Supplementary Material

1 Additional Chest ImaGenome Terminology Descriptions

Table 6: Semantic category of nodes and edges in CXR knowledge graphs. All nodes are mapped to UMLS CUIs in the scene graph jsons. All object nodes have corresponding bounding box coordinates on frontal CXRs except ones with *. All nodes and edges are evaluated with the gold standard dataset except the edges marked with **, which are modifiers of the context edges.

Category ID	type	names
technicalassessment	attribute node	low lung volumes, rotated, artifact, breast/nipple shadows, skin fold
texture	attribute node	opacity, alveolar, interstitial, calcified, lucency
anatomicalfinding	attribute node	lung opacity, airspace opacity, consolidation, infiltration, atelectasis, linear/patchy atelectasis, lobar/segmental collapse, pulmonary edema/hazy opacity, vascular congestion, vascular redistribution, increased reticular markings/ild pattern, pleural effusion, costophrenic angle blunting, pleural/parenchymal scarring, bronchiectasis, enlarged cardiac silhouette, mediastinal displacement, mediastinal widening, enlarged hilum, tortuous aorta, vascular calcification, pneumomediastinum, pneumothorax, hydropneumothorax, lung lesion, mass/nodule (not otherwise specified), multiple masses/nodules, calcified nodule, superior mediastinal mass/enlargement, rib fracture, clavicle fracture, spinal fracture, hyperaeration, cyst/bullae, elevated hemidiaphragm, diaphragmatic eventration (benign), subdiaphragmatic air, subcutaneous air, hernia, scoliosis, spinal degenerative changes, shoulder osteoarthritis, bone lesion
disease	attribute node	pneumonia, fluid overload/heart failure, copd/emphysema, granulomatous disease, interstitial lung disease, goiter, lung cancer, aspiration, alveolar hemorrhage, pericardial effusion
nlp	attribute node	abnormal, normal (with respect to an anatomy/object node)
tubesandlines	attribute node	chest tube, mediastinal drain, pigtail catheter, endotracheal tube, tracheostomy tube, picc, ij line, chest port, subclavian line, swan-ganz catheter, intra-aortic balloon pump, enteric tube
device	attribute node	sternotomy wires, cabg grafts, aortic graft/repair, prosthetic valve, cardiac pacer and wires
majorstructure	object node	right lung, left lung, mediastinum
subanatomy	object node	right apical zone, right upper lung zone, right mid lung zone, right lower lung zone, right hilar structures, right costophrenic angle, left apical zone, left upper lung zone, left mid lung zone, left lower lung zone, left hilar structures, left costophrenic angle, upper mediastinum, cardiac silhouette, trachea, right hemidiaphragm, left hemidiaphragm, right clavicle, left clavicle, spine, right atrium, cavoatrial junction, svc, carina, aortic arch, abdomen, right chest wall*, left chest wall*, right shoulder*, left shoulder*, neck*, right arm*, left arm*, right breast*, left breast*
context	edge	yes (has/present in), no (not have/not present in)
comparison	edge	improved, worsened, no change
severity**	edge	hedge, mild, moderate, severe
temporal**	edge	acute, chronic

2 Chest ImaGenome Construction

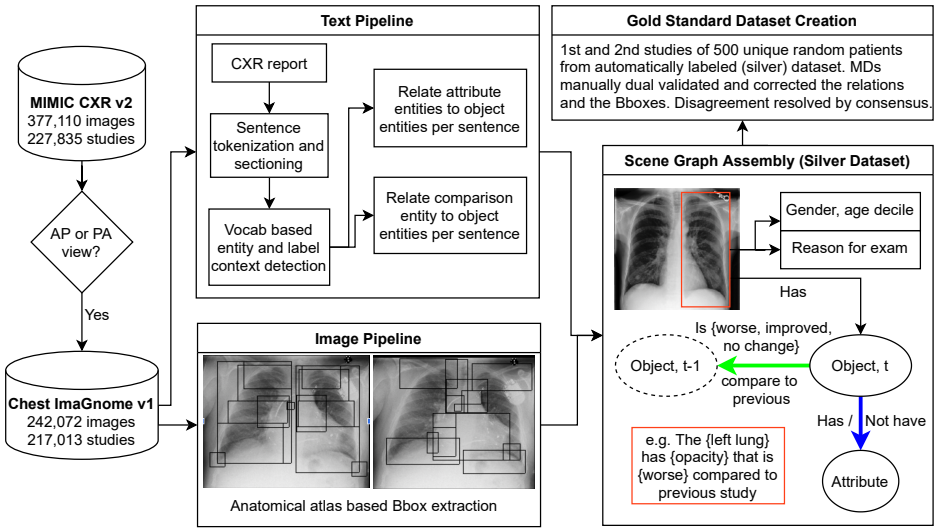


Figure 3: Dataflow for Chest ImaGenome Construction and Evaluation

3 Scene Graph JSON

Below are examples from a scene graph JSON used for explanation of the silver dataset.

3.1 Scene Graph JSON - First Level

```
{
 'chest_imageimage_id': '10cd06e9-5443fef9-9afbe903-e2ce1eb5-dcff1097',
 'viewpoint': 'AP', 'patient_id': 10063856, 'study_id': 56759094,
 'gender': 'F', 'age_decile': '50-60',
 'reason_for_exam': '___F with hypotension.  Evaluate for pneumonia.',
 'StudyOrder': 2, 'StudyDateTime': '2178-10-05 15:05:32 UTC',
 'objects': [ <...list of {} for each object...> ],
 'attributes':[ <...list of {} for each object...> ],
 'relationships':[ <...list of {} of comparison relationships between objects
 from sequential exams for the same patient...> ]
}
```

3.2 Scene Graph JSON - Objects Field

```
{
  'object_id': '10cd06e9-5443fef9-9afbe903-e2ce1eb5-dcff1097_right upper lung zone',
  'x1': 48, 'y1': 39, 'x2': 111, 'y2': 93,
  'width': 63, 'height': 54,
  'bbox_name': 'right upper lung zone',
  'synsets': ['C0934570'],
  'name': 'Right upper lung zone',
  'original_x1': 395, 'original_y1': 532,
  'original_x2': 1255, 'original_y2': 1268,
  'original_width': 860, 'original_height': 736
}
```

3.3 Scene Graph JSON - Attributes Field

```
{
  'right lung': True, 'bbox_name': 'right lung',
```

```
  'synsets': ['C0225706'], 'name': 'Right lung',
  'attributes': [['anatomicalfinding|no|lung opacity',
  'anatomicalfinding|no|pneumothorax',  'nlp|yes|normal'],
  ['anatomicalfinding|no|pneumothorax']],
  'attributes_ids': [['CL556823', 'C1963215;;C0032326', 'C1550457'],
  ['C1963215;;C0032326']],
  'phrases': ['Right lung is clear without pneumothorax.',
  'No pneumothorax identified.'],
  'phrase_IDs': ['56759094|10', '56759094|14'],
  'sections': ['finalreport', 'finalreport'],
  'comparison_cues': [[], []],
  'temporal_cues': [[], []],
  'severity_cues': [[], []],
  'texture_cues': [[], []],
  'object_id': '10cd06e9-5443fef9-9afbe903-e2ce1eb5-dcff1097_right lung'
}
```

3.4 Scene Graph JSON - Comparison Relationships Field

```
{
  'relationship_id': '56759094|7_54814005_C0929215_10cd06e9_4bb710ab',
  'predicate': ''['No status change']'',
  'synsets': ['C0442739'],
  'relationship_names': ['comparison|yes|no change'],
  'relationship_contexts': [1.0],
  'phrase': 'Compared with the prior radiograph, there is a persistent veil
  -like opacity\n over the left hemithorax, with a crescent of air surrounding
  the aortic arch,\n in keeping with continued left upper lobe collapse.',
  'attributes': ['anatomicalfinding|yes|atelectasis',
  'anatomicalfinding|yes|lobar/segmental collapse',
  'anatomicalfinding|yes|lung opacity', 'nlp|yes|abnormal'],
  'bbox_name': 'left upper lung zone',
  'subject_id': '10cd06e9-5443fef9-9afbe903-e2ce1eb5-dcff1097_left upper lung zone',
  'object_id': '4bb710ab-ab7d4781-568bcd6e-5079d3e6-7fdb61b6_left upper lung zone'
}
```

3.5 Scene Graph - Enriched RDF JSON Format

```
{
 <study_id_i> : [
                  [[node_id_1, node_type_1], [node_id_2, node_type_2], relation_name_A],
                  [[node_id_1, node_type_1], [node_id_3, node_type_3], relation_name_B],
                    ...
                ],
 <study_id_i+1>:[
                  [[node_id_1, node_type_1], [node_id_2, node_type_2], relation_name_A],
                  [[node_id_1, node_type_1], [node_id_3, node_type_3], relation_name_B],
                    ...
                ],
}
```

4 Gold Dataset Annotation - Details

The 'gold dataset' is a randomly sampled subset (500 unique patients) from the automatically generated Chest ImaGenome dataset, i.e., the 'silver dataset', that has been manually validated or corrected. The primary purpose of the 'gold dataset' is to evaluate the quality of labels in the 'silver dataset'. For this purpose, we evaluated the Chest ImaGenome dataset along with the 3 components below (A-B). The annotations for each component were collected in stages to reduce the cognitive workload for the annotators. The annotators are all M.D.s with 2 to 10 or more years of clinical experience. One of the annotators is a radiologist trained in the United States, who has over 6 years of radiology experience and specializes in reading imaging exams from the Emergency Department (ED) setting. The annotation tasks were delegated to the annotators according to their clinical experience,

which we think are all more than sufficient for the tasks. Component A and B were annotated by the radiologist and an M.D. and component C was annotated by 4 M.D.'s.

A) Evaluating CXR Knowledge Graph Extraction from Reports

The report knowledge graph for the *first* CXR of the 500 patients was manually reviewed and corrected as necessary for relation extraction between the anatomical locations (objects) and the CXR attributes. From piloting trials, we found that manually annotating multiple targets at a document level lead to a slow and complex task with poor recall. However, sometimes information from prior sentences is necessary to annotate both the anatomical locations and the attributes correctly. Therefore, we set up the annotation task at the sentence level. Sentences from each report are ordered as per the original report, and the phrase boundary for each attribute was marked out for the annotators, where the phrases used for detecting each attribute were curated by consensus between two radiologists from previous work [45].

Since we are targeting a large set of possible anatomical locations (object) to attribute combinations, the annotation was streamlined into the four steps below to minimize the cognitive overload for each step. Steps 1 and 2 are dual annotated by two clinicians (one fully trained radiologist and one M.D.), with disagreements resolved by consensus review. Steps 3 and 4 are single annotated. A random subset of annotations for 500 sentences from step 4 are sampled and dual annotated to estimate inter-annotator agreement. Cleaned results from step 4 constitute the final gold-standard CXR knowledge graph ground truth for the 500 reports.

This annotation component was set up in Excel and was broken down into the following four steps below. In our Excel setup, all sentences from each report are available to the annotators (they can just scroll up or down). The sentences are ordered by 'row_id' sequentially within each report. Unique patients and reports have the same IDs as shown in the figures below.

Step 1 - For each sentence and NLP extracted attribute combination, decide whether the NLP context (affirmed or negated) for the attribute was correct. If not, correct it. Figure 4 shows how this task was set up in Excel. The annotators' task is to make sure the extracted attribute (yellow label_name column) has the correct context given the sentence from the report. This 'context' is used as the relation between the location and the attribute in the final annotated result.

A	C	F	G	H	I	J
inde	subject_id	row_id	section	sentence	context	label_name
0	10020740	55522869|1	finalreport	final report examination: chest (portable ap)		
1	10020740	55522869|2	history	indication: ___ year old man with h/o acute pancreatitis // et tube placement, pna? ards? et tube placement, pna? ards?		
2	10020740	55522869|3	finalreport	impression:		
3	10020740	55522869|4	finalreport	no previous images		
4	10020740	55522869|5	finalreport	there is an {endotracheal tube} in place with its tip approximately 3 cm above the carina	yes	endotracheal tube
5	10020740	55522869|6	finalreport	{nasogastric tube} extends well into the stomach	yes	enteric tube
6	10020740	55522869|7	finalreport	right {subclavian catheter} extends to the level of the carina	yes	subclavian line
7	10020740	55522869|8	finalreport	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion	yes	atelectasis
8	10020740	55522869|8	finalreport	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion	yes	lung opacity
9	10020740	55522869|8	finalreport	mild basilar atelectatic changes without evidence of acute pneumonia or {vascular congestion}	no	vascular congestion
10	10020740	55522869|8	finalreport	mild basilar atelectatic changes without {evidence of acute pneumonia} or vascular congestion	no	pneumonia
11	10020740	55522869|8	finalreport	mild {basilar} atelectatic changes without evidence of {acute} pneumonia or {vascular} congestion	x	abnormal
12	10020740	55522869|9	finalreport	there may well be a small right pleural effusion	yes	lung opacity
13	10020740	55522869|9	finalreport	there may well be a small {right pleural effusion}	yes	pleural effusion

Figure 4: Step 1: Annotate all attributes per sentence.

Step 2 - For each sentence, decide whether the NLP extracted anatomical location(s) were described or implied by the reporting radiologist. If not, remove the location (in yellow column 'bboxes_corrected). If missing, add the location. If unsure (e.g., if lung is mentioned but not sure if it is the right or left lung), the annotator can look in previous sentences from the same report. The task was set up as shown in Figure 5.

B	E	F	G	H
subject_	row_id	section	sentence	bboxes_corrected
10020740	55522869|1	finalreport	FINAL REPORT EXAMINATION: CHEST (PORTABLE AP)	[]
10020740	55522869|2	history	INDICATION: ___ year old man with h/o acute pancreatitis // ET tube placement, PNA? ARDS? ET tube placement, PNA? ARD	[]
10020740	55522869|3	finalreport	IMPRESSION:	[]
10020740	55522869|4	finalreport	No previous images.	[]
10020740	55522869|5	finalreport	There is an endotracheal tube in place with its tip approximately 3 cm above the carina.	['carina', 'trachea', 'neck']
10020740	55522869|6	finalreport	Nasogastric tube extends well into the stomach.	['abdomen', 'mediastinum']
10020740	55522869|7	finalreport	Right subclavian catheter extends to the level of the carina.	['carina', 'mediastinum', 'right clavicle']
10020740	55522869|8	finalreport	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.	['left lower lung zone', 'right lower lung zone', 'right lung', 'left lung']
10020740	55522869|9	finalreport	There may well be a small right pleural effusion.	['right costophrenic angle', 'right lung']

Figure 5: Step 2: Annotate all locations per sentence.

Step 3 - For recall, manually annotate missed objects and/or attributes for sentences with no NLP extractions (a much smaller subset). For this, we used Excel's filtering function to look at all sentences with no automated extractions (empty cells) and de novo added the manual annotations.

Step 4 - Firstly, all rows from steps 1-3 where the annotations differed between the two annotators were reviewed and resolved together by consensus. Then we automatically derived all object-attribute relation combinations for each sentence from steps 1-3's results. The obviously wrong object-to-attribute relations were filtered out for each sentence using the CXR ontology. For the remaining object-to-attribute relations for each sentence, the task was to indicate whether the logical statement of *"object X contains (or does not contain) attribute Y"* is true or false, as shown in Figure 6. Probable relation is still defined to be true for this annotation. Annotating for uncertain relations is beyond the scope of this project. However, for future dataset expansion, we have kept the NLP cues for the certainty for each object-attribute relation in the scene graph JSON.

patient_id	row_id	section	bbox	relation	label_name	sentence
10020740	55522869|5	finalreport	trachea	1	endotracheal tube	There is an endotracheal tube in place with its tip approximately 3 cm above the carina.
10020740	55522869|6	finalreport	abdomen	1	enteric tube	Nasogastric tube extends well into the stomach.
10020740	55522869|6	finalreport	mediastinum	1	enteric tube	Nasogastric tube extends well into the stomach.
10020740	55522869|6	finalreport	neck	1	enteric tube	Nasogastric tube extends well into the stomach.
10020740	55522869|7	finalreport	mediastinum	1	subclavian line	Right subclavian catheter extends to the level of the carina.
10020740	55522869|7	finalreport	right clavicle	1	subclavian line	Right subclavian catheter extends to the level of the carina.
10020740	55522869|8	finalreport	left hilar structures	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lower lung zone	0	pneumonia	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lower lung zone	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lower lung zone	1	abnormal	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion
10020740	55522869|8	finalreport	left lower lung zone	1	atelectasis	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lower lung zone	1	lung opacity	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lung	0	pneumonia	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lung	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lung	1	abnormal	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion
10020740	55522869|8	finalreport	left lung	1	atelectasis	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	left lung	1	lung opacity	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right hilar structures	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lower lung zone	0	pneumonia	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lower lung zone	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lower lung zone	1	abnormal	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion
10020740	55522869|8	finalreport	right lower lung zone	1	atelectasis	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lower lung zone	1	lung opacity	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lung	0	pneumonia	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lung	0	vascular congestion	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.
10020740	55522869|8	finalreport	right lung	1	abnormal	mild basilar {atelectatic changes} without evidence of acute pneumonia or vascular congestion
10020740	55522869|8	finalreport	right lung	1	atelectasis	Mild basilar atelectatic changes without evidence of acute pneumonia or vascular congestion.

Figure 6: Step 4: Annotate all logically correct statements/relations for each sentence.

Since step 4 was single annotated, to estimate the final inter-annotator agreement, we randomly sampled 500 sentences for dual annotations. This annotated result is also shared on PhysioNet.

B) Evaluating Comparison Relation Extraction:

The *second* CXR exam report for the 500 patients was reviewed for comparison relation extraction. The annotation was also set up in Excel and conducted at the sentence level. However, the annotator is also shown the whole previous CXR report for context. Similarly, we split the annotation task up into several steps, where steps 1 and 2 are dual annotated and disagreement resolved via consensus. Steps 3 and 4 were single annotated. A subset of 500 sentences from the final annotations was reviewed by a second annotator for assessing inter-annotator agreement.

Step 1 - Given the previous report and the current report sentence, decide whether the extracted comparison cue(s) (improved, worsened, no change) is/are correct. If not, correct it/them. In this step, the annotators are asked to validate or correct the column 'comparison' in Figure 7.

Step 2 - Building from step 1 for each sentence, given a validated or corrected comparison cue, validate whether all the anatomical location(s) extracted are correct (column 'bbox' in Figure 7). If incorrect or missing, remove or add the correct location(s) to the column.

Step 3 - Building from step 2 for each sentence, given each correct comparison cue and anatomical location relation, decide whether the attributes assigned to the location described or implied in the sentence are correct or not. If not, correct it. Figure 8 illustrates how step 3 was set up, where the annotators' task is to validate or correct the 'label_name' column with respect to the 'bbox', 'relation' and 'comparison' columns for each sentence.

subject_id	row_id	section	sentence	comparison	bbox	StudyOrd
10127462	57192363|4	finalreport	the right hemidiaphragm is elevated		right hemidiaphragm	1
10127462	57192363|5	finalreport	there is mild vascular congestion		right lung	1
10127462	57192363|5	finalreport	there is mild vascular congestion		right hilar structures	1
10127462	57192363|5	finalreport	there is mild vascular congestion		left lung	1
10127462	57192363|5	finalreport	there is mild vascular congestion		left hilar structures	1
10127462	57192363|6	finalreport	there is no pneumothorax or pleural effusion		right lung	1
10127462	57192363|6	finalreport	there is no pneumothorax or pleural effusion		right costophrenic angle	1
10127462	57192363|6	finalreport	there is no pneumothorax or pleural effusion		left lung	1
10127462	57192363|6	finalreport	there is no pneumothorax or pleural effusion		left costophrenic angle	1
10127462	57192363|7	finalreport	cardiac size is top normal accentuated by the projection		cardiac silhouette	1
10127462	56032421|0	prelimread	wet read: ___ ___ ___ 9:03 pm no radiographic evidence for acute process			2
10127462	56032421|1	finalreport	final report chest radiograph			2
10127462	56032421|2	history	indication: status post fusion, elevated temperature, assessment for lung pathology			2
10127462	56032421|3	finalreport	comparison: ___			2
10127462	56032421|4	finalreport	findings: as compared to the previous radiograph, the lung volumes have increased	improved	right lung	2
10127462	56032421|4	finalreport	findings: as compared to the previous radiograph, the lung volumes have increased	improved	left lung	2
10127462	56032421|5	finalreport	the size of the cardiac silhouette is still at the upper range of normal	no change	cardiac silhouette	2
10127462	56032421|6	finalreport	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image	improved	right lung	2
10127462	56032421|6	finalreport	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image	improved	right hilar structures	2
10127462	56032421|6	finalreport	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image	improved	left lung	2
10127462	56032421|6	finalreport	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image	improved	left hilar structures	2
10127462	56032421|7	finalreport	minimal right pleural effusion, causing blunting of the right costophrenic angle		right lung	2
10127462	56032421|7	finalreport	minimal right pleural effusion, causing blunting of the right costophrenic angle		right costophrenic angle	2
10127462	56032421|8	finalreport	minimal areas of atelectasis at the right and left lung bases		right lung	2
10127462	56032421|8	finalreport	minimal areas of atelectasis at the right and left lung bases		right lower lung zone	2
10127462	56032421|8	finalreport	minimal areas of atelectasis at the right and left lung bases		left lung	2
10127462	56032421|8	finalreport	minimal areas of atelectasis at the right and left lung bases		left lower lung zone	2
10127462	56032421|9	finalreport	no pneumothorax		right lung	2
10127462	56032421|9	finalreport	no pneumothorax		left lung	2

Figure 7: Step 1 and 2: Annotate change relations for different anatomical locations

patient_id	study_id	studyOrdi	row_id	section	bbox	relation	label_name	comparison	sentence
10127462	56032421	2	56032421|4	finalreport	left lung	1	low lung volumes	['improved']	findings: as compared to the previous radiograph, the lung volumes have increased
10127462	56032421	2	56032421|4	finalreport	right lung	1	low lung volumes	['improved']	findings: as compared to the previous radiograph, the lung volumes have increased
10127462	56032421	2	56032421|5	finalreport	cardiac silhouette	0	enlarged cardiac silhouette	['no change']	the size of the cardiac silhouette is still at the upper range of normal
10127462	56032421	2	56032421|5	finalreport	cardiac silhouette	1	normal	['no change']	the size of the cardiac silhouette is still at the upper range of normal
10127462	56032421	2	56032421|6	finalreport	left hilar structures	1	abnormal	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	left hilar structures	1	lung opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	left hilar structures	1	pulmonary edema/hazy opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	left lung	1	abnormal	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	left lung	1	lung opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	left lung	1	pulmonary edema/hazy opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right hilar structures	1	abnormal	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right hilar structures	1	lung opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right hilar structures	1	pulmonary edema/hazy opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right lung	1	abnormal	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right lung	1	lung opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image
10127462	56032421	2	56032421|6	finalreport	right lung	1	pulmonary edema/hazy opacity	['improved']	the radiographic evidence of mild pulmonary edema is still present but less severe than on the previous image

Figure 8: Step 3: Annotate change relations for different anatomical locations with respect to attribute

Step 4 - For recall, we used the filtering function in Excel to isolate all sentences with no comparison cue extractions from step 3. Sentences with missing comparison annotations were manually de-novo annotated.

C) Evaluating Anatomy Object Detection for CXR Images:

The first and second CXR images for the same 500 patients were dual validated and corrected for the bounding box objects (i.e., 1000 frontal CXR images altogether). Given the resources we had, we selected 28 anatomical objects (out of 36 available) that are clinically most important for frontal CXRs interpretations. The automatically extracted bounding box coordinates were first plotted on resized and padded 224×224 images. From piloting, we determined that this image size is sufficiently large to annotate the anatomies that we were targeting. The plotted images were displayed one at a time to annotators via a custom Jupyter Notebook that we had set up to allow bounding box coordinates and label annotations. We set up the annotation task on two panels, one for lung-related bounding boxes (Figure 9) and another for mediastinum related and other bounding boxes (Figure 10).

Four M.D.'s were trained to perform this task after reviewing a set of 20-30 training examples with a radiologist. Since the inter-annotator agreement is high (mean IoU > 0.96 for all objects), the final cleaned gold standard bbox coordinates use the average coordinates from two annotators for each bounding box.

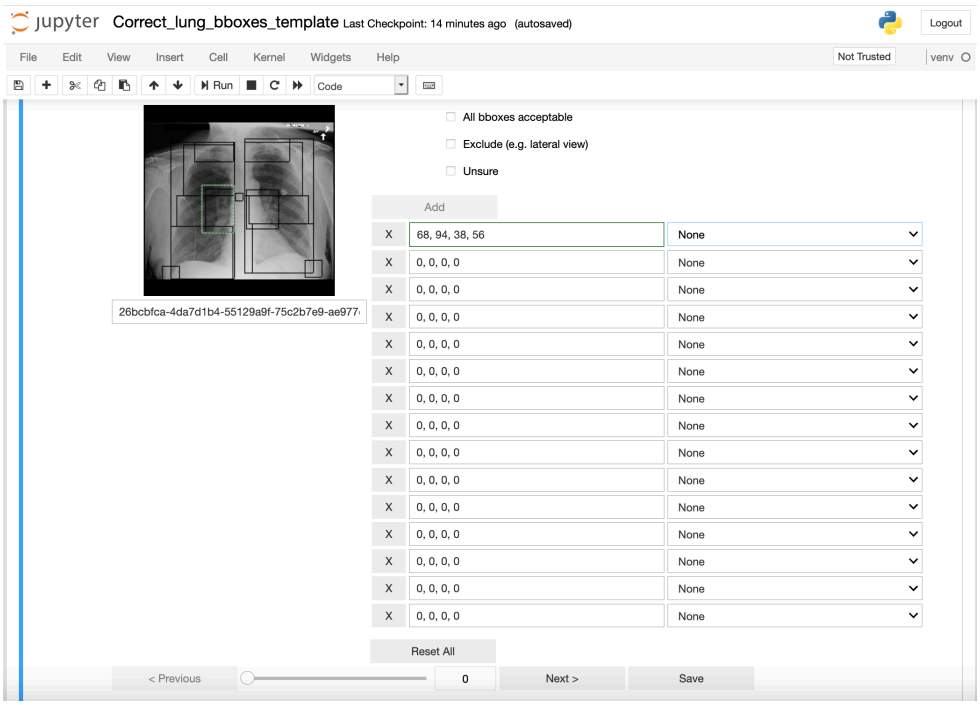


Figure 9: Bbox Annotations - Lung-related Bboxes Panel

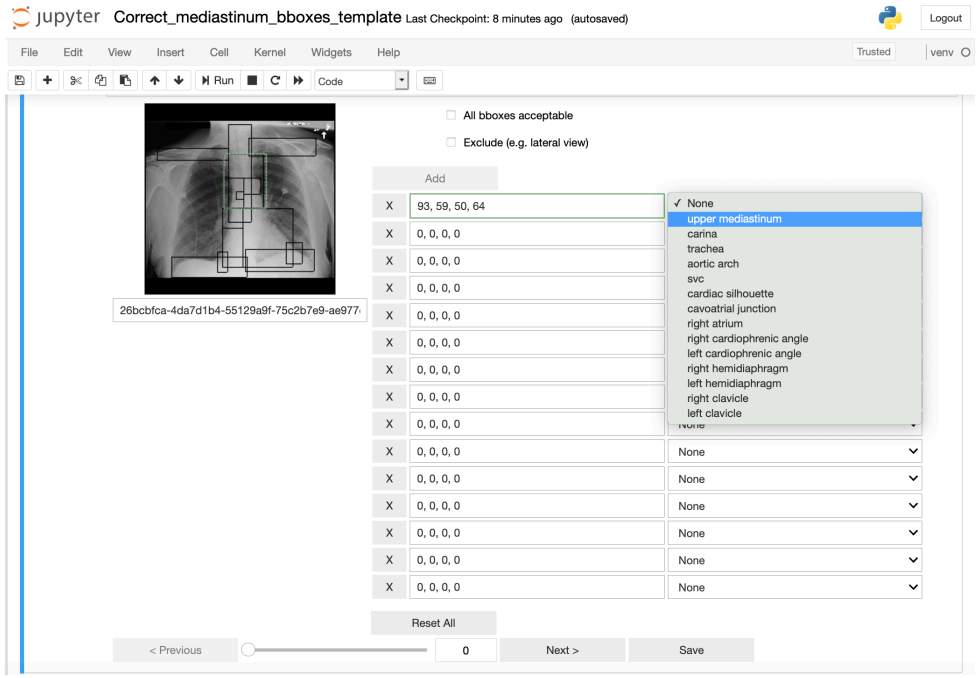


Figure 10: Bbox Annotations - Mediastinum-related and Other Bboxes Panel

5 Dataset Usage Supporting Files

gold_all_sentences_500pts_1000studies.txt contains all the sentences tokenized from the original MIMIC-CXR reports that were used to create the gold standard dataset. We include this file because sentences with no relevant object, attribute or relation descriptions did not make it into the gold standard dataset. We renamed 'subject_id' from MIMIC-CXR dataset to 'patient_id' in Chest Im-

aGenome dataset to avoid confusion with field names for relationships in the scene graphs. Otherwise, the ids are unchanged. Sentences in the tokenized file are assigned to 'history', 'prelimread', or 'finalreport' in the 'section' column. The 'sent_loc' column contains the order of the sentences as in the original report. Minimal tokenization has been done to the sentences.

gold_bbox_scaling_factors_original_to_224x224.csv contains the scaling 'ratio' and the paddings ('left', 'right', 'top', and 'bottom') added to square the image after resizing the original MIMIC-CXR dicoms to 224×224 sizes. These ratios were used to rescale the annotated coordinates for 224×224 images back to the original CXR image sizes.

auto_bbox_pipeline_coordinates_1000_images.txt contains the bounding box coordinates that were automatically extracted by the Bbox pipeline for the different objects for images in the gold standard dataset. It is in a tabular format like with the ground truth for easier evaluation purposes.

object-bbox-coordinates_evaluation.ipynb notebook calculates the bounding box object detection performance using ground truth files from the 4 M.D. annotators , as well as consolidating the final **gold_bbox_coordinate_annotations_1000images.csv**.

Preprocess_mimic_cxr_v2.0.0_reports.ipynb processes the reports (tokenize sentences and sort them into history, prelim or final report sentences) from the original MIMIC-CXR v2.0.0 and save output as **silver_dataset/cxr-mimic-v2.0.0-processed-sentences_all.txt**. Only sentences with object or attribute extractions ended up in the final scene graph jsons in the Chest ImaGenome dataset.

The **semantics** directory contains the object (**objects_detectable_by_bbox_pipeline_v1.txt** and **objects_extracted_from_reports_v1.txt**), attribute (**attribute_relations_v1.txt**) and comoparison (**comparison_relations_v1.txt**) relations labels in the Chest ImaGenome dataset. It also contains **semantics/label_to_UMLS_mapping.json**, which maps all Chest ImaGenome concepts to UMLS CUIs [4].

6 Dataset Evaluation

Table 7 reports anatomical location level object-to-attribute relations extraction performance by the scene graph extraction pipeline. The report numbers are calculated by a combination of notebooks: 'generate_scenegraph_statistics.ipynb', 'object-attribute-relation_evaluation.ipynb' and 'object-bbox-coordinates_evaluation.ipynb'.

Table 7: CXR image object detection evaluation results. * These anatomical locations are extracted by the Bbox pipeline but they are not manually annotated in the gold standard dataset due to resource constraints. ** The mediastinum bounding boxes were not directly annotated due to resource constraints. Mediastinum's bounding box boundary can be derived from the ground truth for the upper mediastinum and the cardiac silhouette.

Bbox name (object)	**Object-attribute relations frequency (500 reports)**	**Relationships F1 (500 reports)**	**Bbox IoU (over 1000 images)**	**% Bboxes corrected (1000 images)**	**% Relations missing Bbox coordinates (over whole dataset)**
left lung	1453	0.933	0.976	9.90%	0.03%
right lung	1436	0.937	0.983	6.30%	0.04%
cardiac silhouette	633	0.966	0.967	9.70%	0.01%
mediastinum	601	0.952	**	**	0.02%
left lower lung zone	609	0.932	0.955	8.60%	2.36%
right lower lung zone	580	0.902	0.968	6.00%	2.27%
right hilar structures	572	0.934	0.976	4.10%	1.91%
left hilar structures	571	0.944	0.971	4.30%	2.28%
upper mediastinum	359	0.940	0.994	1.40%	0.12%
left costophrenic angle	298	0.908	0.929	9.60%	0.63%
right costophrenic angle	286	0.918	0.944	6.90%	0.39%
left mid lung zone	173	0.940	0.967	5.70%	2.79%
right mid lung zone	169	0.830	0.968	5.30%	2.31%
aortic arch	144	0.965	0.991	1.40%	0.62%
right upper lung zone	117	0.873	0.972	5.80%	0.04%
left upper lung zone	83	0.811	0.968	6.40%	0.22%
right hemidiaphragm	78	0.947	0.955	7.90%	0.15%
right clavicle	71	0.615	0.986	2.80%	0.50%
left clavicle	67	0.642	0.983	3.00%	0.51%
left hemidiaphragm	65	0.930	0.944	11.30%	0.14%
right apical zone	58	0.852	0.969	5.40%	1.99%
trachea	57	0.983	0.995	0.90%	0.24%
left apical zone	47	0.938	0.963	6.20%	2.40%
carina	41	0.975	0.994	0.80%	1.47%
svc	19	0.973	0.995	0.70%	0.66%
right atrium	14	0.963	0.979	4.00%	0.18%
cavoatrial junction	5	1.000	0.977	4.30%	0.25%
abdomen	80	0.904	*	*	0.26%
spine	132	0.824	*	*	0.10%

7 Pictorial Overview of Model Architectures

Due to space limitations, we present overview figures for the models designed for Example Tasks 1 and 2 here.

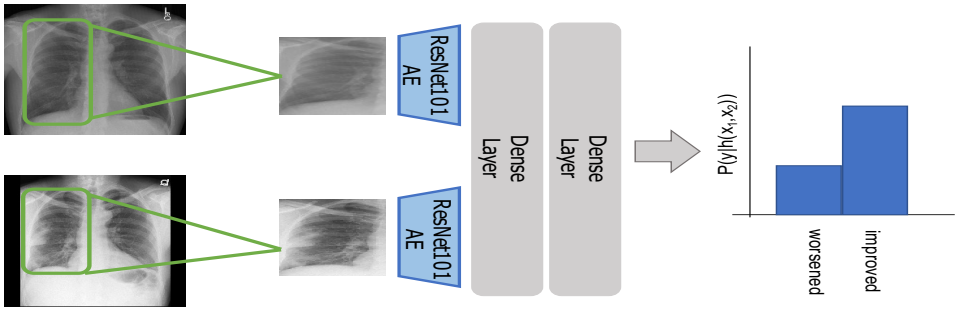


Figure 11: Example Task 1 Model Overview. Given a pair of CXR images, we extract features for the anatomical regions of interest with a pretrained ResNet autoencoder, concatenate representations and pass them through a dense layer and a final classification layer.

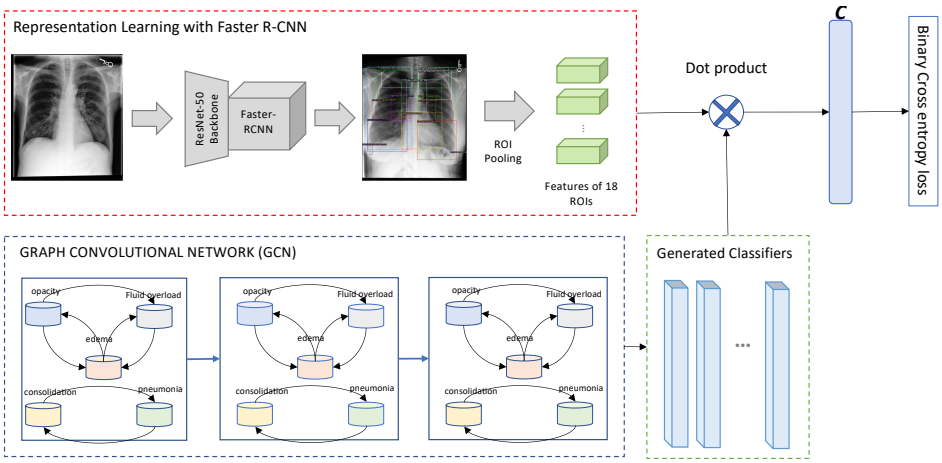


Figure 12: Example Task 2 Model Overview. Given a pair of CXR images, we extract features for the anatomical regions of interest with a pretrained Faster R-CNN and a GCN to learn the label dependencies.

8 Qualitative Evaluation

In Figure 13, we visualize the output from our model for the anatomical finding predictions of costophrenic angles and enlarged cardiac silhouette. In Figure 14, we present an additional example, showing that the model is able to provide accurate localization information as well as predict the correct finding, i.e., showing accurate localization.

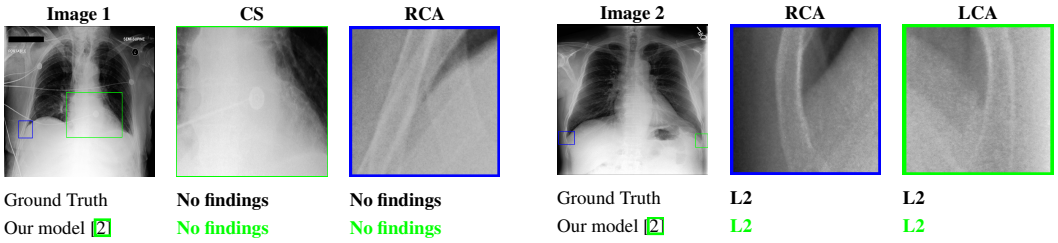
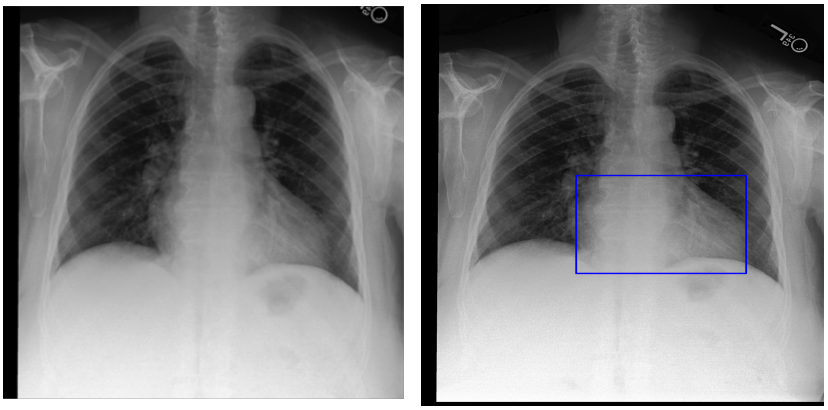


Figure 13: Examples of the prediction results. The overall chest X-ray image is shown alongside two anatomical regions, and predictions are compared against the ground-truth labels.



(a) Original Image (b) Our model [2]

Figure 14: Example image with enlarged cardiac silhouette, showing that the trained model detects the finding in the correct bounding box.