

A scenic landscape featuring a calm lake in the foreground, a dense forest of evergreen trees on the right, and a range of rugged mountains in the background under a blue sky with scattered clouds. The text is overlaid on the center of the image.

# **SinGAN:** **Learning a Generative Model** **from a Single Natural Image**



# Reference paper

---

## Goal

- Generate new realistic natural images.
- High variability.
- Keep global structure and fine textures of the training image.

## Setup

Pyramid of fully convolutional GANs, each responsible for learning the patch distribution at a different scale of the image.

## Applications

- Random Sample
- Super-resolution
- Paint-to-image
- Harmonization
- Editing
- Single image animation

# Related works

## Single image deep models

- Overfit a deep model on a single training sample → Task-specific
- GAN-based model for a single natural image by Shocher et al. → The generation is conditioned on the input image and cannot be used to draw random samples.
- Unconditional single image GANs for texture generation → Cannot generate realistic samples if trained on non-texture images.

## Generative models for image manipulation

- GAN-based methods for many image manipulation tasks → Trained on class specific datasets



**Figure:** Single image texture generation

# Novelty

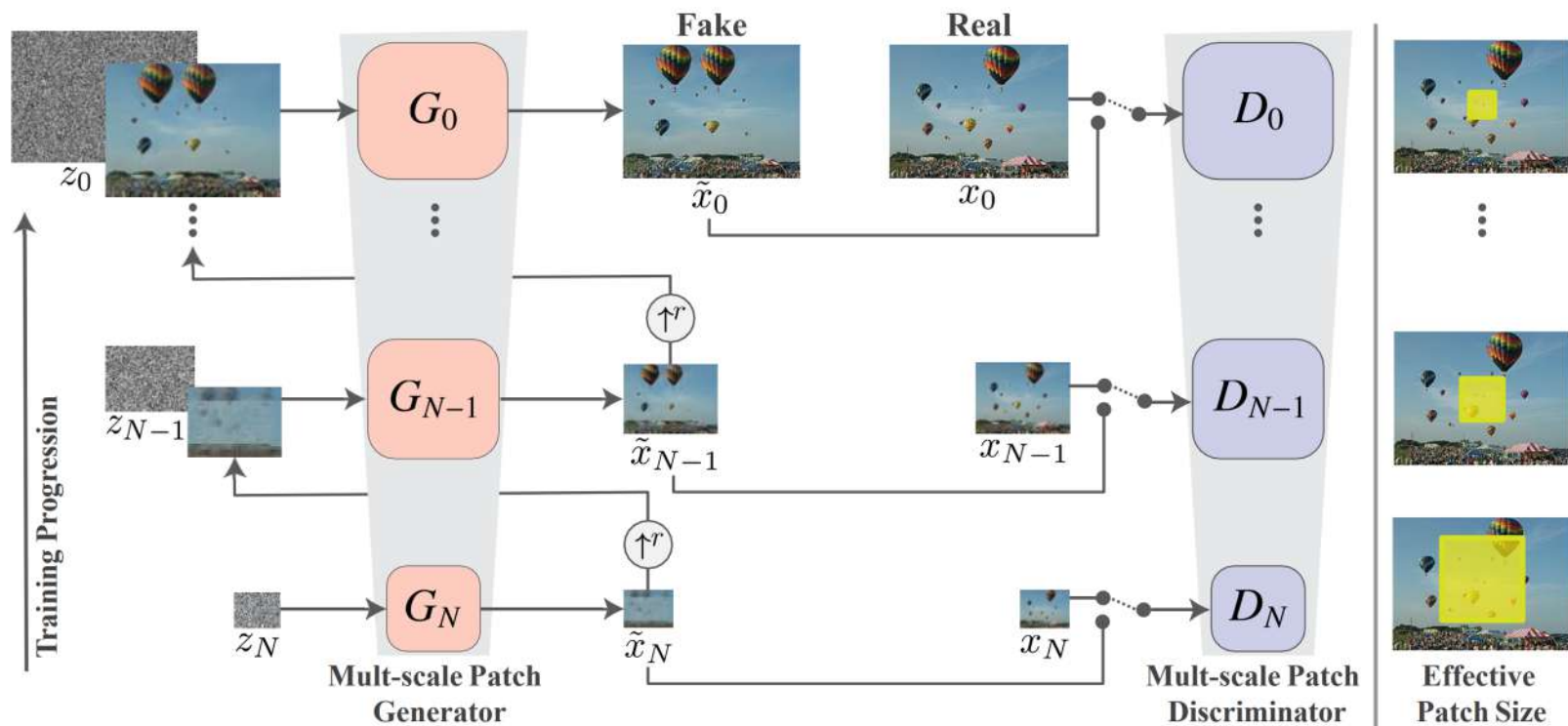
Machine learning has shown a great potential in the medical field, since it allowed to achieve fast and very accurate diagnoses of severe disorders.

There are not large datasets available, so by training a deep network from scratch we would need to face the problem of overfitting.

Use SinGAN to increase the size of the available datasets, by generating new medical images to be used in several machine learning tasks.

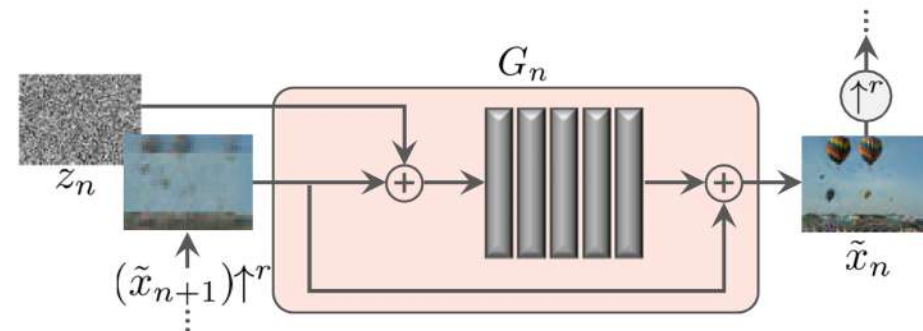


# SinGAN's architecture



Multi-scale  
pipeline

Single scale generation



# Losses in SINGAN

## ➤ Generator loss

$$L_G = -\mathbb{E}[D(\tilde{x})]$$

## ➤ Discriminator loss

$$L_D = D(\tilde{x}) - D(x) + \lambda(\|\nabla_{\hat{x}} D(\hat{x})\|_2 - 1)^2$$

## ➤ Reconstruction loss

$$L_R = (\tilde{x} - x)^2$$

$$\begin{aligned} x &= \text{real image} \\ \tilde{x} &= G(z, x) = \text{fake image} \\ \hat{x} &= \varepsilon x + (1 - \varepsilon)\tilde{x} \end{aligned}$$



# Replicating paper's results:

## Random Sample

---

Baloons in a  
landscape

Training image



Scale 0



Scale 1



Scale 2



Scale 3



Scale 4



Scale 5



Scale 6



Scale 7



# Further example on natural images: Random Sample

Flying birds in  
the sky

Training image



Scale 0



Scale 1



Scale 2



Scale 3



Scale 4



Scale 5



Scale 6



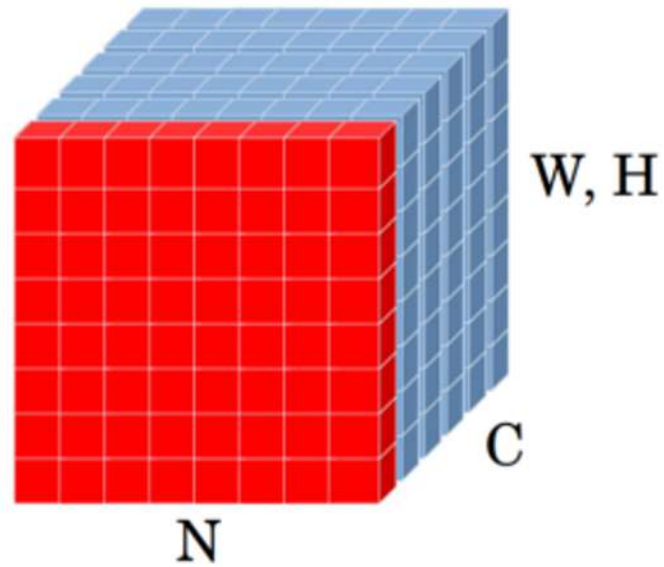
Scale 7





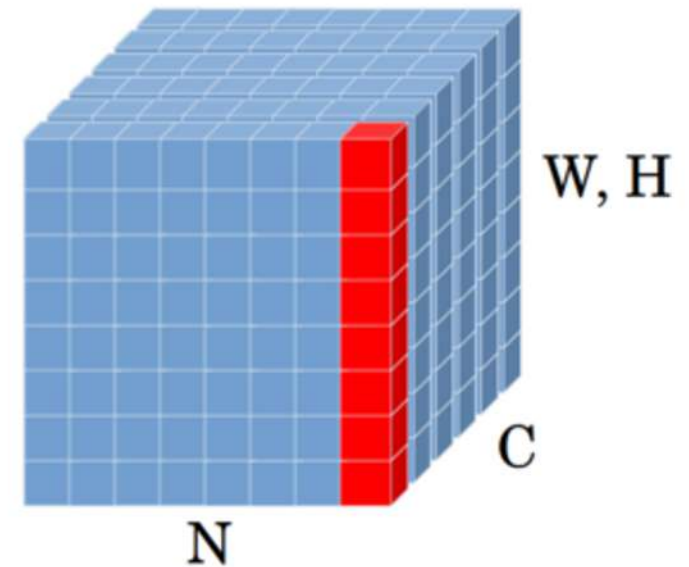
# Changes to the original network

➤ Batch normalization → Instance normalization



Batch Normalization

In BN, we consider one feature map over all the training sample.



Instance Normalization

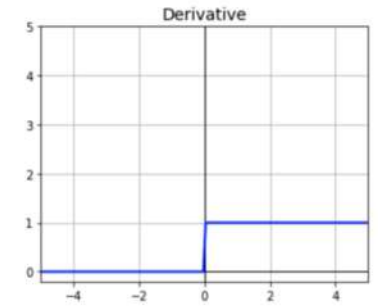
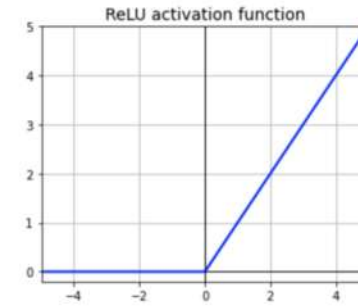
In IN, we consider one training sample and feature map

# Changes to the original network

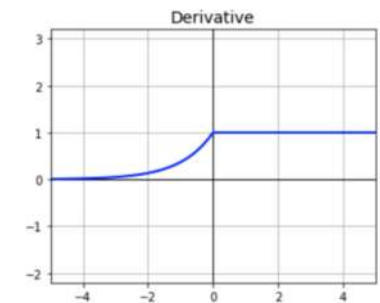
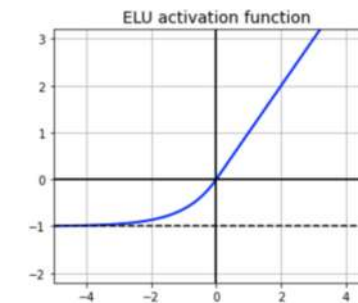
➤ Activation function

➤ Scale

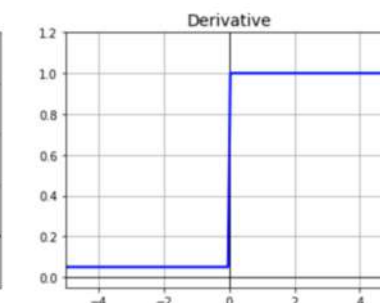
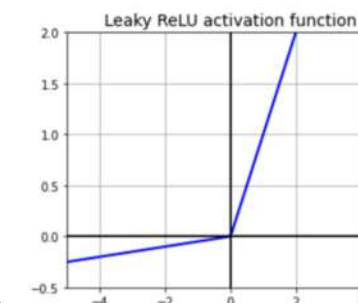
- ReLU



- Exponential ReLU



- Leaky ReLU





# Metric: **SIFID**

- *Fréchet Inception Distance (FID)* is a common metric for GAN evaluation that measures the deviation between the distribution of deep features of generated and real images, by using the activation vector after the last pooling layer of the Inception Network.
- However, since we have a single training image, we are interested in its *internal patch statistics*.
- In **SIFID** we use the internal distribution of deep features at the output of the convolutional layer just before the second pooling layer.

# Experiments in the Medical Field

---

## ➤ Brain MRI

- random sample
- harmonization

## ➤ Dental radiography

- random sample
- editing

## ➤ Chest X-ray

- random sample
- harmonization

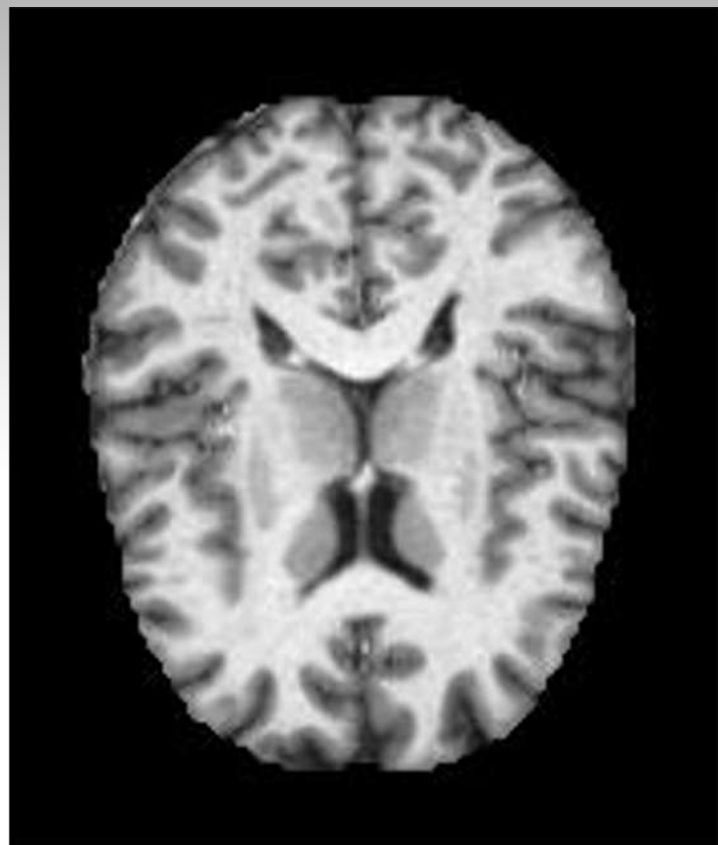
## ➤ Femoral fracture

- editing

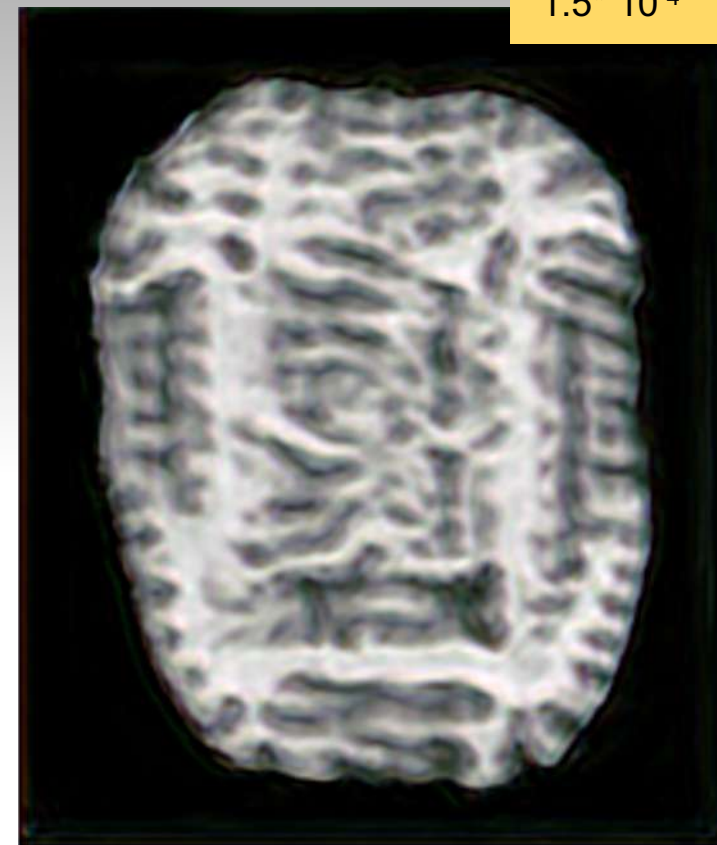


# Brain MRI

## Random samples



Training image



Injection at scale 7

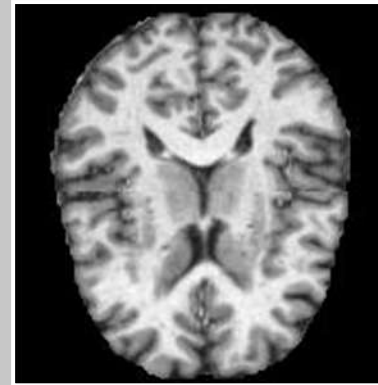
# Brain MRI Harmonization

$1.75 * 10^{-4}$

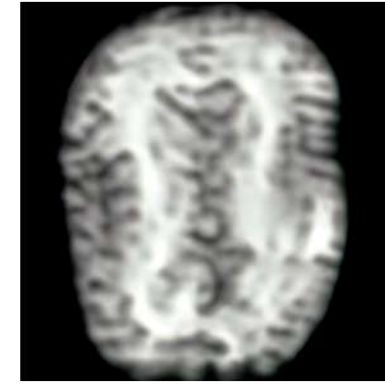


Harmonization with  
injection at scale 3 using  
batch normalization

Input image

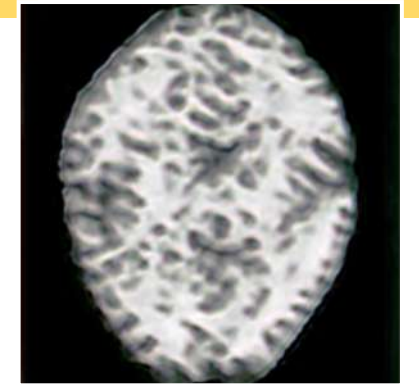


Scale 0



$1.36 * 10^{-4}$

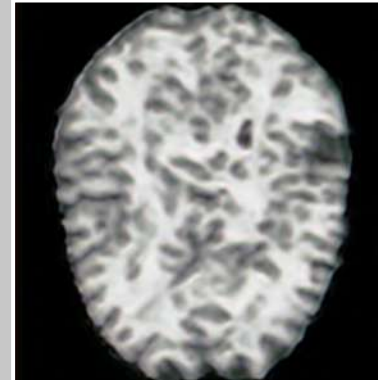
Scale 1



$1.56 * 10^{-4}$

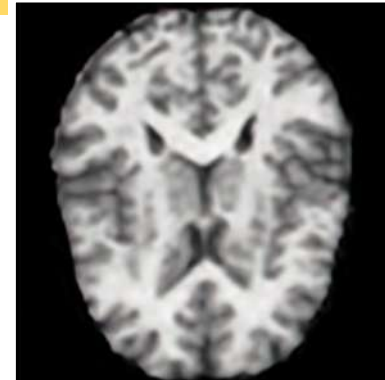
Scale 2

$1.91 * 10^{-4}$



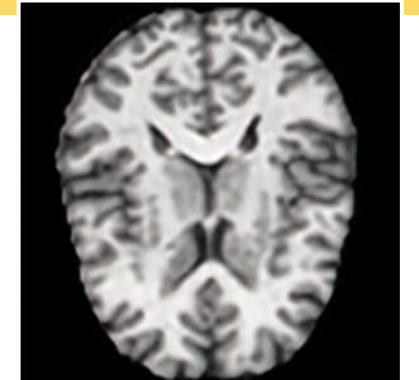
Scale 3

$2.15 * 10^{-4}$



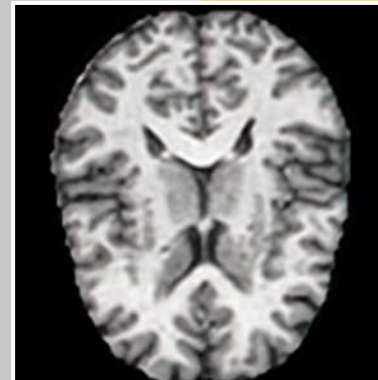
Scale 4

$2.15 * 10^{-4}$



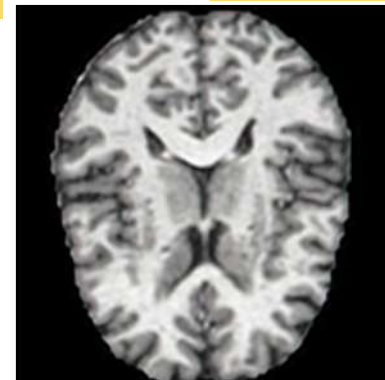
Scale 5

$2.11 * 10^{-4}$



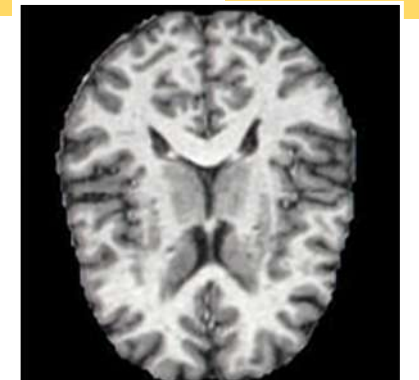
Scale 6

$2.26 * 10^{-4}$



Scale 7

$2.21 * 10^{-4}$





# Chest X-ray

## Random samples

Training image



Scale 0

$2.26 * 10^{-4}$



Scale 1

$2.41 * 10^{-4}$



Scale 2

$1.79 * 10^{-4}$



Scale 3

$2.21 * 10^{-4}$



Scale 4

$1.81 * 10^{-4}$



Scale 5

$2.16 * 10^{-4}$



Scale 6

$2.38 * 10^{-4}$



Scale 7

$2.71 * 10^{-4}$



# Chest X-ray

## - Harmonization

Input image



Scale 0  $2.17 * 10^{-4}$



Scale 1  $2.06 * 10^{-4}$



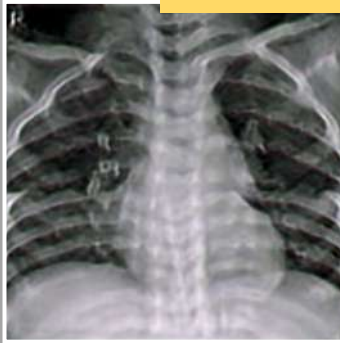
Scale 2  $2.21 * 10^{-4}$



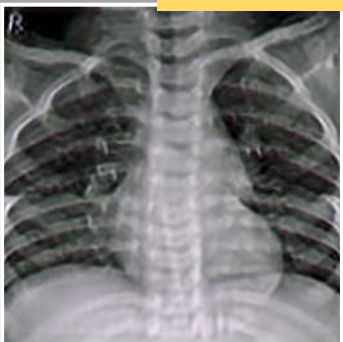
Scale 3  $2.51 * 10^{-4}$



Scale 4  $2.63 * 10^{-4}$



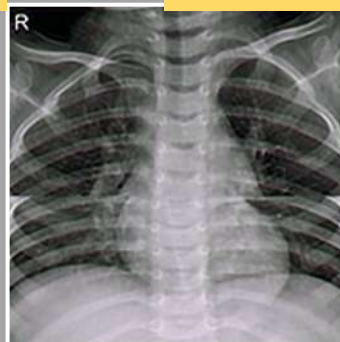
Scale 5  $2.74 * 10^{-4}$



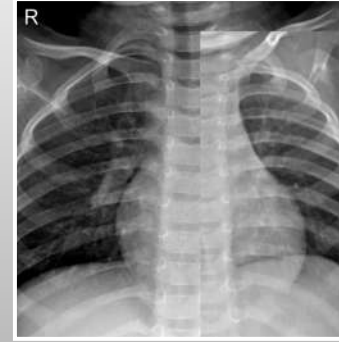
Scale 6  $2.53 * 10^{-4}$



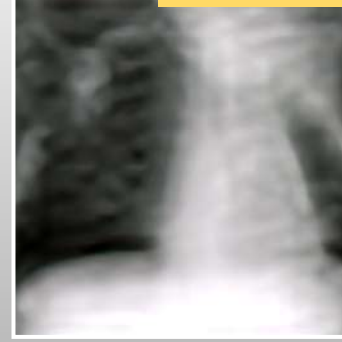
Scale 7  $2.62 * 10^{-4}$



Input image



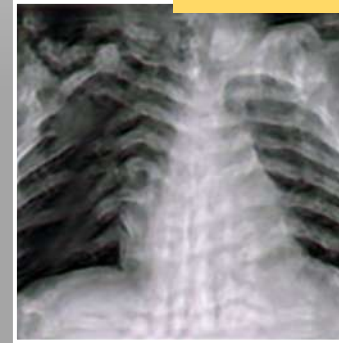
Scale 0  $2.04 * 10^{-4}$



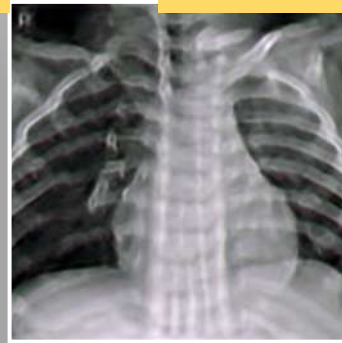
Scale 1  $2.39 * 10^{-4}$



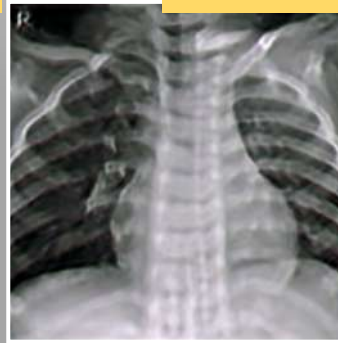
Scale 2  $2.54 * 10^{-4}$



Scale 3  $2.62 * 10^{-4}$



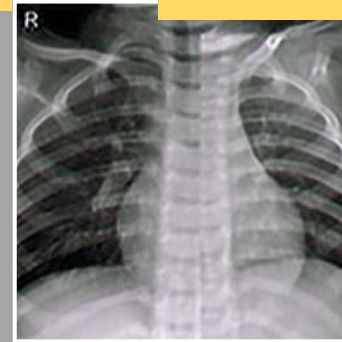
Scale 4  $2.63 * 10^{-4}$



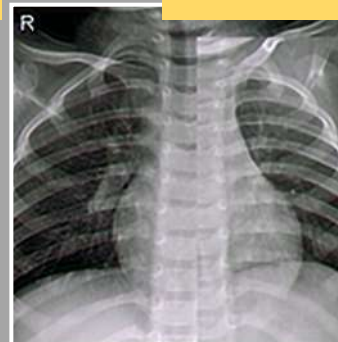
Scale 5  $2.59 * 10^{-4}$



Scale 6  $2.66 * 10^{-4}$



Scale 7  $2.63 * 10^{-4}$



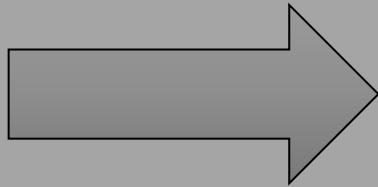


# Femoral Fracture Editing

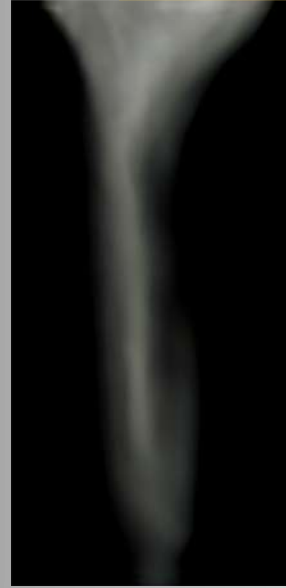
Training image



Input image



Scale 0  $2.14 * 10^{-4}$



Scale 1  $2.68 * 10^{-4}$



Scale 2  $3.04 * 10^{-4}$



Scale 3  $2.87 * 10^{-4}$



Scale 4  $3.04 * 10^{-4}$



Scale 5  $3.02 * 10^{-4}$



Scale 6  $3.06 * 10^{-4}$

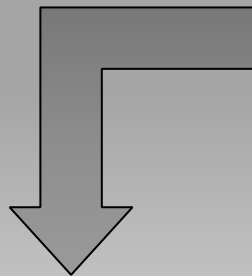


Scale 7  $2.90 * 10^{-4}$



# Dental Radiography

Random samples



Training  
image

scale 0

$4.66 * 10^{-4}$



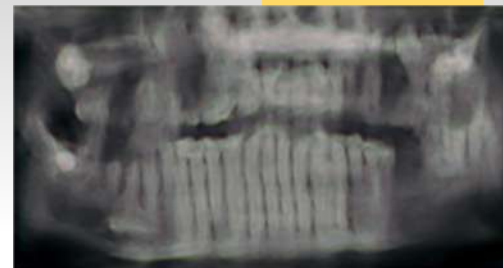
scale 1

$4.64 * 10^{-4}$



scale 2

$4.81 * 10^{-4}$



scale 3

$4.89 * 10^{-4}$



scale 4

$4.56 * 10^{-4}$



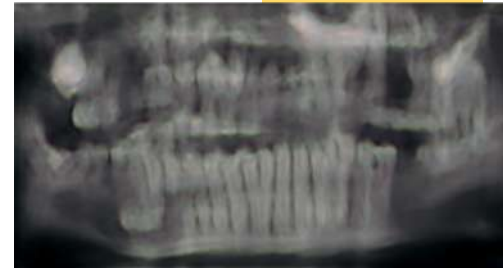
scale 5

$4.57 * 10^{-4}$



scale 6

$4.79 * 10^{-4}$



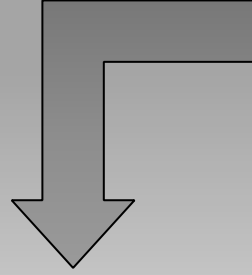
scale 7

$4.39 * 10^{-4}$



# Dental Radiography

## Editing



Input  
image

scale 0

$5.83 * 10^{-4}$



scale 1

$7.21 * 10^{-4}$



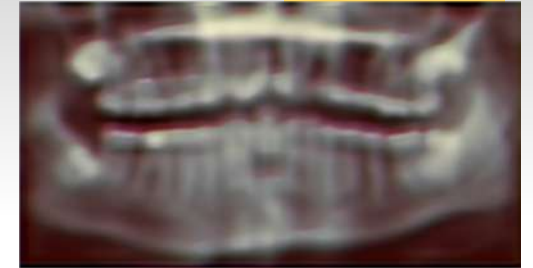
scale 2

$7.73 * 10^{-4}$



scale 3

$7.58 * 10^{-4}$



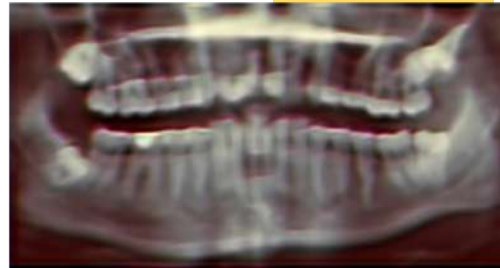
scale 4

$7.33 * 10^{-4}$



scale 5

$6.97 * 10^{-4}$



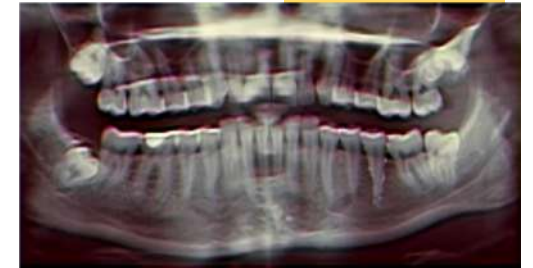
scale 6

$6.15 * 10^{-4}$



scale 7

$4.45 * 10^{-4}$





# Conclusions

---

- Random samples: not satisfactory
- Harmonization - Editing: good results, but in most cases requires manual interventions on the image