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

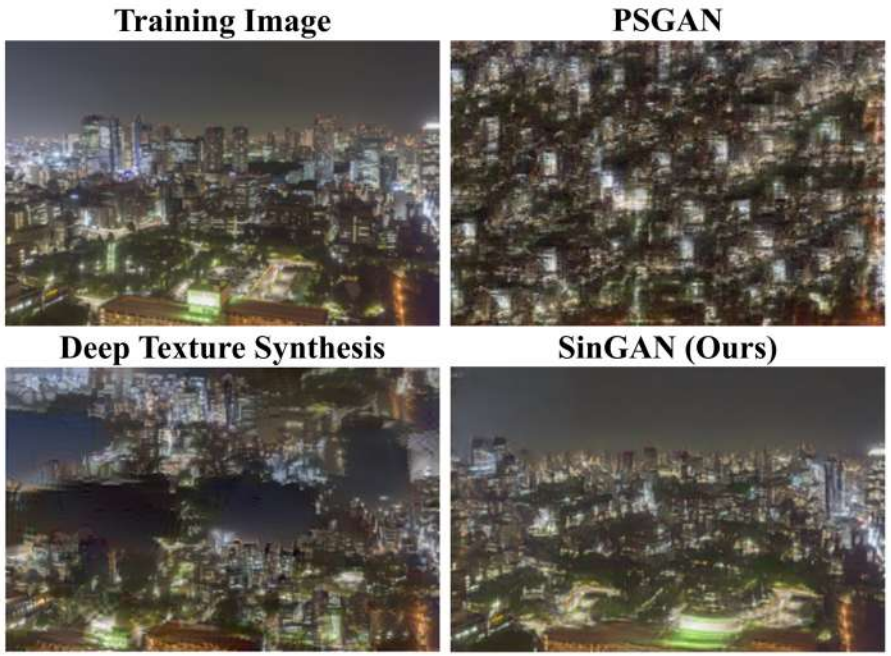


Figure: Single image texture generation

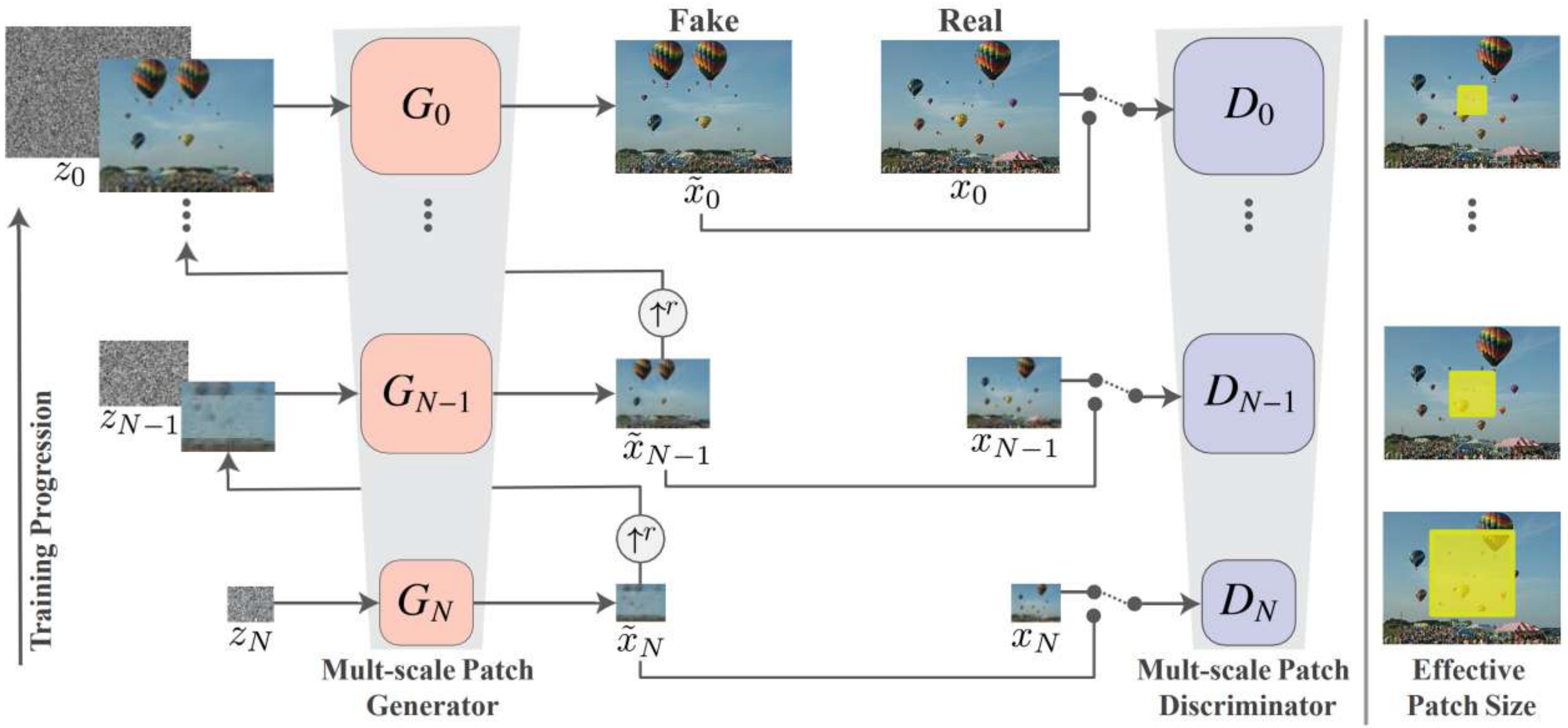
Novelty

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.

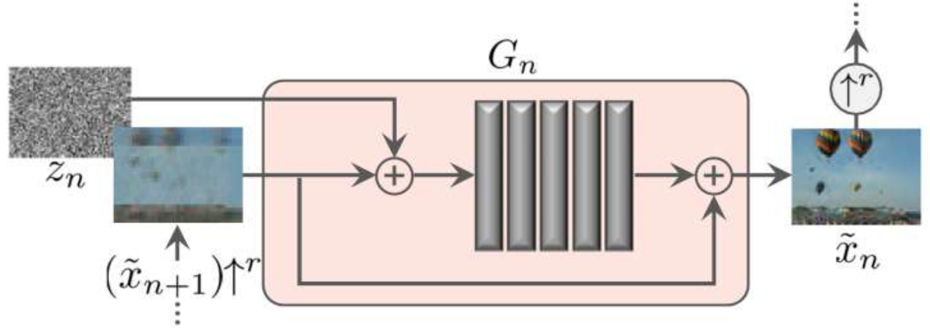
Use SinGAN to increase the size of the available datasets, by generating new medical images to be used in several machine learning tasks.

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

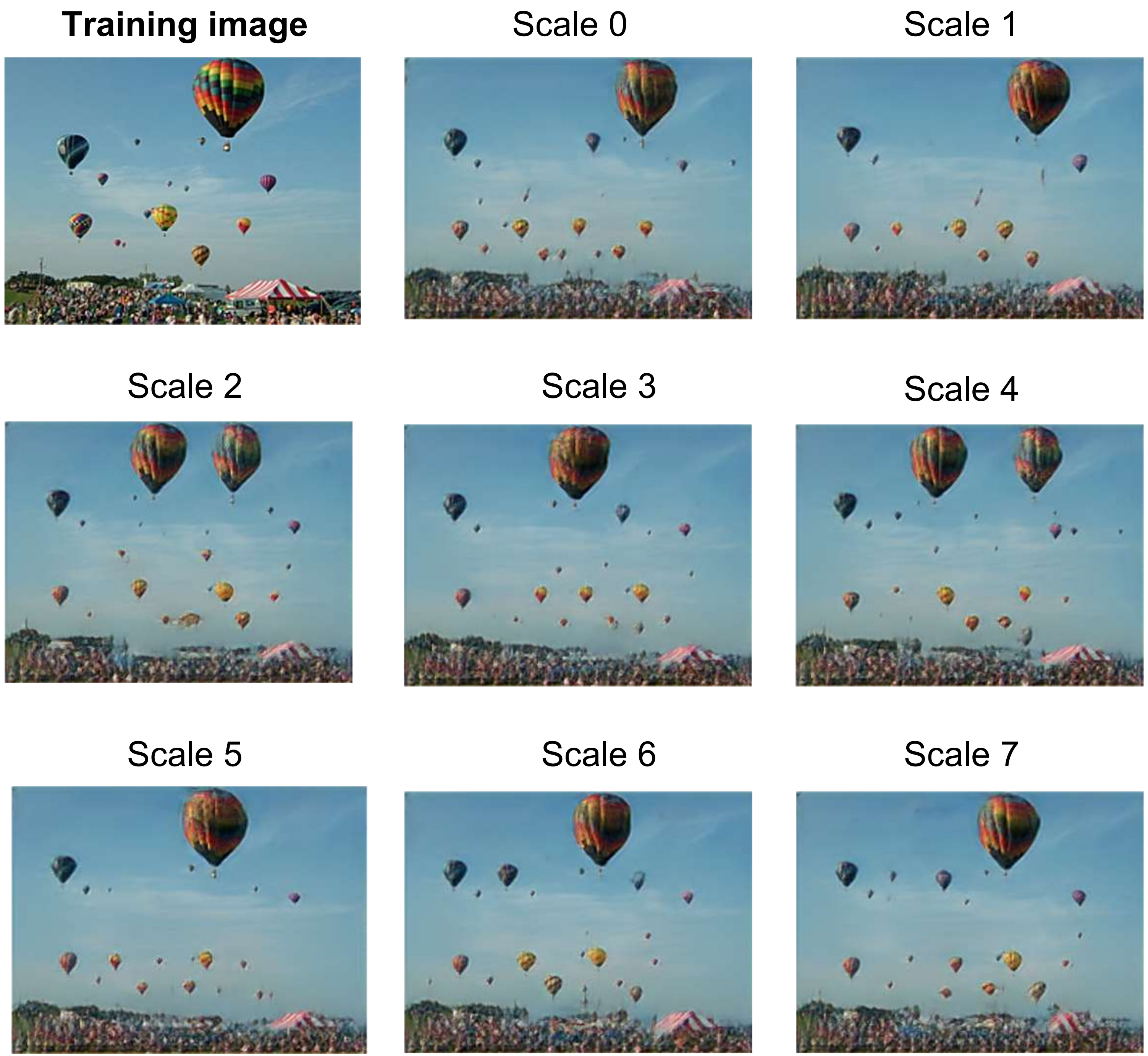
$$L_R = (\tilde{x} - x)^2$$

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

Replicating paper's results:

Random Sample

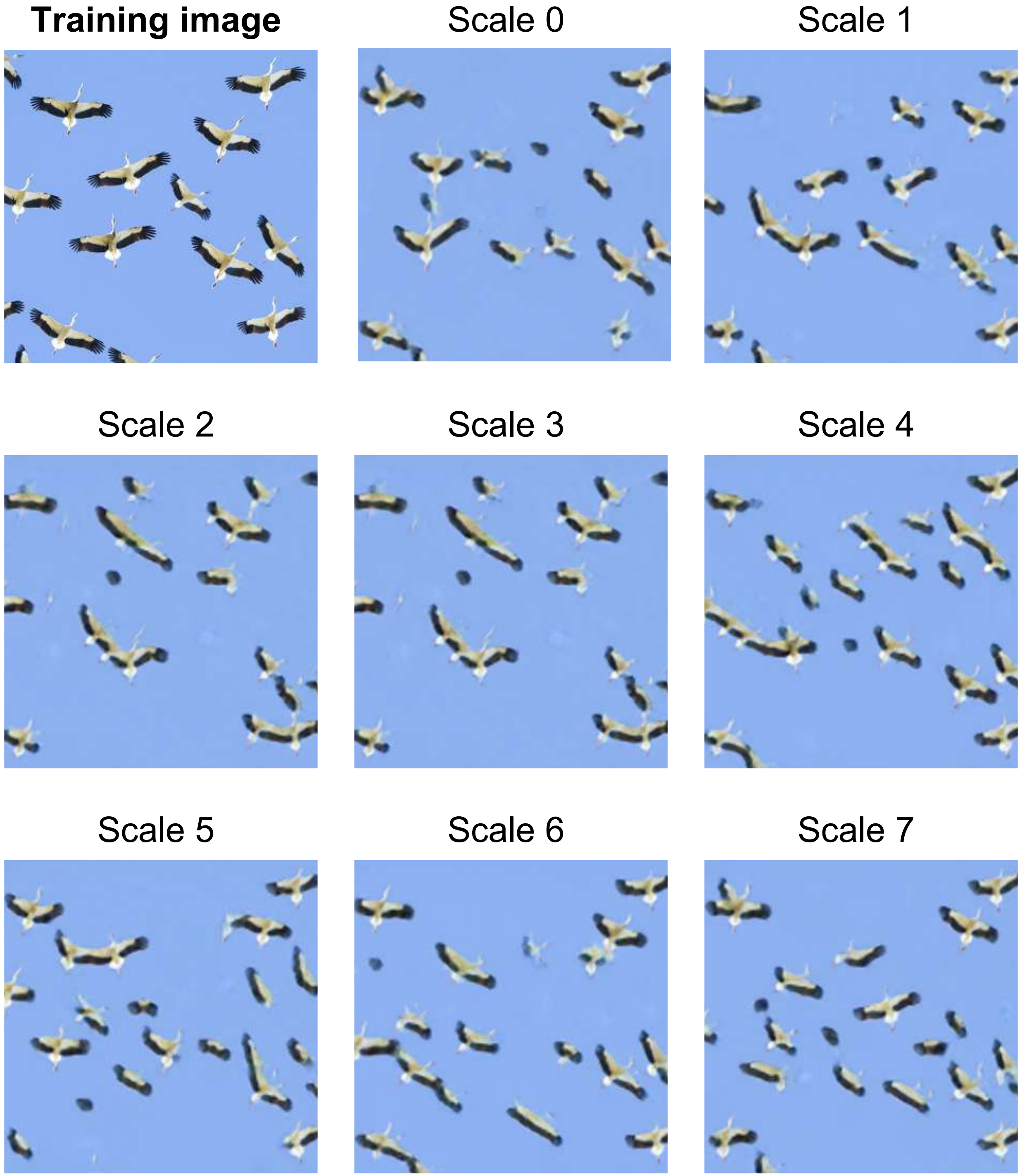
Baloons in a landscape



Further example on natural images:

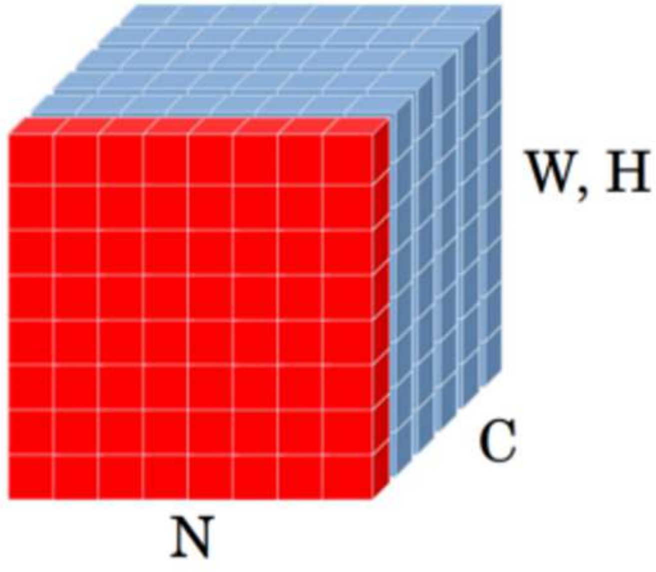
Random Sample

Flying birds in the sky

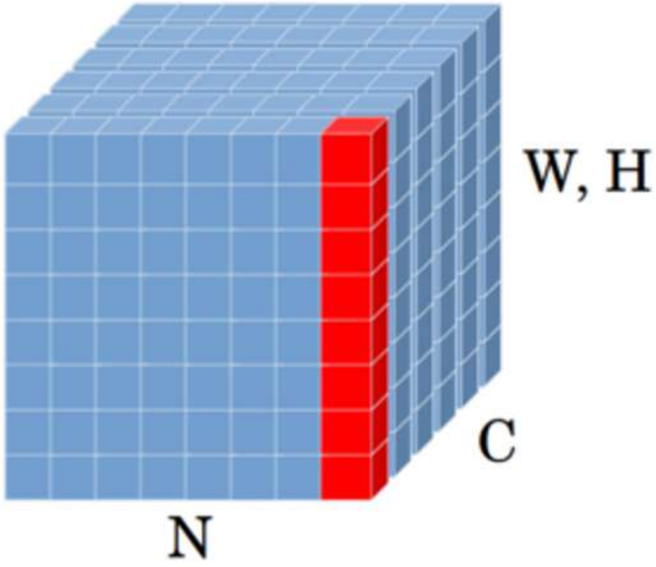


Changes to the original network

➢ **Batch normalization→ Instance normalization**



Batch Normalization



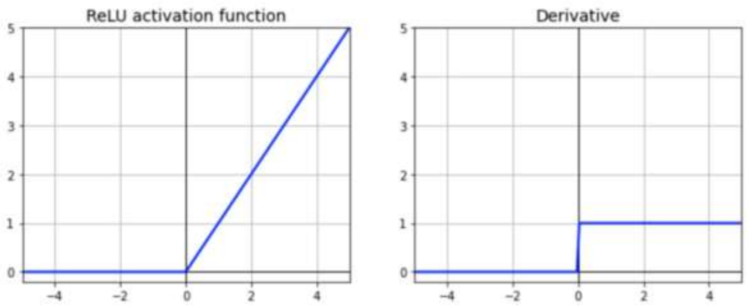
Instance Normalization

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

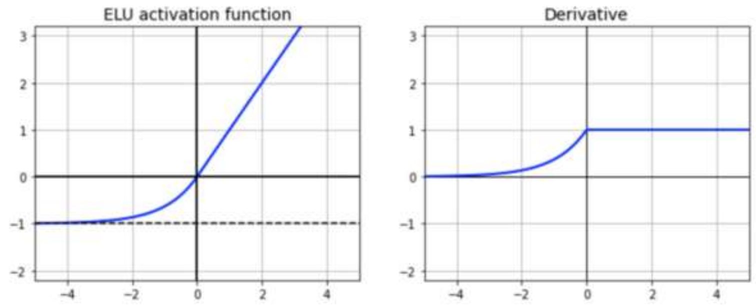
In IN, we consider one training sample and feature map

Changes to the original network

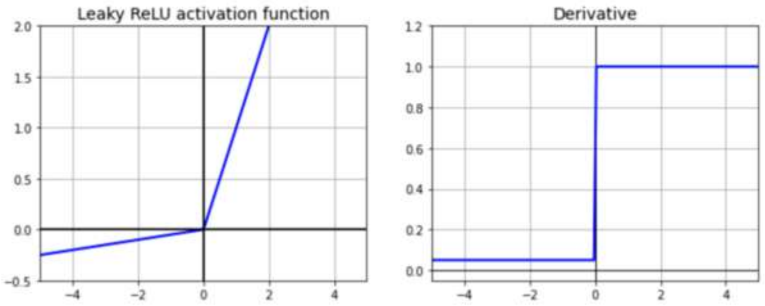
- ➢Activation function
  - ReLU



  - Exponential ReLU



  - Leaky ReLU



- ➢Scale

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

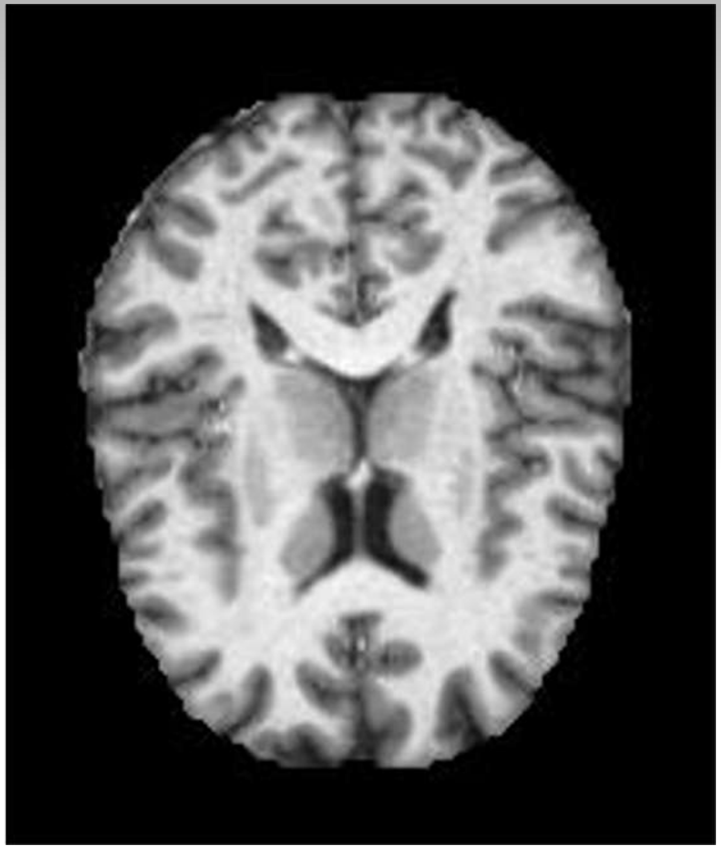
➢ Brain MRI
- random sample
- harmonization

➢ Dental radiography
- random sample
- editing

➢ Chest X-ray
- random sample
- harmonization

➢ Femural fracture
- editing

Brain MRI
Random samples

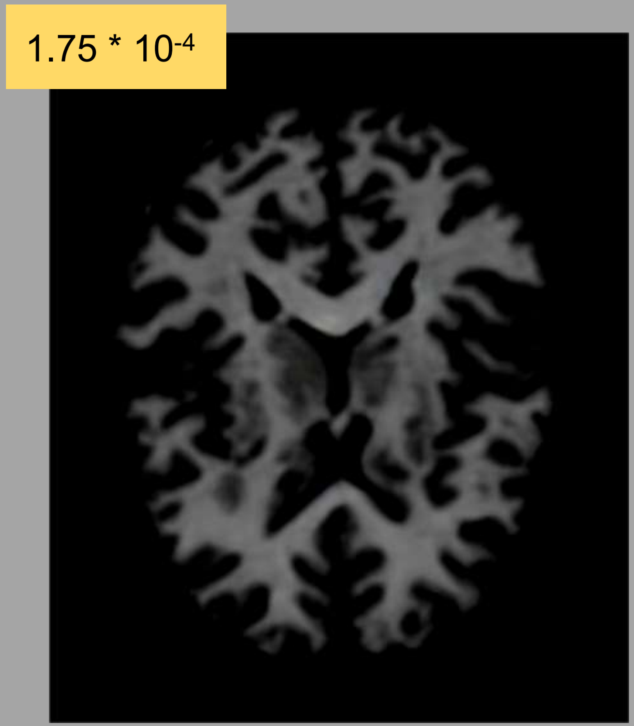


Training image

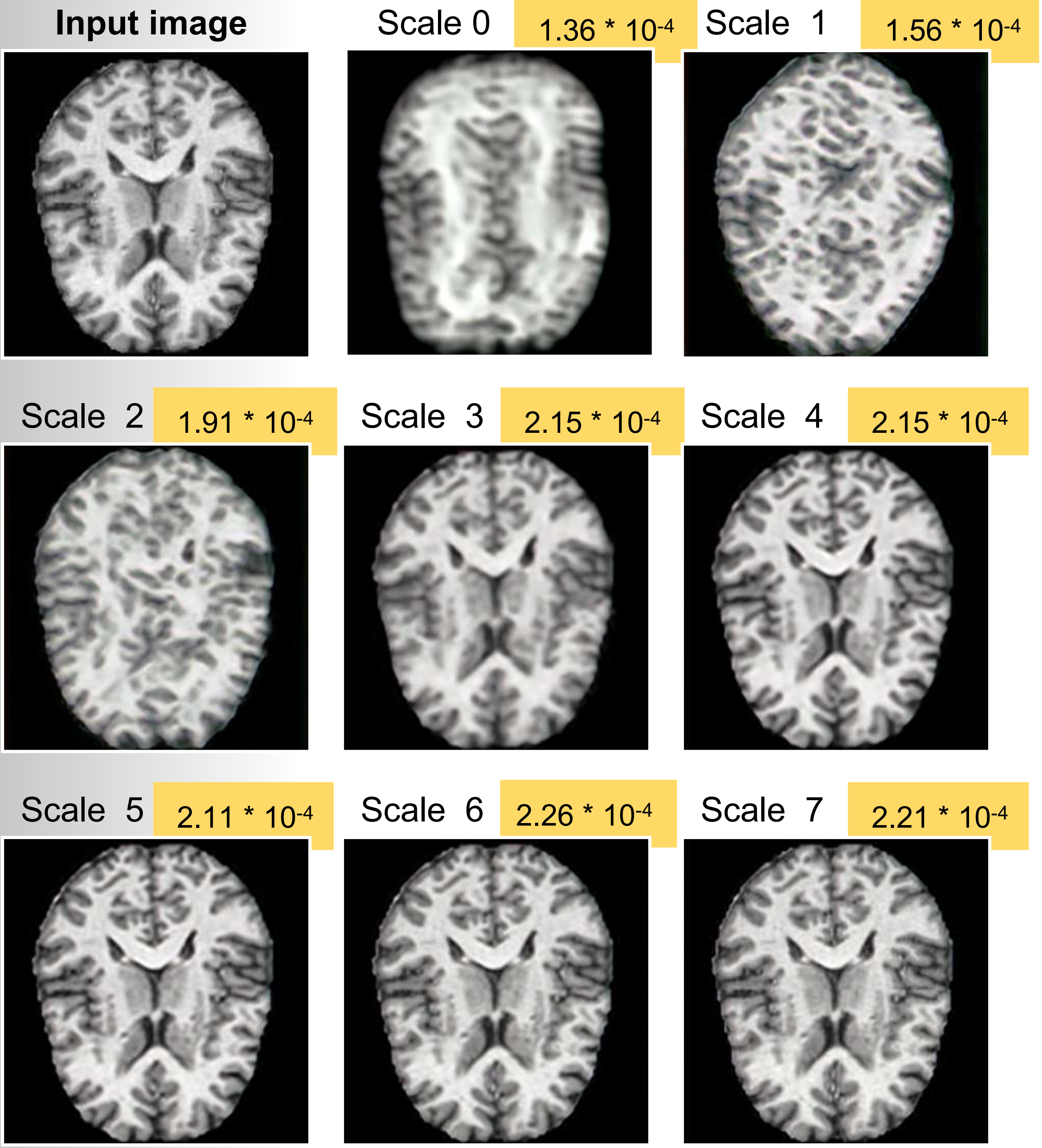
Injection at scale 7

Brain MRI

Harmonization

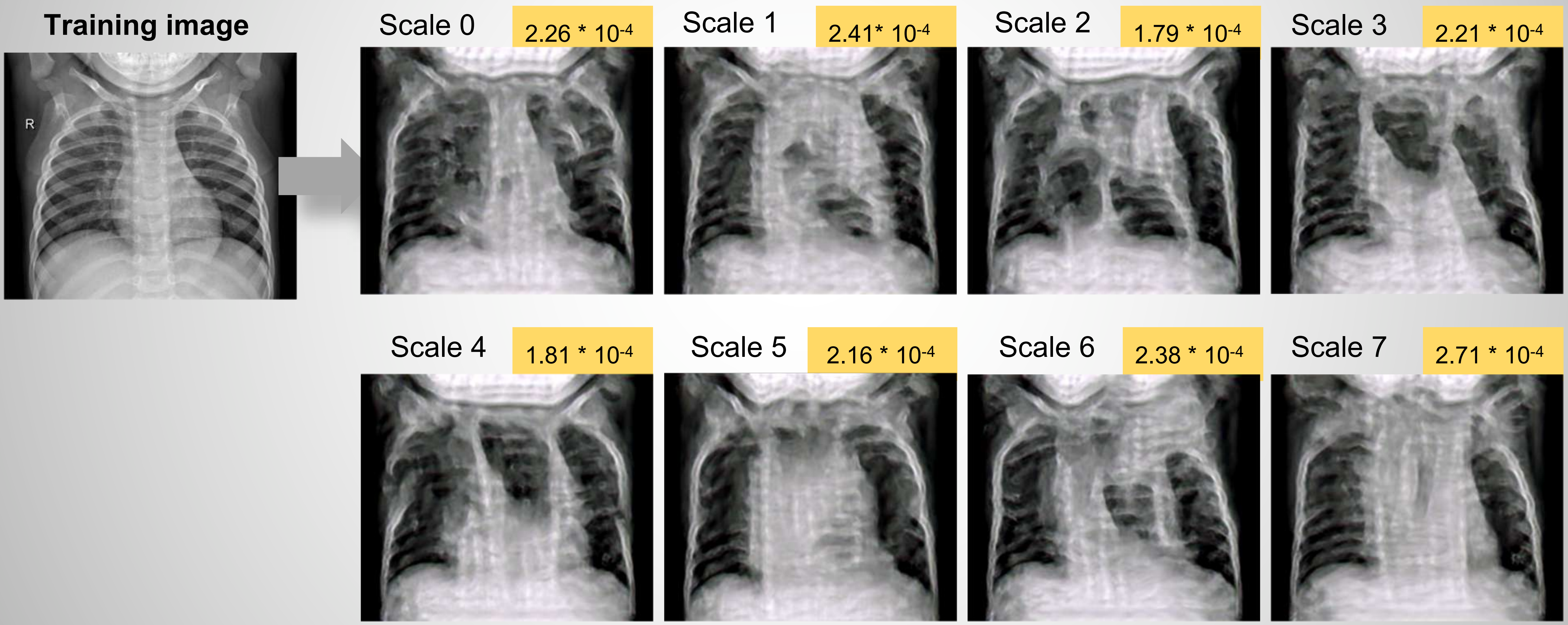


Harmonization with injection at scale 3 using batch normalization

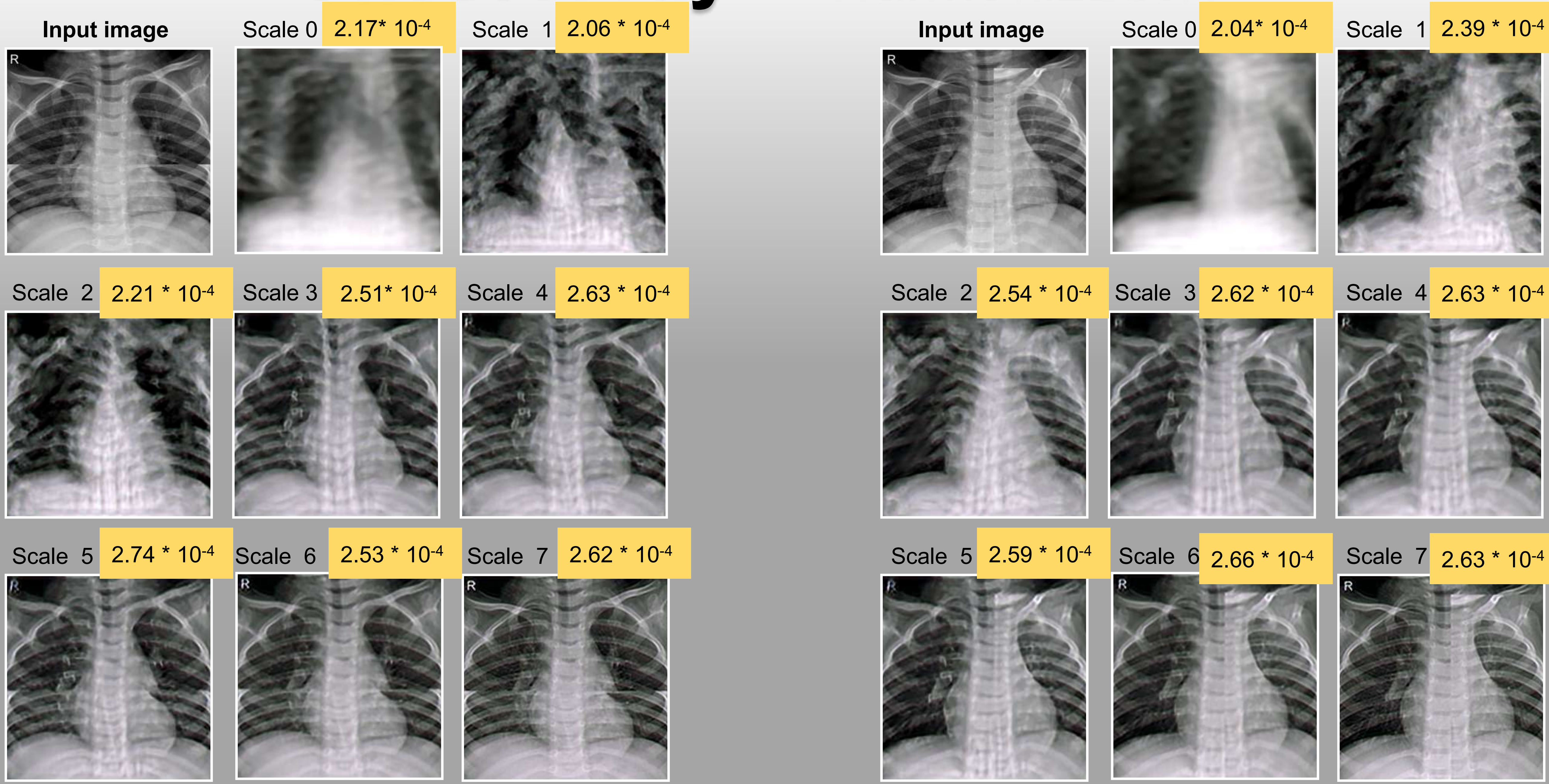


Chest X-ray

Random samples

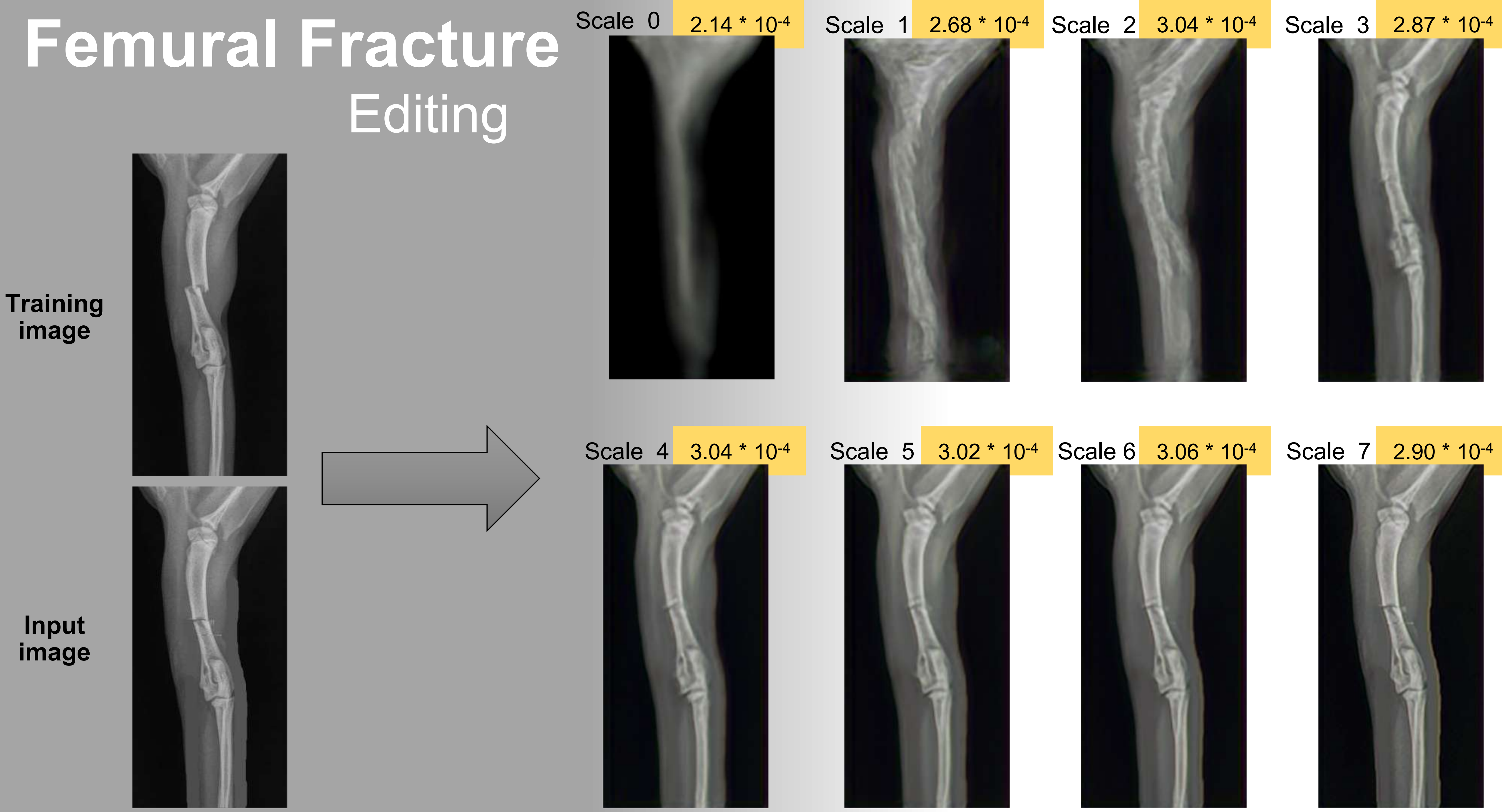


Chest X-ray - Harmonization



Femural Fracture

Editing



Dental Radiography

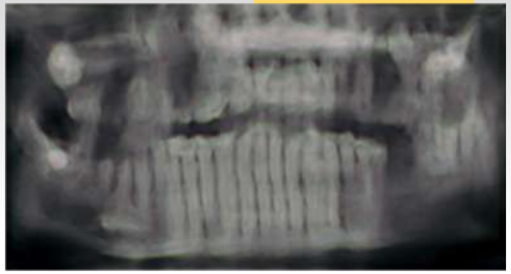
Random samples

Training image

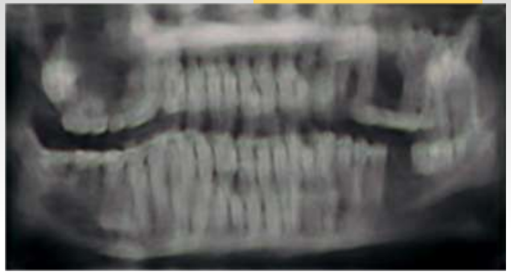
scale 0 $4.66 * 10^{-4}$

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 $4.57 * 10^{-4}$

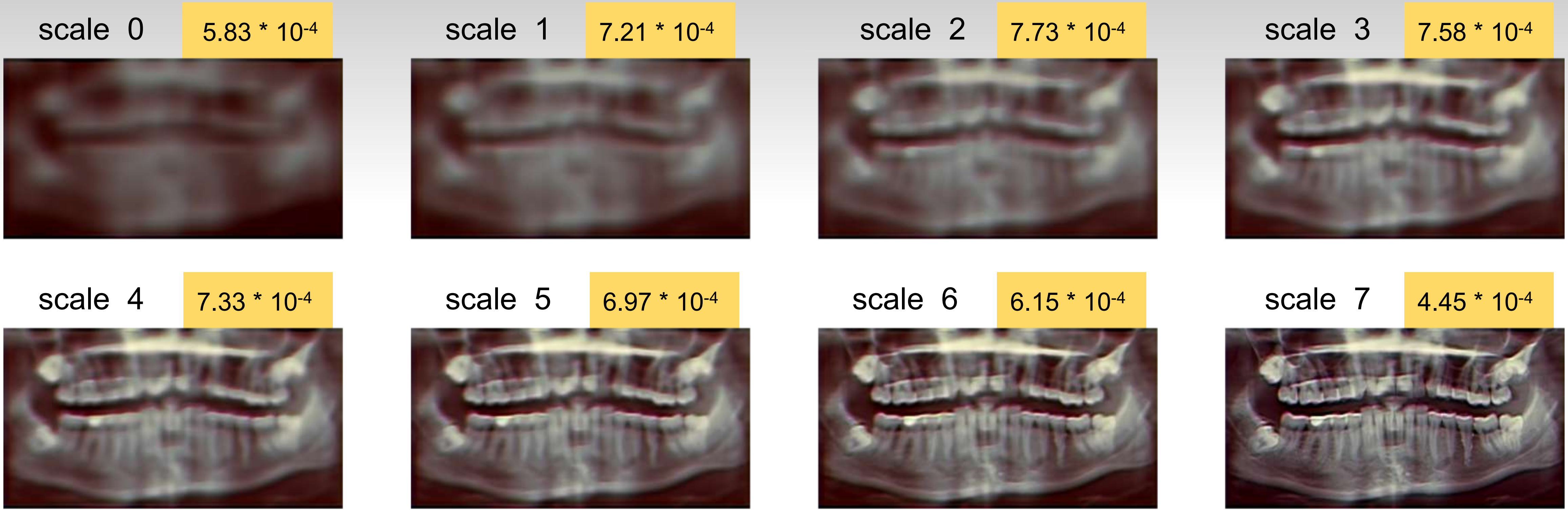
scale 6 $4.79 * 10^{-4}$

scale 7 $4.39 * 10^{-4}$

Dental Radiography

Editing

Input image



Conclusions

- Random samples: not satisfactory

- Harmonization - Editing: good results, but in most cases requires manual interventions on the image