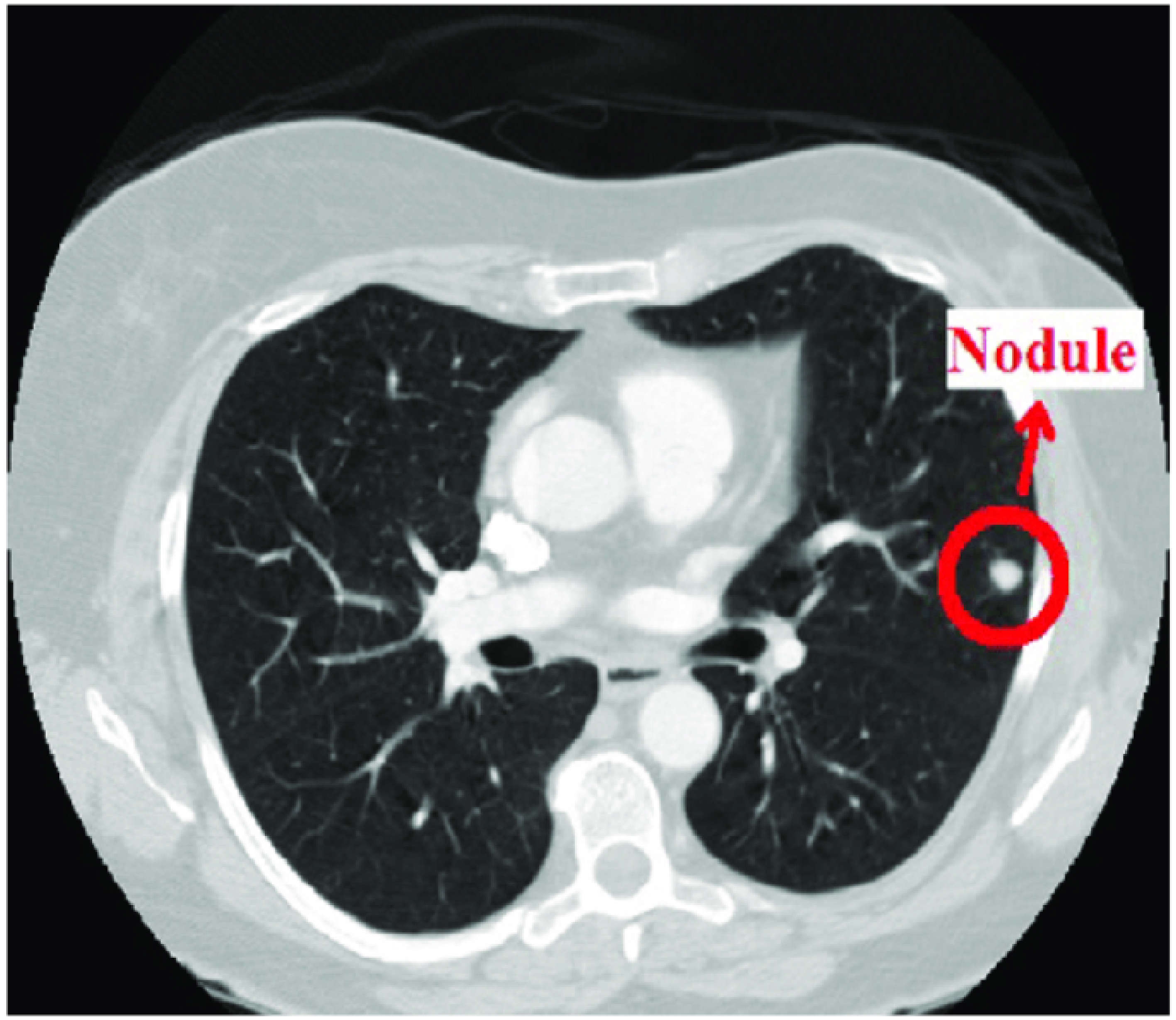
Implementing Cycle-GAN on Lung Nodule Images to Improve Classification Accuracy

Shreyas Puttaraju

Background

The Lung Image Database Consortium (LIDC) is a vast archive of thoracic CT scans of individuals with lung cancer. Each image used in this project is a part of the whole scan around a nodule. Current Deep Learning models are not able to accurately classify the nodule type(Malignant or Benign).



Introduction

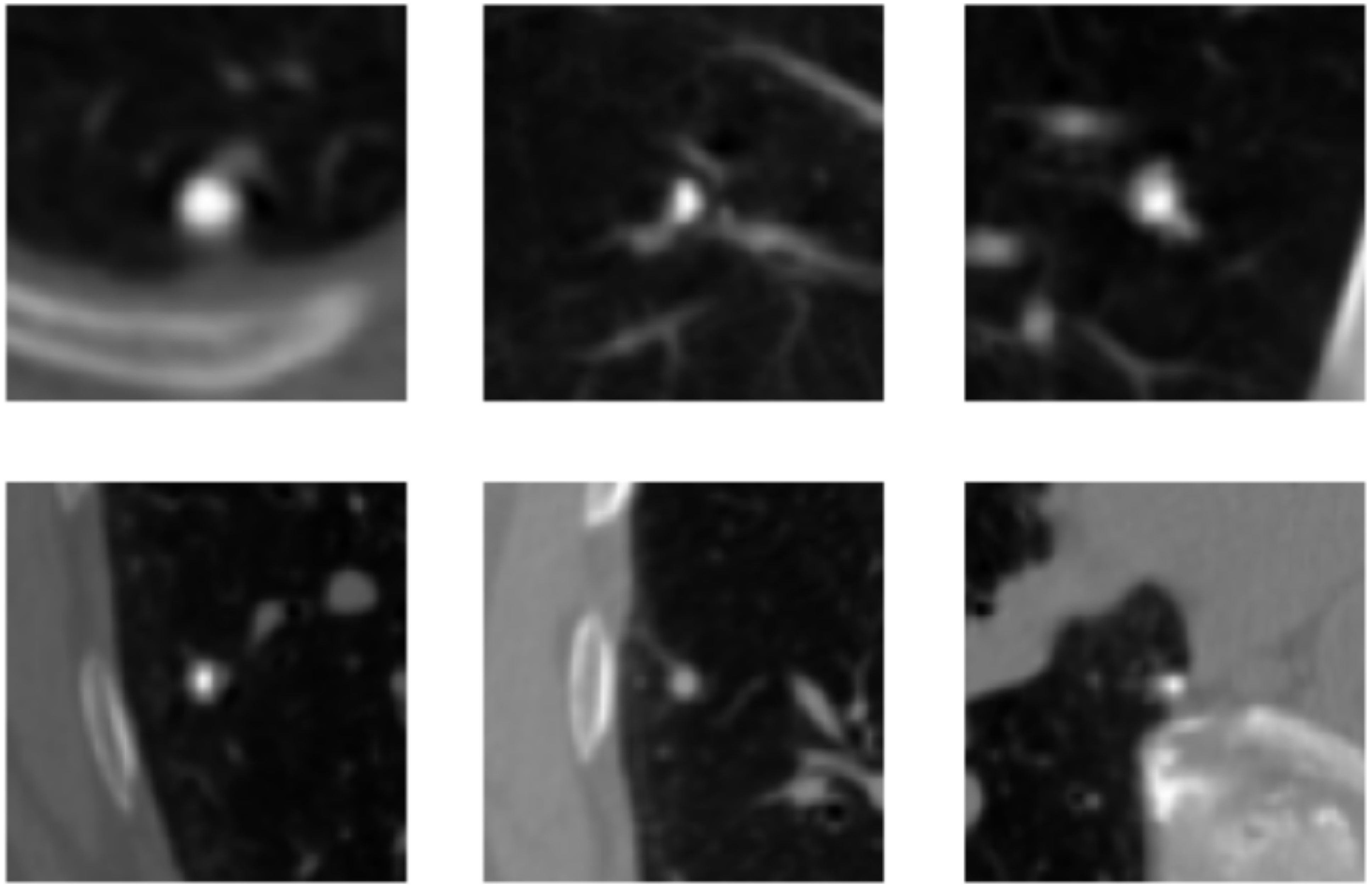
The present neural network models classify few of the Lung nodules with high accuracy but at the same time fail terribly while classifying other Lung nodules. The goal of the project is to get high accuracy score on those nodules which are wrongly classified.

Research Hypothesis

- Can we improve the classification accuracy on the nodules that are hard to classify by using unpaired image-to-image translation techniques?

DATA

- The LIDC dataset contains CT scans of lungs with nodule annotations for nine semantic characteristics from up to four radiologists.
- These semantic features include the likelihood of malignancy which is measured on a five-point ordinal scale with classes 1, 2, 3, 4, and 5.
- From the nodule images, 64 image features describing nodule shape, nodule size, image intensity, and image texture were extracted.
- The LIDC consists of 2669 potential nodules, of which we use 829 nodules with four radiologist annotations and 507 images with three radiologist annotations for training and testing.



Lung Nodule Images

Methodology

1. Grouping the nodules which are easy to classify, and which are hard to classify.

2. Reconstruct the hard nodules such a way that the models can be sort accurately.

3. Classifying the nodule images.

Methodology

1. Grouping the nodules which are easy to classify, and which are hard to classify.
2. Reconstruct the hard nodules such a way that the models can be sort accurately.
3. Classifying the nodule images.

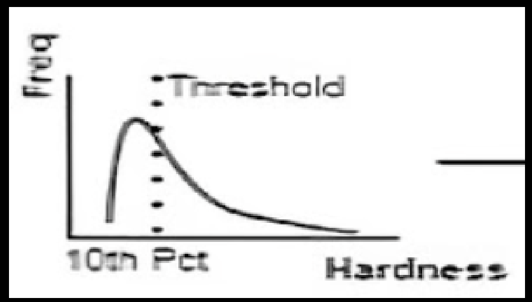
Grouping nodules based on Instance Hardness (Easy and Hard)

- The images are classified into easy and hard based on wMSE(Weighted Mean Squared error).
- The wMSE calculation is based on training an ensemble of classifiers using malignancy ratings aggregated into a single reference truth.
- For the LIDC dataset, four bagging classifiers with decision trees as the base model are trained on 64 image features, and the radiologist ratings are aggregated by majority consensus or mean.
- The wMSE calculation is repeated 20 times with shuffled ratings and the mean value is assigned as the instance-level hardness value for each nodule.

Step 1

Step 2

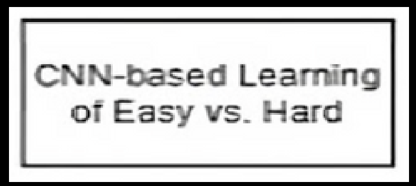
Step 3



Step 4

Step 5

Step 6



- The basis of the wMSE calculation is Selective Iterative Classification(SIC). An ensemble of classifiers is trained in series, similar to a boosting approach, with each model using a variable number of malignancy ratings that are aggregated into a single reference truth.

- The base model of the bagging classifiers is a decision tree with a maximum depth of five.

- The radiologist ratings are aggregated by majority consensus or by mean rating otherwise. For example, if three ratings are being considered, we take the mode if it exists or the mean if it does not.

- The wMSE threshold was set to 0, i.e. if a nodule was correctly classified on all the 4 classifiers they were considered as easy or else they are considered as hard.

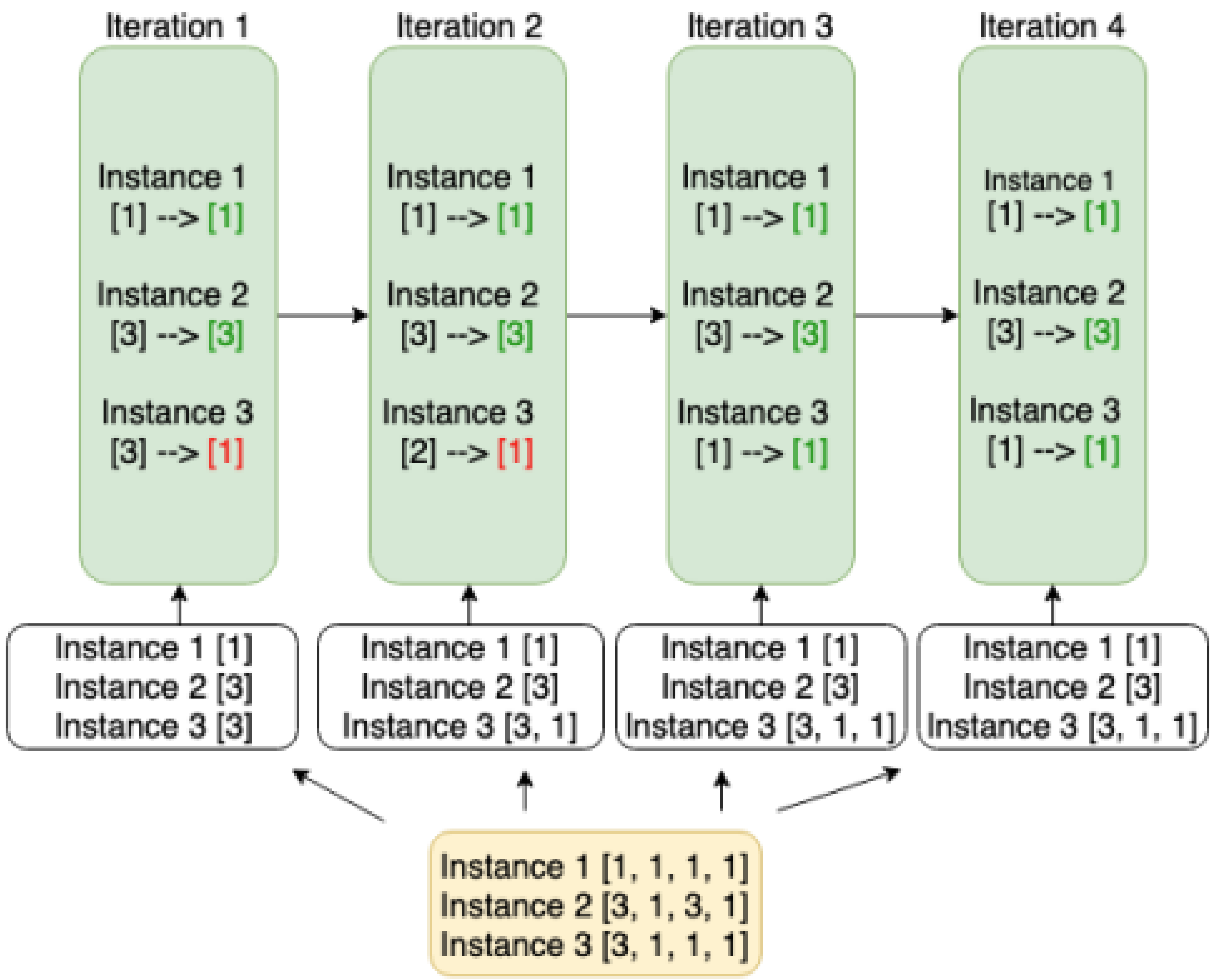


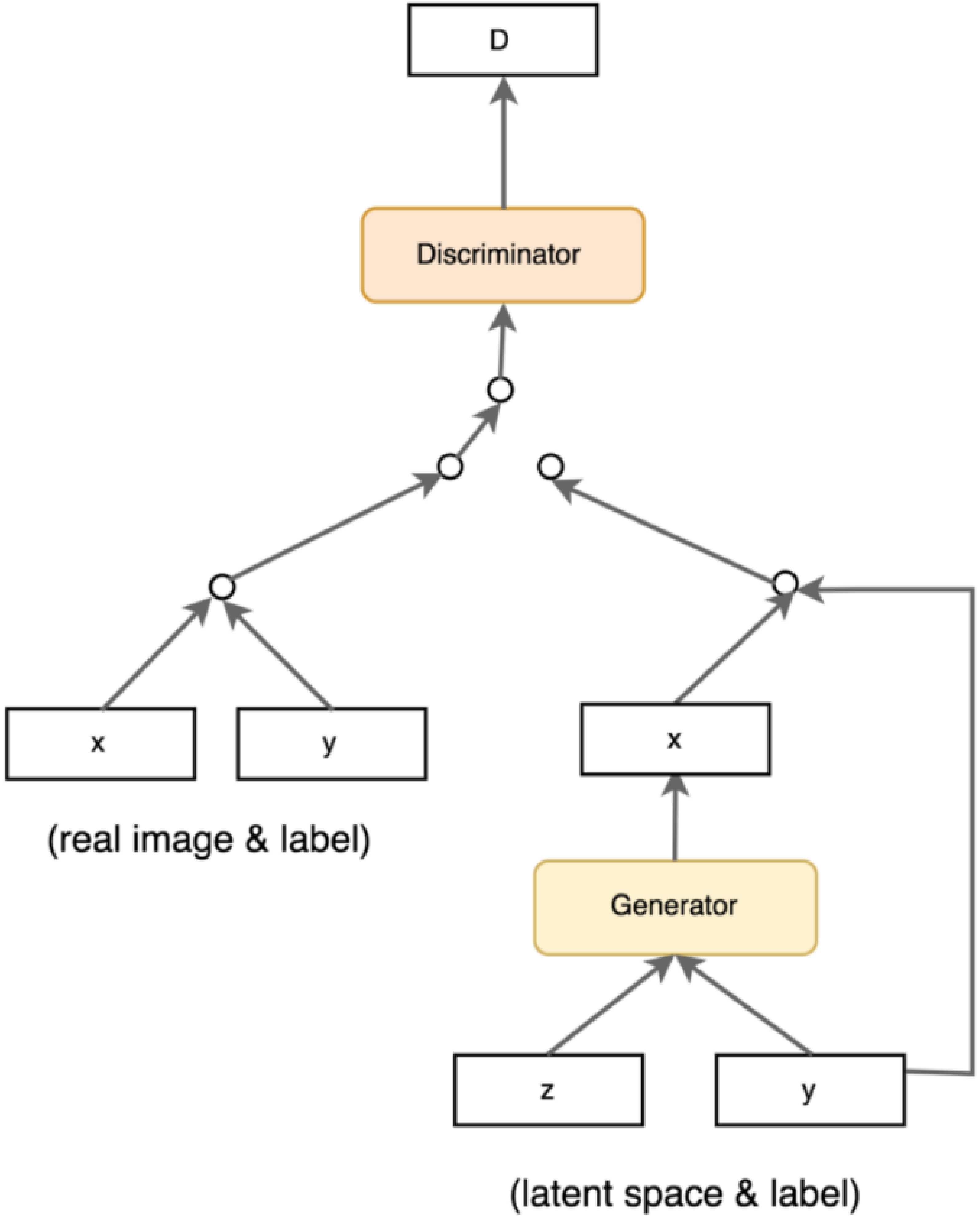
Figure 1: Demonstration of SIC methodology

Methodology

1 Group the nodules which are easy to classify, and which are hard to classify

2 **Reconstruct the hard nodules such a way that the models can be sort accurately.**

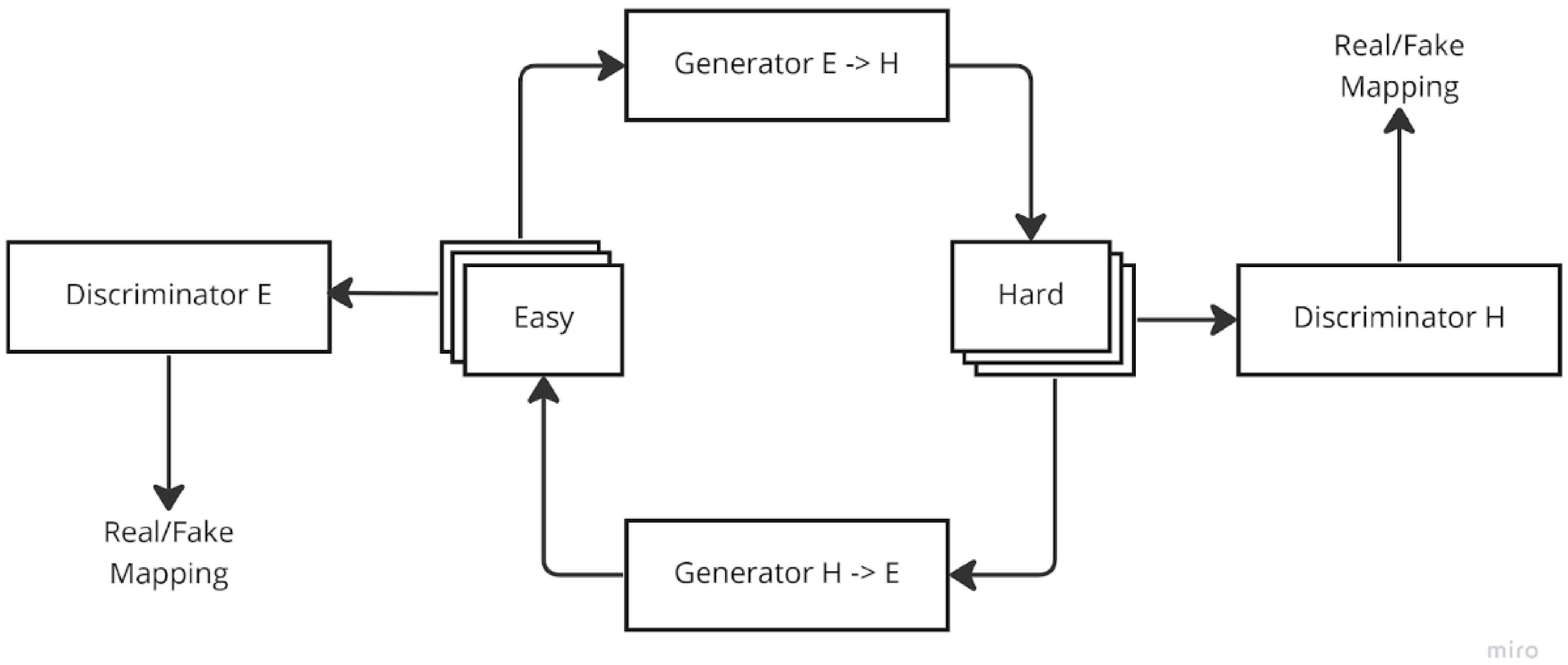
3 Classifying the nodule images.



Gan Background

- A Generative Adversarial Network (GAN) is a deep learning model that trains two neural networks, a generator and a discriminator, to generate new data samples that are indistinguishable from real data.
- The generator network takes in a random noise input and maps it to a synthetic data output through a series of operations in the network.
- A discriminator network is trained to distinguish between real and generated data and used in combination with a generator to improve the quality of generated data.
- Generally, GAN's work on labeled data.

Cycle-GAN



- Cycle-GAN is a GAN architecture for image-to-image translation tasks where the goal is to learn a mapping between two image domains. It uses two generators, one for mapping source images to target images and another for mapping target images back to source images, and two discriminators to ensure that the generated images are realistic.

- Cycle-GAN can be trained without the need for paired training data.

- We train the Cycle-GAN to convert Easy nodules to hard nodules.

Loss Functions

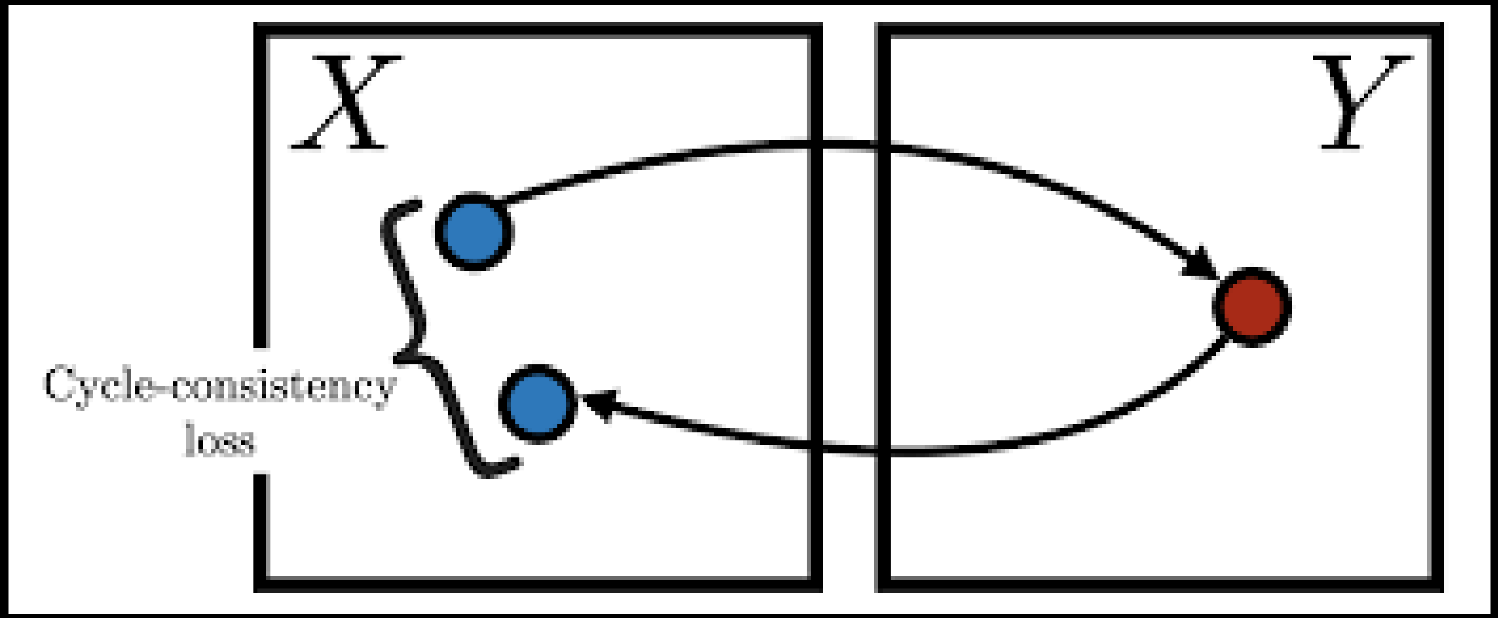
Aadversarial loss

- The adversarial loss encourages the generator network to produce images that are similar to the target domain, while the discriminator network tries to distinguish between the generated and real target images.
- This is achieved through a minimax game between the two networks, where the generator tries to minimize the loss while the discriminator tries to maximize it.
- The adversarial loss is typically calculated as the binary cross-entropy loss between the predictions of the discriminator network and the target labels.
- The target labels are either "real" or "fake", depending on whether the input to the discriminator is a real target image or a generated image.

Loss Functions

Cycle Consistency Loss

- The cycle consistency loss function is used to enforce the property that a source image should remain unchanged after being translated to the target domain and back to the source domain.
- It ensures that the generator network is able to maintain the content of the source image after translation.
- The cycle consistency loss is calculated as the mean absolute difference between the original source image and the reconstructed source image obtained by translating the generated target image back to the source domain using a second generator network.
- The second generator network is trained to translate target images back to source images.



Evaluation Metric

Fréchet Inception Distance

- FID is a commonly used metric for evaluating the quality of generated images in GANs.
- The FID score measures the difference between the statistics of real images and generated images in a feature space defined by a pre-trained convolutional neural network (often called the Inception network).
- The FID score is calculated as the Euclidean distance between the mean vectors and the covariance matrices of the feature representations of the real images and the generated images.
- The FID scores between the range 0-20 is often considered as good image reconstruction.

Methodology

1. Grouping the nodules which are easy to classify, and which are hard to classify.
2. Reconstruct the hard nodules such a way that the models can be sort accurately.
3. **Classifying the nodule images.**

Classification of Generated Images

- We use VGG19 for classification of the images. VGG19 is a deep neural network, consisting of 19 layers (16 convolutional layers and 3 fully connected layers).

- The network is trained on the ImageNet dataset, a large image classification dataset, and is capable of achieving high accuracy on various image classification tasks.

- VGG19 is widely used as a base model for transfer learning, where the pre-trained weights of the network are used as initialization for training on a different task.

Experiments

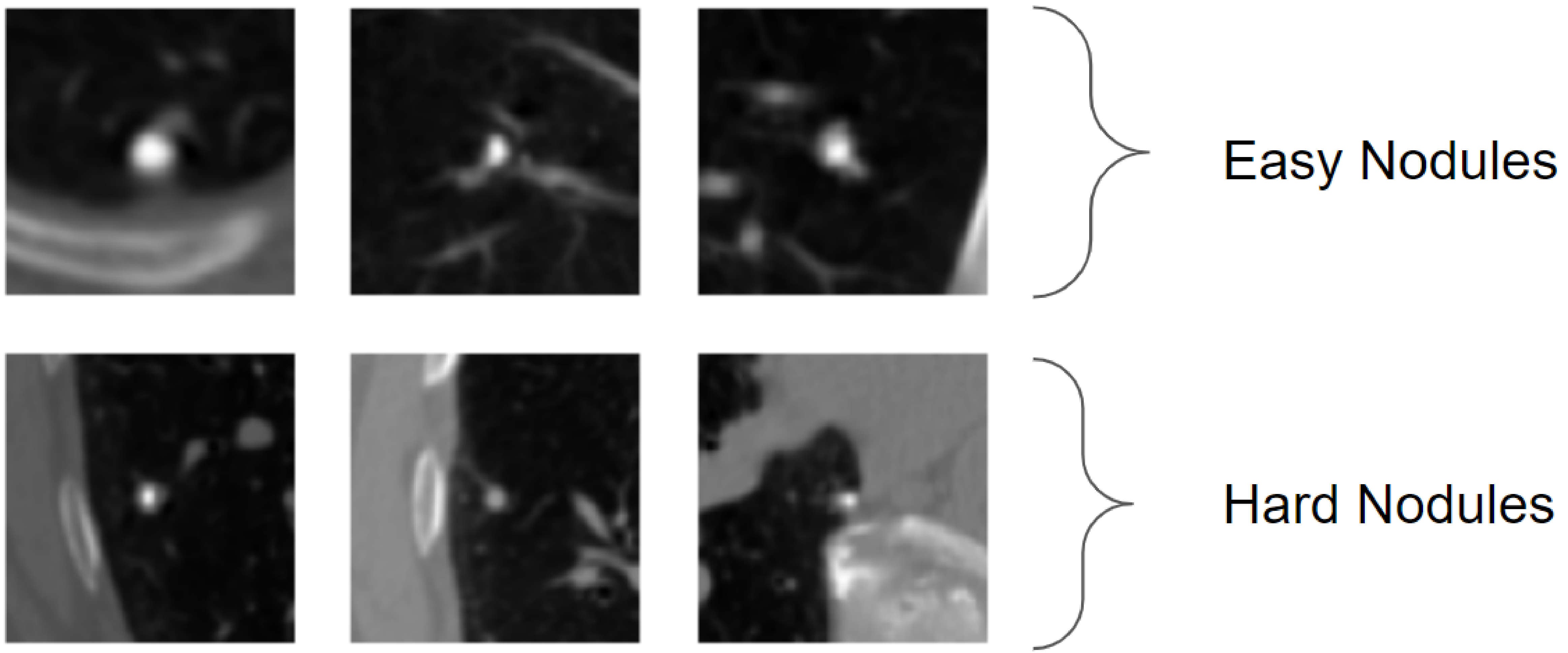
- Trying different image perturbation techniques before feeding the easy images to the Cycle-GAN.
- Experimenting to modify both the losses to optimize the Cycle-GAN.

Results

- At the initial stages when the easy images of LIDC were directly fed to the CycleGAN the FID score we got was about 4000.
- Through different experimentation and trial error of different techniques the FID score right now is about 200.
- CycleGAN is missing minute details in the reconstructed image.

Dataset

Total Number of Images - 1907
Number of Easy Nodules - 656
Number of Hard Nodules - 1251
Sample Images



Limitations

- Quality of generated images: The quality of the generated images is low when compared to supervised learning methods, as the model relies on the adversarial loss function, which will not always produce high-quality images.

- Difficulty in handling fine details: CycleGAN struggles with preserving fine details and textures in the generated images, especially in complex domains.

- Limited control over the output: In some cases, it is challenging to control the specific output style or attribute desired in the generated image, as the model relies on the adversarial loss to guide the image generation process.

Current Work

- Investigating ways to keep the details of the nodules by modifying the CycleGAN architecture.
- Implementing the modified version of CycleGAN “Self-Supervised CycleGAN for Object-Preserving Image-to-Image Domain Adaptation” published at ECCV2020 conference. (6th European Conference On Computer Vision)
- Object-Preserving GAN can over come the limitations of CycleGAN

References

- Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, “Generative adversarial nets,” in Advances in Neural Information Processing Systems 27, pp. 2672–2680. 2014.

- J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired Image-to-Image Translation using CycleConsistent Adversarial Networks,” ArXiv:1703.10593,Mar. 2017.

- Edward Xu, Thiruvarangan Ramaraj, Roselyne Tchoua, Jacob Furst, Daniela Raicu, “Contextualizing Lung Nodule Malignancy Predictions with Easy vs. Hard Image Classification”

Thank you!