Towards a Deep Keypoint Detection Algorithm for the Aesthetic Assessment of Breast Cancer Surgery Outcomes

Tiago Gonçalves [1,2]
tiago.f.goncalves@inesctec.pt
Wilson Silva [1,2]
wilson.j.silva@inesctec.pt
Jaime S. Cardoso [1,2]
jaime.cardoso@inesctec.pt

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

Abstract

Nowadays, a vast majority of breast cancers are curable, creating conditions to the study of the aesthetic outcomes of the treatment, which has an impact on the patient's quality of life. To create a fully automatic and objective system for the aesthetic assessment of breast cancer conservative treatment outcomes, the correct detection of breast keypoints in photographs is crucial. This work compares two recently proposed methods, one that only relies on deep learning, and another that combines deep learning with traditional computer vision techniques. On a breast cancer surgery treatment image dataset, the second method leads to better results.

1 Introduction

Breast cancer is the leading cause of cancer death in women worldwide while being the most frequently diagnosed type of cancer. However, if detected in an early stage and properly treated, a vast majority (estimated 90% of the cases) of breast cancers are curable, leaving, thus, space for the study of the aesthetic outcomes of the treatment [8], which may positively contribute to the improvement of the patient's quality of life. Breast cancer conservative treatment (BCCT) has become the recommended approach for early diagnosed breast cancer because it has similar survival rates as mastectomy, but better cosmetic results [5]. Objective methods based on breast symmetry measurements have been proposed [3, 7], but they have not been selected as the gold standard. Currently, the main goal is to achieve a fully automatic and objective system, so, the correct detection of breast keypoints (essential for the symmetry measurements) is fundamental. This work intends to use deep learning techniques to detect keypoints in photographs of women after being subjected to BCCT.

2 Traditional Computer Vision Methods

Traditional computer vision methods consist of modelling images like graphs and follows a multi-step approach: detection of breast endpoints, which are then used to find breast contours, and detection of nipples. An image can be seen as a graph by considering each pixel a vertex and pairs of neighbouring pixels as being connected by arcs; in this case, we focus on finding images edges of different features of patient body (trunk, breast and areola complex), so images are modelled as weighted graphs and arc weights are assigned based on the gradient magnitude (with small magnitude resulting in higher arc weight). The automatic detection of breast endpoints is done as proposed by Cardoso et al. [6], which assumes that photographs contain only the torso of the patient. As such, the highest point of the trunk contour endpoint 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. After the detection of breast endpoints, obtaining breast contour can be seen as a shortest path problem. The inner region of the breast is free of edges, so, the shortest path between the endpoints is usually the breast contour. This was later extended by Sousa et al., leading to more accurate models [12]. To find the nipples' position, the method proposed by Cardoso et al. is followed [4]. In this work, we used the shortest path algorithm to obtain breast contours (see section 4).

3 Deep Keypoint Detection Algorithm

The methods presented in the previous section have to be performed separately, following, thus, an established pipeline. To improve this, an integrated approach for keypoint detection may be of utmost relevance, since, the computation of all keypoints at the same time may favour the creation of an end-to-end algorithm for the aesthetic assessment of breast cancer surgery outcomes. Based on the works done by Belagiannis et al. [1] and Cao et al. [2], Silva et al. [10] proposed a novel deep neural network (DNN) capable of automatically detecting keypoints in photographs of patients after being subjected to BCCT. The architecture of the proposed DNN (Fig. 1) contains two principal modules: regression and refinement of heatmaps, and regression of keypoints. The first module generates an intermediate representation consisting on a fuzzy localization for the keypoints that are supposed to be detected; this acts as a regularization process of the DNN, and it is done with the help of the segmentation model, U-Net [9]. The second module has as input the multiplication of the image with the refined output of the previous module; it is composed of three blocks: VGG16 [11] (without the fully-connected layers, pre-trained with ImageNet and then fine-tuned in our dataset), Convolutional Layers and Dense Layers. It is used to perform the regression of 72 coordinates, corresponding to the keypoints that make up the breast endpoints, breast contour and nipples. To train such DNN, one needs to have the ground truth for the keypoints and the ground truth for the heatmaps, with the latter being created considering a Gaussian centred at each keypoint, with a pre-defined standard deviation (Fig. 2). The loss function (Eq. 1) is also composed of two different terms: heatmap regression, which works here as a regularization term, and keypoints regression, the main goal.

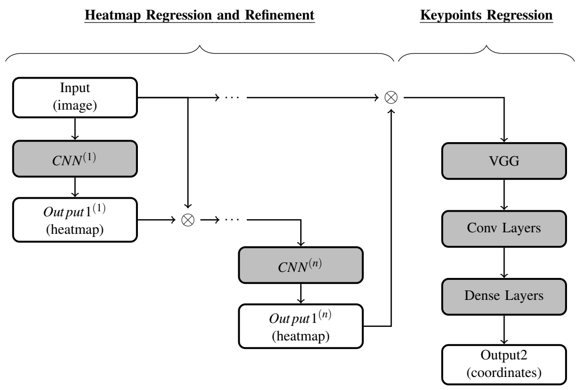

Figure 1: Architecture of the Deep Keypoint Detection Model, from [10].

$$\mathcal{L} = \mathcal{L}_{heatmaps} + \mathcal{L}_{keypoints} \tag{1}$$

The loss for the heatmaps refinement process (Eq. 2) is represented as

$$\mathcal{L} = \sum_{j=1}^{N_h} \lambda_j \mathcal{L}_{heatmap}(j) \tag{2}$$

where N_h is the number of steps, j represents a step in the refinement process, and λ_j is the weight given to that step. Regarding the regression of heatmaps (Eq. 4) and keypoints (Eq. 3), the mean squared error (MSE) was the loss function selected, as follows:

$$\mathcal{L}_{keypoints} = \frac{1}{N_k} \sum_{\forall k} (x_k^{target} - \hat{x}_k)^2 \tag{3}$$

where N_k is the number of coordinates, x_k^{target} is the ground-truth for a single coordinate and $\hat{x}_k$ is the prediction.

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

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

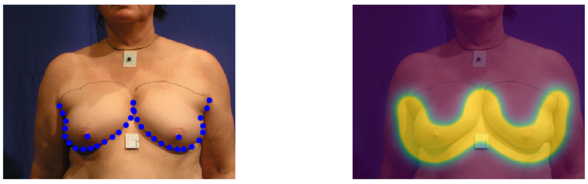

Figure 2: Examples of ground-truth: keypoints and heatmap.

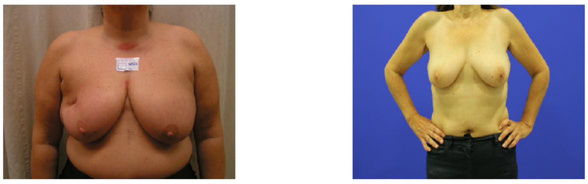

Figure 3: Examples of images of the dataset.

4 Implementation and Results

The dataset considered (745 images) was divided into train (416 images), validation (105 images) and test (224 images) sets (Fig. 3). Images and keypoints were resized to the dimensions of 512×384 and keypoints were divided by the width of the images to be in the interval of [0, 1]. Regarding the deep learning model, data augmentation was employed, in an online setting, to images and keypoints, in order to prevent over-fitting. The model with lower loss on validation set was saved and was used to perform inference on test set. For the experimental evaluation, two different scenarios were tested: the DNN Model, i.e., where all the keypoints are the result of a prediction of the deep model and the Hybrid Model, which receives the predicted endpoints (left, middle-left, middle-right, right) from the DNN Model and applies the shortest path algorithm described in section 2 to extract the breast contour keypoints. Examples of keypoints' predictions for both methods can be seen in Fig. 4 and in Fig. 5. Table 1 presents the average error distance (measured in pixels) for endpoints, breast contours and nipples. **Note: Mean** corresponds to the mean error, **STD** stands for standard deviation and **Max** is the max error value.

Table 1: Average error distance for endpoints, breast contours and nipples, measured in pixels. Best results are highlighted in bold.

Model	Endpoints			Breast Contour			Nipples		
	Mean	STD	Max	Mean	STD	MAX	Mean	STD	Max
DNN Model	**12**	**10**	**77**	5	**2**	**17**	**13**	**8**	**47**
Hybrid Model	**12**	**10**	**77**	**4**	5	48	**13**	**8**	**47**

Figure 4: Examples of keypoints predictions, obtained with the Deep Model. Ground-truth is in blue and prediction is in red.

Figure 5: Examples of keypoints predictions, obtained with the Hybrid Model. Ground-truth is in blue and prediction is in red.

5 Conclusions and Future Work Recommendations

In this work, we did a comparison between the DNN Model and the Hybrid Model, proposed by Silva et al. [10]. The evaluation was performed on a dataset with more images and with more variability regarding lighting conditions, background and patient position (Fig. 3). However, the Hybrid Model surpasses the DNN Model in the breast contour task, as published in the original work. On the other hand, the DNN Model's inference time is much faster, representing, thus, an advantage against the Hybrid Model. Regarding future work, we will focus on integrating the detection of keypoints (which is of utmost importance for the symmetry measurements) with the aesthetic assessment, to create an end-to-end architecture for classification of breast cancer conservative treatment aesthetic outcomes.

References

[1] Vasileios Belagiannis and Andrew Zisserman. Recurrent human pose estimation. In 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), pages 468–475. IEEE, 2017.

[2] Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 7291–7299, 2017.

[3] 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, 40(2):115–126, 2007.

[4] Jaime S Cardoso, Inês Domingues, and Hélder P Oliveira. Closed shortest path in the original coordinates with an application to breast cancer. International Journal of Pattern Recognition and Artificial Intelligence, 29(01):1555002, 2015.

[5] Maria Cardoso, Jaime Cardoso, Helder 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, 126, 12 2015. doi: 10.1016/j.cmpb.2015.11.010.

[6] Maria J Cardoso and Jaime Cardoso. Automatic breast contour detection in digital photographs. 2008.

[7] 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, 16(4):429–435, 2007.

[8] Helder P. Oliveira, Jaime S. Cardoso, Andre Magalhaes, and Maria J. Cardoso. Methods for the Aesthetic Evaluation of Breast Cancer Conservation Treatment: A Technological Review. Current Medical Imaging Reviews, 9(1):32–46, April 2013. ISSN 15734056. doi: 10.2174/1573405611309010006. URL `http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1573-4056&volume=9&issue=1&spage=32`.

[9] 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, pages 234–241. Springer, 2015.

[10] Wilson Silva, Eduardo Castro, Maria J Cardoso, Florian Fitzal, and Jaime S Cardoso. Deep keypoint detection for the aesthetic evaluation of breast cancer surgery outcomes. In 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pages 1082–1086. IEEE, 2019.

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

[12] Ricardo Sousa, Jaime S Cardoso, JF Pinto Da Costa, and Maria João Cardoso. Breast contour detection with shape priors. In 2008 15th IEEE International Conference on Image Processing, pages 1440–1443. IEEE, 2008.