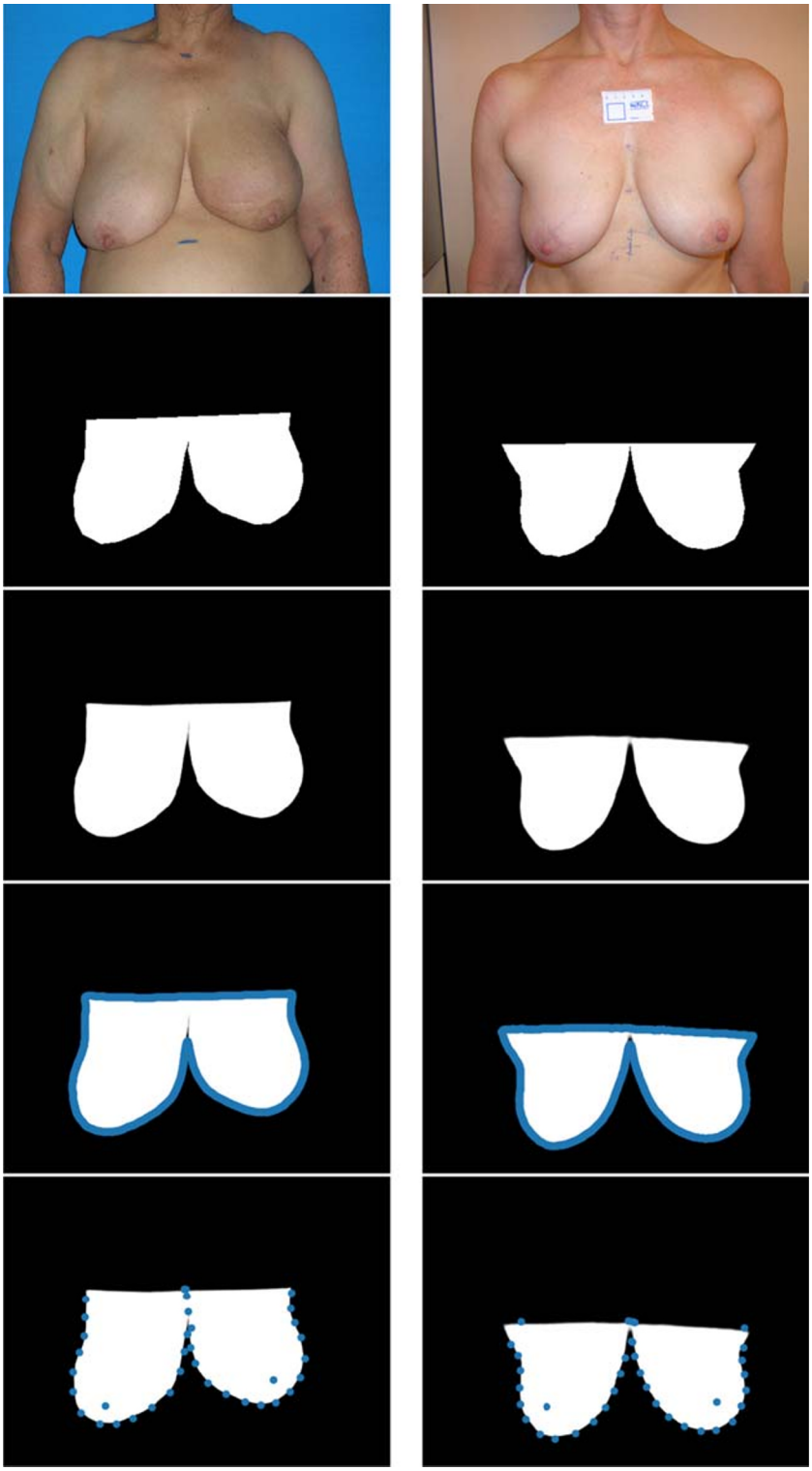


Fig. 22 Chronological scheme (from top to bottom) of the proposed Segmentation-Based Keypoint Detection Algorithm. Each column represents a single image. The first row is the ground-truth image, the second row is the ground-truth mask, the third row is the U-Net++ predicted mask, the fourth row is the set of all the detected contour keypoints and the fifth row is the set of breast keypoints after the post-processing step described in Subsection 5.5

knowledge, the state-of-the-art breast keypoint detection algorithms. Moreover, this novel algorithm achieves lower values of standard deviation and maximum error, which suggests it is even more consistent when compared with the other two. Regarding the study of performance, it can be understood that the Deep Keypoint Detection achieves better execution time, however, it has the highest error for the breast contour. The Segmentation-Based Keypoint Detection Algorithm presents the best balance between time-efficiency and accuracy, being the most accurate model, with a time efficiency comparable to the most time-efficient method (Figs. 23, 24 and 25).

Table 1 Average error distance for endpoints, breast contours and nipples, measured in pixels and average execution time of the models' inferences (on the test set of each cross-validation fold, that has approximately 43 to 45 images). Best results are highlighted in bold. **Note:** STD stands for standard deviation and Max stands for maximum error

Model	Endpoints			Breast Contours			Nipples			Execution Time (s)
	Mean	STD	Max	Mean	STD	Max	Mean	STD	Max	
Silva et al. Deep Keypoint Detection Algorithm	40	33	182	21	8	72	70	39	218	150
Silva et al. Hybrid Keypoint Detection Algorithm	40	33	182	13	14	104	70	39	218	1704
Segmentation-Based Keypoint Detection Algorithm	38	34	195	11	5	34	70	39	218	280

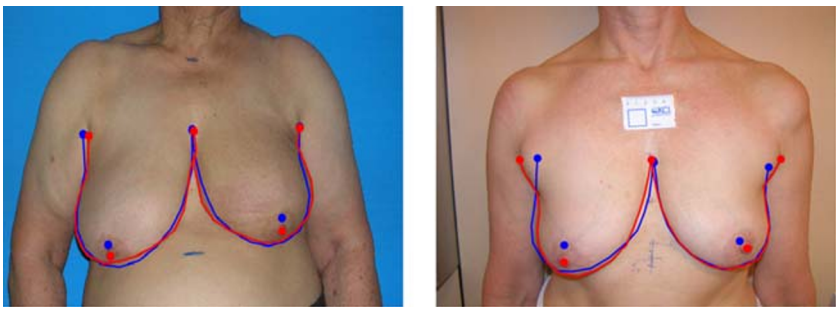


Fig. 23 Example predictions from Silva et al. Deep Keypoint Detection Algorithm. Prediction is in blue and ground-truth is in red

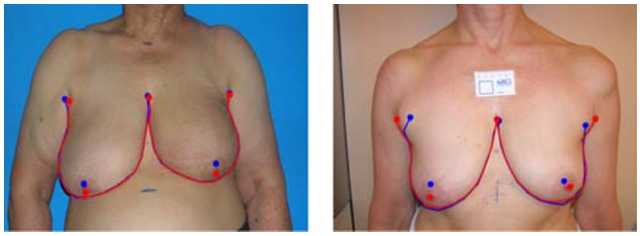


Fig. 24 Example predictions from Silva et al. Hybrid Keypoint Detection Algorithm. Prediction is in blue and ground-truth is in red

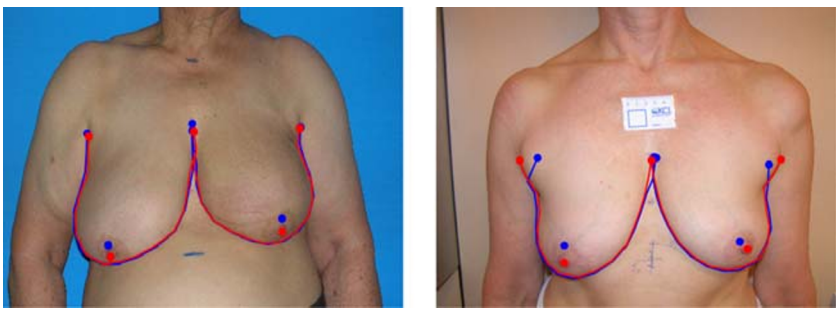


Fig. 25 Example predictions from Segmentation-Based Keypoint Detection Algorithm. Prediction is in blue and ground-truth is in red

7 Conclusions and future work

This work presented a novel algorithm based on the interaction of two deep learning models (a segmentation model, and a keypoint detection model) that can surpass the state-of-the-art algorithms present in the literature. Furthermore, a comparative study regarding algorithms performance has been done to assess which one would fit better a web-based application for the aesthetic assessment of BCCT. Our proposed model revealed itself as the best in terms of executing time, while being very competitive in terms of keypoint prediction. As future work, the next step will be to improve results on nipples detection task and to modify this novel segmentation-based keypoint detection algorithm by integrating all the tasks of its pipeline in a unique DNN with a combined loss function. The integration and full deployment of this algorithm in a web-application are also planned.

Acknowledgements The authors thank Prof. Florian Fitzal for sharing the VIENNA dataset. This work was partially financed by the ERDF - European Regional Development Fund through the Operational Programme for Competitiveness and Internationalisation - COMPETE 2020 Programme and by National Funds through the Portuguese funding agency, FCT - Fundação para a Ciência e a Tecnologia within project "POCI-01-0145-FEDER-028857" and PhD grant number SFRH/BD/139468/2018.

Compliance with Ethical Standards

Conflict of interests The authors declare that they have no conflict of interest.

References

1. Al-Ghazal S, Blamey R, Stewart J, Morgan A. The cosmetic outcome in early breast cancer treated with breast conservation. European Journal of Surgical Oncology (EJSO). 1999;25(6):566–570. https://doi.org/10.1053/ejso.1999.0707.
2. Belagiannis V, Zisserman A. Recurrent Human Pose Estimation. 2016. arXiv: 1605.02914.
3. Buitinck L, Louppe G, Blondel M, Pedregosa F, Mueller A, Grisel O, Niculae V, Prettenhofer P, Gramfort A, Grobler J, Layton R, Vanderplas J, Joly A, Holt B, Varoquaux G. API Design for machine learning software: experiences from the scikit-learn project. 2013. arXiv: 1309.0238.
4. Cao Z, Simon T, Wei SE, Sheikh Y. Realtime multi-Person 2d Pose Estimation using Part Affinity Fields. 2016. arXiv: 1611.08050.
5. Cardoso JS, Cardoso MJ. Breast contour detection for the aesthetic evaluation of breast cancer conservative treatment. Computer recognition systems 2, vol. 45. In: Kacprzyk J., Kurzynski M., Puchala E., Wozniak M., and Zolnierek A., editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2007. pp. 518–525.
6. Cardoso JS, Cardoso MJ. Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment. Artif Intell Med. 2007;40(2):115–126. https://doi.org/10.1016/j.artmed.2007.02.007.
7. Cardoso JS, Domingues I, Oliveira HP. Closed Shortest Path in the Original Coordinates with an Application to Breast Cancer. International Journal of Pattern Recognition and Artificial Intelligence. 2015;29(01):1555002. https://doi.org/10.1142/S0218001415550022.
8. Cardoso JS, Silva W, Cardoso MJ. Evolution, current challenges, and future possibilities in the objective assessment of aesthetic outcome of breast cancer locoregional treatment. The Breast. 2020;49:123–130. https://doi.org/10.1016/j.breast.2019.11.006. http://www.sciencedirect.com/science/article/pii/S0960776193109870987.
9. Cardoso MJ, Cardoso J, Santos AC, Barros H, Oliveira MCd. Interobserver agreement and consensus over the esthetic evaluation of conservative treatment for breast cancer. The Breast. 2006;15(1):52–57. https://doi.org/10.1016/j.breast.2005.04.013.
10. Cardoso MJ, Cardoso JS, Vrieling C, Macmillan D, Rainsbury D, Heil J, Hau E, Keshtgar M. Recommendations for the aesthetic evaluation of breast cancer conservative treatment. Breast Cancer Res Treat. 2012;135(3):629–637. https://doi.org/10.1007/s10549-012-1978-8.
11. Cardoso, et al. AUTOMATIC BREAST CONTOUR DETECTION IN DIGITAL PHOTOGRAPHS. In: Proceedings of the First International Conference on Health Informatics. Funchal, Madeira, Portugal: SciTePress - Science and and Technology Publications; 2008. pp 91–98, https://doi.org/10.5220/0001039500910098.
12. Chen LC, Papandreou G, Schroff F, Adam H. Rethinking Atrous Convolution for Semantic Image Segmentation. 2017. 1706.05587. http://arxiv.org/abs/1706.05587.
13. Chollet F. Xception: Deep Learning with Depthwise Separable Convolutions. 2016. arXiv: 1610.02357.
14. Chollet F et al. Keras. 2015. https://keras.io.
15. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-Scale Hierarchical Image Database; 2009.
16. Fitzal F, Krois W, Trischler H, Wutzel L, Riedl O, Kühbelböck U., Wintersteiner B, Cardoso M, Dubsky P, Gnant M, Jakesz R, Wild T. The use of a breast symmetry index for objective evaluation of breast cosmesis. The Breast. 2007;16(4):429–435. https://doi.org/10.1016/j.breast.2007.01.013.
17. Giordano SH, Buzdar AU, Hortobagyi GN. Breast Cancer in Men p. 11.
18. Gonçalves T, Silva W, Cardoso J. Deep aesthetic assessment of breast cancer surgery outcomes. XV Mediterranean conference on medical and biological engineering and computing – MEDICON 2019. In: Henriques J, Neves N, and de Carvalho P, editors. Springer International Publishing, Cham; 2020. pp 1967–1983, https://doi.org/10.1007/978-3-030-31635-8_236.
19. Gonzalez RC, Woods RE. Digital image processing, 3rd ed. Upper Saddle River, N.J: Prentice Hall; 2008.
20. Harris JR, Levene MB, Svensson G, Hellman S. Analysis of cosmetic results following primary radiation therapy for stages I and II carcinoma of the breast. International Journal of Radiation Oncology*Biology*Physics. 1979;5(2):257–261. https://doi.org/10.1016/0360-3016(79)90729-6. https://linkinghub.elsevier.com/retrieve/pii/0360301679907296.
21. Krois W, Romar AK, Wild T, Dubsky P, Exner R, Panhofer P, Jakesz R, Gnant M, Fitzal F. Objective breast symmetry analysis with the breast analyzing tool (BAT): improved tool for clinical trials. Breast Cancer Research and Treatment. 2017;164(2):421–427. https://doi.org/10.1007/s10549-017-4255-z. http://link.springer.com/10.1007/s10549-017-4255-z.
22. Long J, Shelhamer E, Darrell T. Fully Convolutional Networks for Semantic Segmentation. 2014. arXiv: 1411.4038.

23. Lorensen WE, Cline HE. Marching cubes: A high resolution 3d surface construction algorithm. In: Proceedings of the 14th annual conference on Computer graphics and interactive techniques - SIGGRAPH '87. ACM Press, Not Known; 1987. pp 163–169, https://doi.org/10.1145/37401.37422. http://portal.acm.org/citation.cfm?doid=37401.37422.
24. Noguchi M., Saito Y., Mizukami Y., Nonomura A., Ohta N., Koyasaki N., Taniya T., Miyazaki I. Breast deformity, its correction, and assessment of breast conserving surgery. Breast Cancer Research and Treatment. 1991;18(2):111–118. https://doi.org/10.1007/BF01980973. http://link.springer.com/10.1007/BF01980973.
25. Oliveira HP, Cardoso JS, Magalhaes A, Cardoso MJ. Methods for the aesthetic evaluation of breast cancer conservation treatment: a technological review. Current Medical Imaging Reviews. 2013;9(1):32–46. https://doi.org/10.2174/1573405611309010006.
26. Peng C, Zhang X, Yu G, Luo G, Sun J. Large Kernel Matters – Improve Semantic Segmentation by Global Convolutional Network. 2017. arXiv: 1703.02719.
27. Pezner RD, Lipsett JA, Vora NL, Desai KR. Limited usefulness of observer-based cosmesis scales employed to evaluate patients treated conservatively for breast cancer. International Journal of Radiation Oncology*Biology*Physics. 1985;11(6):1117–1119. https://doi.org/10.1016/0360-3016(85)90058-6. https://linkinghub.elsevier.com/retrieve/pii/0360301685900586.
28. Pezner RD, Patterson MP, Hill L, Vora N, Desai KR, Archambeau JO, Lipsett JA. Breast retraction assessment: an objective evaluation of cosmetic results of patients treated conservatively for breast cancer. International Journal of Radiation Oncology*Biology*Physics. 1985;11(3):575–578. https://doi.org/10.1016/0360-3016(85)90190-7. https://linkinghub.elsevier.com/retrieve/pii/0360301685901907.
29. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015. arXiv: 1505.04597.
30. Silva W, Castro E, Cardoso MJ, Fitzal F, Cardoso JS. Deep keypoint detection for the aesthetic evaluation of breast cancer surgery outcomes. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI'19); 2019.
31. Simonyan K, Zisserman A. Very Deep Convolutional Networks for large-Scale Image Recognition. 2014. arXiv: 1409.1556.
32. Sousa R, Cardoso JS, Pinto da Costa JF, Cardoso MJ. Breast contour detection with shape priors. In: 2008 15Th IEEE international conference on image processing. USA: IEEE, San Diego, CA; 2008. pp. 1440–1443, https://doi.org/10.1109/ICIP.2008.4712036.
33. Stewart BW, Wild C. International Agency for Research on Cancer, World Health Organization: World cancer report 2014. 2014. OCLC:1013966433. http://libweb.iaea.org/library/eBooks/World-Cancer-Report2014.pdf.
34. Street W. 2018.
35. Tsouskas LI, Fentiman IS. Breast compliance: A new method for evaluation of cosmetic outcome after conservative treatment of early breast cancer. Breast Cancer Research and Treatment. 1990;15(3):185–190. https://doi.org/10.1007/BF01806355. http://link.springer.com/10.1007/BF01806355.
36. Van Limbergen E, van der Schueren E, Van Tongelen K. Cosmetic evaluation of breast conserving treatment for mammary cancer. 1. Proposal of a quantitative scoring system. Radiotherapy and Oncology. 1989;16(3):159–167. https://doi.org/10.1016/0167-8140(89)90016-9. http://linkinghub.elsevier.com/retrieve/pii/0167814089900169.
37. van der Walt S, Schönberger JL, Nunez-Iglesias J, Boulogne F, Warner JD, Yager N, Gouillart E, Yu T. scikit-image: image processing in Python. PeerJ. 2014;2:e453. https://doi.org/10.7717/peerj.453. https://peerj.com/articles/453.
38. Wilkes GM, Barton-Burke M. Oncology nursing drug handbook (2018). OCLC:1030285714. 2018.
39. Zeiler MD. ADADELTA: An Adaptive Learning Rate Method. 2012. arXiv: 1212.5701.
40. Zhou Z, Rahman Siddiquee MM, Tajbakhsh N, Liang J. UNEt++: A Nested u-Net Architecture for Medical Image Segmentation. Deep learning in medical image analysis and multimodal learning for clinical decision support, vol. 11045. In: Stoyanov D., Taylor Z., Carneiro G., Syeda-Mahmood T., Martel A., Maier-Hein L., Tavares J. M. R., Bradley A., Papa J. P., Belagiannis V., Nascimento J. C., Lu Z., Conjeti S., Moradi M., Greenspan H., and Madabhushi A., editors. Springer International Publishing, Cham; 2018. pp 3–11, https://doi.org/10.1007/978-3-030-00889-5_1.