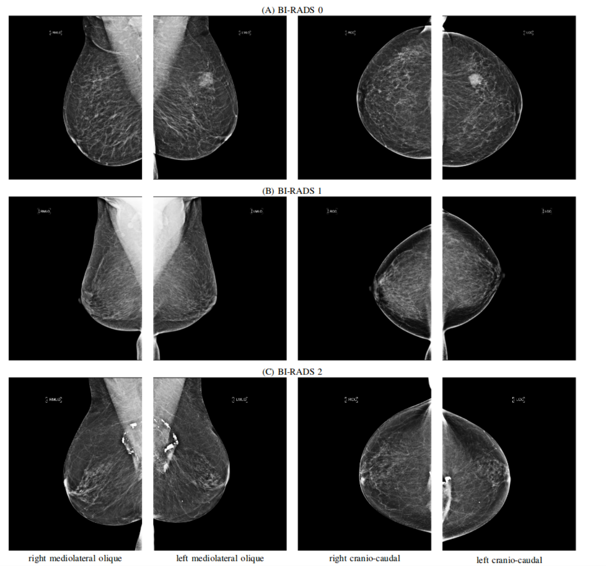
TFML 2019, Kraków

Michał Kuźba

Articles

- Krzysztof J. Geras, Stacey Wolfson, Yiqiu Shen, Nan Wu, S. Gene Kim, Eric Kim, Laura Heacock, Ujas Parikh, Linda Moy and Kyunghyun Cho
  High-Resolution Breast Cancer Screening with Multi-View Deep Convolutional Neural Networks (2017)
- Nan Wu, Krzysztof J. Geras, Yiqiu Shen, Jingyi Su, S. Gene Kim, Eric Kim, Stacey Wolfson, Linda Moy and Kyunghyun Cho
  Breast density classification with deep convolutional neural networks (2018)
- Konrad Żołna, Krzysztof J. Geras and Kyunghyun Cho
  Classifier-agnostic saliency map extraction (2018)

Breast cancer screening



Problem specification

- Multi view
- labelled by doctor decision not the actual cancer development
- large resolution (2600x2000), cannot be reduced
- category ”not clear”
- one channel

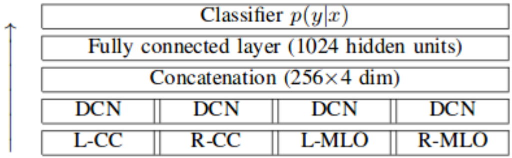
Dataset

- around 200,000 mammographic exams, almost 900,000 images
- orders of magnitude larger than previous datasets
- 4TB, dataset not available publicly, streaming from disk
- no Transfer Learning
- natural distribution (13%, 46%, 41%)

		breast density category				
		0	1	2	3	
BI-RADS	0	1702	9607	12656	1839	25804
	1	9803	40060	37167	5157	92187
	2	8434	35998	34029	4727	83188
		19939	85665	83852	11723	

Architecture

layer	kernel size	stride	#maps	repetition
global average pooling			256	
convolution	3×3	1×1	256	×3
max pooling	2×2	2×2	128	
convolution	3×3	1×1	128	× 3
max pooling	2×2	2×2	128	
convolution	3×3	1×1	128	× 3
max pooling	2×2	2×2	64	
convolution	3×3	1×1	64	× 2
convolution	3×3	2×2	64	
max pooling	3×3	3×3	32	
convolution	3×3	2×2	32	
input			1	



Results

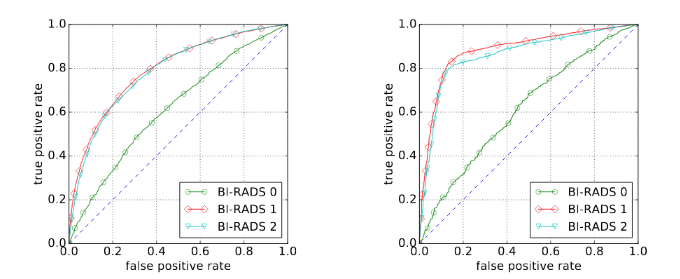


TABLE V

AVERAGE AUC (MACAUC) AS A FUNCTION OF THE CONFIDENCE THRESHOLD $T_{P\%}$. WHEN $P = 30\%$, WE REFER TO THE MACAUC AS A HIGH-CONFIDENCE MACAUC (HC-MACAUC).

$T_{P\%}$	$T_{10\%}$	$T_{20\%}$	$\boldsymbol{T_{30\%}}$	$T_{50\%}$	$T_{100\%}$
macAUC	0.865	0.827	0.811	0.781	0.732

Comparison with human performance

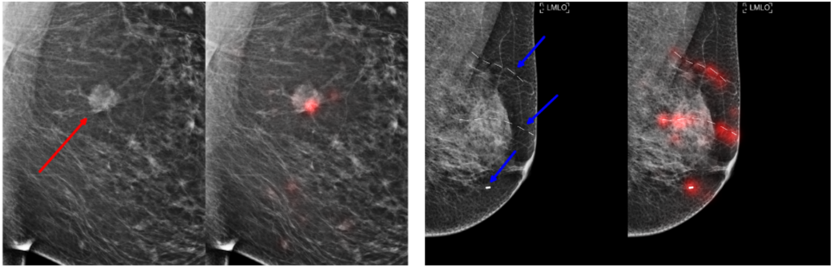
TABLE VI
RESULTS OF OUR READER STUDY COMPARING ACCURACIES OBTAINED BY THE COMMITTEE OF RADIOLOGISTS, OUR NEURAL NETWORK (MV-DCN) AND AN ENSEMBLE OF THE TWO.

	radiologists	MV-DCN	radiologists + MV-DCN
0 vs. others	0.650	0.547	0.653
1 vs. others	0.765	0.757	0.792
2 vs. others	0.699	0.759	0.759
macAUC	0.704	0.688	0.735

Next steps

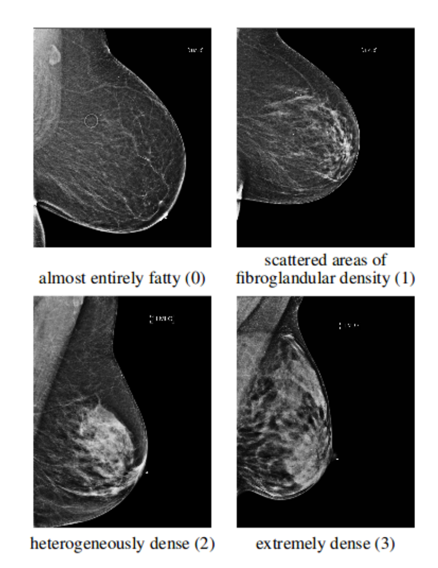
- actual cancer labels
- more complex pipeline, including other information about patients, history, other exams, breast density
- learning where to look

Visualization



(a) BI-RADS 0 (b) BI-RADS 2

Breast density classification



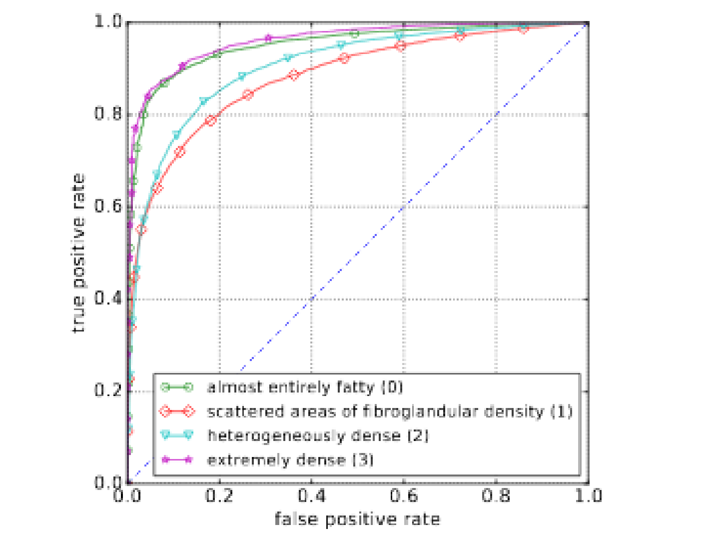
Breast density classification

- easier problem, less data required
- density has ”masking effect” and the risk increases

		breast density category				
		0	1	2	3	
BI-RADS	0	1702	9607	12656	1839	25804
	1	9803	40060	37167	5157	92187
	2	8434	35998	34029	4727	83188
		19939	85665	83852	11723	

Results

- same architecture, Transfer Learning (speeds up)



Classifier-agnostic saliency map extraction

Levels of agnosticism:

- works for specific architecture, e.g. ResNet50
- works for any given classifier
- **works without any classifier; explains the data**

Here agnosticism is not an objective, but a remedy. ImageNet dataset.

Map extraction

- $m : \mathbb{R}^{W \times H \times 3} \rightarrow [0, 1]^{W \times H}$
- $m = argmax_{m'} S(m', f)$
- $S(m, f) = \frac{1}{N} \sum_{n=1}^{N} [l(f((1 - m(x_n)) \cdot x_n), y_n) + R(m(x_n))]$
- $L(m, f) = \frac{1}{N} \sum_{n=1}^{N} [l(f((1 - m(x_n)) \cdot x_n), y_n)]$
- $m = argmax_{m'} \mathbb{E}_f [S(m', f)]$

Algorithm 1: Classifier-agnostic saliency map extraction

input : an initial classifier $f^{(0)}$,
an initial mapping $m^{(0)}$,
dataset D,
number of iterations K

output : the final mapping $m^{(K)}$

Initialize a sample set $F^{(0)} = \left\{ f^{(0)} \right\}$.

for $k \leftarrow 1$ **to** K **do**

$\theta_{f^{(k)}} \leftarrow \theta_{f^{(k-1)}} - \eta_f \nabla_{\theta_f} L(m^{(k-1)}, f^{(k-1)})$

$F^{(k)} \leftarrow F^{(k-1)} \cup \left\{ f^{(k)} \right\}$

$f' \leftarrow \text{Sample}(F^{(k)})$

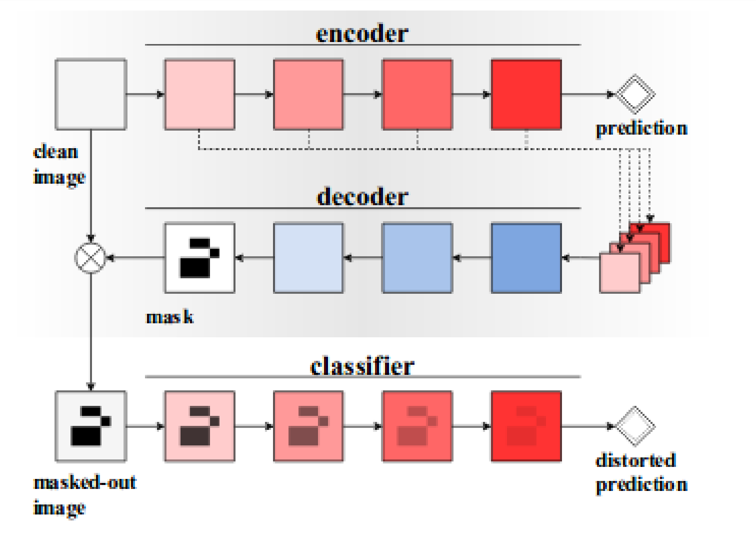
$\theta_{m^{(k)}} \leftarrow \theta_{m^{(k-1)}} + \eta_m \nabla_{\theta_m} S(m^{(k-1)}, f')$

$F^{(k)} \leftarrow \text{Thin}(F^{(k)})$

Modifications

- entropy instead of loss in the score (adversarial artifacts)
- thinning
- regularization - L1 norm of m

Architecture

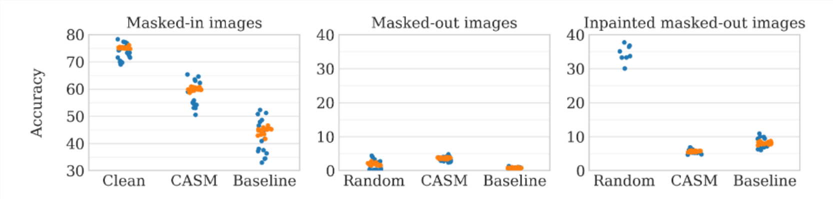


Details

- sharing weights
- discretize masks (>= average value)
- largest connected component

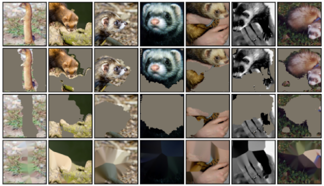
Evaluation

- visualization
- testing on various classifiers
- object localization (weakly supervised localization)



Masks

(a) CASM (**L100**)



(b) CASM (**L**)

(c) Baseline

Figure 2: The original images are in the first row. In the following rows masked-in images, masked-out images and inpainted masked-out images are shown, respectively. Note that the proposed approach (a-b) remove all relevant pixels and hence the inpainted images show the background only. Seven randomly selected consecutive images from validation set are presented here. Please look into the appendix for extra visualizations.

Unkown classes localization

	A	B	C	D	E	F	All
F	*46.5*	*46.4*	*48.1*	*45.0*	*45.7*	41.3	44.9
E, F	*39.5*	*41.2*	*43.1*	*40.3*	39.5	38.7	40.0
D, E, F	*37.9*	*39.3*	*40.0*	38.0	38.0	37.4	38.1
C, D, E, F	*38.2*	*38.5*	39.9	37.9	37.9	37.8	38.1
B, C, D, E, F	*36.7*	36.8	39.9	37.4	37.0	37.0	37.4
-	**35.6**	**36.1**	**39.0**	**37.0**	**36.6**	**36.7**	**36.9**