Stanford MEDICINE

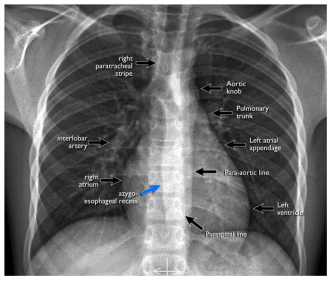
Cloud-Based Petabyte-Scale Medical Image Database for Imaging Research

Curtis P. Langlotz, MD, PhD
Professor of Radiology and Biomedical Informatics
Associate Chair for Information Systems
Department of Radiology, Stanford University

Medical Informatics Director for Radiology
Stanford Health Care

@curtlanglotz

Karpathy, Andrej & Li, Fei Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions, CVPR, 2015

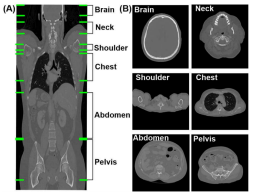


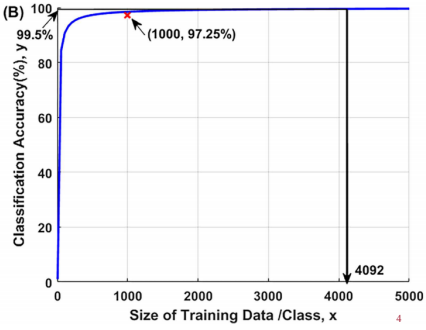
http://www.radiologyassistant.nl/

Medical Imaging: Interesting Challenges

- ~1500 different imaging studies
  - Many distinct imaging modalities (e.g., DR, CT, MR, US, NM)
  - Visualizing a variety of body regions (e.g., head, chest, extremities, abdomen)
  - Each containing dozens of organs and body parts (e.g., liver, pancreas, kidney)
  - Each affected by hundreds of diseases (Total: ~23,000 conditions)
- Typically 12- or 16-bit gray scale, rather than color images.
- 3D volumetric data sets, sometimes incorporating time as 4th dimension.
- Multi-channel (e.g., MR T1-weighing, T2-weighting, flow-sensitive, post-contrast).
- Differential diagnosis varies by anatomic structure, requiring anatomic segmentation.

Radiology Deep Learning Data Needs





Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do, 2016, https://arxiv.org/pdf/1511.06348v2.pdf,

Stanford Radiology Data 2016

Study Type	Study Count	Study Percentage	Estimated Annual Run Rate	Images	Image Percentage
DR = Digital Radiography	2,334,454	51.98%	220,000	3,218,595	0.28%
CT = Computed Tomography	721,182	16.06%	77,000	575,851,840	50.08%
DX = Digital Radiography	2,856	0.06%	600	1,498,911	0.13%
MG = Mammography	206,396	4.60%	21,000	1,926,933	0.17%
MR = Magnetic Resonance	504,275	11.23%	64,000	478,996,296	41.66%
NM = Nuclear Medicine	71,629	1.60%	8,700	547,842	0.05%
OT = Other	1,482	0.03%	325	3,721,808	0.32%
PT = Positron emission tomography (PET)	75,419	1.68%	12,000	55,599,456	4.84%
RF = Radio Fluoroscopy	104,124	2.32%	14,000	1,125,445	0.10%
US = Ultrasound	398,695	8.88%	50,000	20,595,938	1.79%
XA = X-Ray Angiography	70,176	1.56%	10,000	6,725,763	0.58%
Total	4,490,688	100.00%	477,625	1,149,808,827	100.00%

0.5 petabytes

Developing a "Medical ImageNet"

A cloud-based, multi-institutional, petabyte-scale, searchable, open repository of diagnostic imaging studies for developing intelligent image analysis systems

- Data migration/federation/honest broker
- Linkage to EMR and multi-omics
- Cohort discovery tools
- Image viewing software
- Governance
- Image classification and annotation
  - Natural language processing, research data sets, crowd source

Deep Learning / Medical Imaging Opportunities

- Computer-aided detection
- Computer-aided classification
- Work prioritization
- Intelligent imaging utilization
- Automatic report drafting
- Image quality control