Context Encoding Chest X-rays

Davide Belli, S. Hu,
E. Sogancioglu, and B. van Ginneken
davide.belli@student.uva.nl

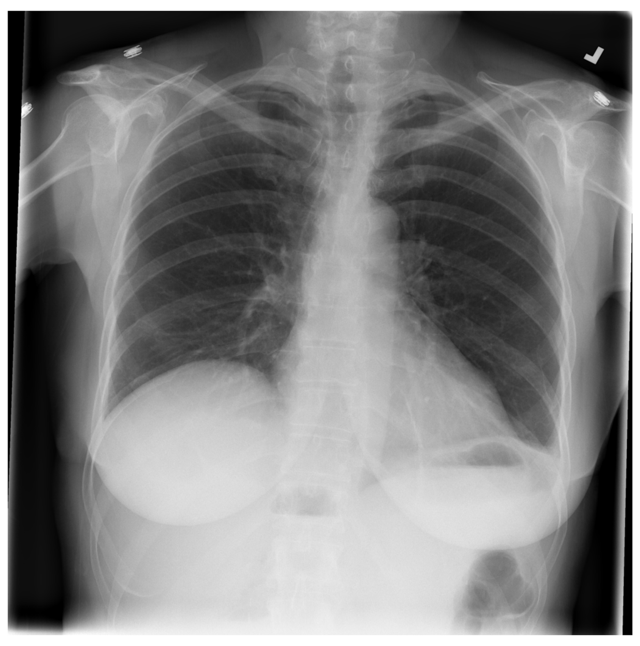
Universiteit van Amsterdam

Amsterdam Machine Learning Lab

Abstract

Chest X-rays are one of the most commonly used technologies for medical diagnosis. We explore how to automate the **abnormality detection task by image inpainting under adversarial training**. We modify a context encoder model for this task and train it over 1.1M 128x128 images from healthy X-rays. By computing and visualizing the pixel-wise difference between the source and the reconstructed images, **we can highlight abnormalities to simplify further detection and classification tasks**.

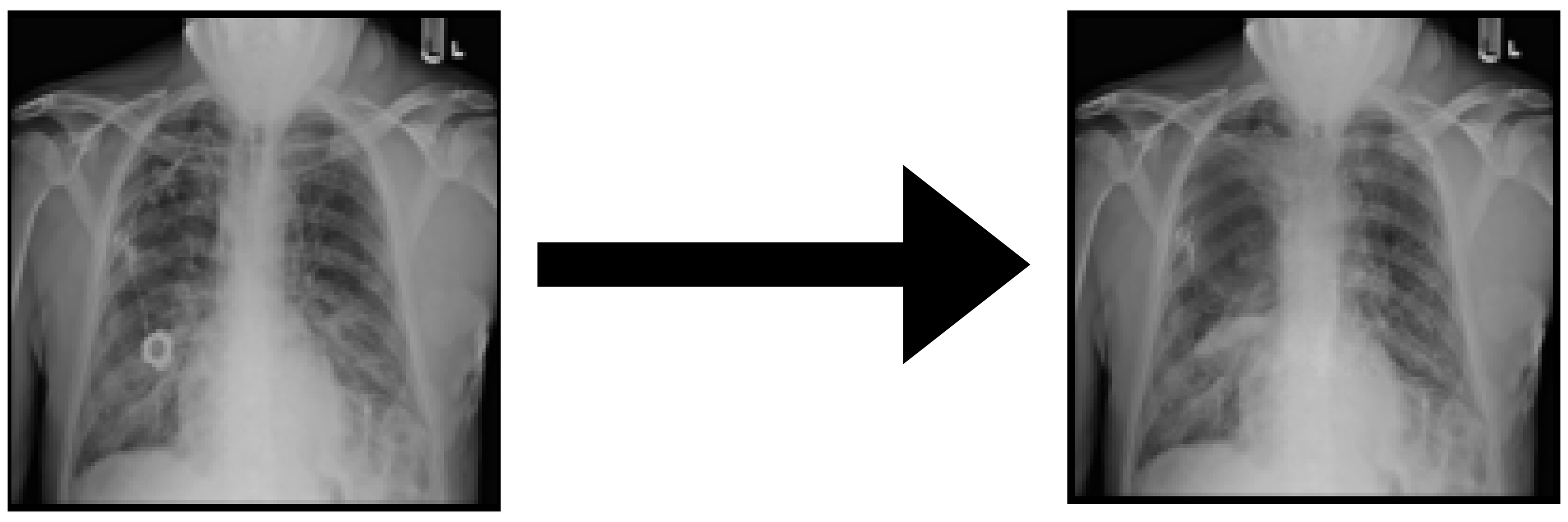
Dataset



- *ChestX-ray14* dataset [1]
- 112K images (1024x1024)
- 14 diseases and 1 healthy classes
- we extract 1M patches from lungs (128x128)

Figure 1: Sample from *ChestX-ray14*.

Task



- the inpainted patch should **only remove abnormalities**
- the inpainted patch should have a **clear content**
- the inpainted patch should be **coherent with the surrounding regions** (e.g. continuity in ribs shapes)

Approach

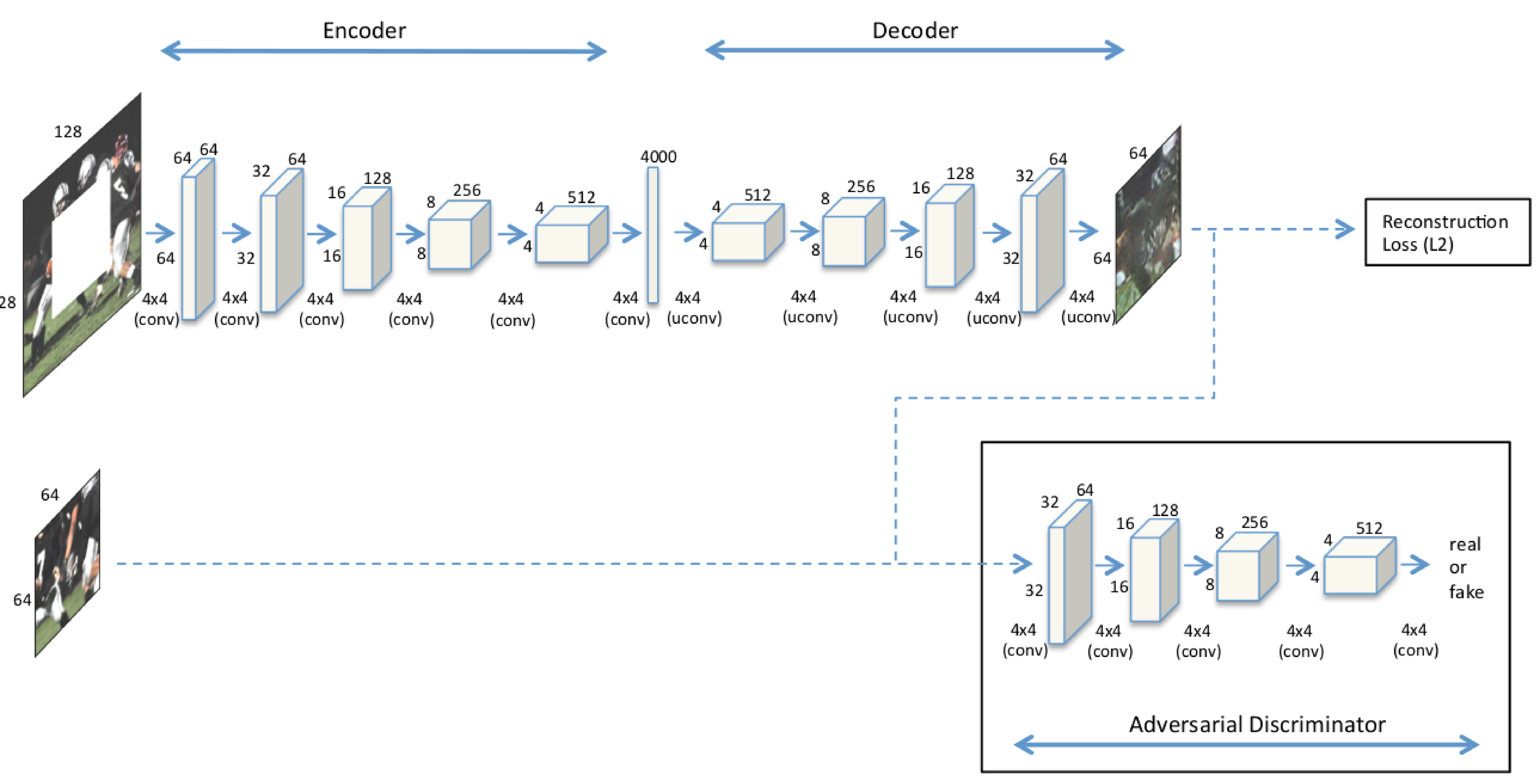


Figure 2: Original context encoder architecture from Pathak et al. [2].

Methods

- Self-supervised **training on healthy images** (while test on both healthy and unhealthy)
- Only train on the **lungs regions**
- Enforce realistic content with **adversarial loss**
- Enforce continuity in borders with a **pixel-wise weighted loss**
- $\mathcal{L} = 0.998 * \mathcal{L}_{L2} + 0.002 * \mathcal{L}_{adv}$

Results - Low Resolution, Full Images

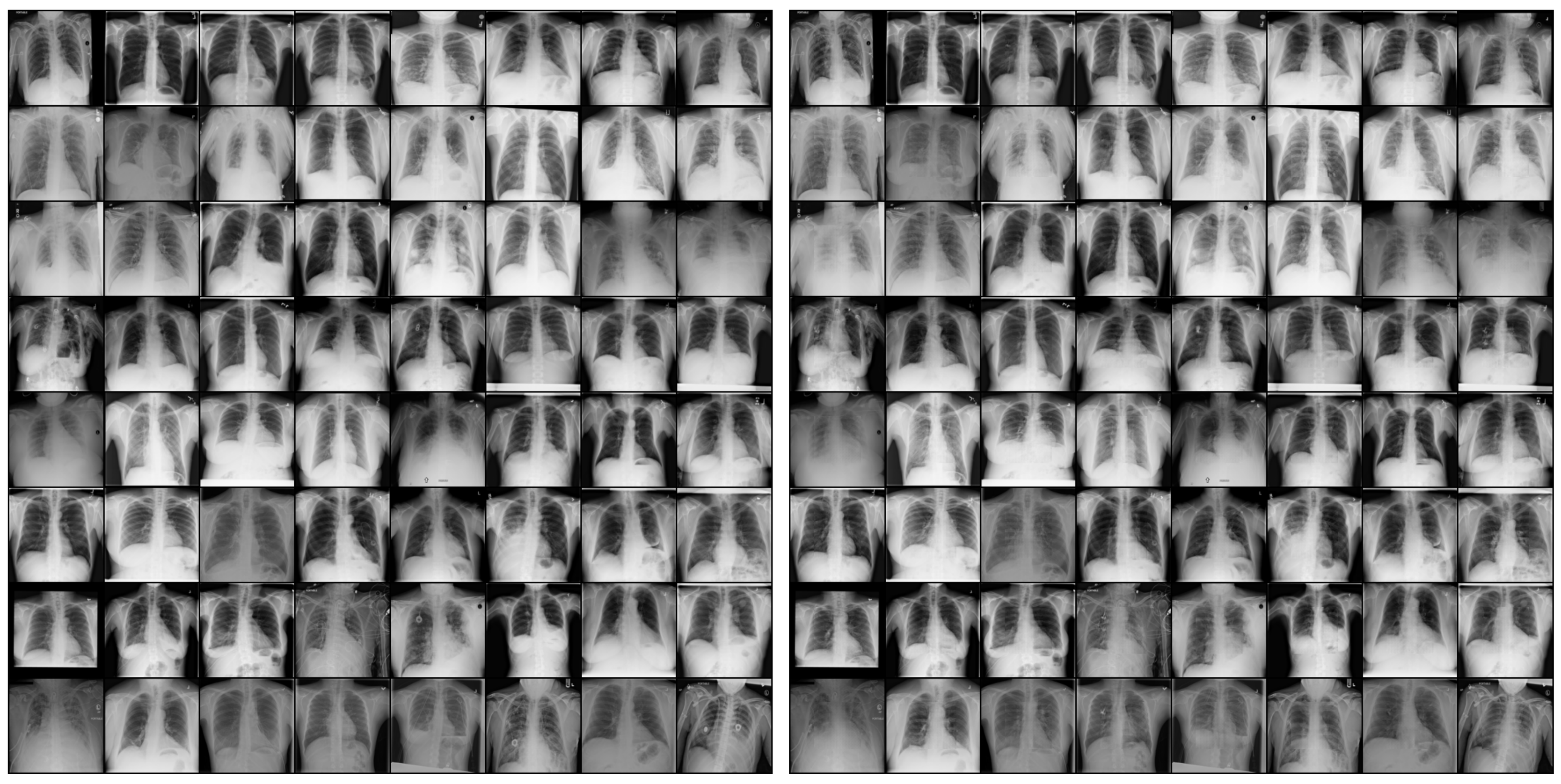


Figure 3: Real images (left) and images inpainted with our technique (right).

Results - High Resolution, Patches

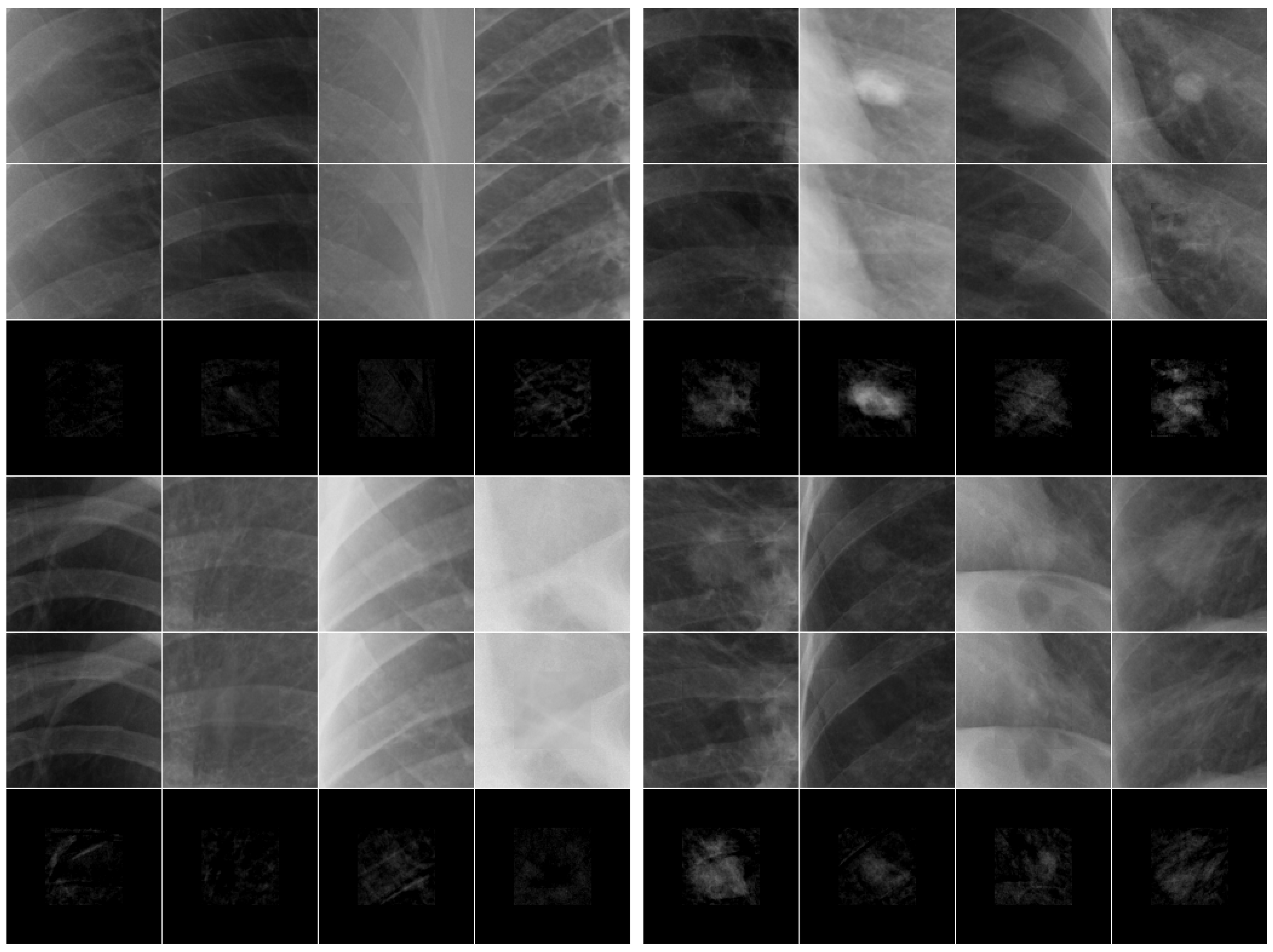


Figure 4: Healthy (left) and unhealthy (right) images from the high-resolution test set. Top rows: original images. Middle rows: healthy inpaintings reconstructed using our technique. Bottom rows: pixel-wise difference between real and inpainted images.

Evaluation

Qualitative evaluation:

- 2AFC (2 Alternatives Forced Choice) observer study with a 20-years experienced professional.
- 59.45 % correct choice

Quantitative evaluation:

Measure	Healthy test set	Unhealthy test set
PSNR	30.88 ± 3.61	27.06 ± 3.25
MSE	79.20 ± 129.16	164.82 ± 122.95
SSIM	0.81 ± 0.09	0.76 ± 0.08

Table 1: Distances between real and inpainted images averaged over healthy and unhealthy test sets.

References

[1] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” 2017.

[2] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, “Context encoders: Feature learning by inpainting,” *CoRR*, vol. abs/1604.07379, 2016.