Tumor recognition using CNN

Zexi Yu, zexi.yu@usask.ca

Overview

- **General Summary:**
  - Objective
  - Audience and benefit
  - General Structure
- **Data Wrangling**
- **Learning Model:**
  - CNN introduction
  - Structure used in this project
- **Result and Future works**

General Summary

- **Objective:**
  - Build a model that will detect whether a lung cancer occurred.
- **Audience and benefit:**
  - Audience: hospitals/governmental health/cancer departments
  - Benefit:
    - reduce the false positive rate that plagues the current detection technology
    - get patients earlier access to life-saving interventions
    - give radiologists more time to spend with their patients.

General Summary

- **General Structure:**
  - Data Wranglings
    - HU transformation
    - Morphology operations
  - Down-sampling
  - CNN

DW: DICOM Reading& HU transform

- **Hounsfield Unit / CT numbers:**
  - Image from wiki

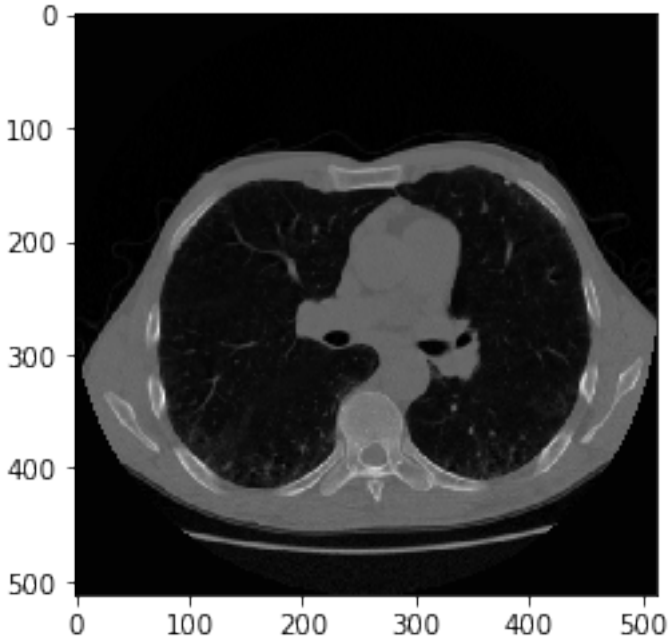
Substance		HU
Air		−1000
Fat		−120 to −90[2]
Soft Tissue, Contrast		+100 to +300
Bone		• +200 in craniofacial bone • +700 in cancellous bone • +3000 in cortical bone
Fluids	Chyle	−30[3]
	Water	0
	Urine	-5 to +15[2]
	Bile	-5 to +15[2]
	CSF	+15
	Blood	+13[4] to +50[5]
	Clotted blood	+50[6] to +75[4][6]
	Mucus	0[7] - 130[8] ("high attenuating" at over 70 HU)[9][10]
Parenchyma	Lung	-700 to −600[11]
	Kidney	+20 to +45[2]
	Liver	60 ± 6[12]
	Lymph nodes	+10 to +20[13]
	Muscle	+35 to +55[2]

DW: DICOM Reading& HU transform

- **DICOM Reading:**
  - Dicom.read_file
- **HU transform:**
  - Dicom.RescaleIntercept
  - Dicom.RescaleSlope

DW: Morphology Operations

- **Goal: Segmentation of nodes in lung**
- **Six Steps**
- **Original Image**

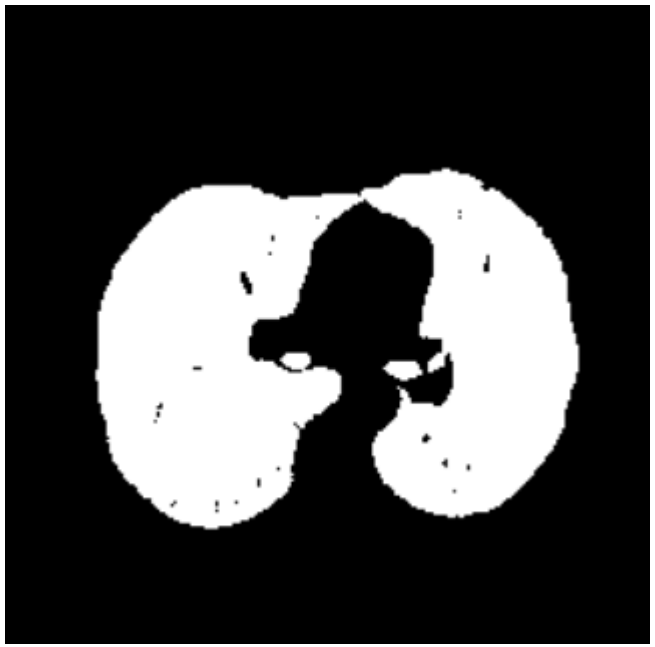


DW: Morphology Operations

- **Step 1: Convert the image into a binary one**

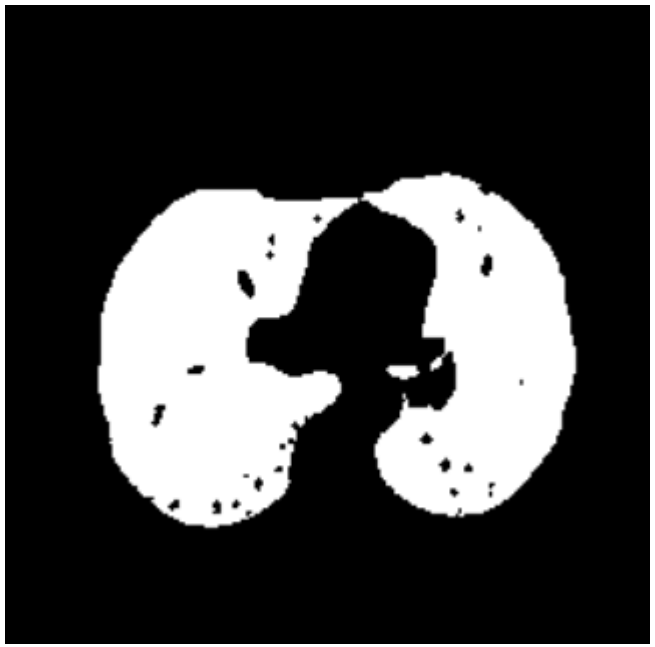
DW: Morphology Operations

- **Step 2: Remove the blobs connected to the border of the image**
- **Step 3: Label the image and keep 2 largest areas**



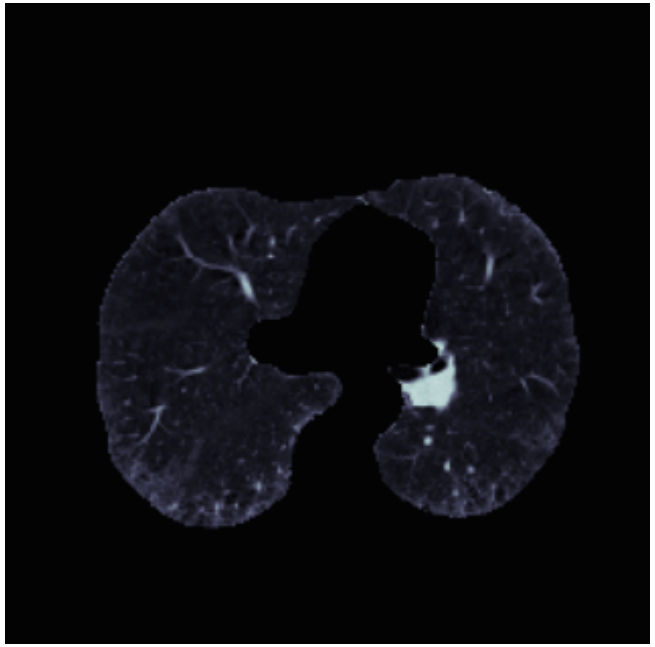
DW: Morphology Operations

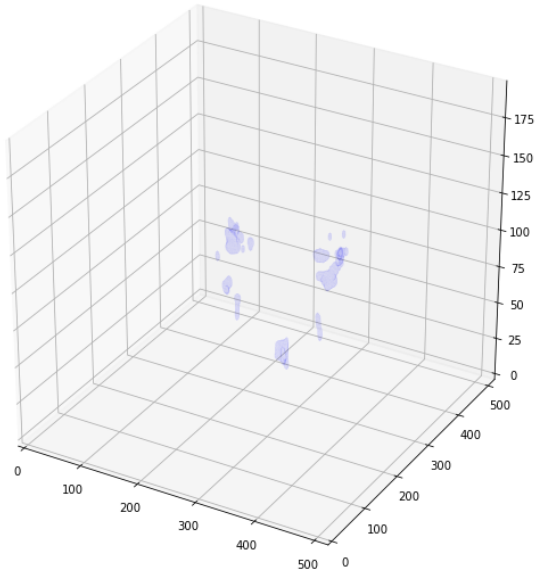
- **Step 4: Erosion operation with a disk of radius 2 – separate the lung nodules**
- **Step 5: Closure operation with a disk of radius 12 – keep nodules attached to lung wall**



DW: Morphology Operations

- **Step 6: Fill in the holes**
- **Superimpose**





DW: Down-sampling

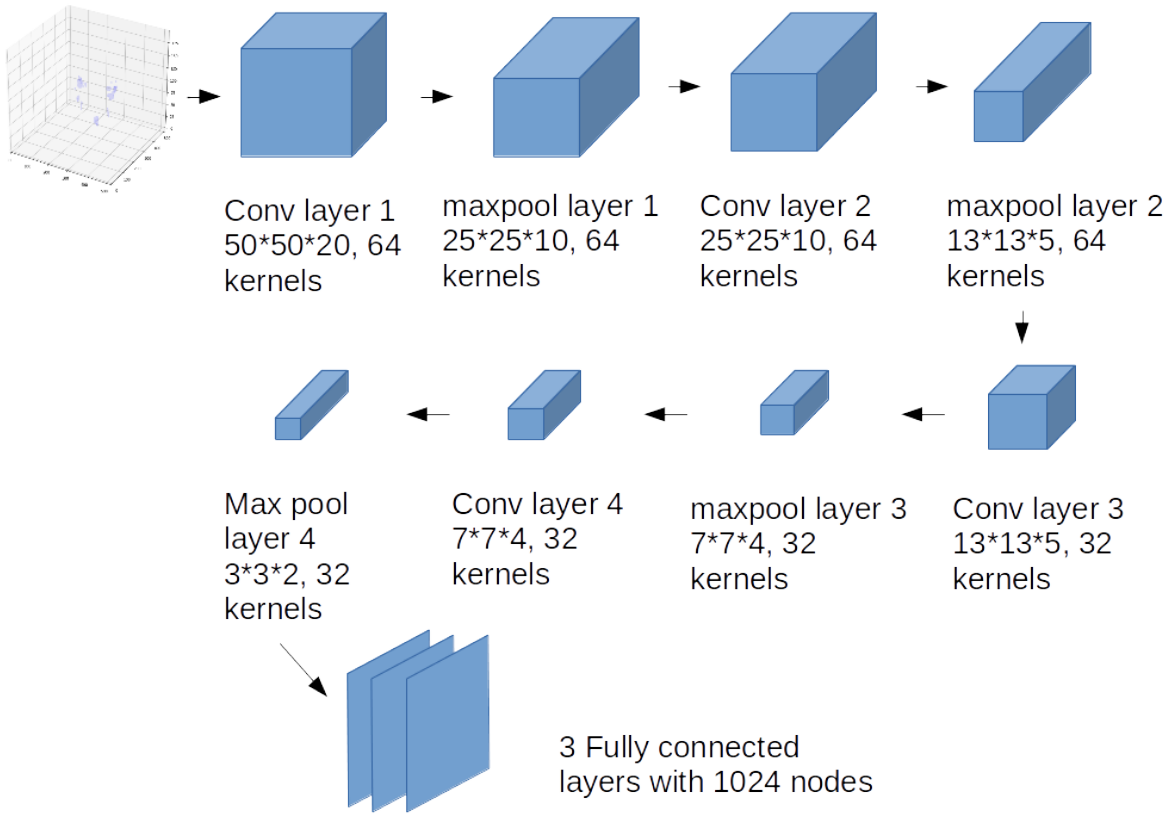
- **Sacrifices due to time and computation limit:**
  - No further node segment
  - Only 1/5 of training set is used
  - Down-sampling

Learning Method: CNN Intro

- **CNN - Convolutional Neural Networks**
  - Suitable for recognition task in Image processing
  - Use convolutional layer to identify image characteristics
  - Fully connected layer for classification
  - Maxpool layer and dropout layer to prevent overfitting

Learning Method: CNN structure

- **General Structure:**



Result

- **Kaggle score is log loss function. Small score is better**
- **Kaggle score 0.63856, beat the benchmark 0.69314**
- **Tend to predict the non-cancer, may have high false negative rate**

Future Work

- **Train the CNN using all the training set instead of 1/5 of them and see if the performance improves.**
- **Manually label the nodes in the training set, perform node segmentation and train and change CNN structure according to the single node images.**