**Kaggle X-ray Pneumonia detection**

**1. Registration**

The official website: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>

* Open the official website
* Register an account on Kaggle
* Join the competition by clicking some buttons on the official website (I just joined so I was able to see the exact button)
* Deadline for registration is Oct, 17th

The time line:

|  |  |
| --- | --- |
| 10/17/2018 | Entry deadline. You must accept the competition rules before this date in order to compete. |
| 10/17/2018 | Team Merger deadline. This is the last day participants may join or merge teams. |
| 10/24/2018 | Stage 1 ends & Model upload deadline\*. |
| 10/25/2018 | Stage 2 begins. New test set uploaded. |
| 10/31/2018 | Stage 2 ends & Final submission deadline. |
| 11/9/2018 | Solutions & Other Winners Obligations due from winners. |
| 11/25-30, 2018 | RSNA 2018 Conference in Chicago, IL |

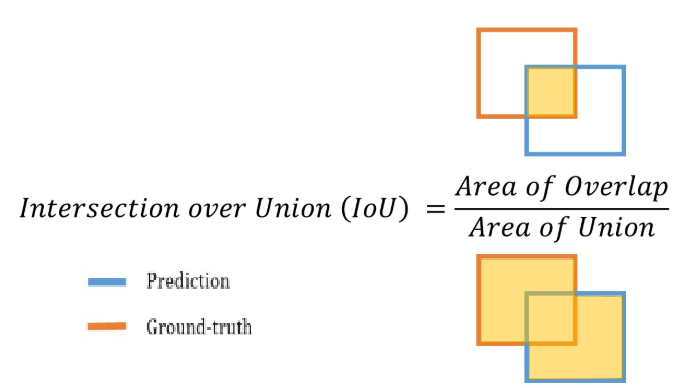
**2. Problem introduction (with kernels)**

The Goal:

To help detecting pneumonia via chest radiograph (CXR), improving the efficiency and reach of diagnostic services

Model Evaluation:

This competition is evaluated on the mean average precision value at different intersection over union (IoU) thresholds.



The IoU threshold value (t) will have range from 0~1, but they only evaluate from 0.4 to 0.75 with a step size of 0.05: (0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75)

At each threshold value t, a precision value (PV) is calculated:

|  |  |
| --- | --- |
| PV(t)= |  |

The average precision of a single image:



The given csv files:

1.stage\_1\_train\_labels.csv:

the training set. Contains patientIds and bounding box / target information

2.stage\_1\_sample\_submission.csv:

a sample submission file in the correct format.

3.stage\_1\_detailed\_class\_info.csv:

provides detailed information about the type of positive or negative class for each image.

The given image files:

1.stage\_1\_test\_images (1000 images)

2.stage\_1\_train\_images (25684 images)

**3. Challenges we need to tackle**

Feels like it’s better to use CNN to train the dataset and apply it to the testing dataset. So we need to learn how to use CNN and maybe some other method for image analysis.

In addition, we need to understand some domain knowledge about CXR and pneumonia because we might need to formatting the image file if necessary.

**4. Initial plan (what API/method/feature engineering, etc. to do)**

I would recommend looking at all the intro files made by current research team. They’ve given some exploratory data analysis as well as some ai work. Below is a summary of the work they’ve done:

|  |  |  |
| --- | --- | --- |
| **Work** | **Link** | **Comments** |
| Exploratory Data Analysis | <https://www.kaggle.com/peterchang77/exploratory-data-analysis> | Very useful to see the overall data structure |
| Lesson 1. Classification of chest vs. adominal X-rays | <https://github.com/mdai/ml-lessons/blob/master/lesson1-xray-images-classification.ipynb> | not very useful for the current competition |
| Lesson 2. Lung X-Rays Semantic Segmentation | <https://github.com/mdai/ml-lessons/blob/master/lesson2-lung-xrays-segmentation.ipynb> | useful if we want to analyze the lung image only instead of the entire image |
| Lesson 3. RSNA Pneumonia Detection Challenge (Kaggle API) | <https://github.com/mdai/ml-lessons/blob/master/lesson3-rsna-pneumonia-detection-kaggle.ipynb> | Very useful to understand how to use Kaggle api and Mask-RCNN |
| Lesson 3. RSNA Pneumonia Detection Challenge (MD.ai API) | <https://github.com/mdai/ml-lessons/blob/master/lesson3-rsna-pneumonia-detection-mdai-client-lib.ipynb> | similar to the previous Lesson 3. |

**5. Discussion**

Domain knowledge discussion about X-ray image, lung opacity, and pneumonia:

<https://www.kaggle.com/zahaviguy/what-are-lung-opacities>

bounding box discussion:

<https://www.kaggle.com/jtlowery/intro-eda-with-dicom-metadata>

CNN segmentation:

<https://www.kaggle.com/jonnedtc/cnn-segmentation-connected-components>

some interesting stats from the dataset:

<https://www.kaggle.com/thomasjpfan/q-a-with-only-pictures>

Lung Opacity Overview:

<https://www.kaggle.com/kmader/lung-opacity-overview>