

GAN (Generative Adversarial Networks) for realistic data augmentation and lesion simulation in x-ray breast imaging

Basel Alyafi¹

Supervisors: Robert Marti, Oliver Diaz

¹Erasmus Mundus Master's in Medical Imaging and Applications (MAIA)
UB, UNICALM, UdG

August 8, 2019



Outline

- 1 Introduction
- 2 Materials
- 3 Methods
- 4 Results
 - Main Results
 - Further Inspection
- 5 Conclusions

Outline

- 1 Introduction
- 2 Materials
- 3 Methods
- 4 Results
 - Main Results
 - Further Inspection
- 5 Conclusions

Breast Cancer

EU Statistics

- WHO¹ Statistics in 2018 (both sexes, all ages)
 - Highest incidence rate.
 - Third deadliest cancer after colon and lung.
- In 2019 [13]
 - 92,000 breast-cancer deaths are expected in women.

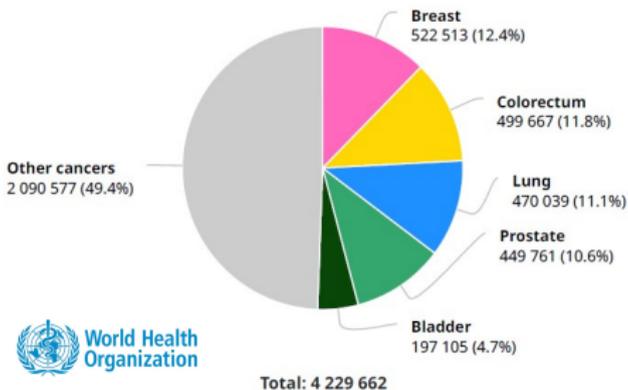


Figure: Statistics of most common cancers in the EU in 2018, both sexes and all ages.

¹World Health Organization

Problem Definition

Computer-Aided Detection (CADe) in Breast Cancer

Low Generalisability in CADe

- Labelled medical datasets are small due to:
 - Annotations are time consuming and costly.
 - Privacy.
- Imbalanced data:
 - Positive (unhealthy) to negative (normal) images. e.g., a few mammograms have lesions.
 - Negative pixels outnumber positive ones. i.e., lesions are relatively small.

Consequences

- Performance drops on clinical cases.
 - Higher False Positive Rate (FPR).

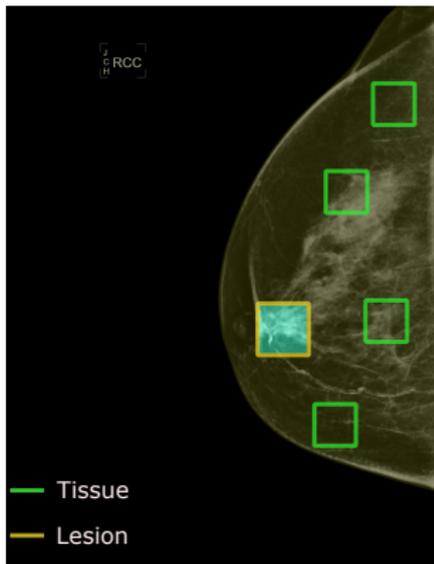


Figure: Tissue patches outnumber lesion ones in a mammogram.

Sampling

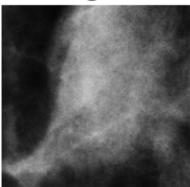
Oversampling

- Replication.
- Features interpolation (SMOTE [7] variations).
- Affine transformations.
e.g., flipping, rotation, scaling, translation.

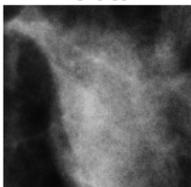
Undersampling

- Create smaller balanced subsets.

original



vertical



horizontal

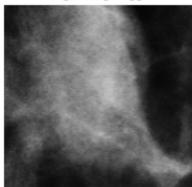


Figure: Horizontal and vertical flipping applied on a mammographic lesion.

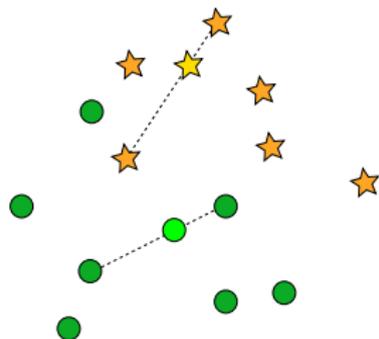


Figure: SMOTE algorithm, the bright sample is located on a line that connects neighbouring samples.

Generative Models

Pixel RNN [14]

- - Explicitly learn data distribution.
- - Pixel by pixel \rightarrow slow.
- RNN to represent long-term dependencies.

Pluses

Neutralise non-pertinent variance sources.

- Shape sources: affine transformations.
- Content sources, e.g., driver + car = car.

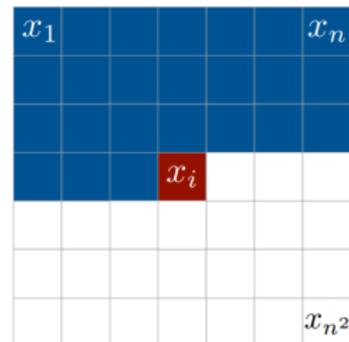


Figure: pixel RNN, conditional dependencies [5].



[2]

Generative Models (2)

Generative Adversarial Networks (GANs)

- Introduced by Goodfellow et al (2014) . [11]
- Minmax problem.
- The Generator (forger)
 - Takes random input.
 - Synthesises images to fool the detective.
- The Discriminator (detective)
 - Captures real samples.
 - Gets penalised when it fails.

GANs Advantages

- + Concurrent.
- + Learn data distribution implicitly.

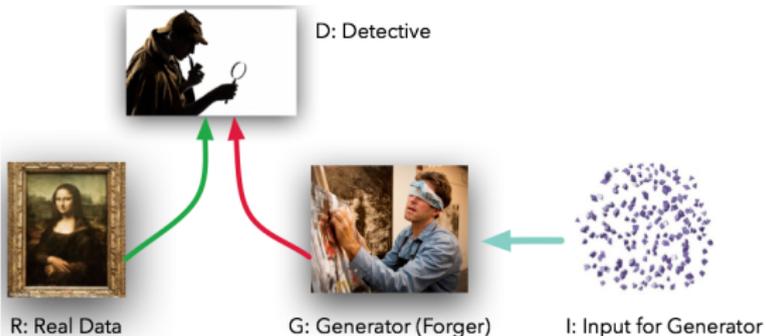


Figure: GANs as a real-life scenario [4].

Related Work

GANs in medical imaging 2018

- Frid-Adar et al, DCGAN² to generate 3-class 2D liver images for augmentation purposes as a function of the training size. [8]
- C. Bowles et al, DCGAN to generate segmented MR³ and CT⁴ brain images + conventional augmentation of synthetic and real images. [6]
- E.Wu et al, conditional infilling to add/remove lesion patches (256 × 256). [18]
- D. Korkinof et al, Progressive GAN to generate full mammograms. [12]
- H. Salehinejad et al, DCGAN to generate chest pathology patches (256 × 256). [16]

²Radford et al [15]

³Magnetic Resonance

⁴Computed Tomography

Objectives

- GANs to generate realistic mammographic patches.
- Synthetic images for imbalanced classification augmentation.

Outline

- 1 Introduction
- 2 Materials**
- 3 Methods
- 4 Results
 - Main Results
 - Further Inspection
- 5 Conclusions

Dataset

OPTIMAM Dataset [9]

- Over 79,000 images.
- 4821 patients.
- Heterogeneous.
- Lesions included
 - Mass (M).
 - Calcification (Cal).
 - Architectural distortion.
 - Focal asymmetry.

Properties

- Detailed explanatory Excel files.
- Processed and unprocessed images.
- Left and right breasts.
- Craniocaudal (CC) and mediolateral (MLO) views.
- **Manufacturers:** e.g., Hologic, Philips, GE.
- **Modalities:** e.g., BioVision, L30, Selenia.

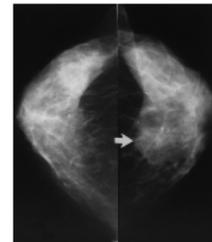
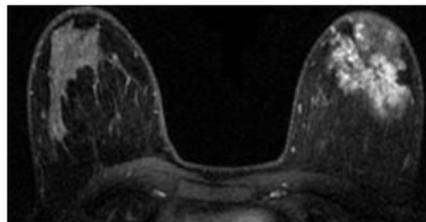
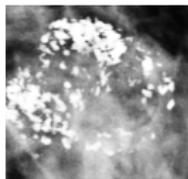
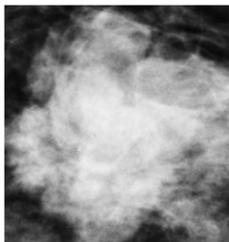


Figure: Left to right, a mass patch (496×519), a calcification patch (409×384 p), architectural distortion [1], focal asymmetry [3].

Data Preparation

Filtering, GT, Mask, Patch extraction

Preprocessing

- Image inclusion criteria
 - Hologic, Lorad Selenia.
 - Processed images only.
- Histogram normalisation.
- Mask generation
 - non-zero thresholding.
- Groundtruth generation
 - Read lesion coordinates from Excel.
 - White rectangles over lesions.
- Patch extraction (128×128)
 - lesion and normal tissue patches.

Outcome

- 4536 lesion (positive) patches (2215 M, 2321 Cal).
- 147,000 normal tissues.

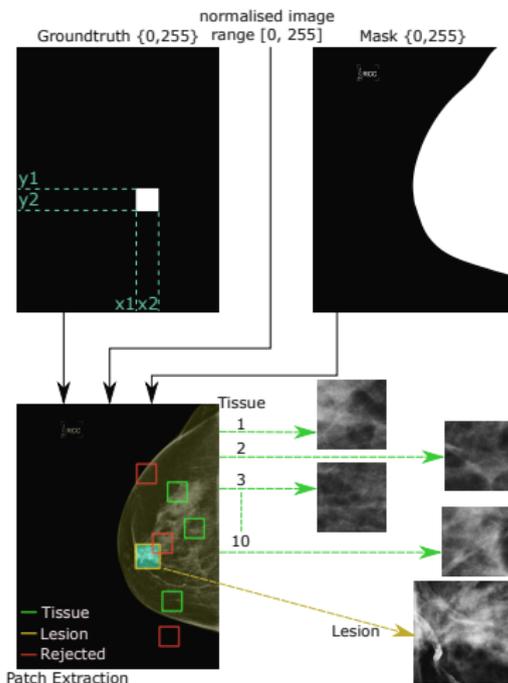


Figure: Data preparation diagram.

Outline

1 Introduction

2 Materials

3 Methods

4 Results

- Main Results

- Further Inspection

5 Conclusions

Deep Convolutional GAN (DCGAN)

Architecture

- Proposed by Radford et al 2016. [15]
- Stable in training.
- Modified to 128×128 output.

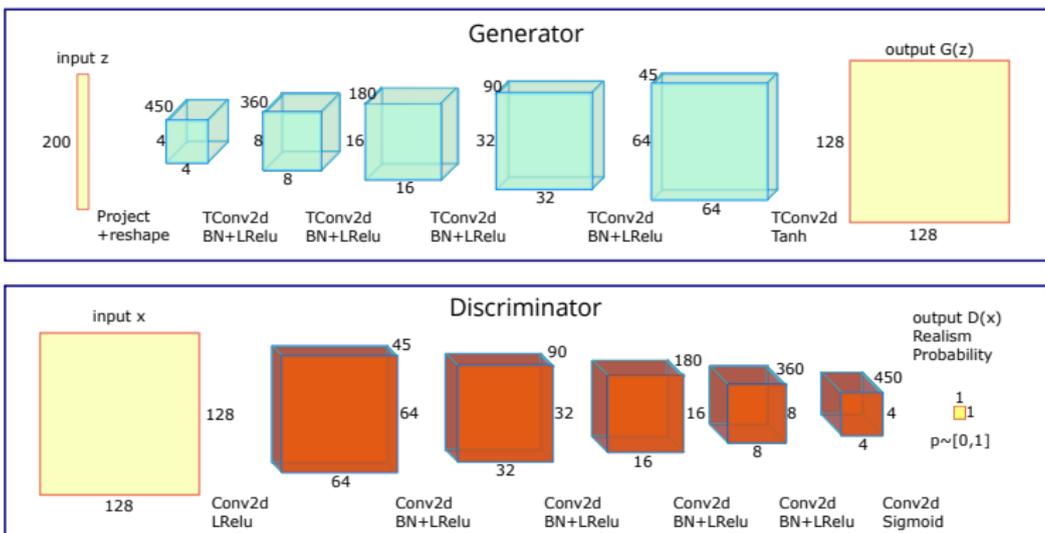


Figure: Generator (top) and Discriminator (bottom) architectures. TConv2d: transpose convolution 2D, BN: 2D batch normalization, LRelu: Leaky Rectified Linear Unit.

Deep Convolutional GAN (DCGAN)

Training

Settings

- Training set: 4536 M + Cal.
- $L_D = -E_{x \in P_x, z \in P_z} [\log(x) + \log(1 - D(G(z)))]$
- $L_G = -E_{z \in P_z} [\log(D(G(z)))]$

Deep Convolutional GAN (DCGAN)

Training

Settings

- Training set: 4536 M + Cal.
- $L_D = -E_{x \in P_x, z \in P_z} [\log(x) + \log(1 - D(G(z)))]$
- $L_G = -E_{z \in P_z} [\log(D(G(z)))]$

Training Steps

- 1 Generate a noise batch $z \sim \mathcal{N}(0, 1)$.
- 2 Forward z through $G \rightarrow G(z)$.
- 3 Forward $G(z)$ and real x through D .
- 4 Calculate L_D .
- 5 Update D .
- 6 Calculate L_G .
- 7 Update G .

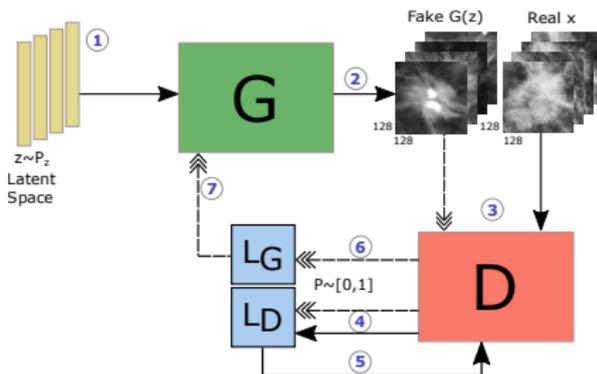


Figure: DCGAN training process.

Deep Convolutional GAN (DCGAN)

Training Techniques and Evaluation

Training Techniques

- One sided label smoothing.
- Conventional augmentation (flipping) on the input.
- D kernels are **larger** than G's.
- **Long** training (1000 epochs).

Deep Convolutional GAN (DCGAN)

Training Techniques and Evaluation

Training Techniques

- One sided label smoothing.
- Conventional augmentation (flipping) on the input.
- D kernels are **larger** than G's.
- **Long** training (1000 epochs).

Frechet Inception Distance (FID)

- FID was used to guide the training (model saving). [10]
- Evaluate images in Inception-v3 feature space.
- The lower the distance the more similar the fake and real images.

Imbalanced Classification

Diagram

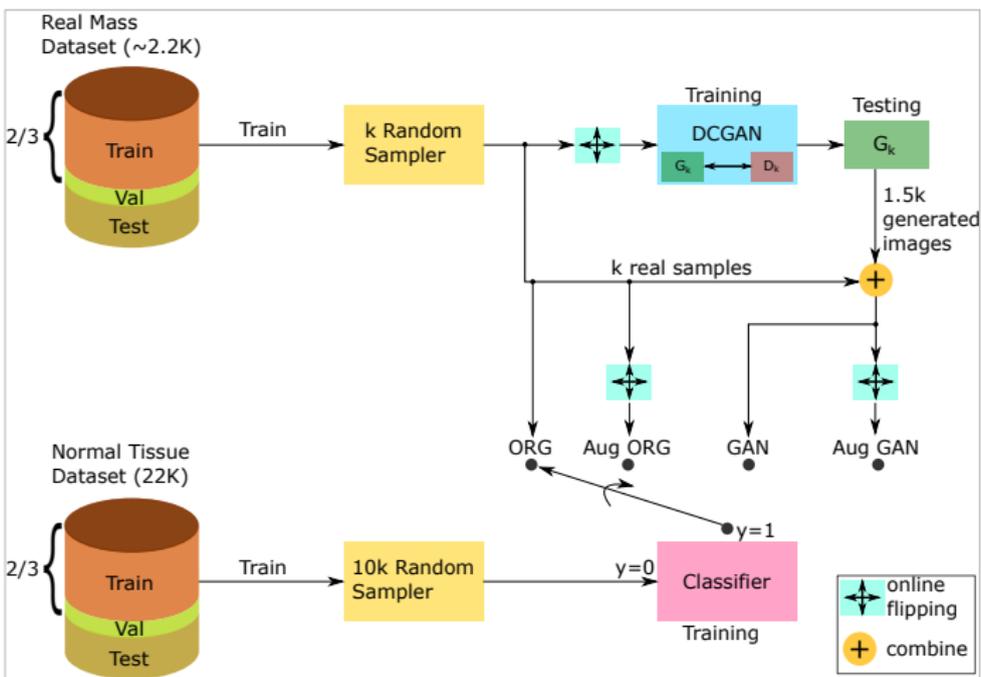


Figure: Imbalanced classification diagram. $k \in \{100, 250, 500, 750, 1000, 1300\}$.

Imbalanced Classification

Settings

Training Dataset

- Mass (Positive):
 $\{ P_k, k \in \{100, 250, 500, 750, 1000, 1300\} \}$
- Normal Tissue (Negative):
 $\{ N_k, k \in \{1000, 2500, 5000, 7500, 10e3, 13e3\} \}$
- Imbalance Ratio (IR) = 10.
- Augmentation Ratio (AR) = 1.5.
- 3 fold CV (60% train, 6.6% validation, 33% test).

Training Modes

- **ORG**: real unbalanced (1:10).
- **Aug ORG**: horizontal then vertical flipping **ORG**.
- **GAN**: **ORG** + fake masses.
- **Aug GAN**: horizontal then vertical flipping **GAN**.

Imbalanced Classification

Settings

Training Dataset

- Mass (Positive):
 $\{P_k, k \in \{100, 250, 500, 750, 1000, 1300\}\}$
- Normal Tissue (Negative):
 $\{N_k, k \in \{1000, 2500, 5000, 7500, 10e3, 13e3\}\}$
- Imbalance Ratio (IR) = 10.
- Augmentation Ratio (AR) = 1.5.
- 3 fold CV (60% train, 6.6% validation, 33% test).

Training Modes

- **ORG**: real unbalanced (1:10).
- **Aug ORG**: horizontal then vertical flipping **ORG**.
- **GAN**: **ORG** + fake masses.
- **Aug GAN**: horizontal then vertical flipping **GAN**.

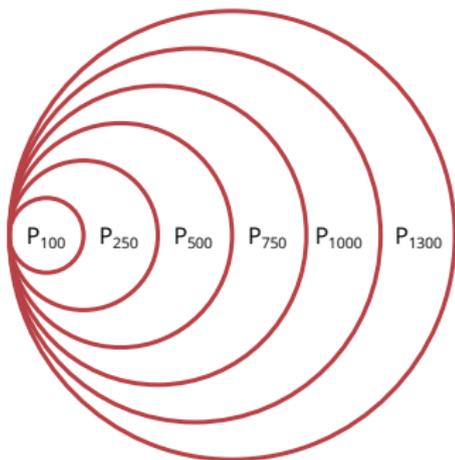


Figure: Subsets overlapping.

Evaluation

- $F1 = 2 \frac{Precision \times Recall}{Precision + Recall}$
- AUC.

Outline

1 Introduction

2 Materials

3 Methods

4 Results
■ Main Results
■ Further Inspection

5 Conclusions

Outline

1 Introduction

2 Materials

3 Methods

4 Results

■ Main Results

■ Further Inspection

5 Conclusions

Lesion Simulation

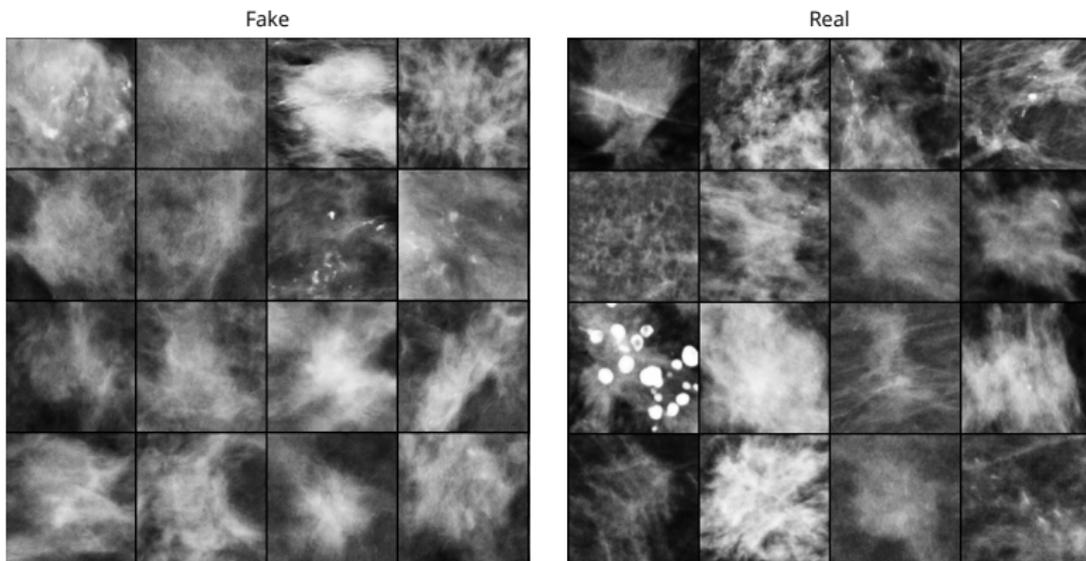


Figure: Fake and real images with mass and calcification. DCGAN trained on 4.5K patches.

FID Plot

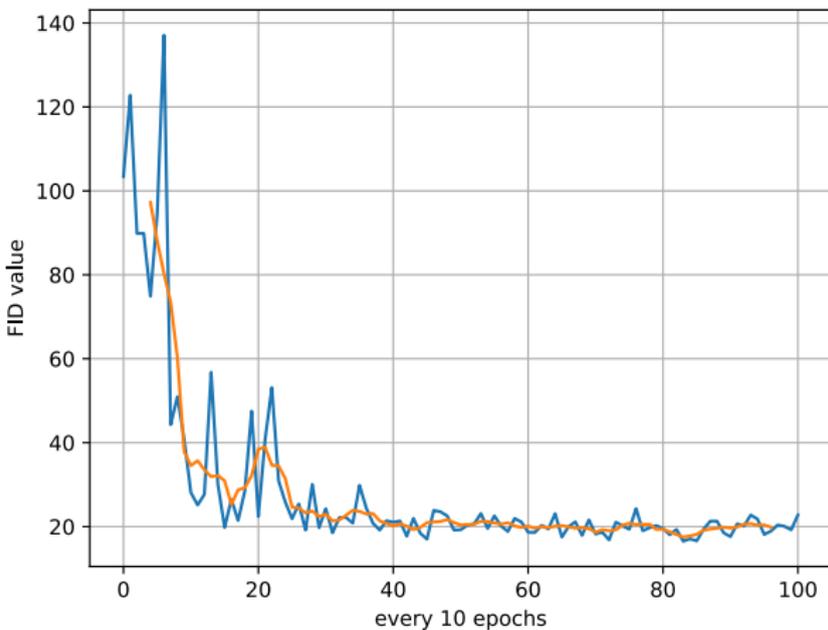


Figure: FID plot for DCGAN trained on 4.5K patches. The orange line is the moving average of the blue one.

Classification Results

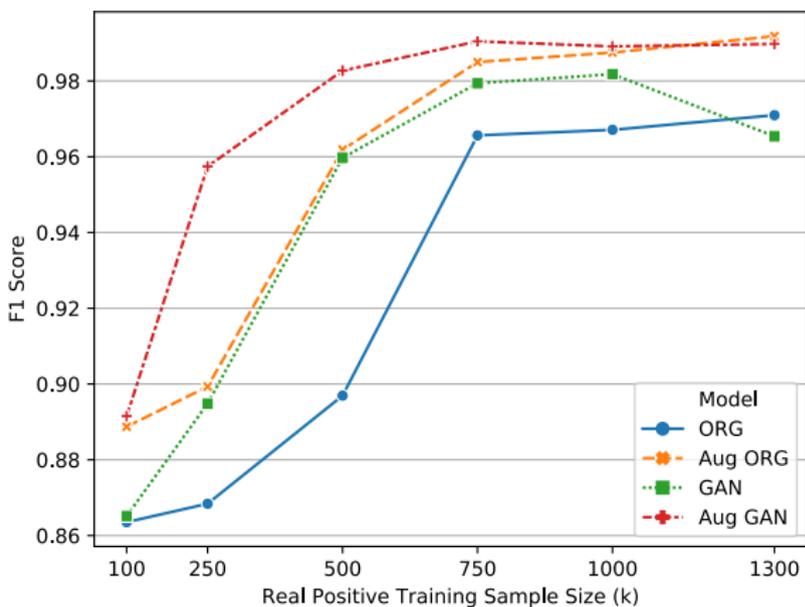


Figure: F1 score for all training modes.

AUC Results

Table: Area Under the ROC Curve (AUC) for different modes and training sizes (k).

Mode	Training Size					
	100	250	500	750	1000	1300
ORG	0.9836	0.9848	0.9896	0.999	0.9989	0.9989
GAN	0.9843	0.9902	0.9984	0.9997	0.9993	0.9987
Aug ORG	0.9877	0.9896	0.9982	0.9998	0.9997	0.9999
Aug GAN	0.9902	0.9984	0.9996	0.9990	0.9998	0.9999

t-SNE Analysis

t-SNE: t-distribution Stochastic Neighbor Embedding. [17]

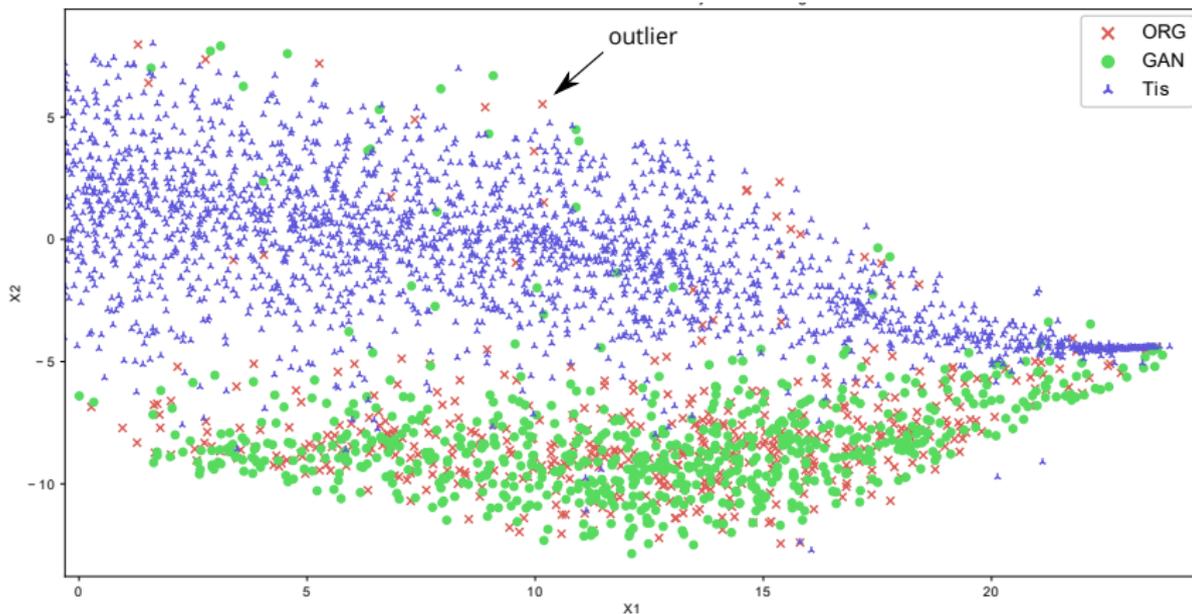


Figure: t-SNE analysis for real and fake masses and real normal tissue patches.

Outline

1 Introduction

2 Materials

3 Methods

4 Results

■ Main Results

■ Further Inspection

5 Conclusions

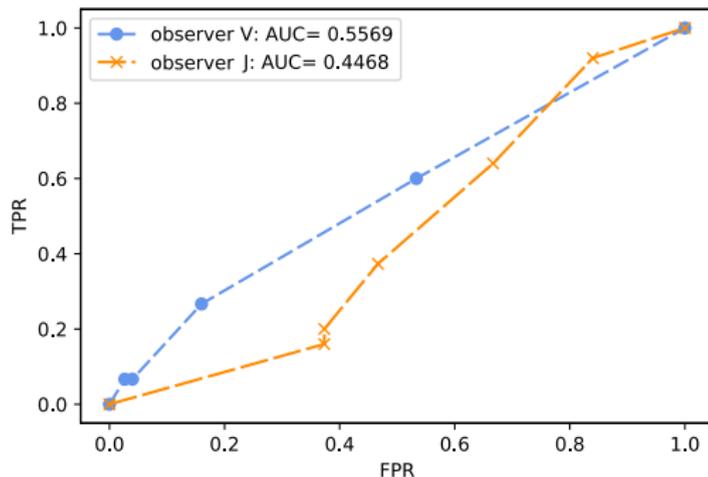
Experts Assessment

Settings

- Sample size
75 real and 75 synthetic
(128 × 128) patches.
- Two doctors were included.

Observers Accuracies

- 48% for observer V.
- 60.7% for observer J.



Clínica Girona



Parc Taulí
Hospital Universitari



MAIA

Outline

1 Introduction

2 Materials

3 Methods

4 Results

- Main Results
- Further Inspection

5 Conclusions

Conclusions

- GANs support **inliers**.
- GANs are sensitive to hyperparameters but powerful.
- Fake synthesis and traditional augmentation are **independent**.
- GANs fill in (**interpolate**) gaps.
- GAN + flipping outperforms GAN alone.
- GAN images are **similar** to real ones.

Limitations and Future Work

- 128×128 patches → larger patches or full mammograms.
- 150 samples assessed by 2 doctors → larger sample and more specialists.

Thanks

Thank you

References I



[Architectural distortion in mammography.](#)

"<https://www.ajronline.org/doi/full/10.2214/AJR.12.10153>".

Accessed on 23 Jun 2019.



[Cats Community.](#)

"<http://www.safecat.org.au/>".

Accessed on 23 Jun 2019.



[Focal asymmetry in mamography.](#)

"<https://pubs.rsna.org/doi/full/10.1148/radiographics.22.1.g02ja2219>".

Accessed on 23 Jun 2019.



[GANs example.](#)

https://cdn-images-1.medium.com/max/1200/1*-gFsbymY9oJUQJ-A3GTfeg.png.

Accessed on 23 Jun 2019.



[Pixel RNN.](#)

"<http://sergeiturukin.com/2017/02/22/pixelcnn.html>".

Accessed on 23 Jun 2019.



[C. Bowles, L. Chen, R. Guerrero, P. Bentley, R. Gunn, A. Hammers, D. A. Dickie, M. Valdés Hernández, J. Wardlaw, and D. Rueckert.](#)

GAN Augmentation: Augmenting Training Data using Generative Adversarial Networks.

Oct 2018.



[N. Chawla, K. Bowyer, L. Hall, and W. Kegelmeyer.](#)

SMOTE: Synthetic minority over-sampling technique.

Journal of Artificial Intelligence Research, 16:321–357, 2002.

cited By 4652.

References II



M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan.

GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321:321 – 331, 2018.



M. D. Halling-Brown, P. T. L. M. N. Patel, L. M. Warren, A. Mackenzie, and K. C. Young.

The oncology medical image database (omi-db).

In *Proc. SPIE 9039 Medical Imaging 2014: PACS and Imaging Informatics: Next Generation and Innovations*, volume 9039, March 2014.



M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter.

GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium.

In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 6626–6637. Curran Associates, Inc., 2017.
<http://papers.nips.cc/paper/7240-gans-trained-by-a-two-time-scale-update-rule>.



I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio.

Generative Adversarial Nets.

In Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 27*, pages 2672–2680. Curran Associates, Inc., 2014.



D. Korkinof, T. Rijken, M. O'Neill, J. Yearsley, H. Harvey, and B. Glocker.

High-Resolution Mammogram Synthesis using Progressive Generative Adversarial Networks.
July 2018.



M. Malvezzi, G. Carioli, P. Bertuccio, P. Boffetta, F. Levi, C. La Vecchia, and E. Negri.

European cancer mortality predictions for the year 2019 with focus on breast cancer. *Annals of Oncology*, 0:1–7, 2019.

References III



A. V. Oord, N. Kalchbrenner, and K. Kavukcuoglu.

Pixel recurrent neural networks.

In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1747–1756, New York, New York, USA, 20–22 Jun 2016. PMLR.



A. Radford, L. Metz, and S. Chintala.

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks.

In *2016 International Conference on Learning Representations (ICLR)*, 2016.



H. Salehinejad, S. Valaee, T. Dowdell, E. Colak, and J. Barlett.

Generalization of Deep Neural Networks for Chest Pathology Classification in X-Rays Using Generative Adversarial Networks.

In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 990–994, April 2018.



L. van der Maaten and G. Hinton.

Visualizing Data using t-SNE.

Journal of Machine Learning Research, 9:2579–2605, Nov 2008.



E. Wu, K. Wu, D. Cox, and W. Lotter.

Conditional Infilling GANs for Data Augmentation in Mammogram Classification.

In D. Stoyanov, Z. Taylor, B. Kainz, G. Maicas, R. R. Beichel, A. Martel, L. Maier-Hein, K. Bhatia, T. Vercauteren, O. Oktay, G. Carneiro, A. P. Bradley, J. Nascimento, H. Min, M. S. Brown, C. Jacobs, B. Lassen-Schmidt, K. Mori, J. Petersen, R. San José Estépar, A. Schmidt-Richberg, and C. Veiga, editors, *Image Analysis for Moving Organ, Breast, and Thoracic Images*, pages 98–106, Cham, 2018. Springer International Publishing.