NIH Clinical Center releases dataset of 32,000 CT images
데이터 소개

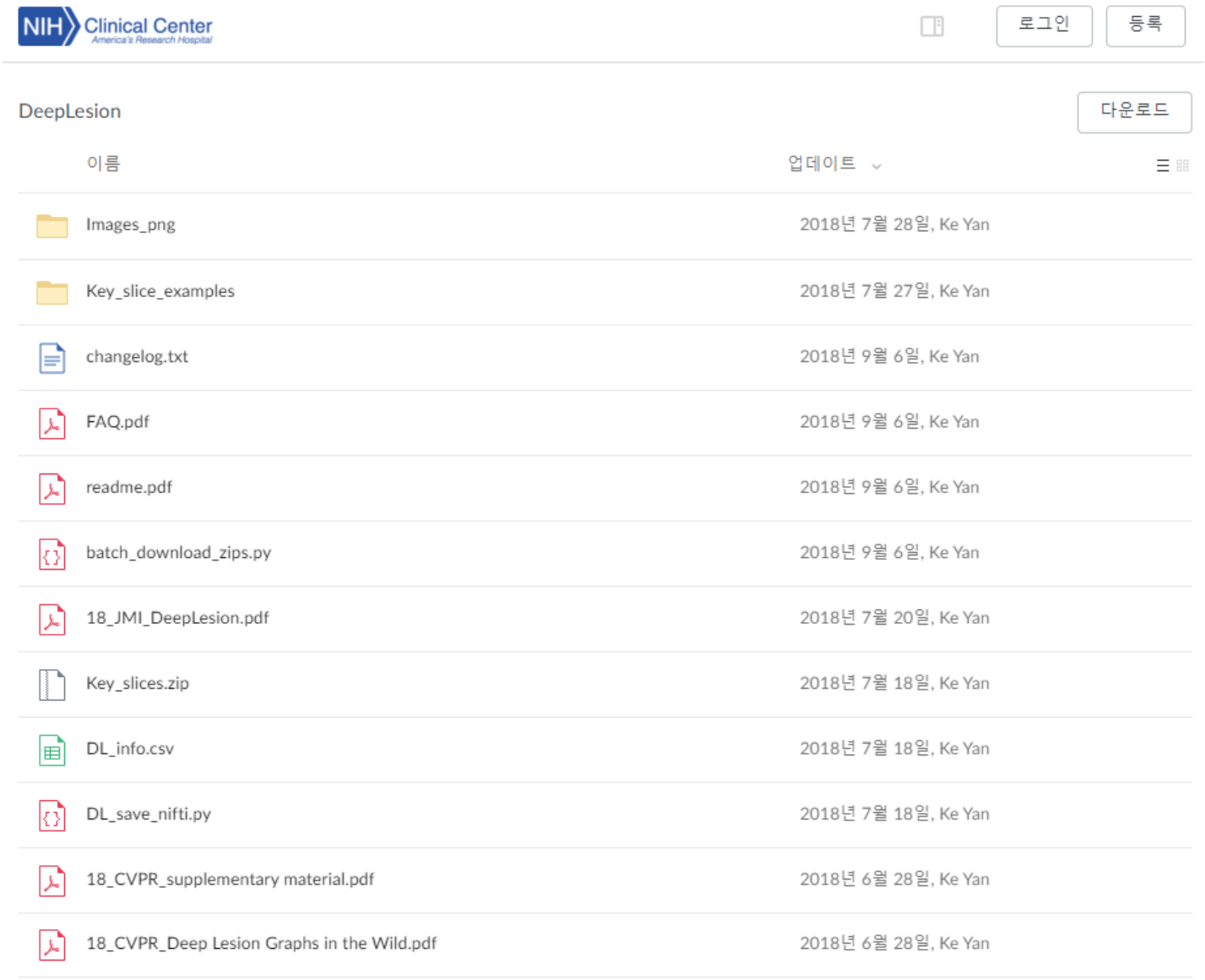
https://nihcc.app.box.com/v/DeepLesion

https://nihcc.app.box.com/v/DeepLesion

CT 이미지가 저장된 디렉토리

CT 이미지에 대한 설명

CT 이미지 다운로드 스크립트



Introduction

- DeepLesion 데이터은 4,427명의 환자로부터 10,594번 CT 스캔 해서 얻은 32,120개의 CT slice를 포함.
- 이미지마다 크기측정과 경계영역을 표시한 1~3개의 병변 (lesions)이 있음.
- 파일명 구조 : {patient index}_{study index}_{series index}_{slice index}.png
- Key_slice.zip : 검토 목적으로 변변에 대한 주석이 있는 key slice
- DL_info.csv : 주석과 메타데이터
- DL_save_nifti.py: 2D 16-bit 이미지에서 3D nifty sub-volumne 으로 변환하는 파이션 코드

* **병변은** 유기체, 일반적으로 인한의 조직 손상이나 비정상적인 변화 질병 이나 외상

Annotations 1/4

- DL_info.csv 파일이 컬럼 정보

1. File name : underscore(_)을 / or ₩로 치환해야 함.
2. Patient index : 1 부터 시작
3. Study index : 각각의 환자마다 1 ~ 26개의 연구가 있음.
4. Series ID
5. Key_slice_index : lesion annotation 의 번호
6. Measurement_coordinates : 병변의 2개 RECIST 직경을 나타내는 8D vector. 처음 4개의 좌표값이 장축

*** 고형 종양 (RECIST)의 반응 평가 기준**

Annotations 2/4

7. Bounding_boxes : 병변의 경계박스(4개의 값)
8. Lesion_diameters_Pixel : 긴축과 짧은 축의 길이 (2개의 값)
9. Normalized_lesion_location : 병변 중심의 상대적인 body 위치
10. Coarse_lesion_type : 병변의 종류, 1 ~ 8으로 뼈, 복부, 종격동, 간, 폐, 신장, 연조직 및 골반과 대응.
11. Possibly_noisy : 병변의 주석이 노이즈가 있으면 1로 설정됨.
12. Slice_range : 데이터셋에서 병변이 포함된 slice의 간격

    예를 들어 첫 번째 병변에서 키 슬라이스는 109이고 슬라이스 범위는 103 ~ 115이며 슬라이스 103 ~ 115가 제공됩니다.

Annotations 3/4

13. Spacing_mm_px_ : x, y, z축에서의 간격(mm per pixel), 두 slice의 물리적인 거리를 나타냄.
14. Image_size
15. DICOM_windows : DICOM file에서 추출한 windowing (min~max)
16. Patient_gender : 환자의 성별, F or M
17. Patient_age : 나이
18. Train_Val_Test : 공식적으로 무작위로 생성된 환자 수준의 데이터 분할, train=1, validation=2, test=3

Annotations 4/4

	A	B	C	D	E	F	G
1	File_name	Patient_index	Study_index	Series_ID	Key_slice_index	Measurement_coordinates	Bounding_boxes
2	000001_01_01_109.png	1	1	1	109	233.537, 95.0204, 234.057, 106.977, 231.169, 101.605, 236.252, 101.143	226.169, 90.0204, 241.252, 111.977
3	000001_02_01_014.png	1	2	1	14	224.826, 289.296, 224.016, 305.294, 222.396, 297.194, 228.978, 297.903	217.396, 284.296, 233.978, 310.294
4	000001_02_01_017.png	1	2	1	17	272.323, 320.763, 246.522, 263.371, 234.412, 305.494, 280.221, 288.118	229.412, 258.371, 285.221, 325.763
5	000001_03_01_088.png	1	3	1	88	257.759, 157.618, 260.018, 133.524, 251.735, 145.571, 265.288, 146.841	246.735, 128.524, 270.288, 162.618

H	I	J	K	L	M
Lesion_diameters_Pixel_	Normalized_lesion_location	Coarse_lesion_type	Possibly_noisy	Slice_range	Spacing_mm_px_
11.9677, 5.10387	0.44666, 0.283794, 0.434454	3	0	103, 115	0.488281, 0.488281, 5
16.019, 6.61971	0.431015, 0.485238, 0.340745	3	0	8, 23	0.314453, 0.314453, 5
62.9245, 48.9929	0.492691, 0.503106, 0.351754	3	0	8, 23	0.314453, 0.314453, 5

N	O	P	Q	R
Image_size	DICOM_windows	Patient_gender	Patient_age	Train_Val_Test
512, 512	-175, 275	F	62	3
512, 512	-175, 275	F	72	3
512, 512	-175, 275	F	72	3

Application

- DeepLesion은 다양한 유형의 병변을 포함하는 대규모 데이터 세트입니다.
- 병변 탐지, 분류, 분할, 검색, 측정, 성장 분석, 다른 병변 간의 관계 마이닝 등에 사용할 수 있습니다.

Limitation

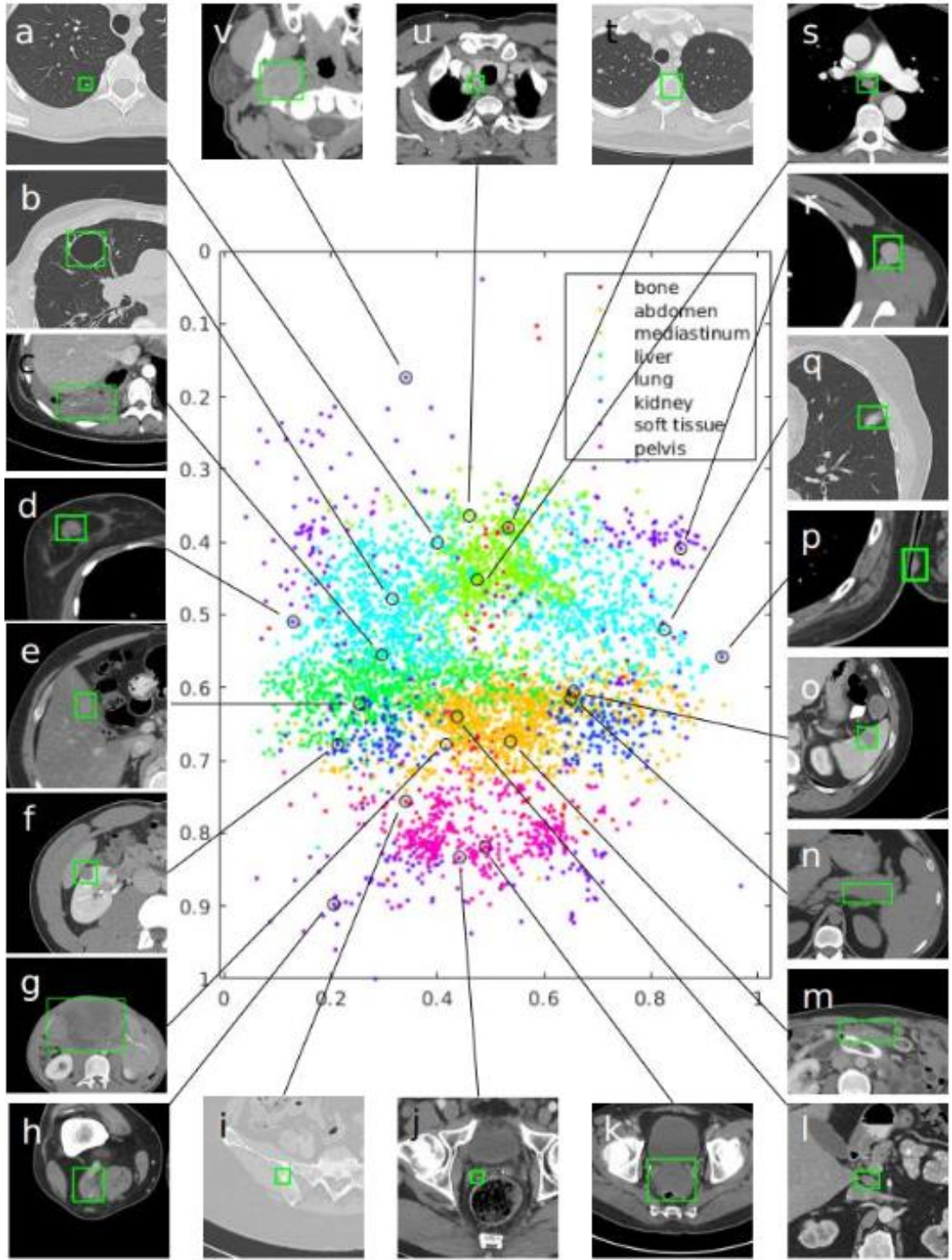
- **2D 직경 측정 및 병변의 경계 상자** 만 포함되어 병변 분할 마스크, 3D 경계 상자 또는 세분화 된 병변 유형이 없습니다. 따라서 일부 응용 프로그램 (예 : 병변 세그먼트)에는 추가 수동 주석이 필요
- **모든 병변이 이미지에 주석이 달린 것은 아닙니다**. 방사선과 의사는 일반적으로 각 연구에서 대표적인 병변만을 표시합니다. 따라서 일부 병변에는 주석이 없습니다.
- 수작업 검사에 따르면 대부분의 책갈피는 비정상적인 발견이나 병변을 나타내지만 북마크의 **작은 부분**은 실제로 정상 크기의 림프절과 같은 정상적인 구조를 측정 한 것입니다.

Data visualization

DeepLesion 데이터 세트의 하위 집합 (15 %) 시각화.

산포도지도의 x 및 y 축은 각각 각 병변의 상대적 body 위치의 x 및 z 좌표에 해당합니다.

따라서,이 맵은 인체의 정면도와 유사합니다



Supplementary Material

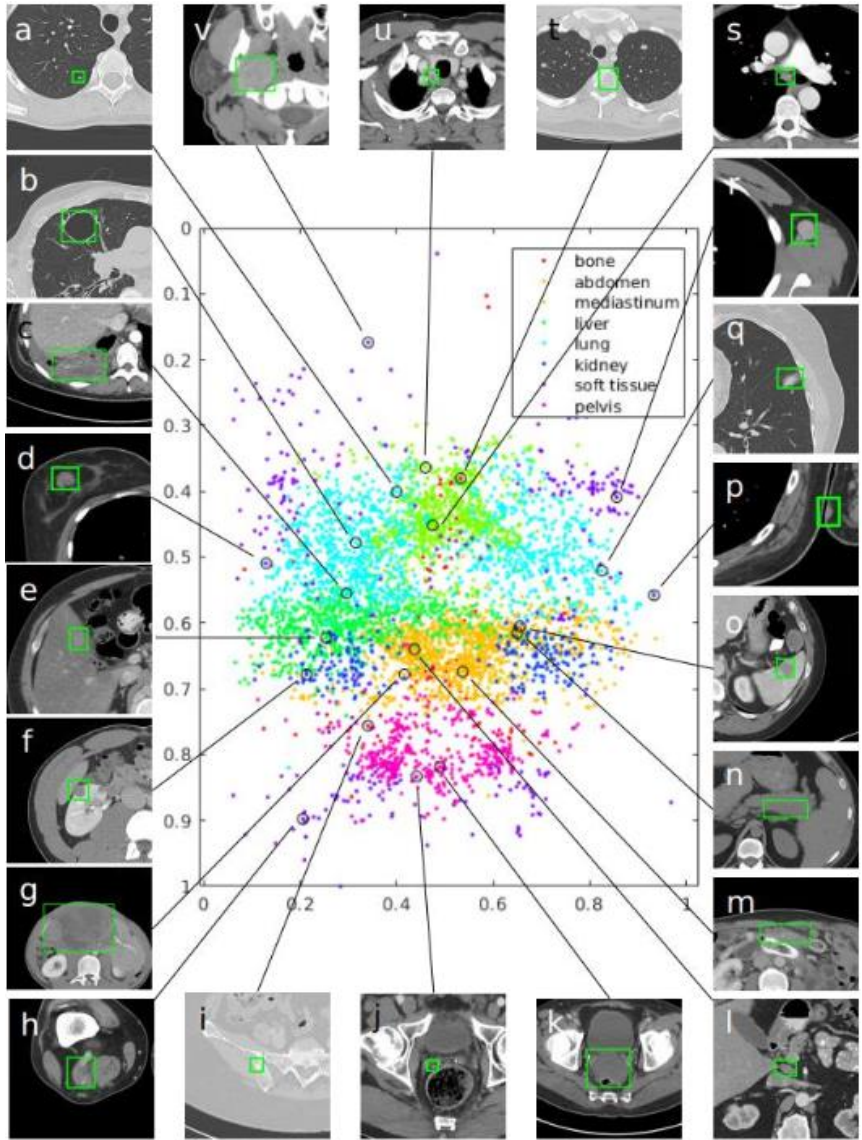
https://nihcc.app.box.com/v/DeepLesion

추가적인 설명 제공

1. DeepLesion 데이터 세트를 시각화하고 몇 가지 세부 사항을 설명
2. Self-supervised body-part regressor의 구현 상세 정보를 제공
3. 내용 기반 병변 조회 결과 보기
4. 환자 내 병변 정합 작업과 환자 내 병변 그래프를 설명

1. DeepLesion Dataset: Visualization and Details

- DeepLesion 데이터 세트의 개요를 제공하기 위해 그림 3에서 유형의 분포와 병변의 상대 Body 위치를 보여주는 scatter 맵을 그림.
- 병변의 상대적인 위치를 나타내는 self-supervised body-part regressor가 실제 위치와 일치하는 것을 알 수 있음.
- 뼈와 연조직 같은 일부 병변 유형이 많은 차이가 있음.
- 폐 / 종격 및 복부 / 간 / 신장과 같은 인접 유형은 개체 간 변동으로 인해 위치가 크게 중복됩니다.



- 그림 1은 병변의 위치와 크기를 얻기 위한 접근법을 보여줍니다.

- 몸체에 병변을 위치시키기 위해 먼저 축 방향 슬라이스에서 몸체의 마스크를 얻은 다음 x 좌표를 얻기 위해 병변 중심의 상대 위치 (0-1)를 계산

- z에 관해서는 자기 감독 된 신체 부분 회귀 분석기 (SSBR)가 사용된다.

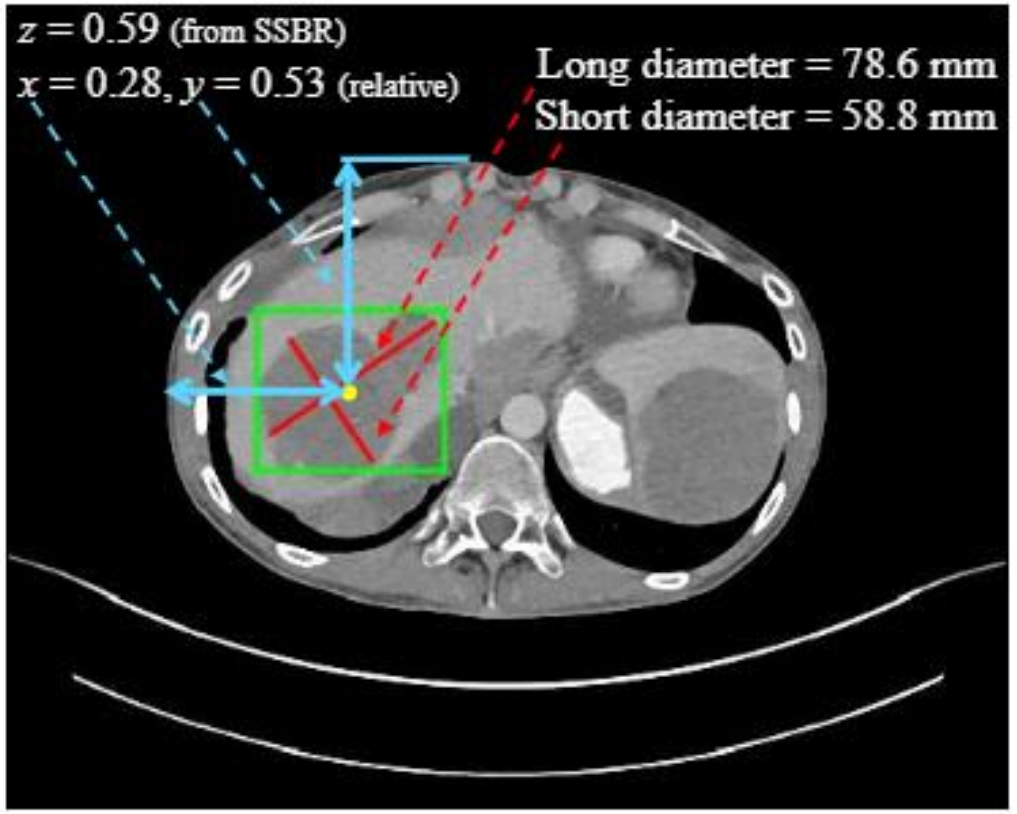


Figure 1. Location and size of a sample lesion. The red lines are the long and short diameters annotated by radiologists during their daily work. The green box is the bounding box calculated from the diameters. The yellow dot is the center of the bounding box. The blue lines indicate the relative x- and y-coordinates of the lesion. The z-coordinate is predicted by SSBR. Best viewed in color.

- 그림 2에서 병변 크기의 분포를 보여준다.

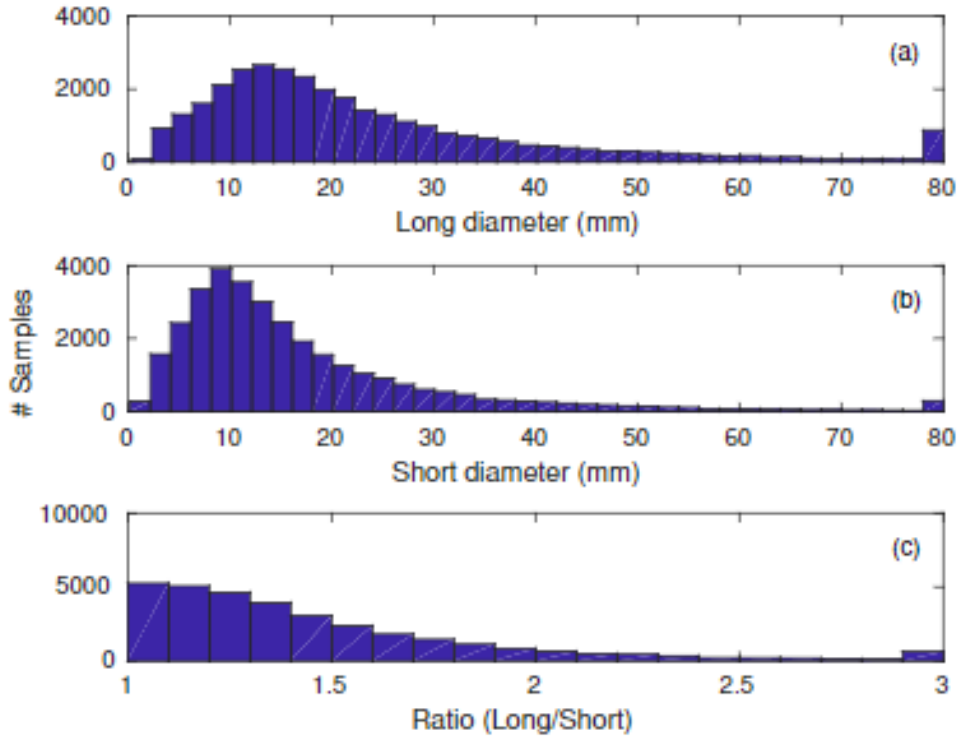


Figure 2. Distribution of the lesion-sizes in DeepLesion. For clarity, values greater than the upper bound of the x-axis of each plot are grouped in the last bin of each histogram.

2. Self-Supervised Body-Part Regressor: Implementation Details

- 42명의 환자를 대상으로 무작위로 800건의 라벨이 없는 CT를 추출
- 슬라이스의 크기 : 128 x 128
- 추가 전처리와 data augment(데이터 늘리기) 없음.
- 미니 배치에서 32volum에서 256개의 slice을 랜덤 선택
- 0.002의 learing rating로 SGD을 사용해서 학습
- 1,500번 반복하면 수렴함.

- 그림 3의 샘플 별변은 학습 된 슬라이스 점수 또는 z 좌표를 정 성적으로 평가하는 데 사용할 수 있습니다.

- SSBR을 정량적으로 평가하기위한 예비 실험 => 새로운 140명의 피험자에서 260 volume으로부터 18,195의 slice을 포함하는 테스트셋을 수집

- 흉부 chest (5903 조각), 복부 abdomen (6744) 또는 골반 pelvis (5548)의 3 가지 등급 중 하나로 분류

- 복부 계급은 간 위 경계에서 시작하여 장골의 상층부에서 끝납니다.
- 분류 정확도는 95.99 %이며 모든 전이 영역 (가슴 - 복부, 복부 - 골반)에서 부분적으로 애매함으로 인해 분류 오류

- SSBR의 효과를 입증되었고, 모든 병원 데이터베이스에 풍부한 레이블이 없는 볼륨에 대해 교육을 받았기 때문에 주석 처리 노력이 필요하지 않습니다.

3. Content-based Lesion Retrieval: More Results

- 병변 검색의 더 많은 예가 그림 4에 나와 있습니다.
- 모든 병변 유형의 전형적인 예를 보여주고 마지막 단계는 실패 사례.
- 검색된 대부분의 병변은 유형, 위치 및 크기가 쿼리와 비슷합니다.
- 더 중요한 것은 대부분의 검색된 병변과 질의문은 트레이닝 라벨에 지정되지 않은 준 신체적으로 유사한 신체 구조에서 비롯된 것입니다.
- 그림 4의 실패 사례는 쿼리와 다른 유형 => 주로 위치, 크기 및 모양이 쿼리와 비슷하기 때문에 검색

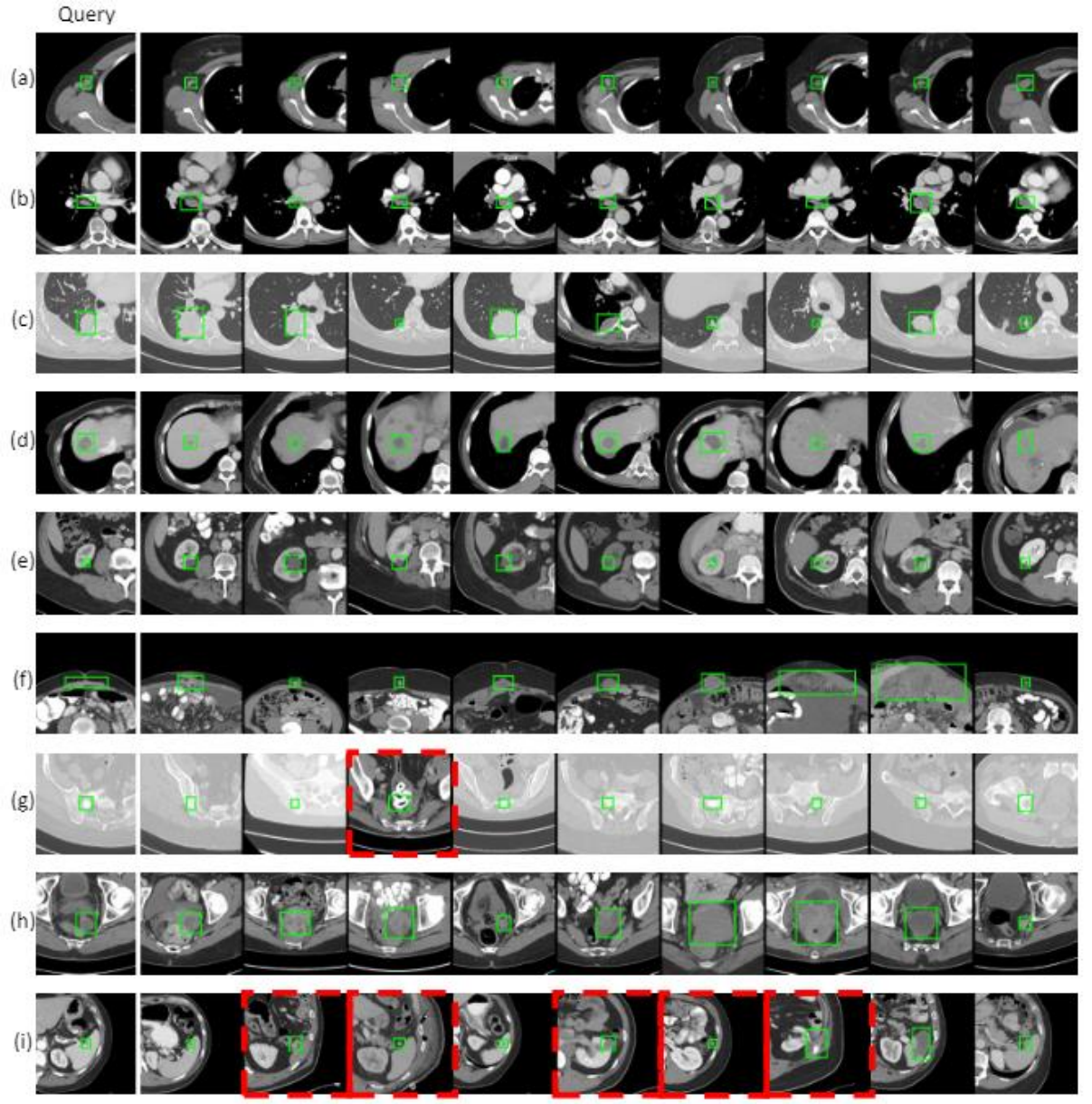


Figure 4. More examples of query lesions (first column) and the top-9 retrieved lesions on the test set of DeepLesion. We constrain that the query and all retrieved lesions must come from different patients. Red dashed boxes indicate incorrect results. The lesions in each row are: (a) Right axillary lymph nodes; (b) subcarinal lymph nodes; (c) lung masses or nodules near the pleura; (d) liver lesions near the liver dome; (e) right kidney lesions; (f) lesions near the anterior abdomen wall; (g) lesions on pelvic bones except the one in the red box, which is a peripherally calcified mass. (h) inferior pelvic lesions; (i) spleen lesions except the ones in red boxes.

4. Intra-Patient Lesion Matching: An Example

- 병변 정합 작업을 직관적으로 보여주기 위해 그림 5에서 표본 환자의 병변을 보여 주며, 그림 6의 표식 그래프와 그림 7의 최종 추출 병변 시퀀스를 보여줍니다.
- 우리는 병변 그래프와 Algo를 보여줍니다. 논문의 1은 다중성에서 병변을 정확하게 일치시키는 데 사용될 수 있습니다

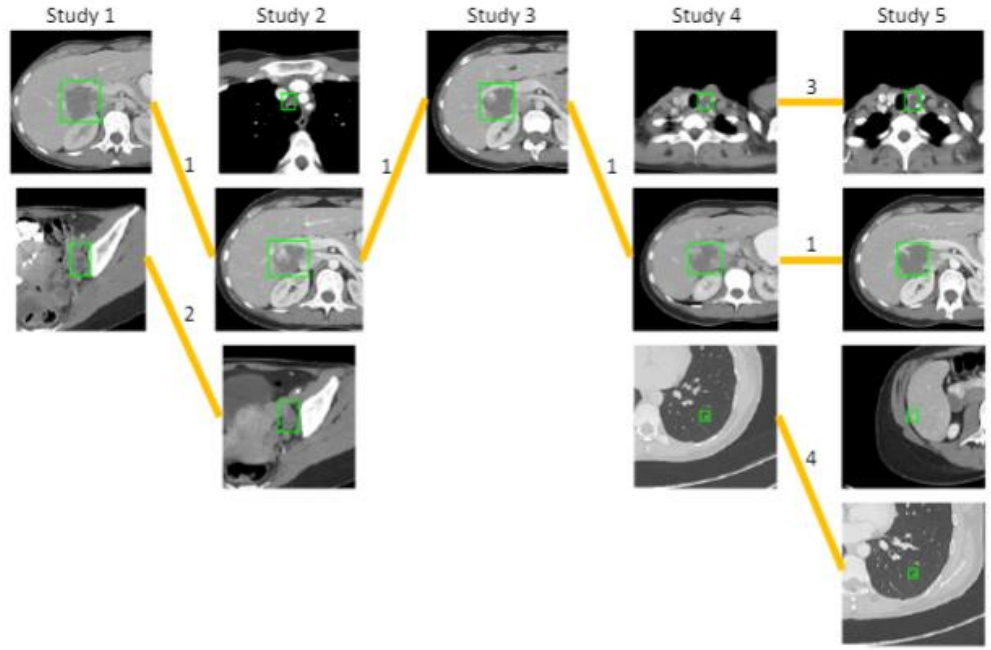


Figure 5. All lesions of a sample patient in DeepLesion. Lesions in each study (CT examination) are listed in a column. Not all lesions occur in each study, because the scan ranges of each study vary and radiologists only mark a few target lesions. We group the same lesion instances to sequences. Four sequences are found and marked in the figure, where the numbers on the connections represent the lesion IDs.

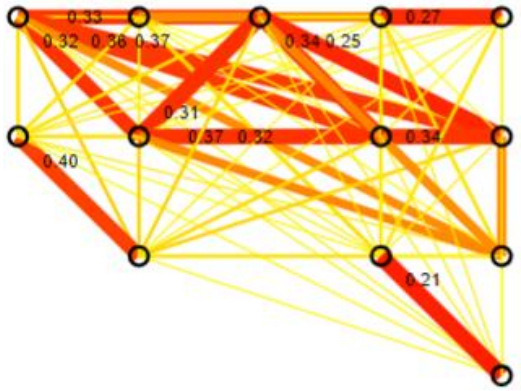


Figure 6. The intra-patient lesion graph of the patient in Fig. 5. For clarity, the lesions in Fig. 5 are replaced by nodes in this figure. The numbers on the edges are the Euclidean distances between nodes. We only show small distances in the figure. Red, thick edges indicate smaller distances. Note that some edges may overlap with other edges or nodes.

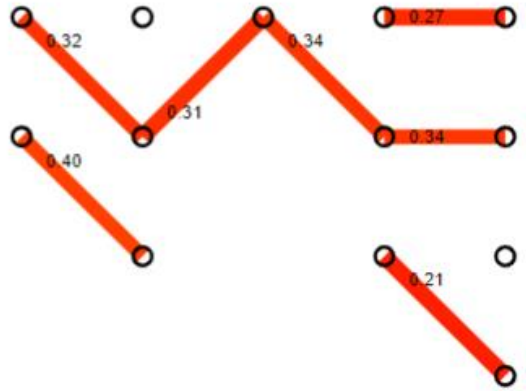


Figure 7. The final lesion sequences found by processing the lesion graph in Fig. 6 using Algo. 1 in the paper. They are the same with the ground-truth in Fig. 5.