ISMI – NODE21 Challenge

Nodule Generation Track

Group 5

Lennart Geertjes - Marco Post - Reza Shokrzad

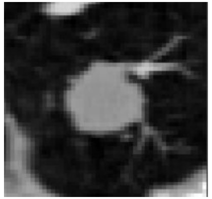
Generation synthetic nodule

- The aim of the task is producing more nodules to enhance the dataset.
- This could improve performance of detection systems.
- This could help detecting cancer in earlier stages.

- RQ: Is using CT patches feasible to generate X-Ray nodules?

Materials

- Applying different sources of data:
  - 1186 CT patches containing nodules and nodule masks.
  - 1031 frontal X-Ray images and lung masks[1].



An example of CT patch

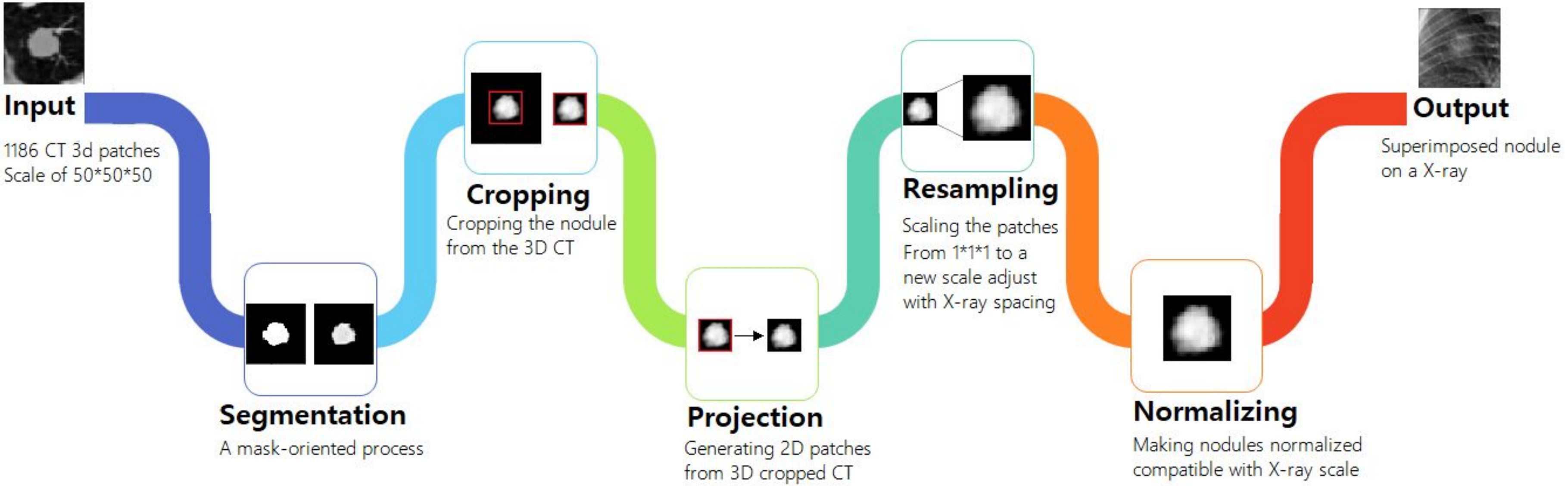
An example of Mask

An example of X-Ray

Method

- Creating a pipeline in Python mostly inspired by Litjens et al. work
    - Feeding the pipeline with 1186 CT in scale of 50*50*50
    - All CT voxels are segmented using the associated nodule mask.
    - The segmented CT is cropped to have 3D nodule, before projecting.
    - The 3D cropped nodule is projected to a 2D image (Campo et al.).
    - The 2D nodule is resampled using interpolation to an isotropic resolution equal to the X-ray.
    - Generated nodule is augmented by four different techniques to increase the training data.
    - The pixel value of nodules is normalized to get compatible with X-ray range.
    - The generated nodule is superimposed onto the X-ray image (randomly chosen location)

Preprocessing



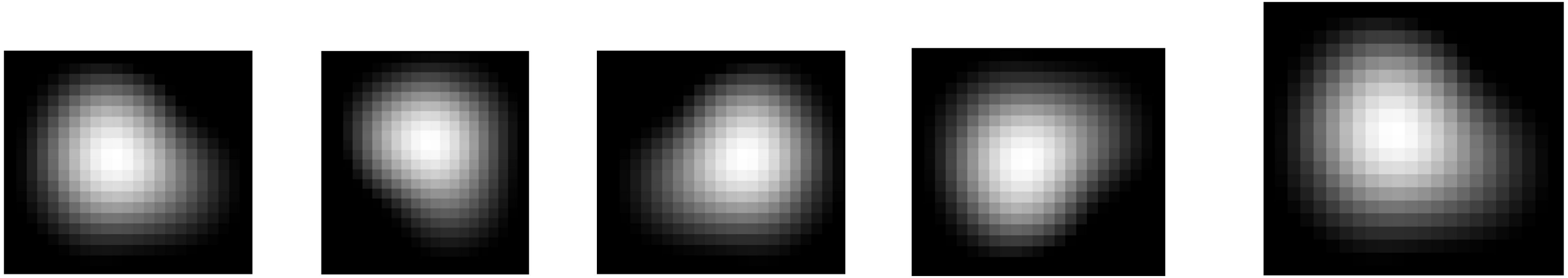
Superimposing Phase

- We applied an empirical mathematical method.
- The method named the Perfect Spherical Nodule Simulation.
- Weighting central pixels of nodules more than borders.
- Methods parameters guarantee proper blending especially near the edges.
- Each pixel value of nodule, after multiplying by c, would add with X-ray corresponding pixel.

$$c[r] = C\left(\frac{4}{D^4}r^4 - \frac{4.2}{D^2}r^2 + 1\right)$$

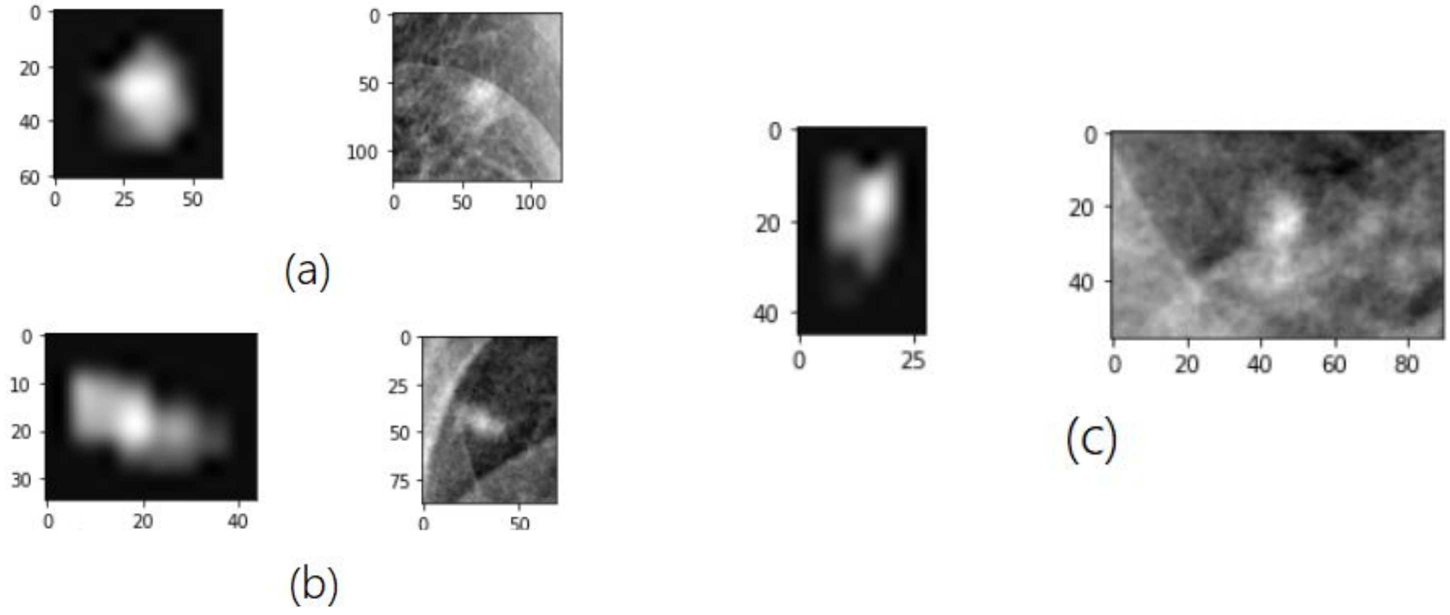
Data Augmentation

- Since new synthetic data needs to be practical, making the options of augmentation methods are limited.
- We applied four variant augmenting methods
- Rotation, Vertical and Horizontal Flipping, and Zooming-in
- Generated 930 nodules



(a) Original Version (b) Rotation 90 (c) Flip Horizontally (d) Flip Vertically (e) Zoom-in

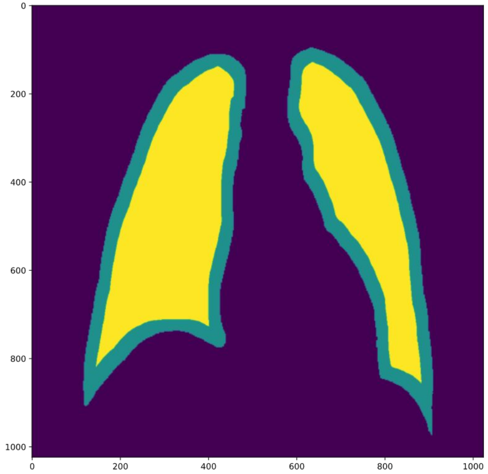
Generated and Superimposed Nodules



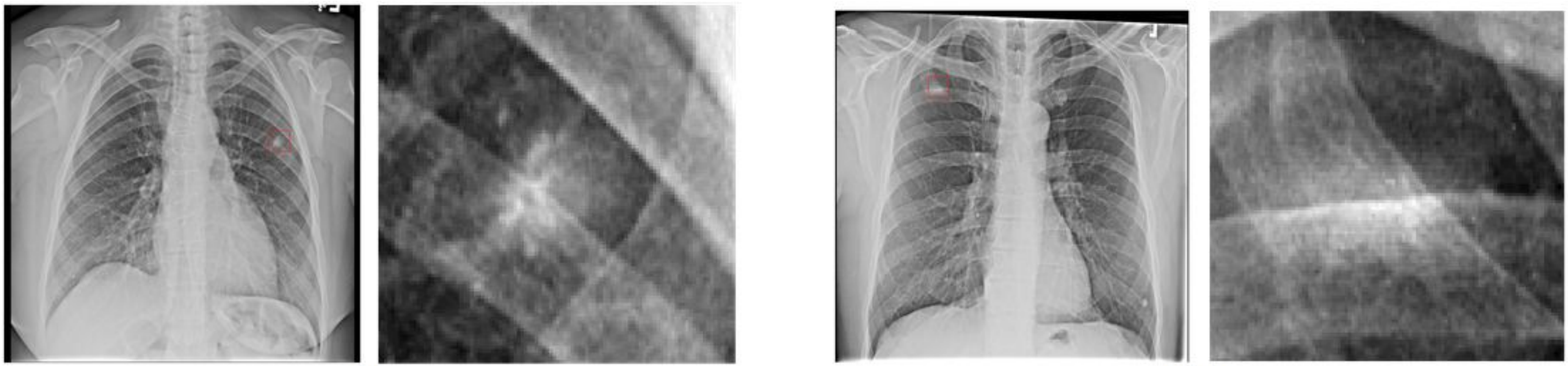
Two examples of projected nodules and superimposing nodules

Lung Mask Preprocessing

- A specific location of X-ray must be chosen to superimpose the 2D nodules.
- The numpy random number generator was applied.
- To prevent inserting nodule outside of lungs, lung mask was used to crop them.



Final Result



(a) (b)

Two examples of superimposed X-rays

Detection Phase

- Detection is necessary to test the performance enhancing capabilities of our newly generated nodule dataset.
- We tested the data on an already existing deep learning detection method.
- We chose Faster RCNN that is quick to implement.
    1. Training the model on only the real data, the outcome of this would serve as a baseline performance.
    2. Training the model on only generated data.
    3. Training the model on dataset which was 50 percent real data and 50 percent generated data.

The Experimental Setup

- Superimposing the projected nodules in different sizes
- ??? iteration to superimpose nodules in 1031? CXR.
- Testing the augmented results with a Faster RCNN? detection algorithm.
- Splitting dataset to training, validation, and test set (90-5-5 %)
- We used a learning rate of 0.005, momentum of 0.9 and a weight decay of 0.0005.
- We trained the model in 10 epochs.

Result

- Each of the systems was tested on 52 x-rays of real patient data.
- the mAP* of the models to detect nodules was calculated.

Dataset	mAP
Real	0.101
Generated	0.0
Mixed	0.039

* mean Average Precision

Discussion

- We could insert different size of nodules.
- We generated much more data than the first time
- We modified the resampling based on the meaning of pixel size (in CT and X-ray)
- We opted a superimposed model in which we added nodule into X-ray instead of computing the mean.
- We modified the normalization process
- We added data augmentation
- We could train a model with our generated data

Future Works

- Choosing a more correlated location on X-ray to inset nodules (it is already randomly chosen)
- Generating synthetic nodule to add behind the heart or diaphragm
- Enhancing the number of data by merging data augmentation techniques
- Applying Gans to produce nodule