DEEP KEYPOINT DETECTION FOR THE AESTHETIC EVALUATION OF BREAST CANCER SURGERY OUTCOMES

Wilson Silva[1,2]*, Eduardo Castro*[1,2]*, Maria J. Cardoso*[2,3,4]*, Florian Fitzal*[5] *and Jaime S. Cardoso*[1,2]

[1] Faculdade de Engenharia, Universidade do Porto, Porto, Portugal
[2] INESC TEC, Porto, Portugal
[3] Faculdade de Ciências Médicas, Universidade NOVA de Lisboa, Lisbon, Portugal
[4] Breast Unit, Champalimaud Foundation, Lisbon, Portugal
[5] Department of Surgery, Medical University, Vienna, Austria

ABSTRACT

Breast cancer high survival rate led to an increased interest in the quality of life after treatment, particularly regarding the aesthetic outcome. Currently used aesthetic assessment methods are subjective, which make reproducibility and impartiality impossible. To create an objective method capable of being selected as the gold standard, it is fundamental to detect, in a completely automatic manner, keypoints in photographs of women's torso after being subjected to breast cancer surgeries. This paper proposes a deep and a hybrid model to detect keypoints with high accuracy. Our methods are tested on two datasets, one composed of images with a clean and consistent background and a second one that contains photographs taken under poor lighting and background conditions. The proposed methods represent an improvement in the detection of endpoints, nipples and breast contour for both datasets in terms of average error distance when compared with the current state-of-the-art.

***Index Terms*—** Keypoint Detection, Deep Neural Networks, Aesthetic Evaluation, Breast Cancer.

1. MOTIVATION

Breast cancer is the most frequently diagnosed cancer and the leading cause of cancer death in women worldwide [1]. Nevertheless, breast cancer is an increasingly treatable disease, with 10-year survival rate now exceeding 80%. This high survival rate led to an increased interest in the consequences of treatment and, in particular, in its aesthetic outcome.

Breast cancer conservative treatment (BCCT) has become the recommended treatment for early breast cancer with identical [2] oncological outcomes than mastectomy and with a better cosmetic result. Even so, the aesthetic outcome depends on several factors, such as: patient's breast shape, tumour's size and location, and surgical and radiotherapy techniques. Current techniques for the aesthetic evaluation of BCCT outcomes involve, at least partially, subjective assessment made by one or several observers. However, as professionals involved in the treatment are often present in the panel of observers, impartiality is not guaranteed. Furthermore, even specialists tend to disagree regarding assessments result and, therefore, method's reproducibility is questionable.

In order to overcome the reproducibility issue of the subjective assessment, some objective methods for the assessment of BCCT were introduced, in which the systems developed by Fitzal *et al.* [3] and by Cardoso and Cardoso [4] represent the most relevant works. Other works, like the ones presented by Kim *et al.* [5], and Kim *et al.* [6], focused on introducing valuable objective measures for the aesthetic assessment of BCCT but did not present a complete system for the final aesthetic evaluation. None of the aesthetic evaluation systems is entirely automatic (they require manual annotation of some keypoints), they all apply to only the classic conservative treatment (leaving out the new surgical techniques) and they have limited performance. Hence, none of them was selected as the gold standard.

Referred works point out the relevance of symmetry measurements in the aesthetic assessment, which makes the correct detection of keypoints fundamental. Thus, a first step towards achieving the goal of an entirely automatic and objective framework, capable of being selected as a gold standard, is the successful detection of fiducial points. In this sense, this work aims to use deep learning techniques to detect keypoints in photographs of women after being subjected to BCCT.

2. RELATED WORK AND TRADITIONAL BASELINE SYSTEM

In this work we implemented a "traditional" computer vision pipeline based on the current state-of-the-art in the field to

This work was partially funded by the Project "NanoSTIMA: Macro-to-Nano Human Sensing: Towards Integrated Multimodal Health Monitoring and Analytics-/NORTE-01-0145-FEDER-00001" financed by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement, and through the European Regional Development Fund (ERDF), and also by Fundação para a Ciência e Tecnologia (FCT) within PhD grant numbers SFRH/BD/139468/2018 and SFRH/BD/136274/2018.

serve as a baseline. This framework will be presented here along with the related work.

The system uses a multi-step approach: first breast endpoints are detected, which are then used to find the contour for each breast. Finally, nipples are detected. Common to all steps is the modelling of images as graphs.

2.1. Graph Concepts

An image can be seen as a graph by considering each pixel a vertex and pairs of neighbooring pixels as being connected by arcs. Given a graph, $G = (V, A)$, we say G is weighted if for each arc, (v_i, v_j), there is an associated weight, $w(v_i, v_j)$. A path from v_1 to v_n is a sequence of vertices $v_1, v_2, ..., v_n$ such that (v_i, v_{i+1}) is an arc in the graph for all $i \in \{1, 2, ..., n-1\}$. The path's cost is given by $\sum_{i=1}^{n-1} w(v_i, v_{i+1})$.

For this application we are interested in finding image edges of different features of the patients body (trunk, breast and aerola complex). As such, images are modeled as weighted graphs and arc weights are assigned based on the gradient magnitude (with small magnitude resulting in higher arc weight).

2.2. Endpoints, Breast Contour and Nipple Detection

The automatic detection of the breast contour endpoints was firstly presented by Cardoso *et al.* [7]. The proposed method, which was used in our baseline model, assumes photographs contain only the torso of the patient, as shown in Figure 2a. The highest point of the trunk contour in each side is assumed to be an external breast contour endpoint and the internal endpoint is set as the midpoint between the external ones.

First authors compute the stable paths [8] between the middle and bottom row. Paths with a cost higher than half of the maximum are discarded. Among the remaining paths, the closest to the center in each side was considered to be the trunk contour. This contour was then extended by finding the longest shortest path which does not contain a long sequence of consecutive pixels with low gradient magnitude.

Similarly to the previous method, breast contour detection can also be tackled by solving the shortest path problem. The first work on automatic detection of fiducial points in photographs of women after being subjected to breast cancer conservative treatment [9] does precisely this. The inner region of the breast is essentially free of edges. As such, the shortest path between the endpoints is often the breast contour. This work was later extended by Sousa *et al.* which showed that the introduction of shape priors leads to more accurate models [10]. They showed non-parametric priors worked better than parametric ones.

The breast contour detection step is done by the Cardoso *et al.*'s method with the changes proposed by Sousa *et al.*. The introduced shape prior is a non-parametric mask which corresponds to the union of all breast regions subtracted by their intersection after an initial endpoint-based scalling and aligning step.

Finding the nipples' position is the final task in the proposed framework. We followed the method proposed in [11]. In this approach, nipple candidates with high response to the Harris corner detector are found on each breast. For each candidate the shortest closed path around the point is computed. If the candidate is a nipple this path will probably correspond to the areola complex contour. We now have a set of pairs candidate/contour, from which we are going to extract one feature from the candidate (Harris corner quality factor) and three from the associated shortest path (the average magnitude of radial derivative, the shape factor and the equivalent diameter). These are used to train an SVM classifier. In this work the same methodology was used but closed shortest paths were computed in polar coordinates and the dataset was automatically labelled (and not manually) based on the distance to the true nipple position.

3. PROPOSED MODEL

In the previous section, methods for the automatic detection of endpoints, breast contour and nipples were presented. Nevertheless, the three tasks are performed separately in all of those works. In order to improve the state-of-the-art in the aesthetic classification of BCCT [4], an integrated approach for keypoints detection could be of major importance. The computation of all keypoints at the same time may favour the creation of an end-to-end algorithm for the aesthetic evaluation of breast cancer surgery outcomes. Moreover, the context information could also be useful to ease the detection of some keypoints.

Deep Neural Networks (DNN) offer a valuable framework to achieve this integrated learning. However, as in biomedical applications we usually deal with small datasets, DNN tend to overfit severely. There are some approaches that allow to mitigate the effect of overfitting, for example, transfer learning and learning an intermediate representation. This latter idea was explored in other domains in the works done by Belagiannis *et al.* [12] and Cao *et al.* [13]. In both works, an intermediate representation consisting on confidence maps in relation to the location of the keypoints was created. An additional interesting idea also explored in those works was an iterative process of refinement.

Based on the ideas previously mentioned, we have built a DNN to automatically detect keypoints in photographs of patients after being subjected to BCCT (see Figure 2a). As shown in Figure 1, the architecture of the proposed DNN comprises two main modules: regression and refinement of heatmaps, and regression of keypoints.

The first module is what we call as Heatmap Regression and Refinement. Here the goal is to generate an intermediate representation consisting on a fuzzy localization for the keypoints we want to detect. This is done in order to help the

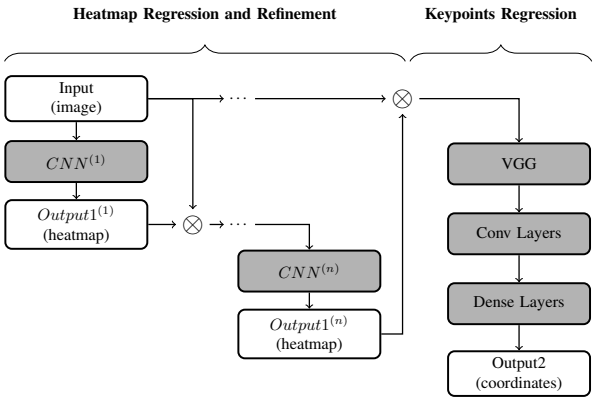
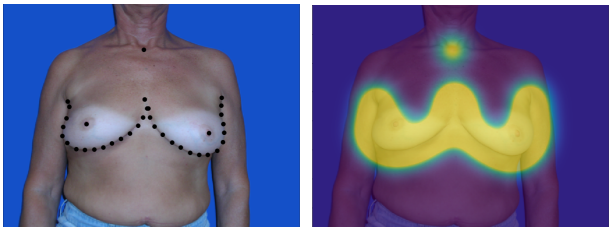


Fig. 1: Proposed iterative DNN architecture.



(a) Keypoints (b) Heatmap

Fig. 2: Example of Ground Truth

regularization process of the DNN. Heatmaps are obtained using the well-known segmentation model, U-Net [14] - referred as CNN. Figure 2b presents an example with an image from the dataset and the respective ground truth heatmap super-imposed.

The second module has as input the multiplication of the image with the refined output of the previous module ($Output1^{(n)}$). The regression of the keypoints is composed of three blocks: VGG16 (without the fully-connected layers), Conv Layers and Dense Layers. The first block, VGG16 [15], is pre-trained with ImageNet and then fine-tuned in our dataset. After VGG16, four convolutional layers are added to further increase image processing and decrease size of feature maps before the dense layers. Finally, three dense layers are used to regress the 74 coordinates, corresponding to the keypoints that make up the breast contour, endpoints, nipples and supra-sternal notch (Figure 2a). The proposed fully supervised learning scheme requires not only the ground truth for the keypoints but also a ground truth for the heatmaps, which is created considering a Gaussian centered at each keypoint, with a pre-defined standard deviation. Regarding the learning process, we have two different terms in the loss function: heatmap regression, which works here as a regularization term, and keypoints regression, our goal. Thus, the loss function is a linear combination between these two terms (Eq. 1).

$$\mathscr{L} = \mathscr{L}_{heatmaps} + \mathscr{L}_{keypoints}. \quad (1)$$

In relation to the regression of the keypoints, the mean squared error (MSE) was the loss function selected (Eq. 2). N_k is the number of coordinates, x_k^{target} the ground truth for a single coordinate and $\hat{x}_k$ the prediction.

$$\mathscr{L}_{keypoints} = \frac{1}{N_k} \sum_{\forall k} (x_k^{target} - \hat{x}_k)^2. \quad (2)$$

The heatmaps were also learnt using MSE. However, the heatmaps undergo an iterative process of refinement. Thus, the complete process is defined by Eq. 3, where j represents a step in the refinement process and λ_j represents the weight given to that step.

$$\mathscr{L}_{heatmaps} = \sum_{j=1}^{N_h} \lambda_j \mathscr{L}_{heatmap}(j). \quad (3)$$

Finally, the loss for the heatmap in each step is defined as follows

$$\mathscr{L}_{heatmap}(j) = \frac{1}{N_p} \sum_{\forall p} (x_p^{target} - \hat{x}_p)^2, \quad (4)$$

where N_p corresponds to the number of pixels in the image, and x_p^{target} and $\hat{x}_p$ to the ground truth and prediction for the pixel values, respectively.

4. EXPERIMENTAL EVALUATION

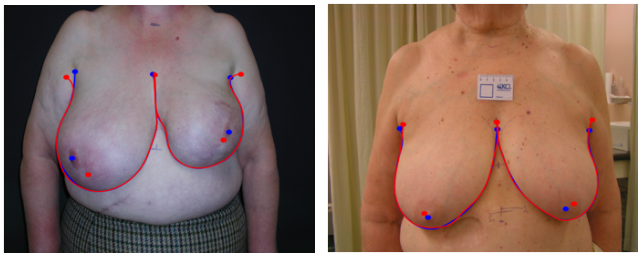
To assess the performance of the proposed method and compare it to the baseline algorithm two datasets were considered. The first was the PORTO dataset, which is the standard dataset used in previous works. This is composed of 120 photographs of patients submitted to BCCT. In these images the torso of the patient is shown in front of a clean and uniform background. The second was obtained by joining three smaller sets of photographs: the 120 images of PORTO dataset, 30 other photographs obtained in similar conditions (TSIO dataset) and 71 additional images captured in poorer lighting conditions and without the concern of having a consistent and distinct background (VIENNA dataset, see Figure 3). For each image, 37 ground truth points were available (4 endpoints, 30 points along the breast contours, 2 nipples and the supra-sternal notch). However, for comparison with the traditional baseline, the supra-sternal notch was not considered. For both datasets, training and test sets were obtained using 5-fold cross-validation.

For the baseline method, the trunk contour extension parameters were optimized by grid searching on the training data. For nipple detection 10 candidates were considered per breast. SVM hyper-parameters were optimized by grid-search using 3-fold cross validation. Regarding the DNN model, hyper-parameters controlling dropout rate and strength of data

120 images dataset - Average error distance (pixels)									
	Endpoints			Breast Contour			Nipples		
Model	**mean**	**std dev**	**max**	**mean**	**std dev**	**max**	**mean**	**std dev**	**max**
Traditional model	50	41	286	11	19	121	63	83	440
DNN model	**30**	**17**	**88**	18	**7**	**43**	**56**	**29**	**130**
Hybrid model	**30**	**17**	**88**	**7**	9	55	**56**	**29**	**130**
221 images dataset - Average error distance (pixels)									
	Endpoints			Breast Contour			Nipples		
Model	**mean**	**std dev**	**max**	**mean**	**std dev**	**max**	**mean**	**std dev**	**max**
Traditional model	81	97	569	36	74	377	123	183	847
DNN model	**38**	**29**	**188**	18	**8**	**56**	**57**	**32**	**195**
Hybrid model	**38**	**29**	**188**	**12**	19	121	**57**	**32**	**195**

Table 1: Average error distance for endpoints, breast contours and nipples measured in pixels.

Fig. 3: VIENNA dataset examples



(a) PORTO dataset example (b) VIENNA dataset example

Fig. 4: Test set examples (221 images dataset). The ground-truth is in blue and the predictions in red.

augmentation were optimized using 5-fold cross validation. In relation to the number of iterations for the heatmap refinement, different numbers of iterations were tested but the one that led to the best results was 3. Weights for each sub-loss function were defined as $\lambda_{j=1} = 0.1$, $\lambda_{j=2} = 0.2$ and $\lambda_{j=3} = 0.4$. Iterative refinement of the keypoints regression module was also investigated but no promising results were obtained. The results for the first and second datasets are depicted in Table 1[1] and were computed considering the average error across the five folds. Test images resolution is variable, with the minimum resolution being (1224×1632 pixels) and the maximum resolution being (2592×3888 pixels).

As shown in Table 1, the DNN approach led to an improvement on the results obtained in two tasks for the first dataset and all three tasks for the second. This method is also faster in inference which can be an important attribute for clinical practice. The baseline algorithm's inference time is in the order of the seconds whereas DNN model's inference time is almost instantaneous (it takes a few milliseconds). Another important difference between the two approaches is that the error on the DNN is more regular, while in the baseline some examples are missed by a wide margin.

Noticing that the inaccuracy in the breast contour of the baseline method was mainly due to a poor estimation of the endpoints, we tested a Hybrid model which used the endpoints detected by the DNN model along with the breast contour algorithm of the baseline solution. A better mean error was obtained when compared to the other two approaches. Figures 4a and 4b show results obtained using the Hybrid approach for examples of PORTO and VIENNA datasets.

[1]Code at `https://github.com/wjsilva19/k_detection`.

5. CONCLUSION

The currently used method to evaluate the aesthetic result of BCCT is subjective and, consequently, it is neither impartial nor reproducible. Given that the main factor determining the aesthetic result is symmetry, keypoints detection is of major importance in the development of a future objective method.

In this work, we propose a DNN model able to improve the state-of-the-art methods in the detection of keypoints in photographs of women's torso after being subjected to BCCT. Moreover, we also propose a Hybrid model consisting on the detection of the endpoints, nipples and supra-sternal notch using the DNN model and finding the breast contour using a shortest path approach. The models were evaluated in two datasets: the first composed only of images with a clean and uniform background and the second with several images taken under poor lighting conditions and without a consistent and clean background. In both datasets, the DNN model surpassed a baseline solution in all tasks but the breast contour. The Hybrid approach obtained the best results in terms of breast contour detection. Future work will focus on computing the shortest path with a neural network and integrating the keypoints detection task with the aesthetic assessment in an end-to-end architecture for classification of breast cancer treatment aesthetic outcomes.

6. REFERENCES

[1] Hélder P. Oliveira, Jaime S. Cardoso, André Magalhães, and Maria J. Cardoso, "Methods for the aesthetic evaluation of breast cancer conservation treatment: a technological review," *Current Medical Imaging Reviews*, vol. 9, no. 1, pp. 32–46, 2013.

[2] Maria J. Cardoso, Jaime S. Cardoso, Hélder P. Oliveira, and Pedro Gouveia, "The breast cancer conservative treatment. cosmetic results - bcct.core - software for objective assessment of aesthetic outcome in breast cancer conservative treatment: a narrative review," *Computer methods and programs in biomedicine*, vol. 126, pp. 154–159, 2016.

[3] F. Fitzal, W. Krois, H. Trischler, L. Wutzel, O. Riedl, U. Kühbelböck, B. Wintersteiner, Maria J. Cardoso, P. Dubsky, M. Gnant, et al., "The use of a breast symmetry index for objective evaluation of breast cosmesis," *The breast*, vol. 16, no. 4, pp. 429–435, 2007.

[4] Jaime S. Cardoso and Maria J. Cardoso, "Towards an intelligent medical system for the aesthetic evaluation of breast cancer conservative treatment," *Artificial Intelligence in Medicine*, vol. 40, no. 2, pp. 115–126, 2007.

[5] Min Soon Kim, William N Rodney, Gregory P Reece, Elisabeth K Beahm, Melissa A Crosby, and Mia K Markey, "Quantifying the aesthetic outcomes of breast cancer treatment: assessment of surgical scars from clinical photographs," *Journal of evaluation in clinical practice*, vol. 17, no. 6, pp. 1075–1082, 2011.

[6] Min Soon Kim, Gregory P Reece, Elisabeth K Beahm, Michael J Miller, E Neely Atkinson, and Mia K Markey, "Objective assessment of aesthetic outcomes of breast cancer treatment: measuring ptosis from clinical photographs," *Computers in Biology and Medicine*, vol. 37, no. 1, pp. 49–59, 2007.

[7] Jaime S. Cardoso, Luis F. Teixeira, and Maria J. Cardoso, "Automatic breast contour detection in digital photographs," in *Proceedings of the International Conference on Health Informatics (HEALTHINF)*, 2008, vol. 2, pp. 91–98.

[8] Jaime S. Cardoso, "Stable text line detection," in *Proceedings of IEEE Workshop on Applications of Computer Vision (WACV) 2009*, 2009.

[9] Jaime S. Cardoso and Maria J. Cardoso, "Breast contour detection for the aesthetic evaluation of breast cancer conservative treatment," in *Computer Recognition Systems 2*, pp. 518–525. Springer, 2007.

[10] Ricardo Sousa, Jaime S. Cardoso, JF Pinto Da Costa, and Maria J. Cardoso, "Breast contour detection with shape priors," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*. IEEE, 2008, pp. 1440–1443.

[11] Jaime S. Cardoso, Ines Domingues, and Helder P. Oliveira, "Closed shortest path in the original coordinates with an application to breast cancer," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 29, 2015.

[12] V. Belagiannis and A. Zisserman, "Recurrent human pose estimation," in *International Conference on Automatic Face and Gesture Recognition*. IEEE, 2017.

[13] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh, "Realtime multi-person 2d pose estimation using part affinity fields," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[14] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.