

# 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

NIH Clinical Center America's Research Hospital		로그인	등록
DeepLesion		다운로드	
이름	업데이트		
Images_png	2018년 7월 28일, Ke Yan		
Key_slice_examples	2018년 7월 27일, Ke Yan		
changelog.txt	2018년 9월 6일, Ke Yan		
FAQ.pdf	2018년 9월 6일, Ke Yan		
readme.pdf	2018년 9월 6일, Ke Yan		
batch_download_zips.py	2018년 9월 6일, Ke Yan		
18_JMI_DeepLesion.pdf	2018년 7월 20일, Ke Yan		
Key_slices.zip	2018년 7월 18일, Ke Yan		
DL_info.csv	2018년 7월 18일, Ke Yan		
DL_save_nifti.py	2018년 7월 18일, Ke Yan		
18_CVPR_supplementary material.pdf	2018년 6월 28일, Ke Yan		
18_CVPR_Deep Lesion Graphs in the Wild.pdf	2018년 6월 28일, Ke Yan		

# 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 병변에 대한 주석이 있는 key slice
- DL\_info.csv : 주석과 메타데이터
- DL\_save\_nifti.py: 2D 16-bit 이미지에서 3D nifty sub-volume 으로 변환하는 파이썬 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

