

CheXNet

**BBox\_List\_2017:** Contains the bounding box for some images that indicates where the disease is. This information was not used when training.

**Images:** Contains different folders that contains the Images of the dataset in different resolutions.

- images: Contains the images in the original size which is 1024x1024
- images\_jpg\_256: Contains the images in a resolution of 256x256
- images\_jpg\_400: Contains the images in a resolution of 400x400

The script named batc\_ddownload\_zips.py can be used to download the dataset again and the file named data.zip contains the Images compressed.

**Data\_Entry\_2017\_v2020.csv:** Contains more extensive information of each Image, being this information:

- Image Index
- Finding Labels
- Follow-up
- Patient ID
- Patient Age
- Patient Gender
- View Position
- OriginalImage[Width,Height]
- OriginalImagePixelSpacing[x,y]

**train\_list.txt / val\_list.txt / test\_list.txt:** Contains for each Image the diseases that it has present (there are some Images that do not contain any disease). This information is given with the name of the file followed by 14 numbers that can be zero or one, indicating, in order, the presence of the following diseases:

- Atelectasis
- Cardiomegaly
- Effusion
- Infiltration
- Mass
- Nodule
- Pneumonia
- Pneumothorax
- Consolidation
- Edema
- Emphysema
- Fibrosis
- Pleural\_Thickening
- Hernia