Summer Review 7

DeepLesion and Chest X-ray NIH Dataset (shared thhrough Box)

Blog Post Link
Deep Lesion Paper Link

Chest X ray Dataset Paper Link

Reviewed by : Arshdeep Sekhon

[1]Department of Computer Science, University of Virginia
https://qdata.github.io/deep2Read/

Introduction : CT

- CT images: computed tomography scan
- computer-processed combinations of many X-ray measurements taken from different angles to produce cross-sectional (tomographic) images (virtual "slices") of specific areas of a scanned object, allowing the user to see inside the object without cutting.
- Digital geometry processing is used to further generate a three-dimensional volume of the inside of the object from a large series of two-dimensional radiographic images taken around a single axis of rotation.
- many types of CT: X-ray CT, PET, PST
- MRI is better than CT
- CT used for: head, lungs, cardiac, angiography
- more detailed than conventional x-rays
- slices

Lesion

- Damage/Abnormal tissue of brain tissue
- large to small, single to multiple, harmless to dangerous
- many types: 120 different named brain tumors alone
- typically cause unknown
- symptom: type, location and size

CT and Lesion

CT scans can be helpful in diagnosing some types of brain tumors, especially those near or involving bone. They can also show swelling, bleeding, and bone and tissue calcification.

Label Collection

- collected using hospital picture archiving and communication systems (PACS)
- stored CT scans with markings by radiologists about size, location

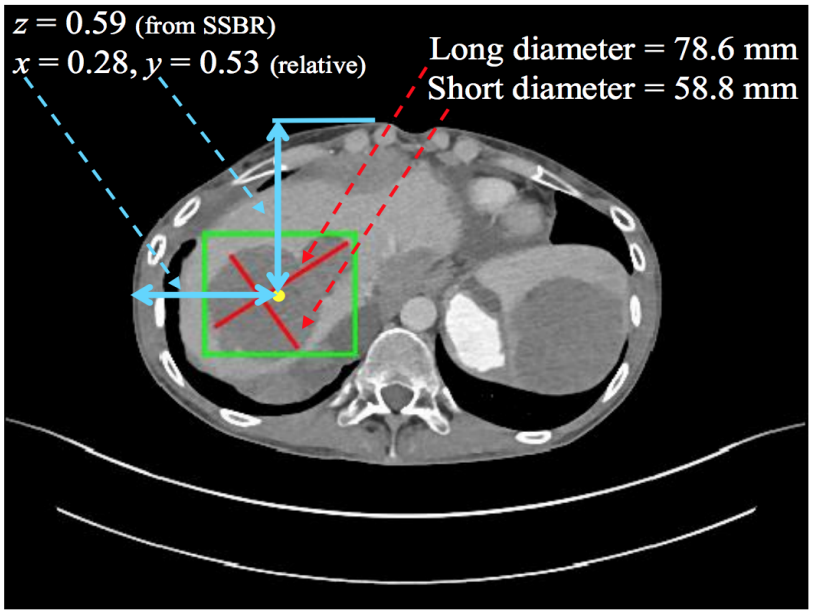

Figure 8. *Location and size of a sample lesion. The red lines are the long and short diameters annotated by radiologists during their daily work. The green box is the bounding box calculated from the diameters. The yellow dot is the center of the bounding box. The blue lines indicate the relative x- and y-coordinates of the lesion. The z-coordinate is predicted by SSBR. Best viewed in color.*

-

Label Collection

- $32,120$ axial CT slices of size 512×512
- from $10,594$ studies of 4,459 unique patients.
- $1-2$ lesions ine ach image
- generate a box tightly around the two diameters and add a 5-pixel padding in each direction to capture the lesions full spatial extent.
- Take 3 nearest slices : 3 channel image

Label Collection

- Annotated data: size and x, y location
- approximate z location
- size and location determines type of lesion
- Sample a small subset of data (2315)
- Mark lesions broadly into categories : Coarse labels: lung abdomen, etc
- Use labels to predict the rest: pseudao labels

Types of tasks

- Metric Learning : If the same type of lesion in other patients
- If the same type of lesion within patients

Chest X Ray data

- short-list eight common thoracic pathology keywords : Atelectasis, Cardiomegaly, Effusion, Infiltration, Mass, Nodule, Pneumonia and Pneumathorax
- $108,948$ frontal-view X-ray images (from $32,717$ patients) and each image is labeled with one or multiple pathology keywords or Normal otherwise.
- typical $= 3000 \times 2000$ pixels
- resized as 1024×1024 bitmap images