Segmentation of Breast Cancer Masses in Digital Mammograms: A Convolutional Network

Master's thesis

Erick Cobos Tandazo

Advisor:
Dr. Hugo Terashima Marín
Supervisors:
Dr. Santiago Enrique Conant Pablos   Dr. José Gerardo Tamez Peña

Centro de Sistemas Intelligentes
Tecnologico de Monterrey

December, 2016

Table of Contents

Thesis objective

The main goal is to segment breast cancer lesions using convolutional networks, an end-to-end learnable model.

- Obtain and process the mammographic database.
- Develop software to handle the database and train new deep learning models.
- Develop and train modern, fine-tuned convolutional networks.
- Test the viability of convolutional networks for breast cancer research through different experiments.
- Propose ideas for future research in the area.

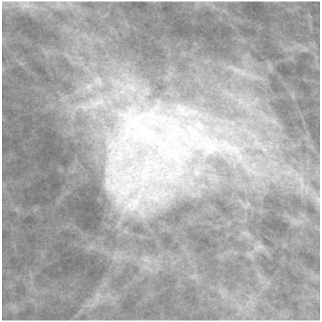
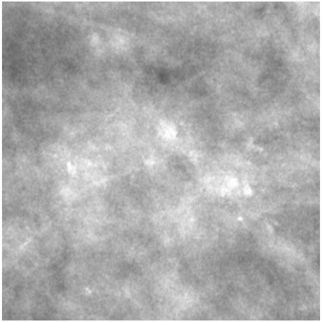
Table of Contents

Why breast cancer?

Breast cancer is the most commonly diagnosed cancer among women.
For women, death rates are the second highest of any cancer. However, survival rate when detected early is close to 100%.
On-going project at the institution.

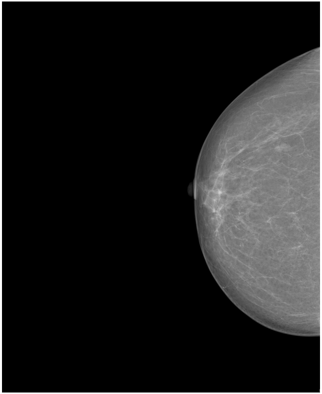
Breast cancer signs

Masses vs. microcalcifications

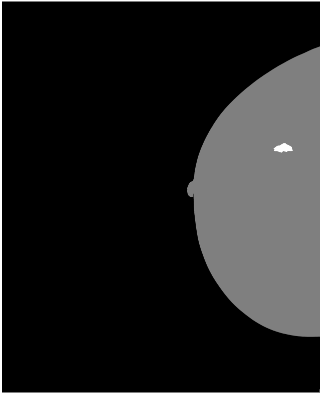
 

Detection vs. diagnosis

Lesion segmentation



(a) Mammogram



(b) Segmentation

Table of Contents

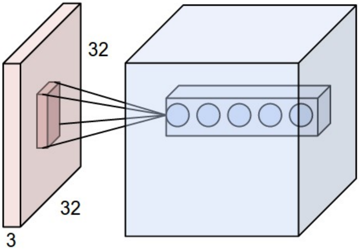
Why convolutional networks?

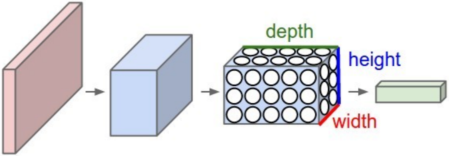
- Convnets have showed great results in image classification tasks.
- Convnets learn which features are important for the classification.
- We don't need experts to carefully handcraft and select features.

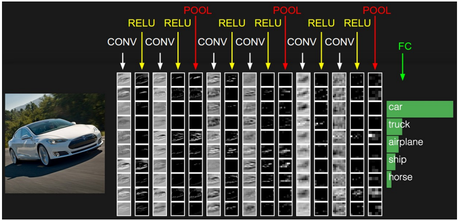
Cons: Need processing power, data.

Convolutional networks

Learnable filters: spatially local and simple in early layers, global and complex in deeper layers.



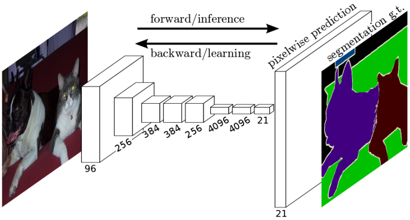


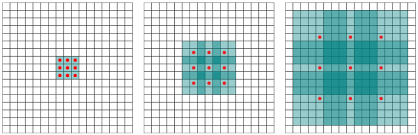


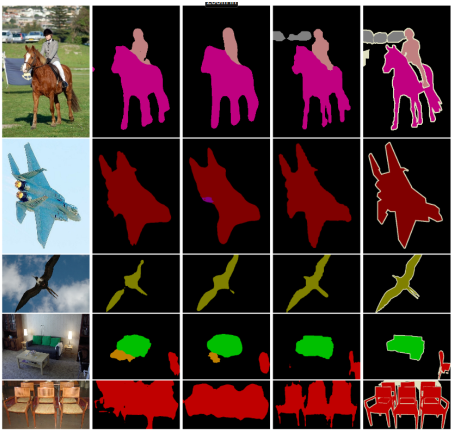
[Karpathy et al., 2016]

Segmentation

Solved by upsampling, deconvolution or dilated convolutions.





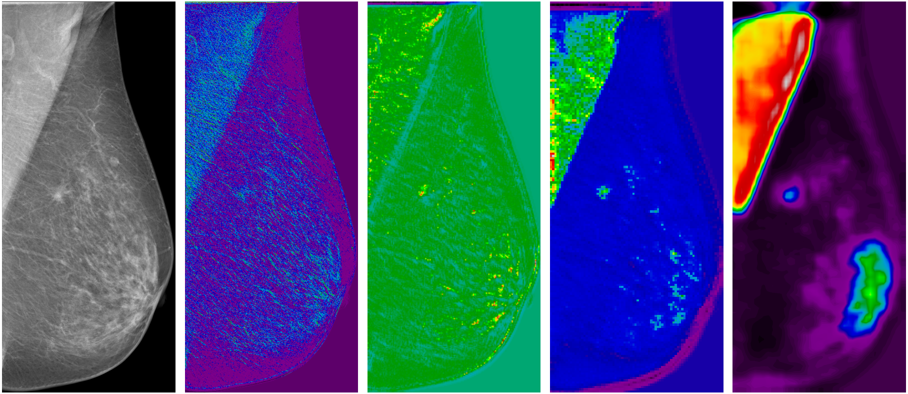


(a) Image (b) FCN-8s (c) DeepLab (d) Our front end (e) Ground truth

[Long et al., 2015], [Yu and Koltun, 2016]

Example

Selected feature maps from different layers.



(a) Input (1408×624)

(b) Layer 1 (704×312)

(c) Layer 4 (352×156)

(d) Layer 7 (176×78)

(e) Output (1408×624)

Table of Contents

Literature review

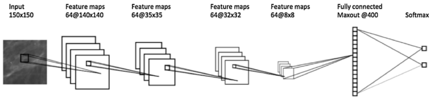
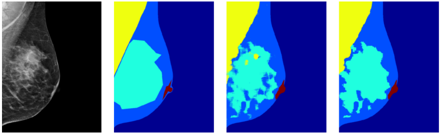


Table 2 – Summary of results in terms of AUC in the test set. Best results are shown in bold typeface and (*) signals scores with no evidence of differences from the highest ($\rho < 0.1$).

Representation	Standalone	Combined with HCfeats
CNN3	**0.82 ± 0.03**	0.82 ± 0.03 (*)
CNN2	0.76 ± 0.05	0.78 ± 0.04
HGD	0.78 ± 0.04	**0.83 ± 0.04**
HOG	0.77 ± 0.03	0.81 ± 0.03 (*)
DeCAF	0.79 ± 0.05	0.82 ± 0.03 (*)
HCfeats	0.77 ± 0.02	–

(a) Mass diagnosis (4 layers, 3.4M params)

Layer	1	2	3	4	5	6	7 - Output
Stage	conv+relu+max	conv+relu+max	conv+relu+max	dropout	full+relu	full+relu	full
# channels	16	16	16	16	128	16	4
Filter size	7×7	5×5	5×5	-	-	-	-
Pooling size	3×3	3×3	3×3	-	-	-	-
Pooling stride	2	2	2	-	-	-	-
Dropout factor	-	-	-	0.5	-	-	-
Spatial input size	61×61	27×27	11×11	3×3	3×3	1×1	1×1



	Pectoral muscle	Fibroglandular tissue	Nipple	Breast tissue (fat, skin)	Average over four tissues
Raw DNN output	0.78	0.60	0.47	0.80	0.66
Post-processed DNN output	0.85	0.61	0.56	0.81	0.71
Raw DNN output, with (x, y) coordinates	0.78	0.60	0.56	0.77	0.68
Post-processed, with (x, y) coordinates	0.79	0.61	0.57	0.77	0.69

Table 2. Average Dice coefficients obtained using the proposed methods for different breast tissues, and average coefficients over the four tissues.

(b) Tissue segmentation (6 layers, 34K params)

[Arevalo et al., 2016], [Dubrovina et al.,2015]

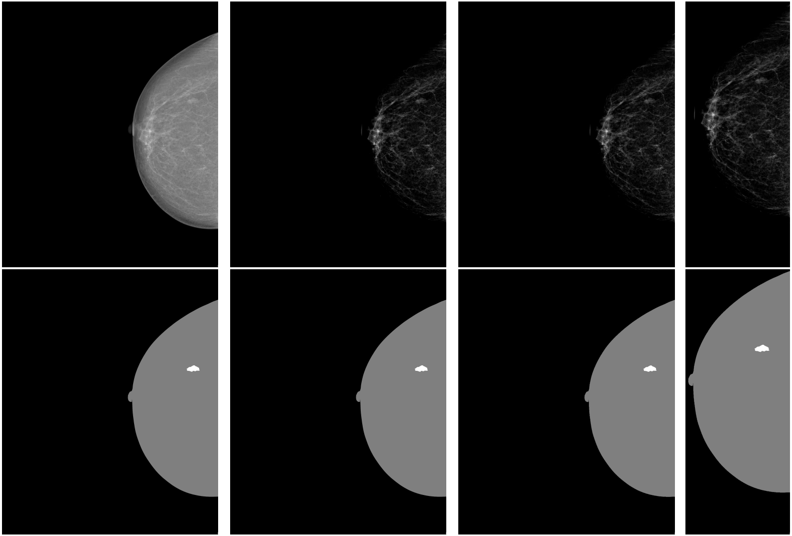
Database

Our mammographic database provided by the Breast Cancer Digital Repository Consortium consists of 256 digital mammograms from 63 Portuguese patients with at least one breast mass.
We augment the data set for training.

	Patients		**Mammograms**		**Masses**	
	Training	**Test**	**Training**	**Test**	**Training**	**Test**
Fold 1	50	13	189	67	106	33
Fold 2	50	13	209	47	112	27
Fold 3	50	13	204	52	110	29
Fold 4	51	12	209	47	110	29
Fold 5	51	12	213	43	118	21
Average	50.6	12.6	204.8	51.2	111.2	27.8

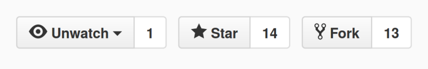
[Moura et al., 2013]

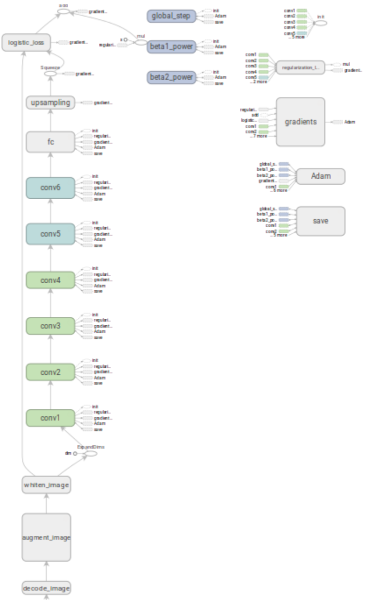
Preprocessing



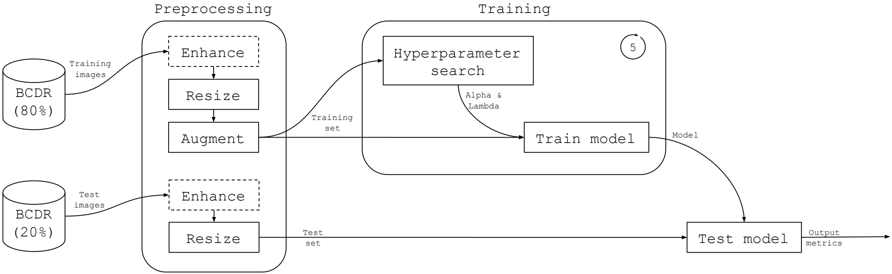
(a) Original image (b) Enhancement (c) Downsampling (d) Final

Software


Unwatch
1
Star
14
Fork
13


global_step
beta1_power
beta2_power
logistic_loss
upsampling
fc
conv6
conv5
conv4
conv3
conv2
conv1
whiten_image
augment_image
decode_image
gradients
Adam
save

Overview of the solution



Model 1

Simple architecture (206K parameters). Three configurations tested: no enhancement + no weighted loss, no enhancement + weighted loss and enhancement + weighted loss.

Layer	Filter	Stride	Pad	Dilation	Volume	Parameters
`INPUT`	-	-	-	-	$52 \times 52 \times 1$	-
`CONV -> RELU`	5×5	2	2	1	$26 \times 26 \times 32$	832
`CONV -> RELU`	3×3	1	1	1	$26 \times 26 \times 32$	9 248
`CONV -> RELU`	3×3	2	1	1	$13 \times 13 \times 64$	18 496
`CONV -> RELU`	3×3	1	1	1	$13 \times 13 \times 64$	36 928
`CONV -> RELU`	3×3	1	2	2	$13 \times 13 \times 96$	55 392
`CONV -> RELU`	3×3	1	2	2	$13 \times 13 \times 96$	83 040
`CONV`	5×5	1	6	3	$13 \times 13 \times 1$	2 401
`BILINEAR (x4)`	-	-		-	$52 \times 52 \times 1$	-

Model 2

Modelled on a VGG network, winner of the 2014 ImageNet. Does not use dilation. Tested with no enhancement + weighted loss.

Layer	Filter	Stride	Pad	Volume	Parameters
INPUT	-	-	-	$112 \times 112 \times 1$	-
CONV -> Leaky RELU	6×6	2	2	$56 \times 56 \times 56$	2 072
CONV -> Leaky RELU	3×3	1	1	$56 \times 56 \times 56$	28 280
MAXPOOL	2×2	2	0	$28 \times 28 \times 56$	-
CONV -> Leaky RELU	3×3	1	1	$28 \times 28 \times 84$	42 420
CONV -> Leaky RELU	3×3	1	1	$28 \times 28 \times 84$	63 588
MAXPOOL	2×2	2	0	$14 \times 14 \times 84$	-
CONV -> Leaky RELU	3×3	1	1	$14 \times 14 \times 112$	84 784
CONV -> Leaky RELU	3×3	1	1	$14 \times 14 \times 112$	113 008
CONV -> Leaky RELU	3×3	1	1	$14 \times 14 \times 112$	113 008
MAXPOOL	2×2	2	0	$7 \times 7 \times 112$	-
FC -> Leaky RELU	7×7	1	3	$7 \times 7 \times 448$	2 459 072
FC	1×1	1	0	$7 \times 7 \times 1$	449
BILINEAR (x16)	-	-	-	$112 \times 112 \times 1$	-

Model 3

Modelled on the Residual network, winner of the 2015 ImageNet. Deeper, fewer parameters and better resolution (4x). Tested with enhancement + weighted loss.

Layer	Filter	Stride	Pad	Dilation	Volume	Parameters
INPUT	-	-	-	-	$128 \times 128 \times 1$	-
CONV -> LRELU	6×6	2	2	1	$64 \times 64 \times 32$	1 184
CONV -> LRELU	3×3	1	1	1	$64 \times 64 \times 32$	9 248
CONV -> LRELU	3×3	2	1	1	$32 \times 32 \times 64$	18 496
CONV -> LRELU	3×3	1	1	1	$32 \times 32 \times 64$	36 928
CONV -> LRELU	3×3	1	2	2	$32 \times 32 \times 128$	73 856
CONV -> LRELU	3×3	1	2	2	$32 \times 32 \times 128$	147 584
CONV -> LRELU	3×3	1	2	2	$32 \times 32 \times 128$	147 584
CONV -> LRELU	3×3	1	2	2	$32 \times 32 \times 128$	147 584
CONV -> LRELU	3×3	1	4	4	$32 \times 32 \times 256$	295 168
CONV	8×8	1	14	4	$32 \times 32 \times 1$	16 385
BILINEAR (x4)	-	-		-	$128 \times 128 \times 1$	-

Training

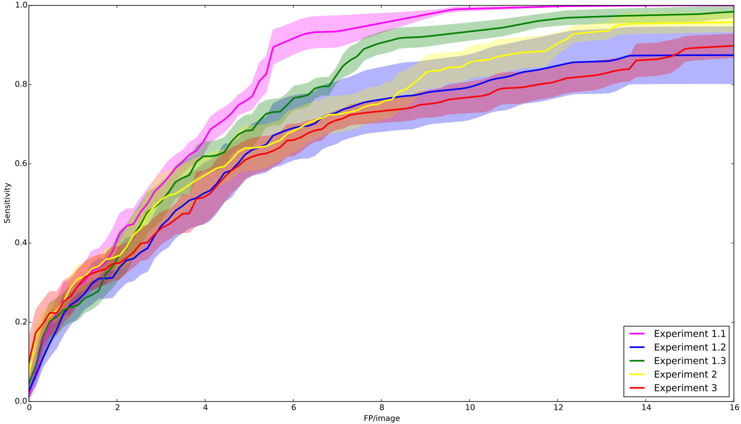
Hyperparameter search: For each fold, we train 20 networks with different learning rate and regularization parameters to select the best combination. We hold out 20% of the training set for evaluation.

Training: We trained the final models for 30 epochs using ADAM. Convergence was achieved.

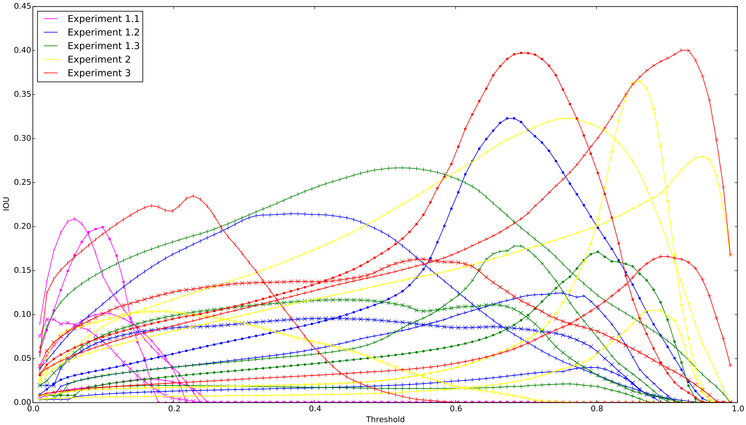
Evaluation: We use 5-fold cross-validation to generate our results and variability estimates.

FROC curves

Free-response ROC curves plot the sensitivity of the system against the number of false predictions per image. Shading shows one standard error of the mean.



Intersection-over-union as the threshold is varied. Each line is one fold.

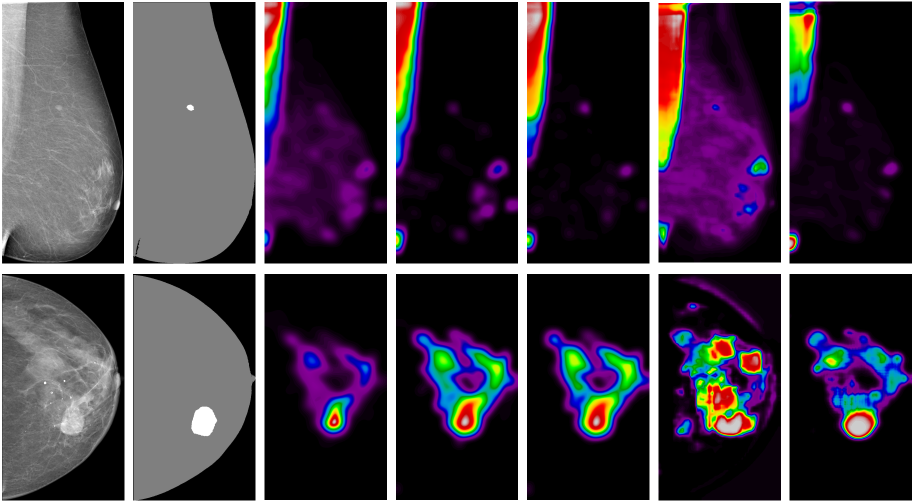


Peak IOU

IOU value for the best threshold per fold.

	Fold 1	**Fold 2**	**Fold 3**	**Fold 4**	**Fold 5**	**Average $\pm$ SEM**
Experiment 1.1	0.03	0.09	0.21	0.20	0.10	0.13 ± 0.03
Experiment 1.2	0.04	0.10	0.21	0.32	0.12	0.16 ± 0.04
Experiment 1.3	0.02	0.12	0.27	0.17	0.18	0.15 ± 0.04
Experiment 2	0.11	0.10	**0.32**	0.37	0.28	0.24 ± 0.05
Experiment 3	**0.17**	**0.17**	0.23	**0.40**	**0.40**	0.27 ± 0.05
Average	0.07	0.11	0.25	0.29	0.22	

Qualitative results



Original Label Exp. 1.1 Exp. 1.2 Exp. 1.3 Exp. 2 Exp. 3

Qualitative results

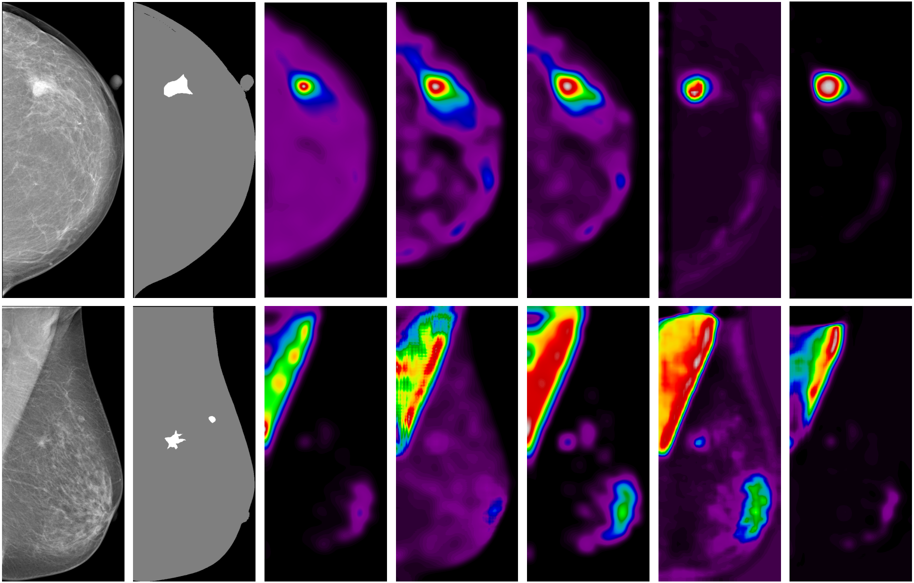


Table of Contents

Conclusion

- Convolutional networks can be used as an end-to-end trainable system to perform segmentation of mammograms.
- Small architectural variations are irrelevant as long as the architecture is flexible enough.
- Using a weighted loss function improves performance whereas enhancement helps only slightly.
- Hyperparameter fine-tuning is necessary to find a good minimum.
- The bottleneck for learning seems to be the small data sets.
- Training small-to-medium networks is feasible with the available resources.
- Convolutional networks remain a promising option for future medical imaging research.

Contributions

- First reported use of convolutional networks for breast cancer lesion segmentation.
- Small, modern architectures tailored to segmentation.
- Mammographic data set and preprocessing tools.
- Software for designing, training and evaluating convolutional networks.
- Support to other deep learning projects in the campus.

Future work

- Create a high-quality, large data set.
- Curate data by deleting obvious lesions, benign lesions and chest muscle, labelling other kind of lesions and breast tissue or joining images from different sources.
- Validate results using different architectures and bigger data sets.
- Use standard computer vision features as input to the network.
- Incorporate current trends to newer architectures.
- Develop standard evaluation metrics for these kind of predictions.

?