Segmentation of Thermal Breast Images Using Convolutional and Deconvolutional Neural Networks

Shuyue Guan, Nada Kamona, Murray Loew
Department of Biomedical Engineering
The George Washington University Medical Center
Washington DC, USA
{frankshuyueguan, nkamona, loew}@gwu.edu

***Abstract* — Breast cancer is the second leading cause of death for women in the U.S. Early detection of breast cancer has been shown to be the key to higher survival rates for breast cancer patients. We are investigating infrared thermography as a noninvasive adjunctive to mammography for breast screening. Thermal imaging is safe, radiation-free, pain-free, and non-contact. Segmentation of breast area from the acquired thermal images will help limit the area for tumor search and reduce the time and effort needed for manual hand segmentation. Autoencoder-like convolutional and deconvolutional neural networks (C-DCNN) are promising computational approaches to automatically segment breast areas in thermal images. In this study, we apply the C-DCNN to segment breast areas from our thermal breast images database, which we are collecting in our clinical trials by imaging breast cancer patients with our infrared camera (N2 Imager). For training the C-DCNN, the inputs are 132 gray-value thermal images and the corresponding manually-cropped breast area images (binary masks to designate the breast areas). For testing, we input thermal images to the trained C-DCNN and the output after post-processing are the binary breast-area images. Cross-validation and comparison with the ground-truth images show that the C-DCNN is a promising method to segment breast areas. The results demonstrate the capability of C-DCNN to learn essential features of breast regions and delineate them in thermal images.**

Keywords—breast thermogram segmentation; convolutional neural networks; deep learning; deconvolutional neural networks; infrared thermography; computer aided diagnosis; autoencoder

I. INTRODUCTION

Breast cancer will be diagnosed among about 12% of U.S. women during their lifetime, making it the second leading cause of death for U.S. women [1], [2]. Early detection of breast cancer via computer-aided detection (CAD) systems has shown to improve outcomes of breast cancer treatment and increase patients' survival times [3]. If the tumor is detected and localized early, the 5-year relative survival rate is more than 94% [4].

X-ray mammography is the gold standard for breast cancer detection, which nevertheless has a substantial false-positive rate, requires exposure to radiation, often is uncomfortable, and is less effective for dense breasts cases. Alternatively, thermography has been shown to be a promising noninvasive adjunct to mammography for breast cancer screening [5]. Studies show thermography being a reliable indicator of increased risk of breast cancer at early stages [6]. We are conducting clinical trials to image patients diagnosed with breast cancer with our infrared camera.

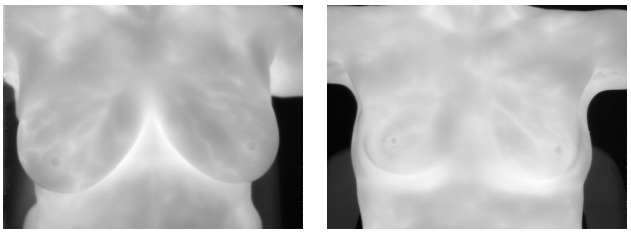


Fig. 1. Full thermal raw images of two patients, including the neck, shoulder, abdomen, background and chair.

As seen in Fig. 1, thermal images contain various parts in addition to the region of interest required for automatic detection of abnormalities in the breast. These parts include the background, chair, neck, shoulders and lower abdomen. Removing all these unnecessary parts and extracting the breast region is a crucial preprocessing step for CAD systems of breast cancer. Automatic segmentation of the breast region will limit the area for tumor search and reduce processing time. Furthermore, automatic segmentation reduces the time and effort required in manual segmentation, and potentially minimize human error.

The accurate segmentation of the thermal breast images is still a difficult task [7] because the breast thermogram has inherent limitations such as low contrast, low signal to noise ratio [8], lack of clear edges and no definite shape. Previous studies for breast thermogram segmentation mostly used histogram analysis, threshold-based techniques, edge-based techniques, and region-based techniques, for example, edge detection by Hough transform feature curve (parabola) extraction [9], edge detection by interpolation of curves [10], Snake algorithm [11], detection of edge and boundary curves [12], anisotropic diffusion filter-based edge detection [13] and automated segmentation algorithm based on ellipse detection (our lab).

Recently, deep-learning is becoming a state-of-the-art method to segment images. The SegNet [14], for example,

was trained to segment urban street images to parts of sky, building, road marking, pavement etc. For medical images applications, deep neural networks have applied to segmentation of retinal vessels, brain tissues in MRI and liver lesion in CT [15]–[17]. However, we are aware of no study that has been using deep-learning based segmentation to segment breast thermograms. This study fills this gap by providing a deep-learning model to automatically segment breast area from the whole thermal breast image.

Compared with the ground truth data (manually cropped breast regions), our segmentation results show an average Intersection-over-Union (IoU) of 0.94 for training and testing by the same patient and 0.83 for training and testing by different patients. Therefore, our proposed deep-learning model can be used to segment breast regions and training samples including the same patient can improve segmentation performance.

II. Materials

A. Breast Thermography

Tumor growth is accompanied by angiogenesis, or the formation of new blood vessels. This growing vascular network supplies the developing tumor with nutrients and oxygen and remove waste products. The increased blood flow in the tumor region results in an increased local temperature of that region compared to the temperature of the surrounding tissues. Previous work indicates that differences of as little as 0.1 K can be clinically important [6]. Therefore, thermography has the potential to detect those elevated skin temperatures that arise from the increased blood flow.

We are using the N2 Imager (N2 Imaging Systems, Irvine, Calif.) borrowed from the U.S Army Night Vision Laboratory. The camera detects wavelengths in the long infrared region in the electromagnetic spectrum, capturing the range 8-12 microns, and humans at normal body temperature radiate thermal infrared at wavelengths around 10 microns. It has a 640x480 array of 17-micrometer pixels, and a stated thermal resolution of 18.6 mK.

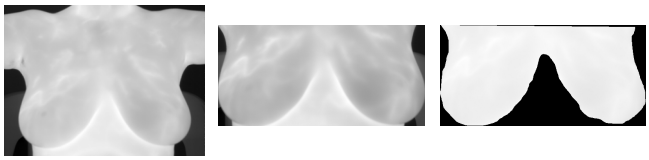
Fig. 2. Our breast infrared thermography system.

Preliminary clinical trials are currently being conducted in collaboration with the Breast Clinic at the GW Medical Faculty Associates. Patients diagnosed with breast cancer are imaged with the infrared camera for a total time of 15 minutes to observe cool down of the breast tissue [6]. This dynamic thermography monitors the temporal behavior of breast thermal patterns, which in our case is the cool down of breast tissue over time. The patient sits still with both arms raised on two arm supports, with the camera positioned approximately 25 inches away from the patient (frontal view). Imaging starts immediately after the patient undresses, capturing images every minute as the patient's skin cools down. Fig. 2 shows the environment for our breast image acquisition.

B. Image Collection and Pre-processing

We collected data from 11 breast cancer patients, with 15 images for each case, at a rate of one image per minute. We chose 15 minutes for our screening time similar to proposed IR image acquisition protocols found in literature, which typically takes 4 to 15 minutes [18]. The 15 minutes allow enough time to observe cooling of the breasts since we are allowing natural cool down of the breast tissue instead force-cooling with a fan, and without causing discomfort for the patient by sitting still for a long period of time.

Initially, images were manually cropped by removing the upper and lower regions (neck and abdomen). All breast IR images were converted to be 8-bit gray-scale. Then, a trained student **manually** traced the breast curvature and cropped the breast region from the rest of the body to form the **ground truth breast region images** for training and testing the segmentation model (Fig. 3). In practice, these truth breast region images were set to be binary values, where (in gray-scale) 0 (black) is for background and 255 (white) for breast areas.



a) b) c)

Fig. 3. Preprocessing of the raw IR images: a) original raw IR image, b) manual rectangular crop to remove shoulders and abdomen, and c) is the hand-trace of the breast contour to generate the manual segmentation (ground truth).

III. Methods

A. Segmentation Architecture

The convolutional neural network (CNN) is a neural network incorporating many convolutional layers. It is the commonly used deep-learning method for image processing because using convolutional layers instead of fully connected layers reduces the total number of weights needed for input images. Also, convolutional layers could keep local structures of images. CNN usually transforms an image to a vector, while deconvolutional neural networks (DCNN) apply the opposite transformation of CNN by transforming a vector to an image. In this study, we used the deep-learning based segmentation model: C-DCNN [19], which connects a CNN and DCNN together. Firstly, convolutional layers transform a 2-D image to a feature vector; then deconvolutional layers convert a vector to image. Overall, the model transforms images to images. In our experiments, the segmentation model

converts infrared breast images to segmented breast region images. Such architecture is also named autoencoder [20] – CNN is the encoder and the feature vectors are "codes" of inputted images. For deep-learning segmentation model, the "codes" could be smaller images or other shapes of data besides vectors.

The details of the C-DCNN structure is shown in TABLE I It consists of six convolutional layers, two max-pooling, two up-sampling layers, three fully-connected (FC) layer and one flatten layer. The activation function for each layer is the ReLU function [21] except the last one for the output being tanh function.

TABLE I. C-DCNN SEGMENTATION ARCHITECTURE FOR THERMAL BREAST IMAGES

Layer	Shape
input: gray-scale image	400x200x1
Conv_3-1 + ReLU Normalization	400x200x1
Conv_3-64 + ReLU	400x200x64
MaxPool_2 Normalization	200x100x64
Conv_3-128 + ReLU	200x100x128
MaxPool_2 Normalization	100x50x128
Flatten	640000
FC 200 + ReLU	200
FC 200 + ReLU	200
FC 640000 + ReLU	640000
Reshape to 100x50x128	100x50x128
Normalization Up-sampling	200x100x128
Conv_3-128 + ReLU	200x100x128
Normalization Up-sampling	400x200x128
Conv_3-64 + ReLU Normalization	400x200x64
output: Conv_3-1 + tanh	400x200x1

The notation Conv_3-64 means there are 64 convolutional neurons (units), with each unit having a filter size of 3×3-pixel (height × width) in the layer. MaxPool_2 is a max-pooling layer with the filter size 2×2-pixel window, stride 2; up-sampling layers have the same size. FC_200 is a fully-connected layer containing 200 units. Normalization is the batch normalization layer, which normalizes the activations of the previous layer at each batch and helps accelerate deep network training [22] . The output layer uses the tanh function, which maps the output value to the range of [-1, 1].

Data shapes from input to output are symmetric. The CNN (encoder) transforms an image to a 200-length vector (code) and the D-CNN (decoder) transforms the vector back to an image. The 8-bit gray-scale input images were scaled from [0, 255] to [-1,1] to match the value range required for the neural network input. Similarly, the neural network segmented output image is then rescaled back to uint-8 [0, 255].

B. *Experiment Designs*

Exp.1 Since all image samples were from 11 breast cancer patients, with 15 samples for each patient, the first experiment randomly selected 12 samples from each patient for the training set and the remaining 3 samples for the testing set. In total, there are 132 breast infrared images along with 132 manual segmented regions for training the segmentation model, and 33 breast infrared images for testing.

Exp.2 In this leave-one-case-out experiment, one patient is left out of the training set by taking all 15 samples of the patient as the testing set, while 150 images of all other 10 patients are used for training. The goal of this experiment is to evaluate the performance of the segmentation model on brand new breast shapes (cases) that were not seen during training. Fig. 4 shows the two proposed experiments.

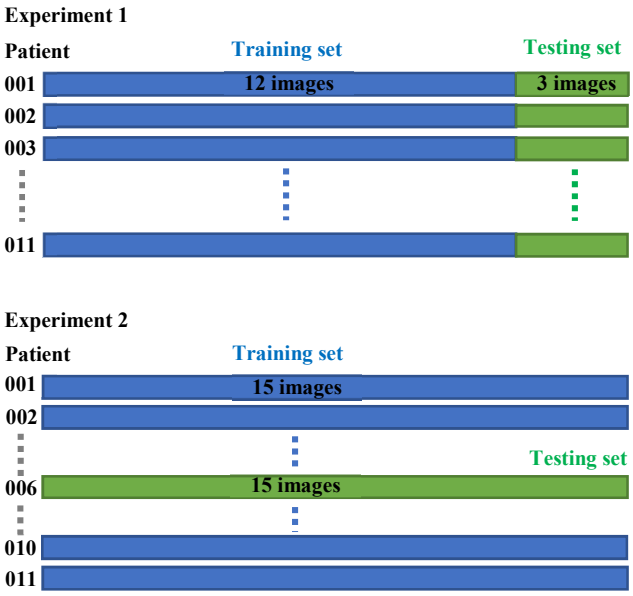


Fig. 4. Training and testing data for Experiment 1 and 2

C. *Testing and Evaluation Criterion*

After training the segmentation C-DCNN model, we put the testing breast infrared (IR) images in the model; as we discussed in the Method section, the outputs are predicted breast area images in gray-scale. Since the truth breast region images are binary, to compare the predicted images with truth data, we applied the Otsu's [23] algorithm to automatically convert gray-scale segmentation images to be binary segmentation images.

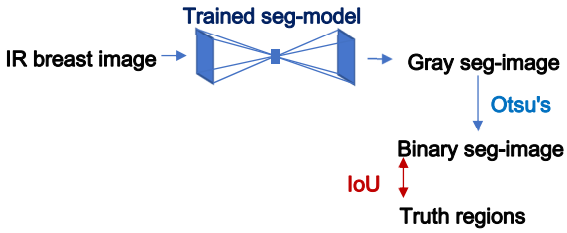


Fig. 5. The evaluation processes

We compared the binary segmentation images with truth region images by computing their Intersection-over-union (IoU), also called the Jaccard Similarity (See Fig. 5). The IoU of two binary images is the ratio of overlapped area divided by area of union. Therefore, for two binary images I_1 and I_2 of the same size, the IoU is:

$$\mathrm{IoU}(I_1, I_2) = \frac{I_1 \cap I_2}{I_1 \cup I_2}$$

If the two images have the same breast region, IoU will be 1. For all the testing results, we computed their IoU with ground truth manual segmented regions to evaluate the segmentation C-DCNN model.

IV. Results and Evauluations

Our implementation of neural networks was on the Keras API backend on TensorFlow [24]. The development environment for Python was Anaconda3. We set 1000 epochs for training and the batch-size is 6. The loss-function for training is mean square error (MSE) between predicted segmentation images and truth images, and optimizer is Nadam [25] using default parameters (except the learning rate changed to 1e-4).

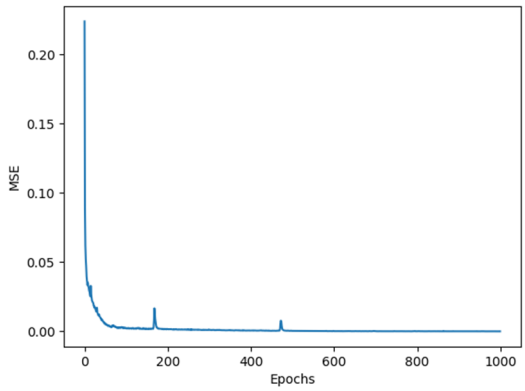


Fig. 6. The training curve

One training curve of MSE is shown in Fig. 6; it shows that MSE decreased fast at the beginning and converged close to 0 after 600 epochs. In Experiment 1, the model was trained by 132 samples and testing on 33 samples once. In Experiment 2, we trained the C-DCNN model from scratch for 11 times by leave-one-case-out training and testing. On average, each epoch took 7 seconds by using single the *GTX-1080 Ti* GPU.

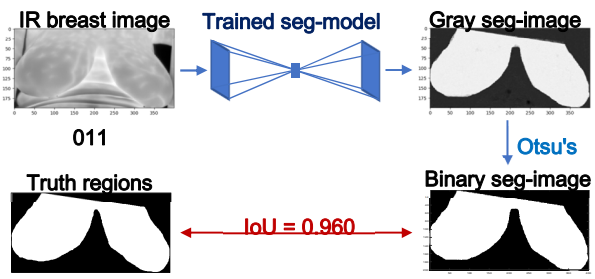


Fig. 7. Segmentation results of one patient from Exp.1

Fig. 7 shows sample segmentation result of one patient from Experiment 1, where 33 test samples come from 11 patients, with 3 testing breast images for each patient. In Fig. 8, bars show the ranges among 3 samples for the same patient because breast images may be slightly different due to patient movement during image acquisition and breast temperature change over time. The overall average IoU is about 0.9424 with 0.0248 standard deviation.

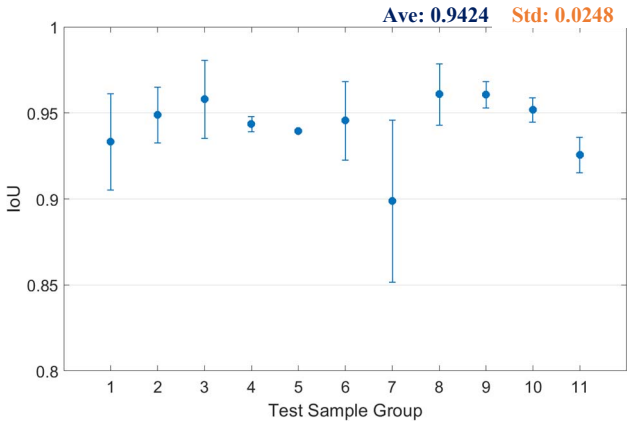


Fig. 8. Results of Exp.1. The blue dots are the average IoU for each patient and bars show the range among 3 samples.

In Experiment 1, training and testing sets contain images of the same patient. That is, IR breast images having the same region are in both training and testing sets. A possible explanation for the good segmentation performance of the C-DCNN might be that it **memorized the breast region** for each patient but have not learned how to segment breast regions. Therefore, Experiment 2 evaluated the trained segmentation model by different patients' IR breast images from training to avoid the memorization.

Fig. 9 shows two result from Experiment 2. In the first row, the segmentation model was trained by 10 patients' IR breast images without patient 002 (leave-one-case-out). All 15 test images come from the 002 patients. The predicted segmentation breast region seems to be synthesized by several trained breast areas from other patients and the segmentation result of patient 002 is not as accurate as the results from Experiment 1, however, the predicted breast area still covers most of the ground truth breast area. The second row shows a better example for patient 007.

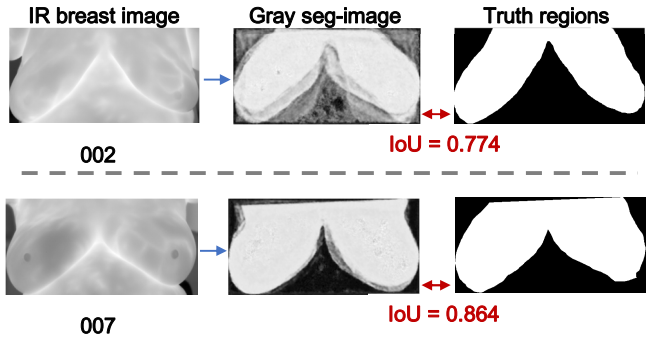


Fig. 9. Segmentation results of two patients from Exp.2

Fig. 10 shows that the average IoUs in most cases are better than 0.8, and the overall average IoU is about 0.8340 with 0.0809 standard deviation. This is relatively high because of the wide variety of breast shapes and contours among different patients. Low average IoU for each case means the breast region (shape and contour) is quite different from other cases in training set. Higher IoU cases mean there are similar breast areas in the training set (See Discussion section).

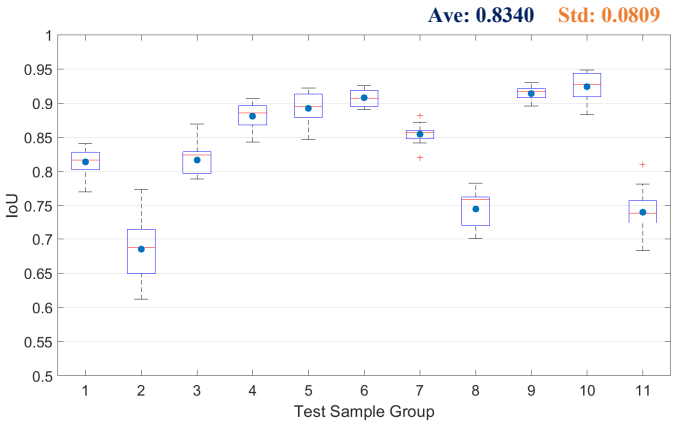


Fig. 10. Results of Exp.2. The blue dots are the average IoU of each patient among its 15 testing samples, the red lines are medians and the bars show the ranges.

V. Discussion

A. Comparsion of the Two Experiments

The average IoU of Exp.1 is 0.9424 and Exp.2 is 0.8340; the results illustrate that if the segmentation model was trained by some IR breast images from one patient, it can have a better performance to segment the breast region in other samples from the same patient (Fig. 11).

In the top part of Fig. 11, one IR breast image of patient 001 (p.001) was put into the two trained segmentation models: one model was trained without p.001's samples (Exp.2) and another one was trained with some p.001's samples (different from the input one) (Exp.1). Both the outputs and IoUs demonstrate that the outcomes by training with or without the same patient's samples can have big differences, with results from the Exp.1 (with the same patient's samples) being better. For p.001, the output from Exp.1 looks very similar to the ground truth region, however, the predicted segmentation area from Exp.2 looks like breast regions from other patients used in the training set.

On the contrary, in the bottom part of Fig. 11, the segmentation outputs for p.009 from the two experiments are very similar. This is because another patient (p.010) has a similar IR breast region (breast shape and contour) as p.009. The segmentation model was trained by similar-looking breast regions. Therefore, if training samples includes breast images very similar to the test image, the segmentation outcomes become better. Such results indicate that including more IR breast images of various breast shapes and sizes in the training process of the C-DCNN segmentation model will greatly improve segmentation results of breast regions of any new coming breast images.

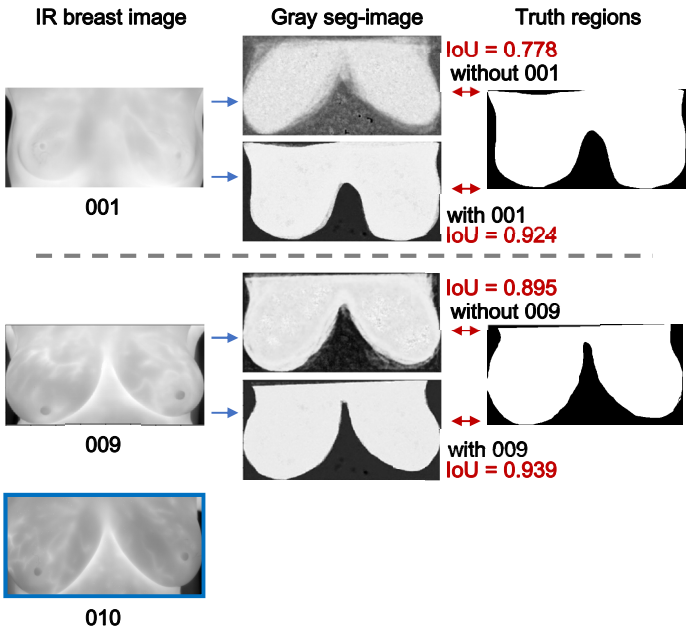


Fig. 11. Comparison of results from the two experiments (*first row*: Exp.2, *second row*: Exp.1). The second column (Gray seg-image) shows output of segmentation models. The third column is the ground truth breast region of the patient's testing samples. (*Top part*: p.001, *bottom part*: p.009).

B. Problems and Future Works

Our lab is currently developing automated segmentation algorithms based on ellipse detection and Snake algorithm, and future work involves comparing results from deep-learning based segmentation with from other segmentation algorithms. In addition, both Otsu's thresholding and IoU computation used in this study have limitations. Although Otsu's thresholding converts gray-scale images to binary automatically, it cannot guarantee getting the best breast region and possibly affecting the outcomes. IoU is used to compare two binary regions but results are subject to image size. There could exist multiple ways to segment the breast region from IR images, suggesting some limitations to the manual hand segmentation. For instance, a predicted segmentation area with low IoU value by the C-DCNN segmentation model might still be a reasonable way to segment the breast region even if it does not match the manual segmentation. Hence, a better evaluation metric needs to be developed to assess the quality of the breast segmentation by our developed model. In future studies, we will consider applying other thresholding and region or contour comparison methods.

For future works, one approach to improve outcomes is to combine deep-learning based segmentation with other methods for pre/post-processing, such as the contrast limited adaptive histogram equalization (CLAHE). Since the histogram equalization globally changes the images, it may not be achieved by CNN because the convolutional operations are localized and could play the roles of various image filters. From Exp.2 we know that more varieties and number of

training images could benefit the C-DCNN segmentation model, thus we will collect more patients and volunteers' samples for future training. Also, we could try to train deep-learning segmentation models other than C-DCNN, such as SegNet [14] or U-Net [26].

VI. Conclusion

Through cross-validation and comparison with the ground-truth images, the results demonstrate the capability of C-DCNN to learn essential features of breast regions and delineate them in thermal images, showing that the C-DCNN is a promising method to segment breast regions. Adding training samples from the same patient can improve segmentation performance from 0.83 to 0.94 based on our evaluation criterion (IoU). The outcomes by training without the same patient is still acceptable. This fact reveals a variety of breast shapes in training samples help improve performance of the segmentation model. To the deep-learning model, greater number of training samples could also improve its performance. To the same patient, the segmentation model can create more regions instead of manually cropping after training by existing samples.

References

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2016," *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 7–30, Jan. 2016.

[2] C. E. DeSantis, S. A. Fedewa, A. Goding Sauer, J. L. Kramer, R. A. Smith, and A. Jemal, "Breast cancer statistics, 2015: Convergence of incidence rates between black and white women," *CA. Cancer J. Clin.*, vol. 66, no. 1, pp. 31–42, Feb. 2016.

[3] V. M. Rao, D. C. Levin, L. Parker, B. Cavanaugh, A. J. Frangos, and J. H. Sunshine, "How Widely Is Computer-Aided Detection Used in Screening and Diagnostic Mammography?," *J. Am. Coll. Radiol.*, vol. 7, no. 10, pp. 802–805, Oct. 2010.

[4] C. DeSantis, R. Siegel, and A. Jemal, "Breast Cancer Facts & Figures 2015-2016.," p. 44.

[5] E. Y.-K. Ng, "A review of thermography as promising non-invasive detection modality for breast tumor," *Int. J. Therm. Sci.*, vol. 48, no. 5, pp. 849–859, May 2009.

[6] L. Jiang, W. Zhan, and M. H. Loew, "Modeling static and dynamic thermography of the human breast under elastic deformation," *Phys. Med. Biol.*, vol. 56, no. 1, pp. 187–202, Jan. 2011.

[7] T. B. Borchartt, A. Conci, R. C. F. Lima, R. Resmini, and A. Sanchez, "Breast thermography from an image processing viewpoint: A survey," *SIGNAL Process.*, vol. 93, no. 10, SI, pp. 2785–2803, Oct. 2013.

[8] Q. Zhou, Z. Li, and J. K. Aggarwal, "Boundary Extraction in Thermal Images by Edge Map," in *Proceedings of the 2004 ACM Symposium on Applied Computing*, New York, NY, USA, 2004, pp. 254–258.

[9] H. Qi, W. Snyder, J. F. Head, and R. L. Elliott, "Detecting breast cancer from infrared images by asymmetry analysis," 2000, vol. 2, pp. 1227–1228 vol.2.

[10] N. Scales, C. Kerry, and M. Prize, "Automated image segmentation for breast analysis using infrared images," in *The 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2004, vol. 1, pp. 1737–1740.

[11] N. EDDIEY.-K, "Segmentation of Breast Thermogram : Improved Boundary Detection with Modified Snake Algorithm," 2006.

[12] P. Kapoor, S. V. A. V. Prasad, and S. Patni, "Image Segmentation and Asymmetry Analysis of Breast Thermograms for Tumor Detection," *Int. J. Comput. Appl.*, vol. 50, no. 9, pp. 40–45, Jul. 2012.

[13] S. S. Suganthi and S. Ramakrishnan, "Anisotropic diffusion filter based edge enhancement for segmentation of breast thermogram using level sets," *Biomed. Signal Process. Control*, vol. 10, pp. 128–136, Mar. 2014.

[14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[15] M. Melinščak, P. Prentašić, and S. Lončarić, "Retinal Vessel Segmentation Using Deep Neural Networks," *VISAPP 2015 10th Int. Conf. Comput. Vis. Theory Appl. Proc. Vol1*, p. 577, May 2015.

[16] P. Moeskops *et al.*, "Evaluation of a deep learning approach for the segmentation of brain tissues and white matter hyperintensities of presumed vascular origin in MRI," *NeuroImage Clin.*, vol. 17, pp. 251–262, Jan. 2018.

[17] P. F. Christ *et al.*, "Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields," *undefined*, 2016. [Online]. Available: /paper/Automatic-Liver-and-Lesion-Segmentation-in-CT-Using-Christ-Elshaer/59d5a686c82c9c73ee5f6ef42b8be0c245dc84b1. [Accessed: 06-Oct-2018].

[18] L. Silva *et al.*, "A New Database for Breast Research with Infrared Image," *J. Med. Imaging Health Inform.*, vol. 4, pp. 92–100, Mar. 2014.

[19] H. Noh, S. Hong, and B. Han, "Learning Deconvolution Network for Semantic Segmentation," presented at the Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1520–1528.

[20] P. Baldi, "Autoencoders, unsupervised learning, and deep architectures," in *Proceedings of ICML workshop on unsupervised and transfer learning*, 2012, pp. 37–49.

[21] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proceedings of the 27th international conference on machine learning (ICML-10)*, 2010, pp. 807–814.

[22] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," *ArXiv150203167 Cs*, Feb. 2015.

[23] T. Hastie, R. Tibshirani, and J. Friedman, "Unsupervised Learning," in *The Elements of Statistical Learning*, Springer, New York, NY, 2009, pp. 485–585.

[24] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *ArXiv160304467 Cs*, Mar. 2016.

[25] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ArXiv14126980 Cs*, Dec. 2014.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, 2015, pp. 234–241.