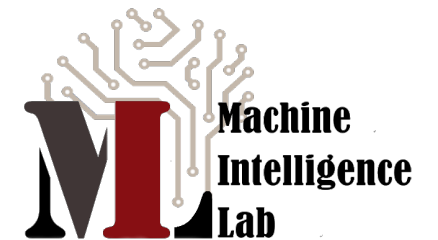




Assignment #2

[Due: Nov. 28, 2020]



TA Yoon Jee Seok (윤지석): wltjr1007@korea.ac.kr

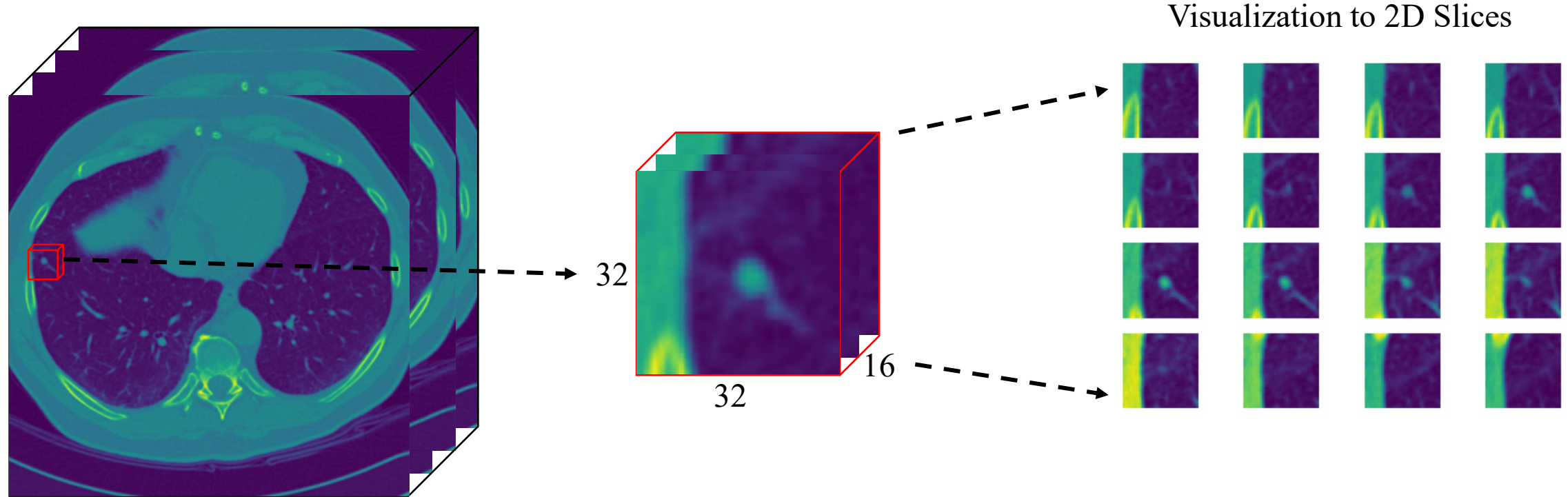
TA Kang Eun Song (강은송): eunsong1210@korea.ac.kr

Task Description

- Objective: Given a lung nodule CT image, classify benign vs. malignant
- Dataset: CT image patches from LUNA16 Challenge <https://luna16.grand-challenge.org/>
 - Input: 3,000 3D patches with a lung nodule in the center point
 - Size: $32cm \times 32cm \times 16cm$, Nodule Diameter: $8.31 \pm 4.76cm$
 - 3,000 patches from 888 subjects
 - 1,500 benign + 1,500 malignant patches
 - Split into 800 train subjects, 88 test subjects
 - Train: 1,400 benign + 1,400 malignant patches, Test: 100 benign + 100 malignant patches
 - Test data will not be available to students
 - Label (Boolean): Benign (False) vs. Malignant (True)

Dataset Example

[Malignant Nodule]



No preprocessing is performed, except for resampling voxels to $1\text{cm} \times 1\text{cm} \times 1\text{cm}$ spacing.
Perform your own preprocessing.

Download Data

https://koreaoffice-my.sharepoint.com/:f/g/personal/wltjr1007_korea_edu/EksffXy1lbxHqWK8yil3sZ8B4uZeFP7nTXiZ_1yo9VnCKQ?e=GpJBsj

How to load the training data:

```
import numpy as np  
dat = np.load("/content/trn_dat.npy") #Should contain (2800,32,32,16) float32 numpy array  
lbl = np.load("/content/trn_lbl.npy") #Should contain (2800,) boolean numpy array
```

Project Requirement

- Build the best model based on given train dataset “train_data.npy”
- Report scores for train, validation split
 - Randomly split the dataset into train (90%), validation (10%) split
- Analyze and discuss your models and results in markdown cells
- TAs will assess your best model with a test dataset, and grade your project accordingly
 - This test dataset will not be available to students

Implementation Requirement

- Use the Google Colab (<https://colab.research.google.com/>)
- You may use any library of your choice (e.g. Scikit-learn, Tensorflow, PyTorch, ...)
 - However, fully automated libraries are not allowed
- **(IMPORTANT) Submit two .ipynb files and model weight files**
 - In each of the .ipynb file, write your name, ID, and the link to your Colab project
 - First .ipynb file: “assignment2_project_yourID.ipynb”
 - This should contain your code, analysis, discussion etc.
 - Second .ipynb file: “assignment2_evaluation_yourID.ipynb”
 - This should contain code for loading, preprocessing, evaluation, etc.
 - Follow further instructions in “assignment2_evaluation_template.ipynb”
 - <https://colab.research.google.com/drive/1s2BsQw7DJaprT0qyJQuYHiEJ-KlEWMgQ>

Grading Factors

- Evaluation scores on test dataset
 - Test dataset will not be available to students
 - Scores will be ranked and used as a reference for grading
- Analysis, discussion etc.
 - Write in English
 - Be thorough and precise

TA's note: Beware of plagiarism (we use proprietary plagiarism checkers in addition to BlackBoard plagiarism checker). Ask the TAs if you have any questions.