PETUUM

Carnegie Mellon University
School of Computer Science

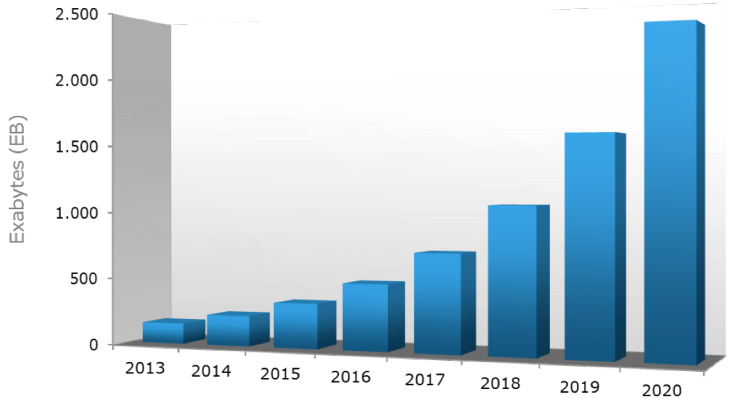
Medical Image Report Generation and Beyond

Zhiting Hu, Pengtao Xie, Xiaodan Liang, and Eric Xing

February 25, 2019

A Tsunami of Healthcare Data

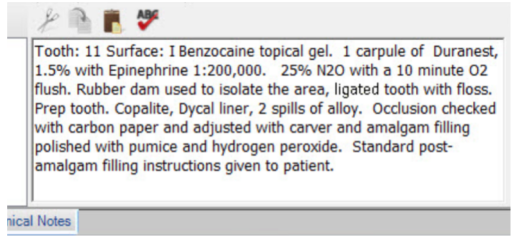
Volume



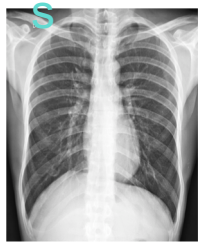
- 153 exabytes (one exabyte = one billion gigabytes) were produced in 2013
- An estimated 2,314 exabytes will be produced in 2020

Complexity

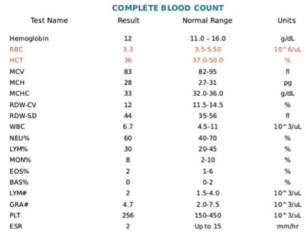
Notes



Images

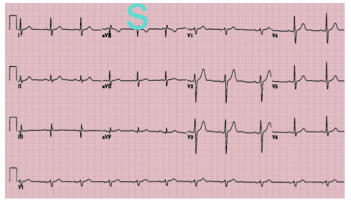


Lab values



Vital signs

Tests



Genomics

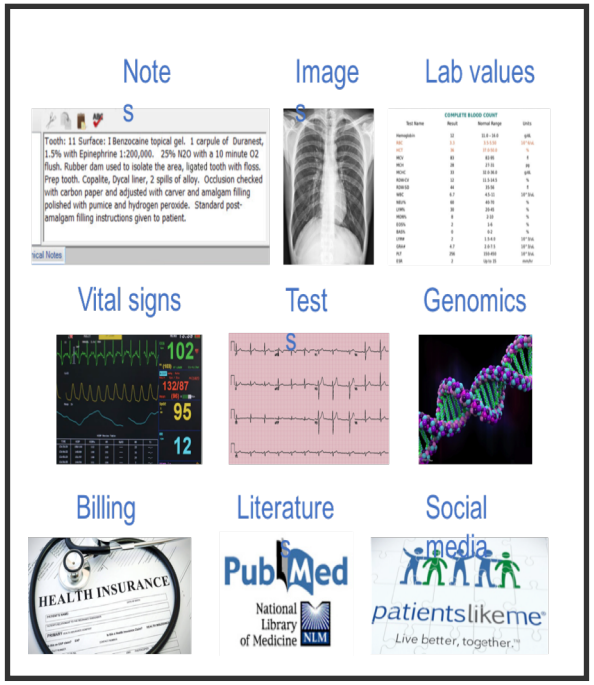
Billing

Literatures

Social media

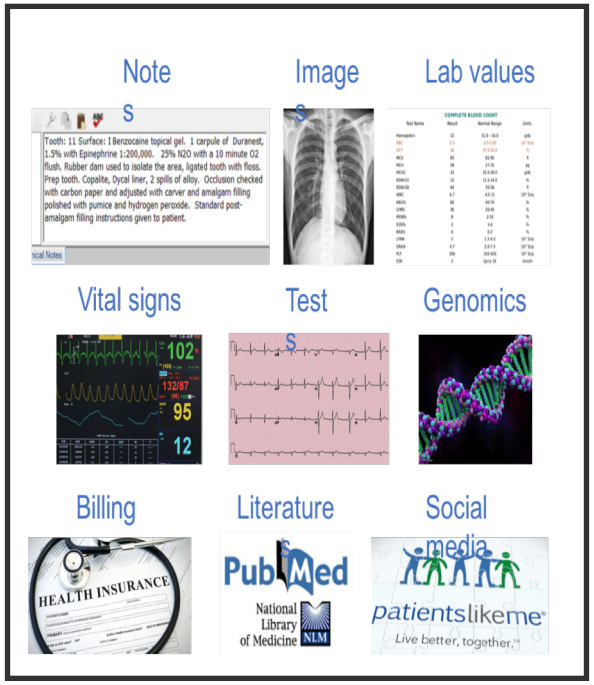
Machine Learning for Healthcare

Clinical Data

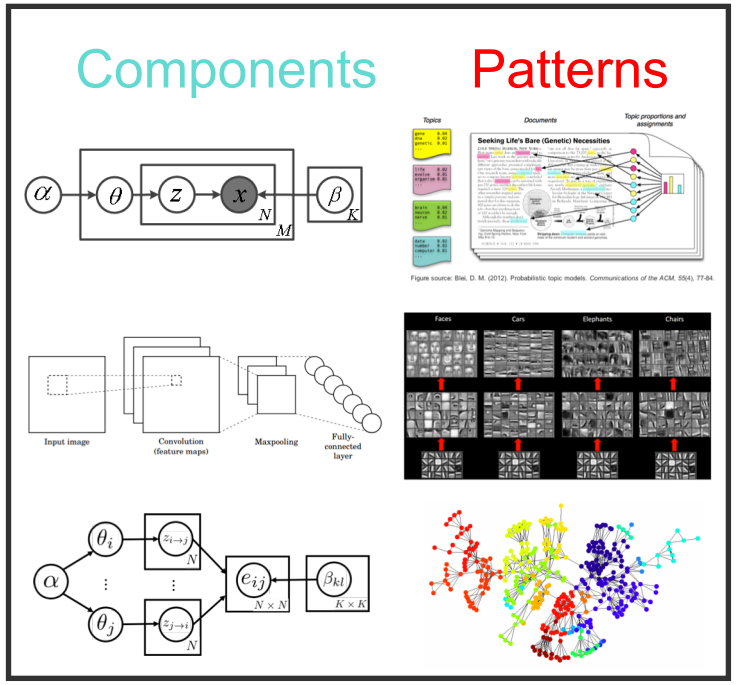


Machine Learning for Healthcare

Clinical Data



Machine Learning



Machine Learning for Healthcare

Clinical Data

Note s

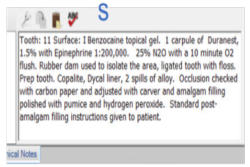
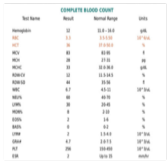


Image s

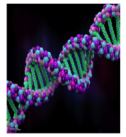
Lab values



Vital signs

Test s

Genomics



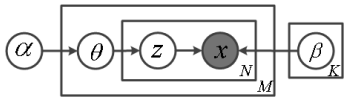
Billing

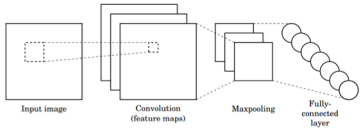
Literature

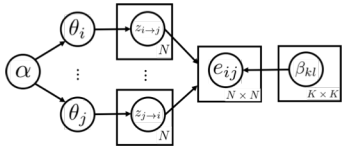
Social media

Machine Learning

Components







Patterns

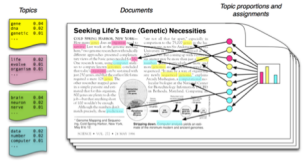
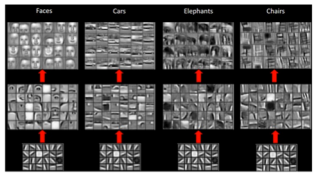
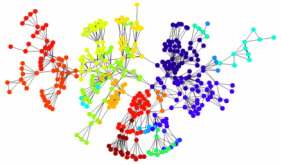


Figure source: Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, *55*(4), 77-84.





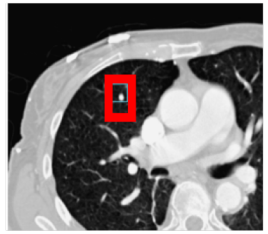
Actionable Insights

Extracted information

Recommended diagnosis and treatment

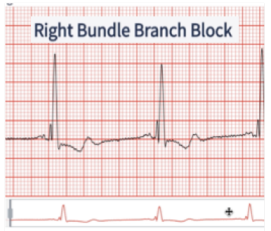
Suggested ICD codes

Detected lung nodule



Predicted mortality rate

Detected arrhythmia



Medical Imaging and Report

- Medical imaging is widely used in clinical practice

Medical Imaging and Report

- Medical imaging is widely used in clinical practice
- Specialized physicians read medical images and write text reports

Medical Imaging and Report

- Medical imaging is widely used in clinical practice
- Specialized physicians read medical images and write text reports

- Knowledge demanding:
  - 1) normal anatomy of, e.g., thorax, basic physiology of chest diseases; 2) analyzing radio graph; 3) evaluating evolution; 4) correlation with other diagnostic results; …

Medical Imaging and Report

- Medical imaging is widely used in clinical practice
- Specialized physicians read medical images and write text reports

- Knowledge demanding:
  - 1) normal anatomy of, e.g., thorax, basic physiology of chest diseases; 2) analyzing radio graph; 3) evaluating evolution; 4) correlation with other diagnostic results; …

- Time consuming:
  - 5-10 mins per image
  - 100s of images per day

General Image-to-Text Problems at a Glance

- Traditionally:
  - Labeling (classification to known labels)
  - Tagging (ROI, bounding-boxes, etc)
  - Simple description (one sentence …)

General Image-to-Text Problems at a Glance

- Traditionally:
  - Labeling (classification to known labels)
  - Tagging (ROI, bounding-boxes, etc)
  - Simple description (one sentence …)

- In need:
  - Full textual summary
  - Report
    - different image genre
    - a full image collection, not just a single image
    - videos

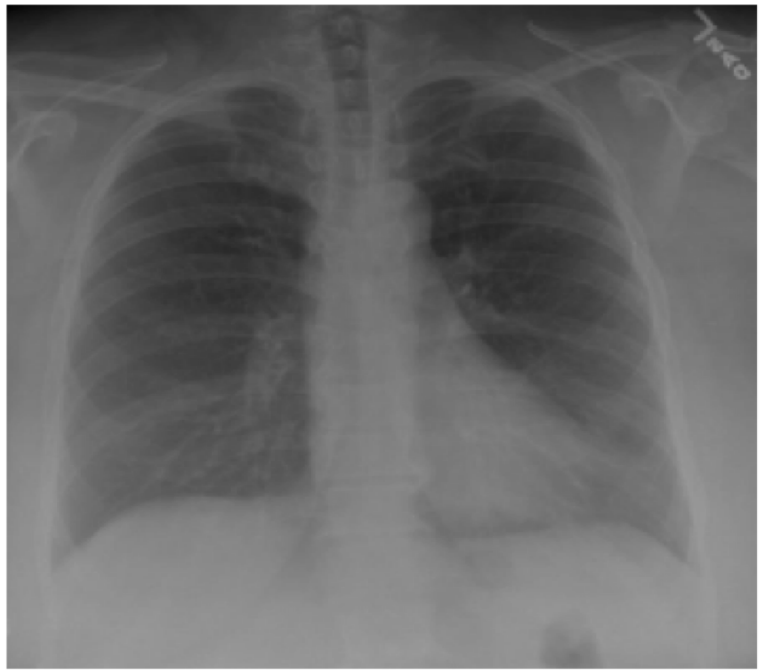
Outline

- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Outline

- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Chest X-ray Report

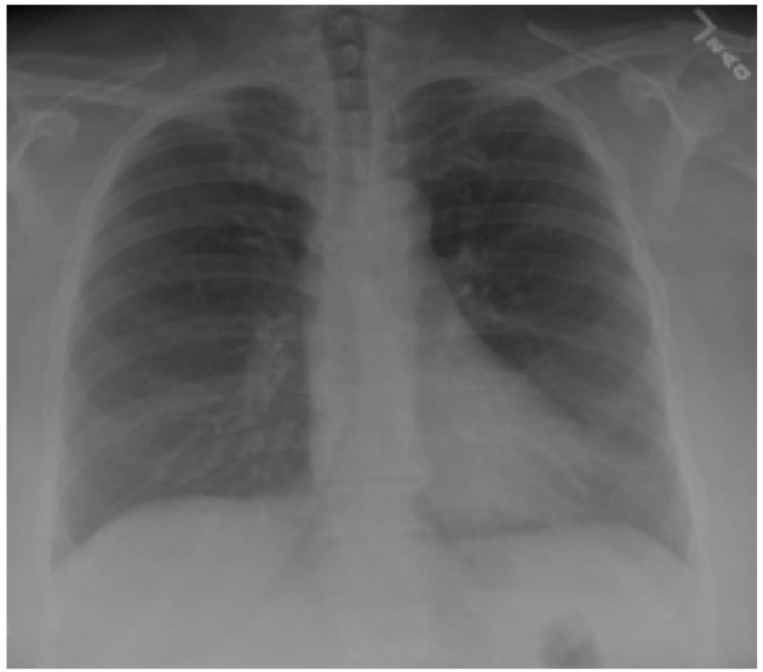


Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

Chest X-ray Report

- A paragraph consists of
  - *Findings:* radiology observations regarding the body area examined
  - *Impression:* most prominent observation or conclusion

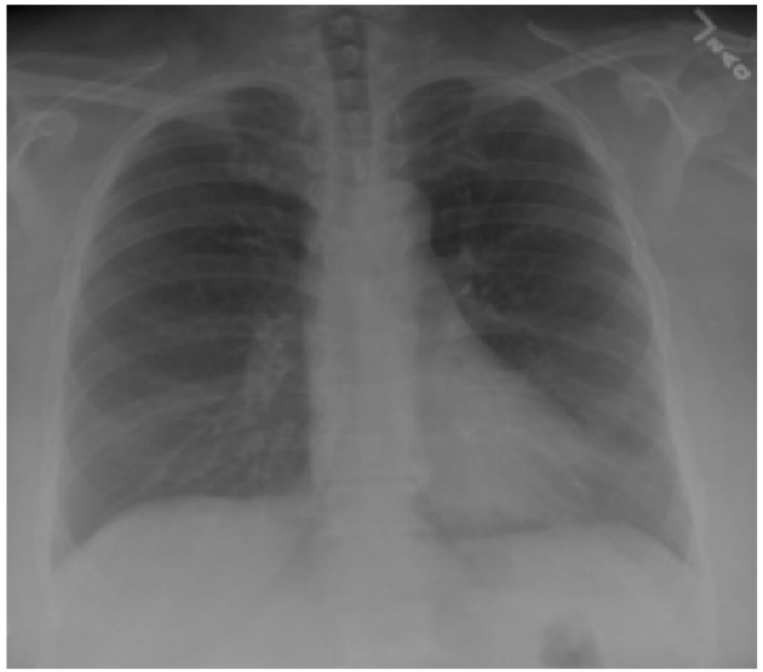


Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

Solution Overview

Key challenges



Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

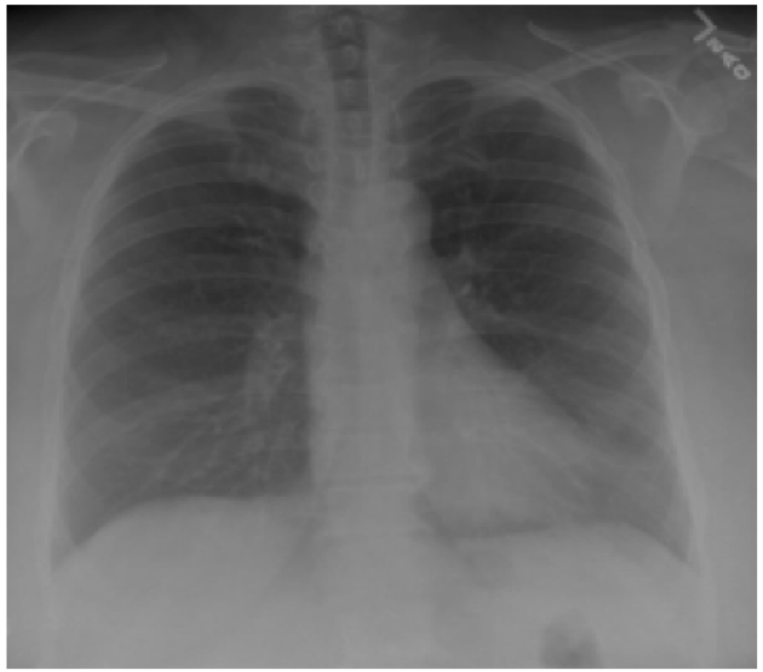
Impression:
No acute cardiopulmonary abnormality.

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)



Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

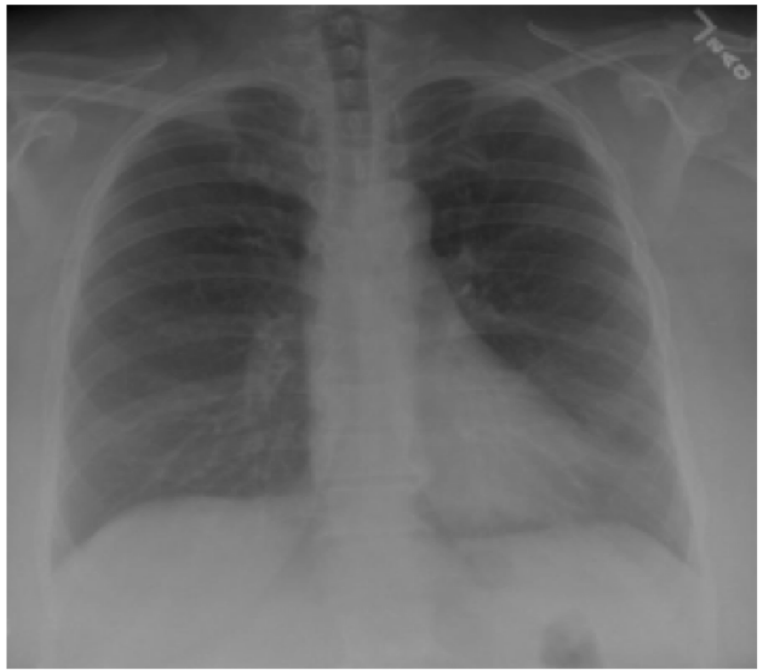
Impression:
No acute cardiopulmonary abnormality.

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs



Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs
- Each sentence has visual evidence

Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs
- Each sentence has visual evidence

Solutions

Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs
- Each sentence has visual evidence

Solutions

<--- Multi-task tag classification to inform lesions

Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

MTI Tags: degenerative change

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs
- Each sentence has visual evidence

Solutions

<--- Multi-task tag classification to inform lesions

<--- Hierarchical text generation: sentence/word

Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

MTI Tags: degenerative change

[Jing et al., 2018]

Solution Overview

Key challenges

- Identify abnormal regions (lesions)
- Reports are usually lengthy paragraphs
- Each sentence has visual evidence

Solutions

<--- Multi-task tag classification to inform lesions

<--- Hierarchical text generation: sentence/word

<--- Visual-semantic co-attention

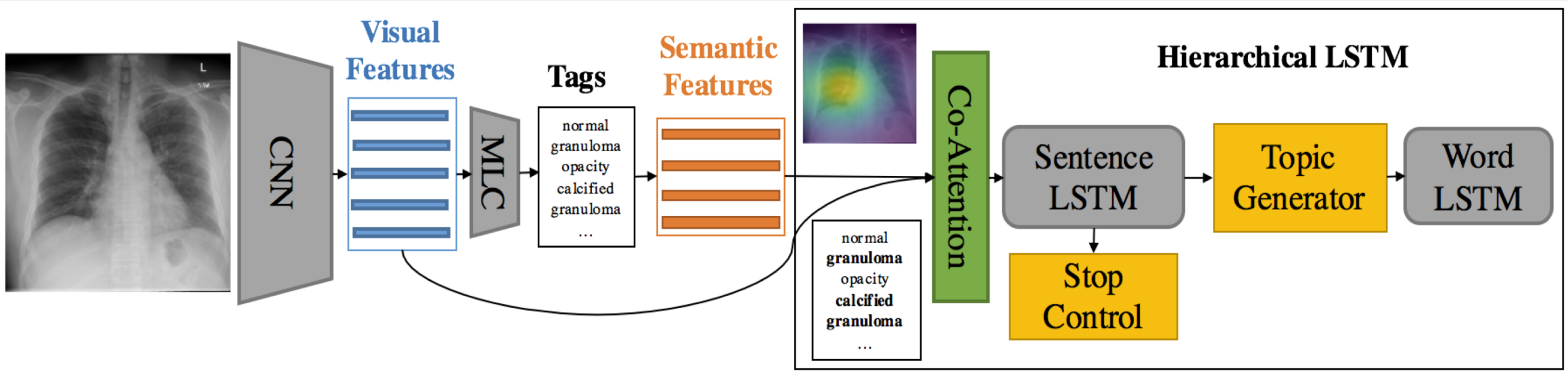
Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.

Impression:
No acute cardiopulmonary abnormality.

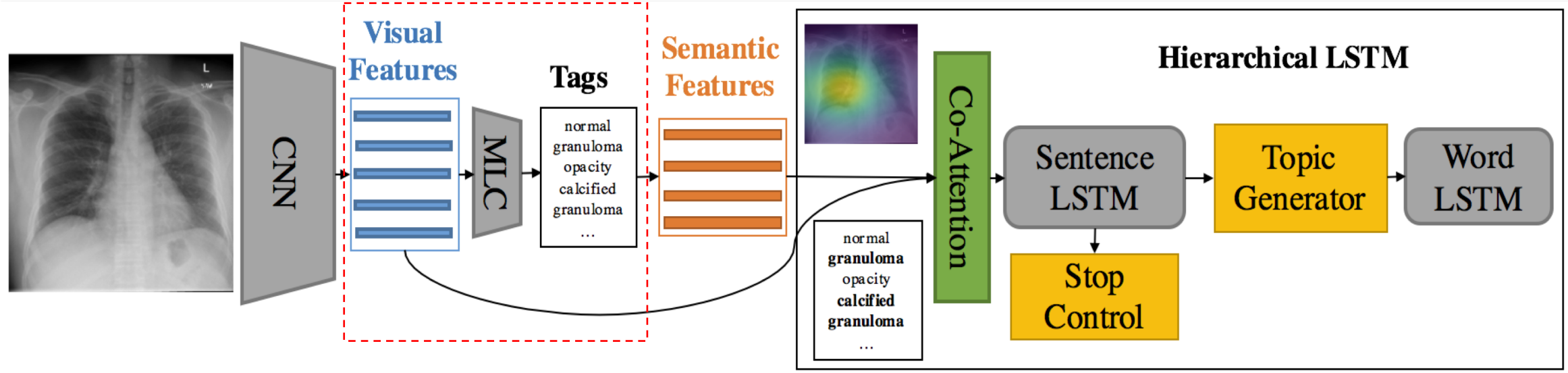
MTI Tags: degenerative change

[Jing et al., 2018]

Model Architecture

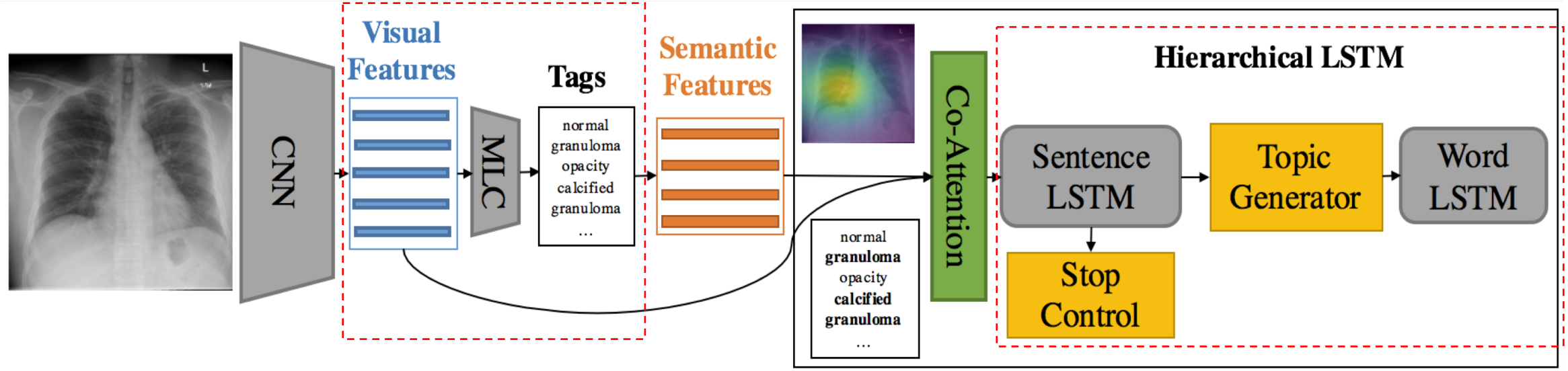


Model Architecture



(I) Lesion tag classification

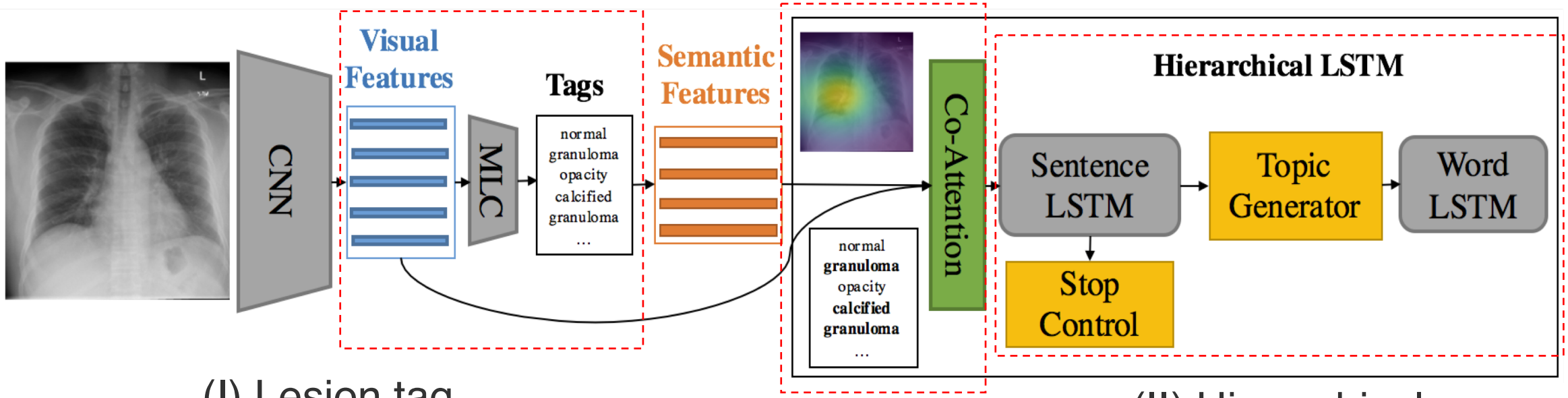
Model Architecture



(I) Lesion tag classification

(II) Hierarchical text generation

Model Architecture

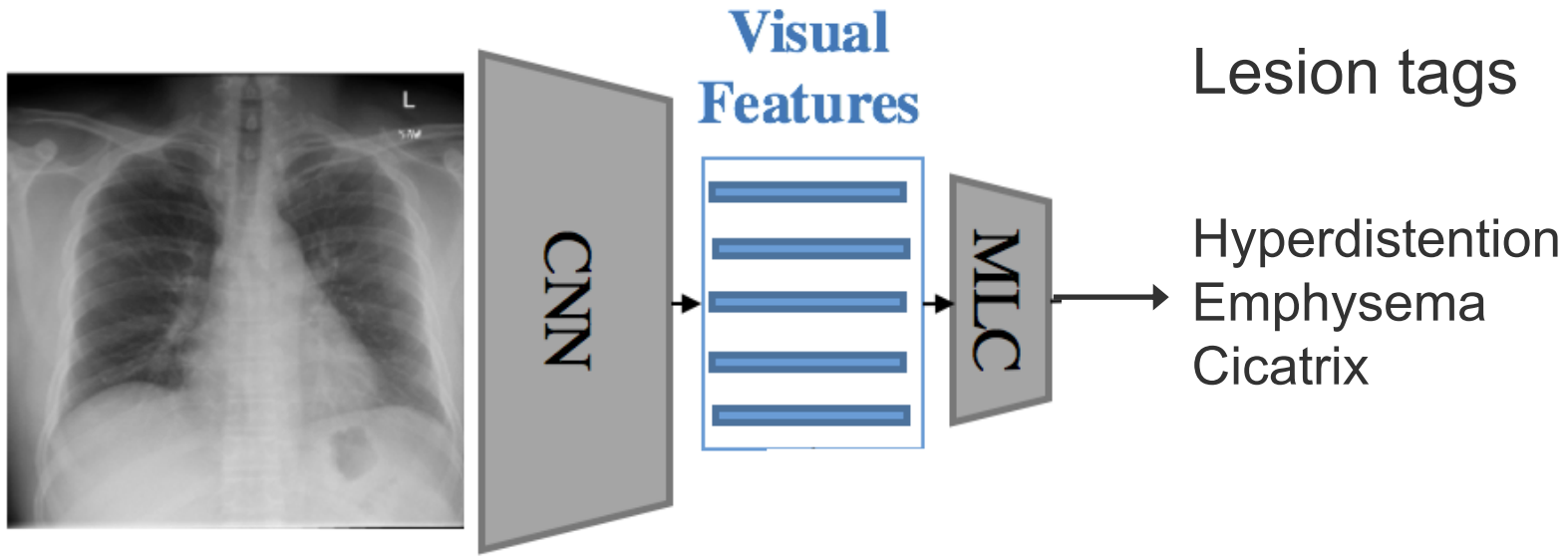


(I) Lesion tag classification

(III) Visual-semantic co-attention

(II) Hierarchical text generation

(I) Lesion Tag Classification

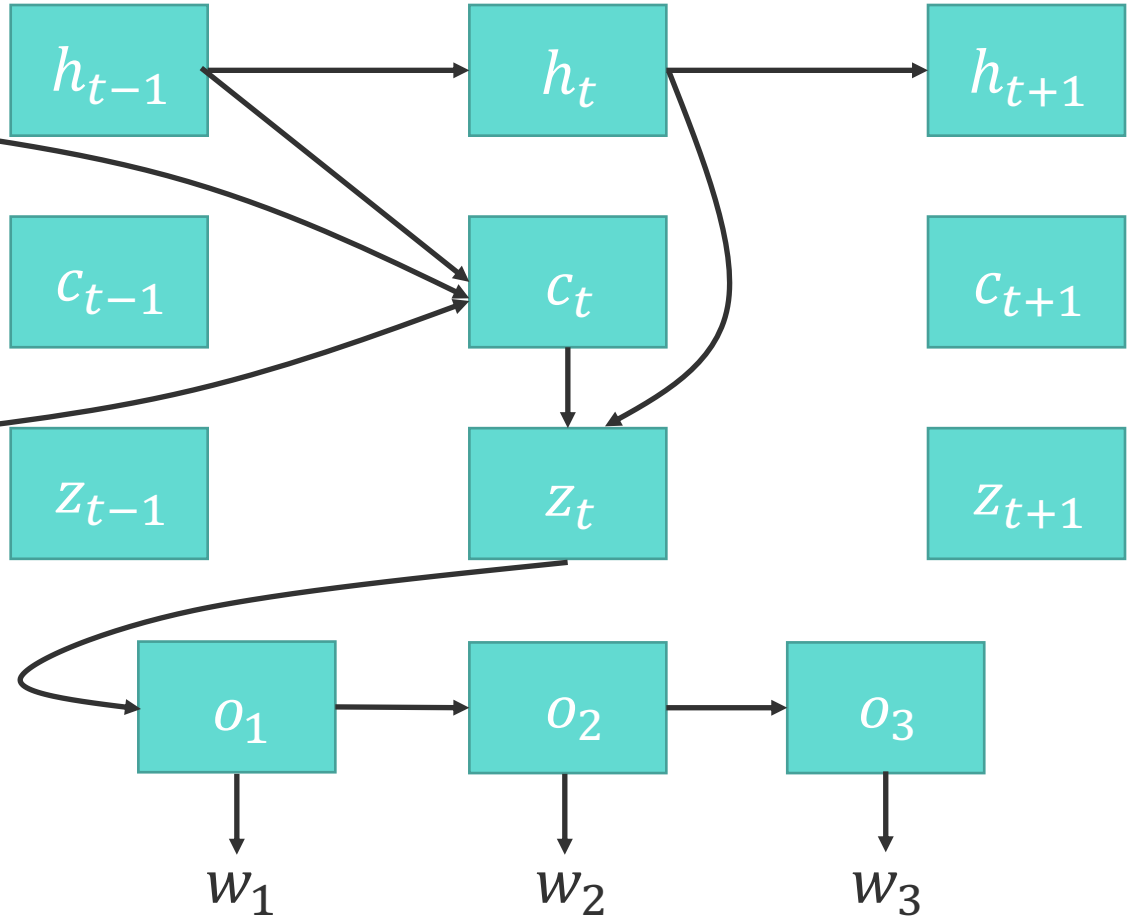


Large amount of *(image, tags)* data available for training

(II) Hierarchical Text Generation

Predicted Tags

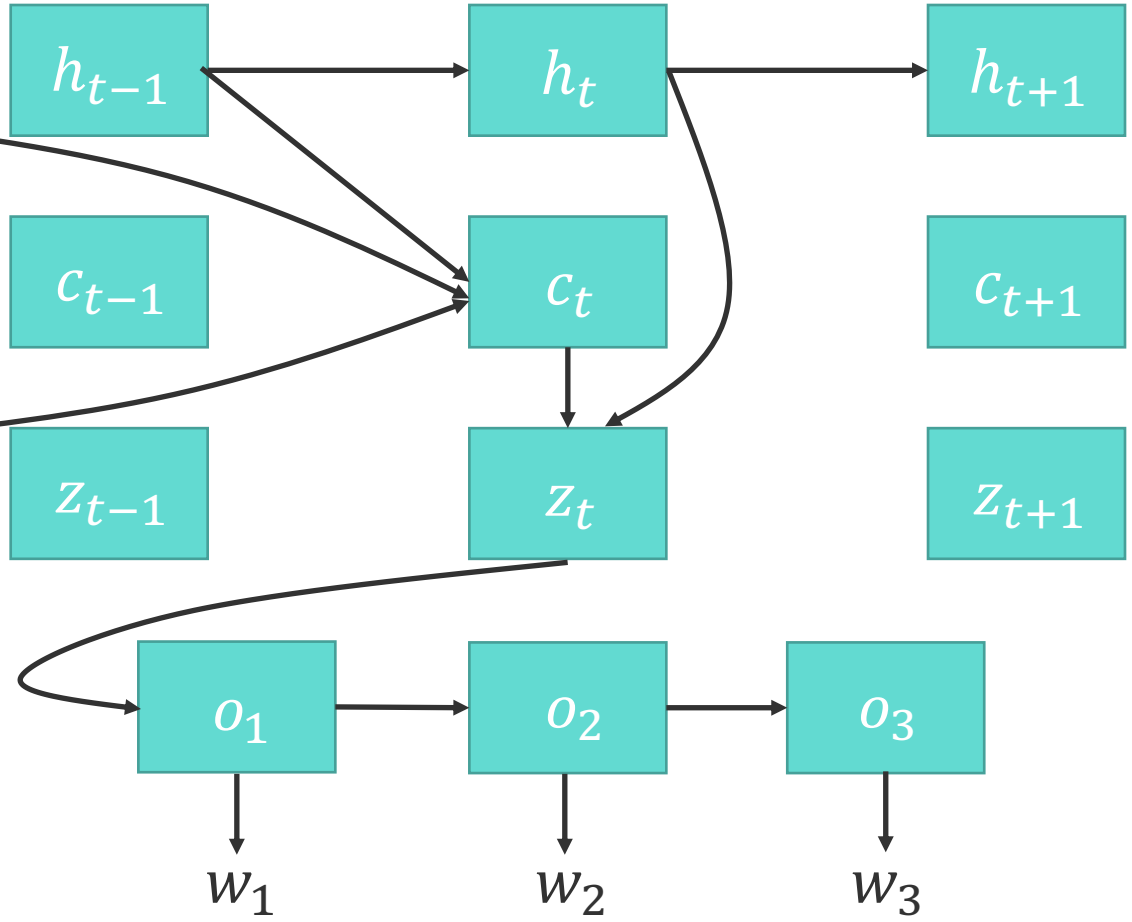
- Hyperdistention
- Emphysema
- Cicatrix



(II) Hierarchical Text Generation

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix

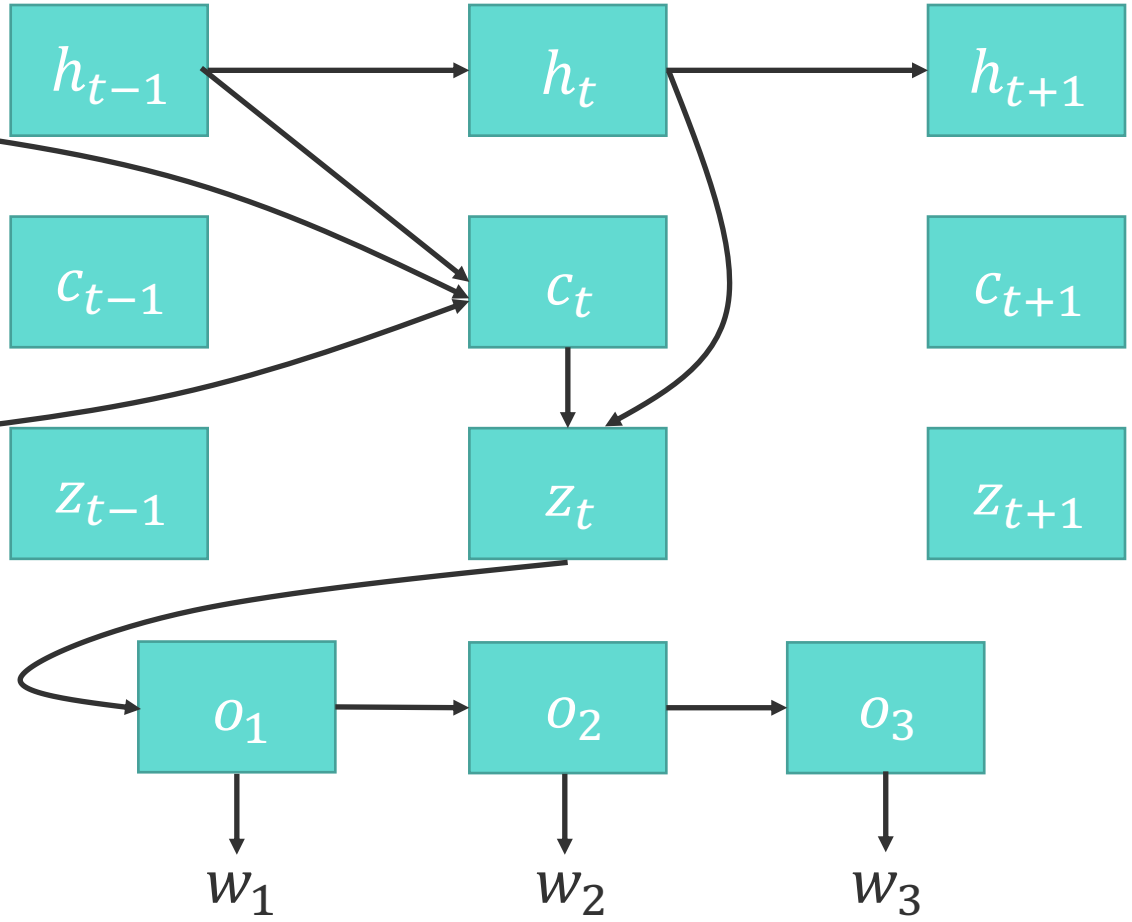


Sentence-level LSTM generates sentence-level hidden vectors

(II) Hierarchical Text Generation

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix



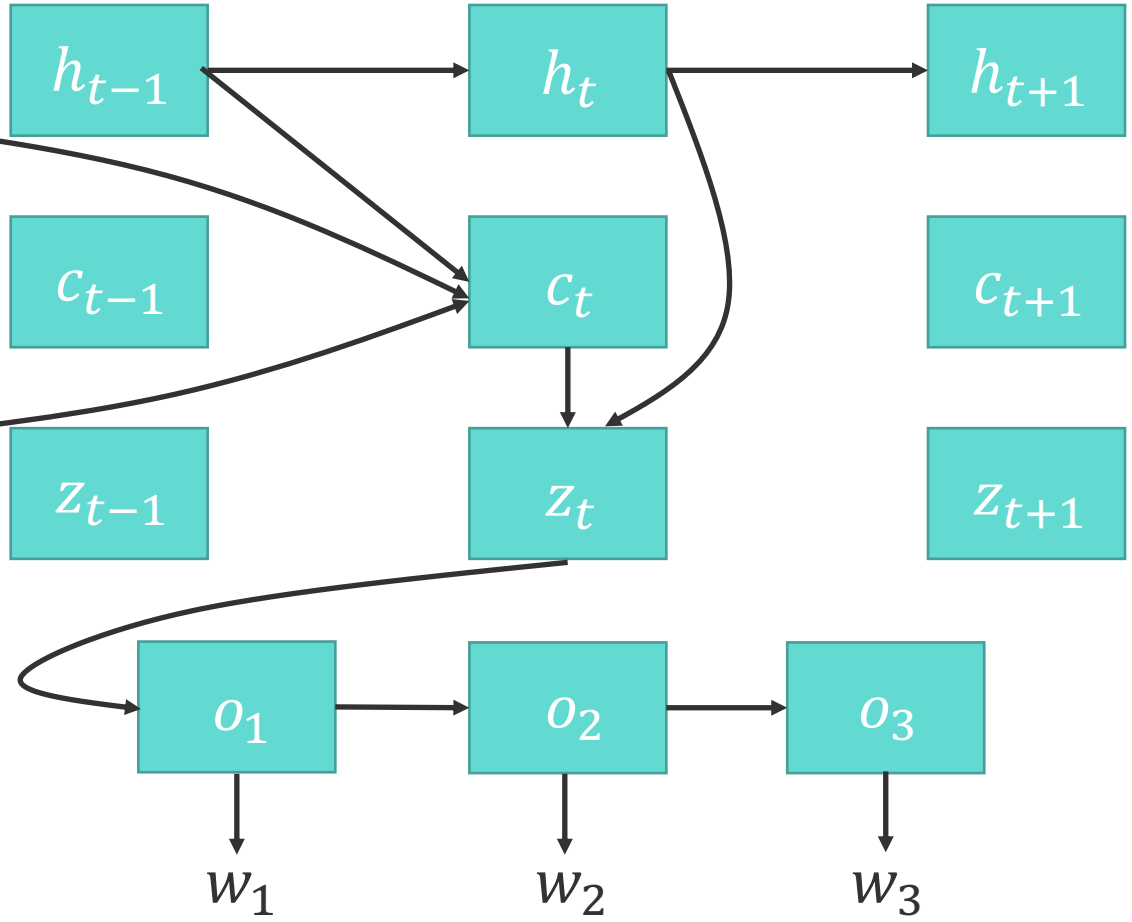
Sentence-level LSTM generates sentence-level hidden vectors

Visual-semantic co-attention module calculates contextual vectors

(II) Hierarchical Text Generation

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix



Sentence-level LSTM generates sentence-level hidden vectors

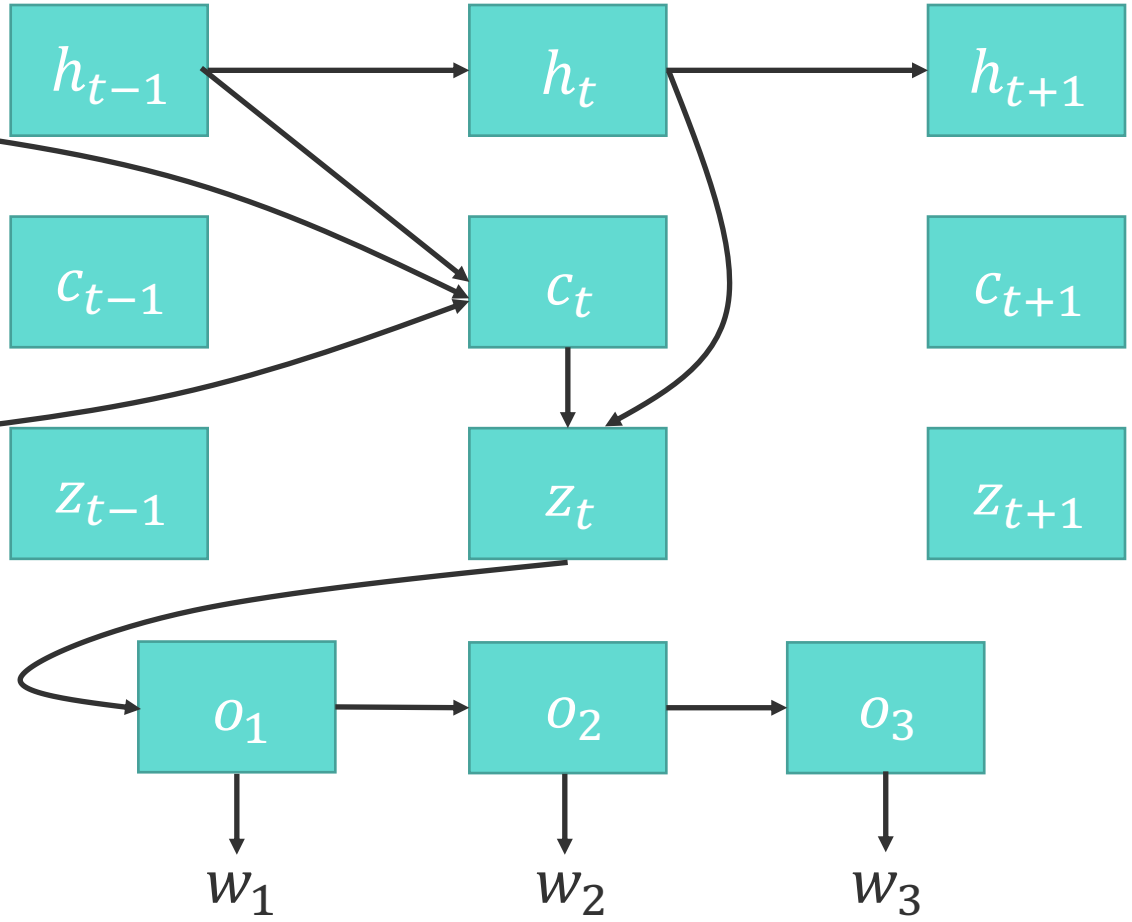
Visual-semantic co-attention module calculates contextual vectors

Topic layer calculates sentence topic vectors

(II) Hierarchical Text Generation

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix



Sentence-level LSTM generates sentence-level hidden vectors

Visual-semantic co-attention module calculates contextual vectors

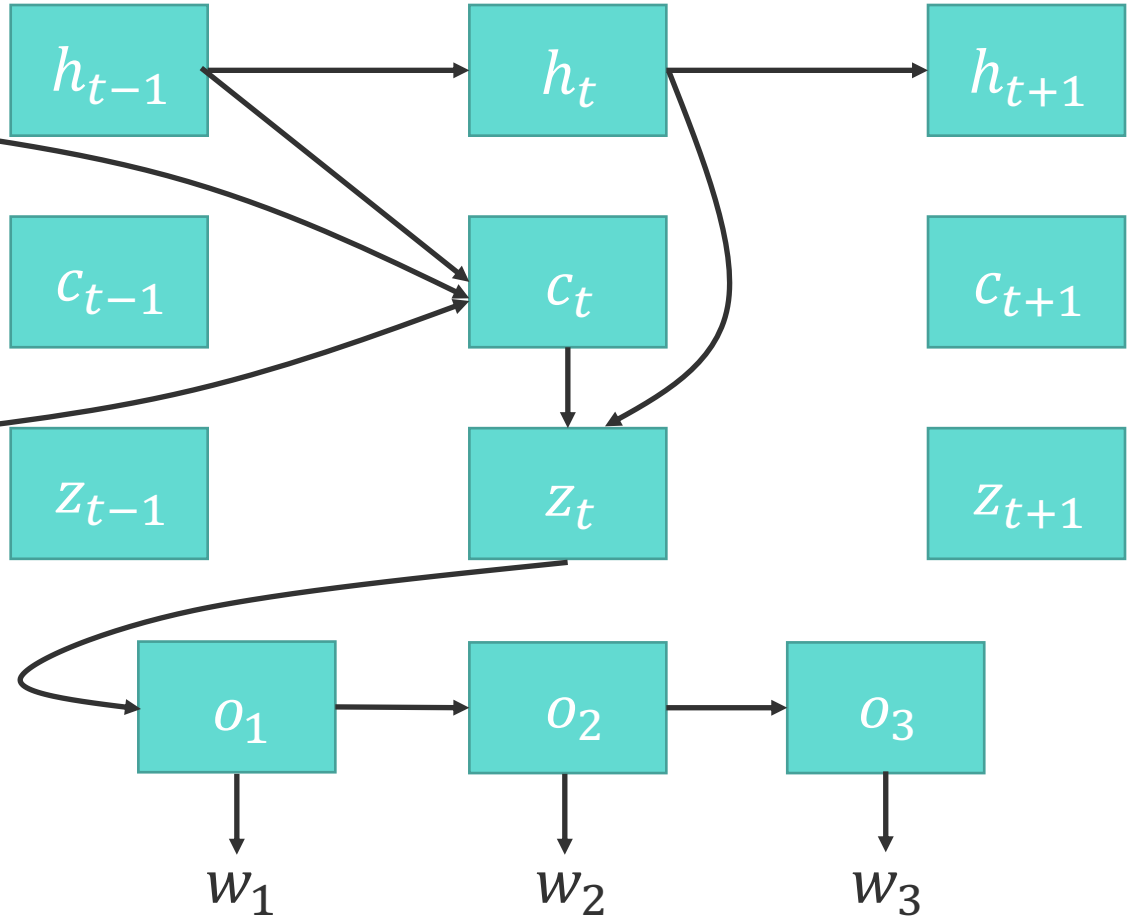
Topic layer calculates sentence topic vectors

Word-level LSTM generates word-level hidden vectors

(II) Hierarchical Text Generation

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix



Sentence-level LSTM generates sentence-level hidden vectors

Visual-semantic co-attention module calculates contextual vectors

Topic layer calculates sentence topic vectors

Word-level LSTM generates word-level hidden vectors

Generated words

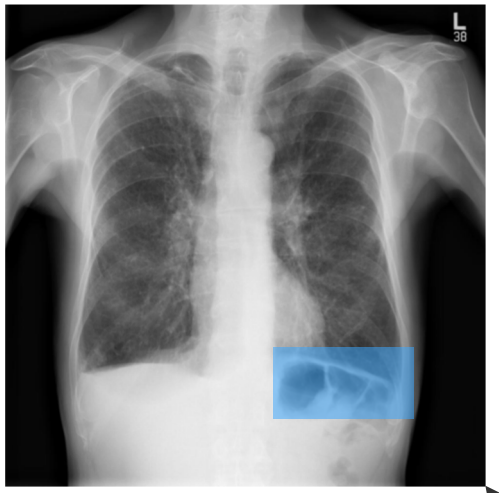
(III) Visual-Semantic Co-Attention

Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix

… There is chronic pleural-parenchymal scarring within the lung bases. No lobar consolidation is seen. …

(III) Visual-Semantic Co-Attention



Predicted Tags

- Hyperdistention
- Emphysema
- Cicatrix

Visual attention

Semantic attention

… There is chronic pleural-parenchymal scarring within the lung bases. No lobar consolidation is seen. …

Experiments

- Image-report data
  - Indiana University Chest X-ray Collection (IU X-Ray)
  - 7,470 image-report pairs
  - 5.7 sentences/image

Experiments

- Image-report data
  - Indiana University Chest X-ray Collection (IU X-Ray)
  - 7,470 image-report pairs
  - 5.7 sentences/image

- Image-tag data
  - NIH Chest X-ray images
  - 108,948 image-tags pairs
  - 14 lesion tags

Evaluation of Report Quality

	Methods	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE	CIDER
Baselines	CNN-RNN [1]	0.316	0.211	0.140	0.095	0.159	0.267	0.111
	LRCN [2]	0.369	0.229	0.149	0.099	0.155	0.278	0.190
	Soft ATT [3]	0.399	0.251	0.168	0.118	0.167	0.323	0.302
	ATT-RK [4]	0.369	0.226	0.151	0.108	0.171	0.323	0.155
Our methods	No-Attention	0.505	0.383	0.290	0.224	0.200	0.420	0.259
	Semantic-Only	0.504	0.371	0.291	0.230	0.207	0.418	0.286
	Visual-Only	0.507	0.373	0.297	0.238	0.211	0.426	0.300
	Co-Attention	**0.517**	**0.386**	**0.306**	**0.247**	**0.217**	**0.447**	**0.327**

Comparison with the state-of-the-art image captioning methods

Evaluation of Clinical Correctness

- Presence/absence of lesions in the report

Evaluation of Clinical Correctness

- Presence/absence of lesions in the report
- Human evaluators
  - manually read the report
  - check whether a lesion exists

Evaluation of Clinical Correctness

- Presence/absence of lesions in the report
- Human evaluators
  - manually read the report
  - check whether a lesion exists

Generated report:

Normal cardiomediastinal silhouette. Interval improvement in lung volumes bilaterally. Improved aeration of the right and left lung bases. Bilateral small pleural effusions and left base atelectatic change, with interval improvement. Visualized XXXX of the chest XXXX are within normal limits.

Evaluation of Clinical Correctness

- Presence/absence of lesions in the report
- Human evaluators
  - manually read the report
  - check whether a lesion exists

Generated report:

Normal cardiomediastinal silhouette. Interval improvement in lung volumes bilaterally. Improved aeration of the right and left lung bases. Bilateral small pleural effusions and left base atelectatic change, with interval improvement. Visualized XXXX of the chest XXXX are within normal limits.

→

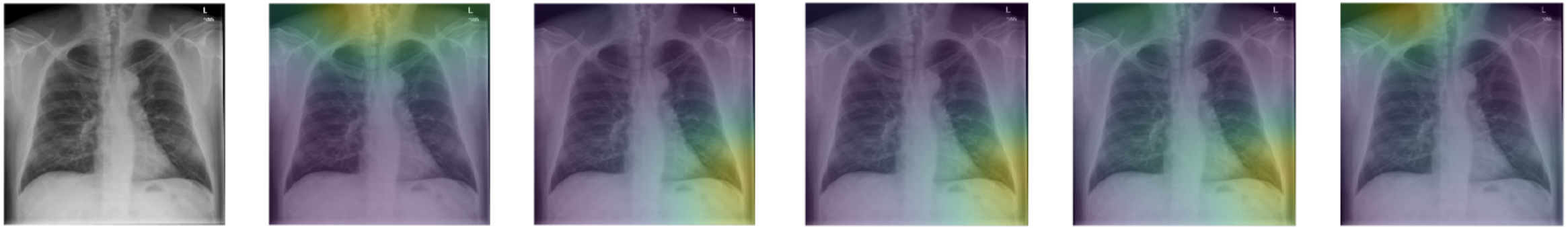
Lesions:

- Has effusion and atelectasis
- No other lesions

Evaluation of Clinical Correctness (cont'd)

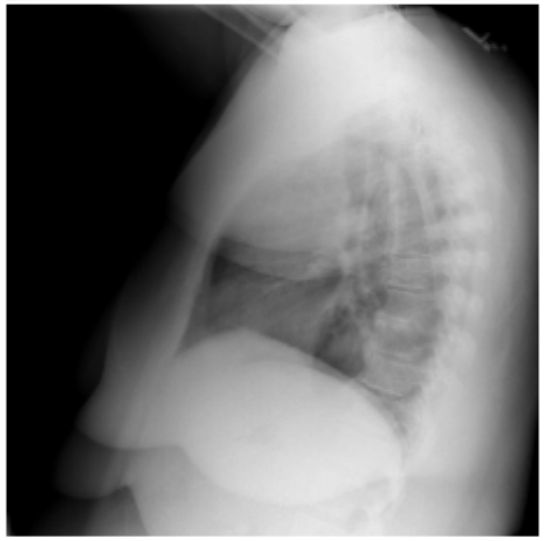
	No-Attention	Visual-Only	Semantic-Only	**Co-Attention**
Macro-F1	0.49	0.51	0.75	**0.79**

Visualization of Co-attention



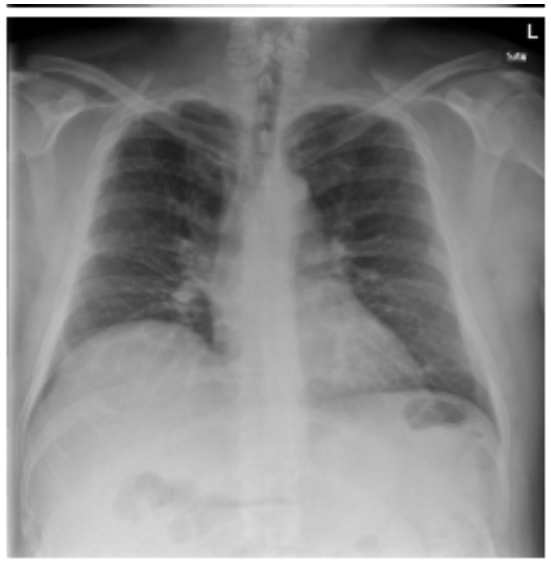
degenerative change; obstruction	normal; degenerative change; nodule; calcified granuloma; hyper expansion; granulomatous disease; granuloma; **pneumonia; scarring; sternotomy**	**normal**; degenerative change; **nodule**; calcified granuloma; hyper expansion; granulomatous disease; **granuloma**; pneumonia; scarring; sternotomy	**normal; degenerative change; nodule**; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy	**normal; degenerative change; nodule**; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy	**normal; degenerative change; nodule**; calcified granuloma; hyper expansion; granulomatous disease; granuloma; pneumonia; scarring; sternotomy
	No acute intrathoracic abnormality.	No bony abnormality.	No focal areas of pulmonary consolidation.	Breast motion.	There is an age indeterminate deformity of a mid-thoracic vertebral body.

Generated Examples

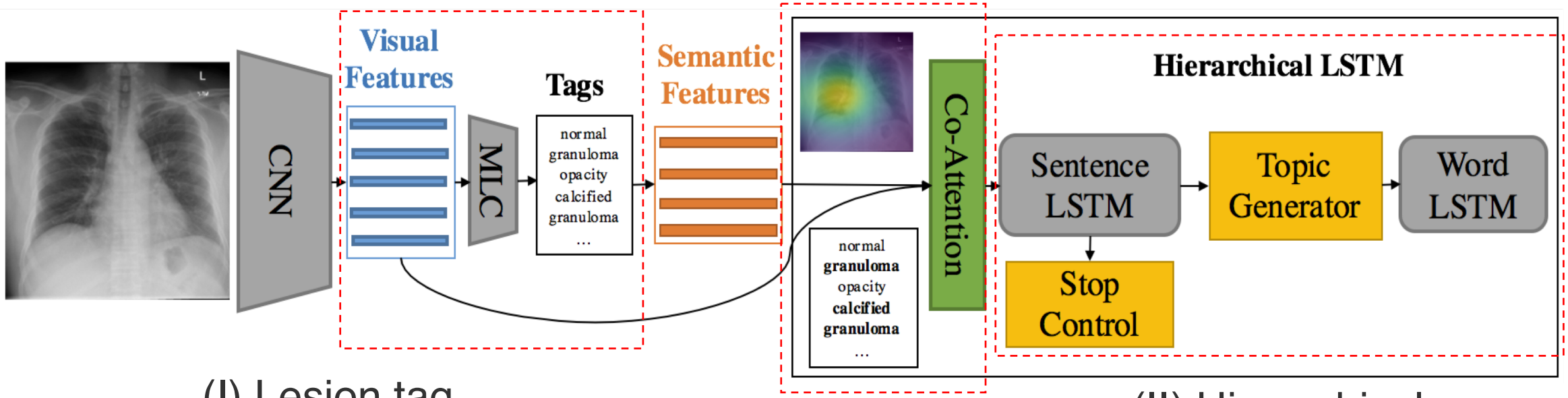
	Ground Truth	Ours-CoAttention
	No active disease. The heart and lungs have in the interval. Both lungs are clear and expanded. Heart and mediastinum normal.	No active disease. The heart and lungs have in the interval. Lungs are clear and expanded. Cardiomediastinal silhouette is within normal limits. No pleural effusion or pneumothorax is seen. No pleural effusion. No cavitary or pneumothorax.
	No evidence of active disease. The lungs are clear. There is no focal airspace consolidation. No pleural effusion or pneumothorax. Heart size and mediastinal contour are within normal limits. There are multilevel degenerative changes of the spine.	No acute cardiopulmonary findings. Heart size is not enlarged. No focal airspace consolidation suspicious pulmonary opacity large pleural effusion or pneumothorax. No focal areas of consolidation. Degenerative changes of the spine. This is moderate exam of the hydropneumothorax. Lungs are clear. There is no focal airspace consolidation pleural effusion or pneumothorax.

(The underlined sentences describe detected abnormalities)

Failure Cases

	Ground Truth	Generated
	No acute cardiopulmonary abnormality. Normal heart size mediastinal contours. Eventration of the right hemidiaphragm. No focal airspace consolidation. No pleural effusion or pneumothorax.	No acute cardiopulmonary abnormality. Stable appearance of the thoracic aorta. The right lateral lower lobe is noted in the right lower right midlung. No large pleural effusion or focal airspace disease. Mild interstitial opacities. Atherosclerotic calcifications bony structures bilaterally. There is no pleural effusion or pneumothorax developed in the right lower lobe.
	No acute cardiopulmonary abnormality. Heart size appears within normal limits . Pulmonary vasculature appears within normal limits. Overlying the middle cardiac silhouette representing a hiatal hernia. No focal consolidation pleural effusion or pneumothorax. No acute bony abnormality.	No active disease. The heart and lungs have in the interval. Nipple and lateral lucency in the lungs suggestive of focal airspace disease. The lungs are hyperexpanded consistent with emphysema in the left lower lobe. This is most at the upper lobes. This may indicate hypoventilated irregularities or effusions. The lungs are otherwise grossly clear. Resolution of by normal pleural effusion.

Recap: Model Architecture



(I) Lesion tag classification

(III) Visual-semantic co-attention

(II) Hierarchical text generation

Outline

- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Improving Abnormality Description

- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated

Improving Abnormality Description

- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description

- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated
- A pure **generation-based** model tends to **overfit** to normal findings
  - Make it easier to generate fluent, natural-looking sentences

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description

- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated
- A pure **generation-based** model tends to **overfit** to normal findings
  - Make it easier to generate fluent, natural-looking sentences

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description

- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated
- A pure **generation-based** model tends to **overfit** to normal findings
  - Make it easier to generate fluent, natural-looking sentences
- **Solution:** alleviate the burden of generating natural sentences

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description

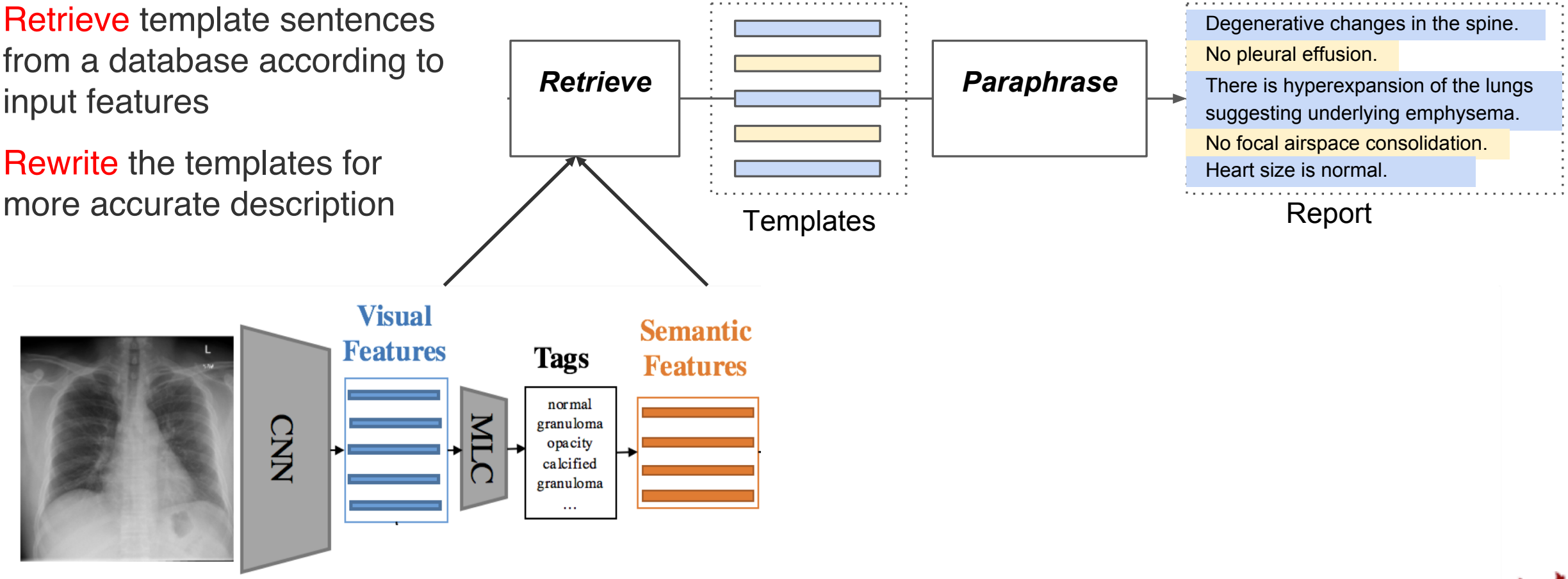
- Normal v.s. abnormal findings
  - Normal findings: dominate the reports; general, templated descriptions
  - Abnormal findings: relatively rare, but critical; more specifically stated
- A pure **generation-based** model tends to **overfit** to normal findings
  - Make it easier to generate fluent, natural-looking sentences
- Solution: alleviate the burden of generating natural sentences
  - Method: retrieval + generation

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description - I: Retrieval + Generation

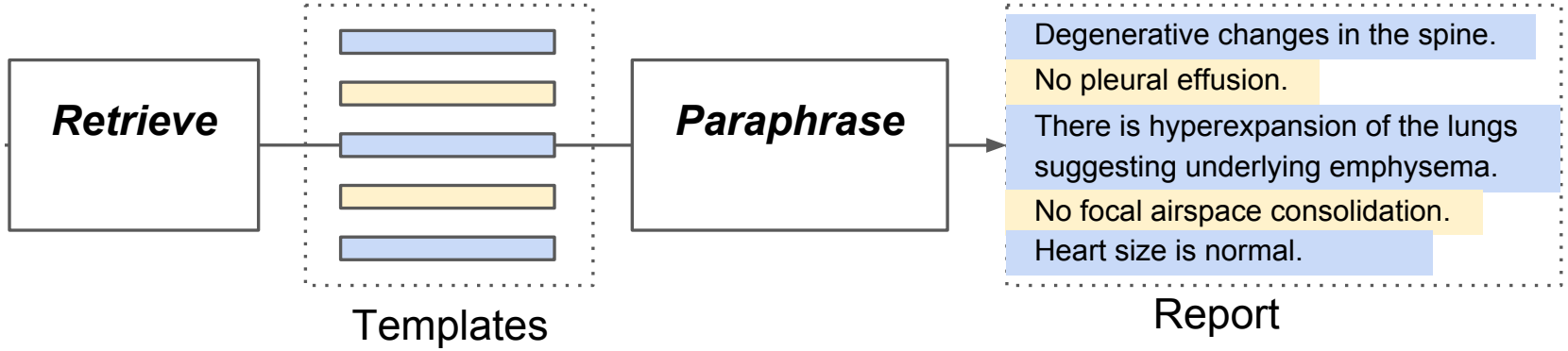
- Retrieve template sentences from a database according to input features
- Rewrite the templates for more accurate description



[Li et al., 2018; 2019]

Improving Abnormality Description - I: Retrieval + Generation

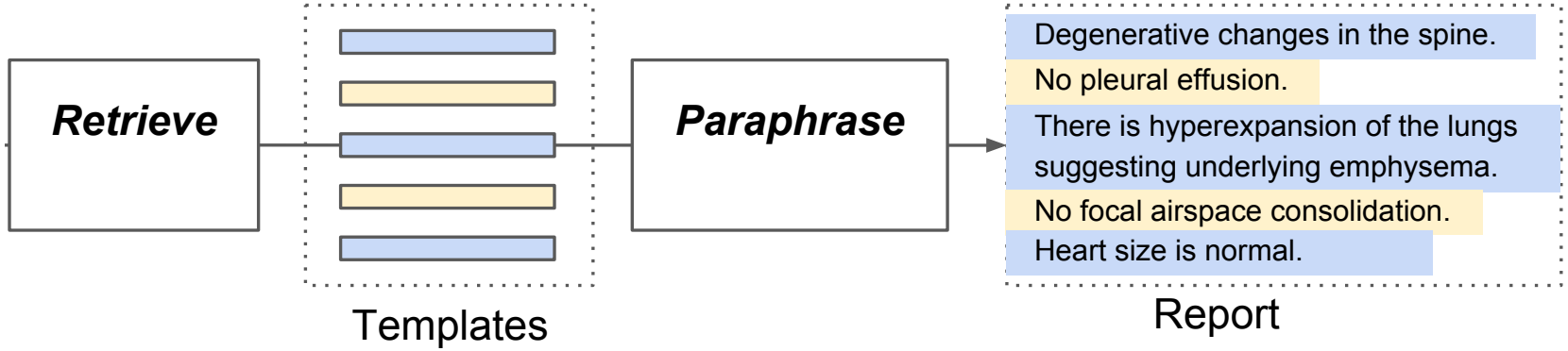
- Retrieve template sentences from a database according to input features
- Rewrite the templates for more accurate description



Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description - I: Retrieval + Generation

- Retrieve template sentences from a database according to input features
- Rewrite the templates for more accurate description

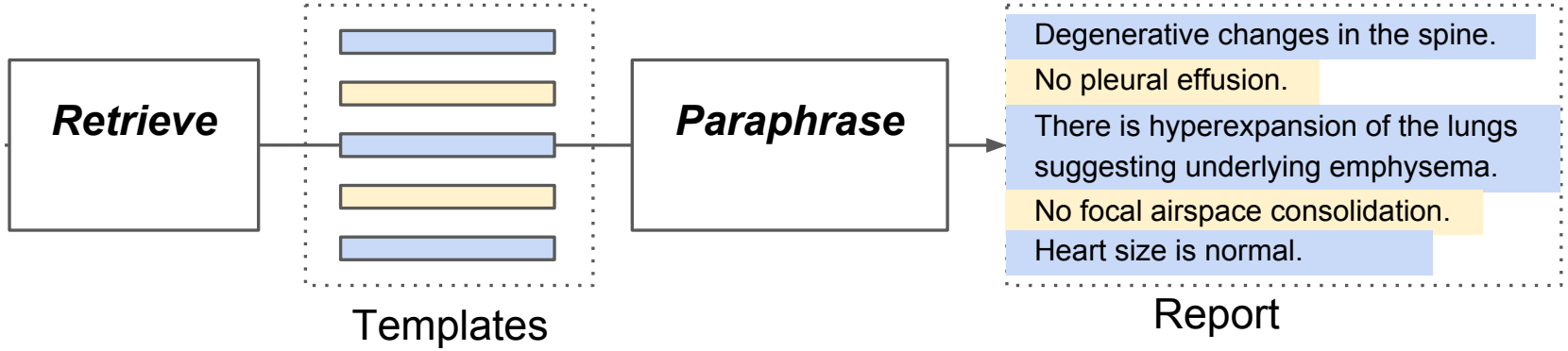


G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

Improving Abnormality Description - I: Retrieval + Generation

- Retrieve template sentences from a database according to input features
- Rewrite the templates for more accurate description



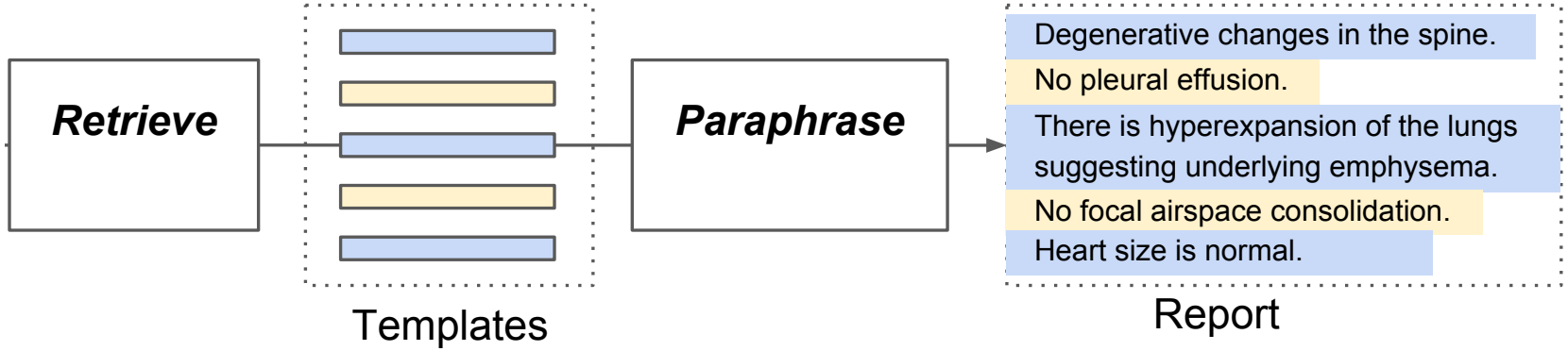
Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

R: The heart size is normal. There is mild effusion. No acute bony abnormalities.

Improving Abnormality Description - I: Retrieval + Generation

- Retrieve template sentences from a database according to input features
- Rewrite the templates for more accurate description



Ground Truth: The heart size and mediastinal contours appear within normal limits. There is blunting of the right lateral costophrenic sulcus which could be secondary to a small effusion versus scarring. No focal airspace consolidation or pneumothorax. No acute bony abnormalities.

G: The heart size is normal. No pleural effusion or pneumothorax. No acute bony abnormalities.

R: The heart size is normal. There is mild effusion. No acute bony abnormalities.

R+G: The heart size is normal. There is blunting of costophrenic sulcus suggesting a small effusion. No acute bony abnormalities.

Improving Abnormality Description - II

- Structured medical knowledge

Improving Abnormality Description - II

- Structured medical knowledge
  - Common types of abnormality:
    - presence of abnormal attributes
    - absence of typical attributes
    - abnormal change of object shape or location

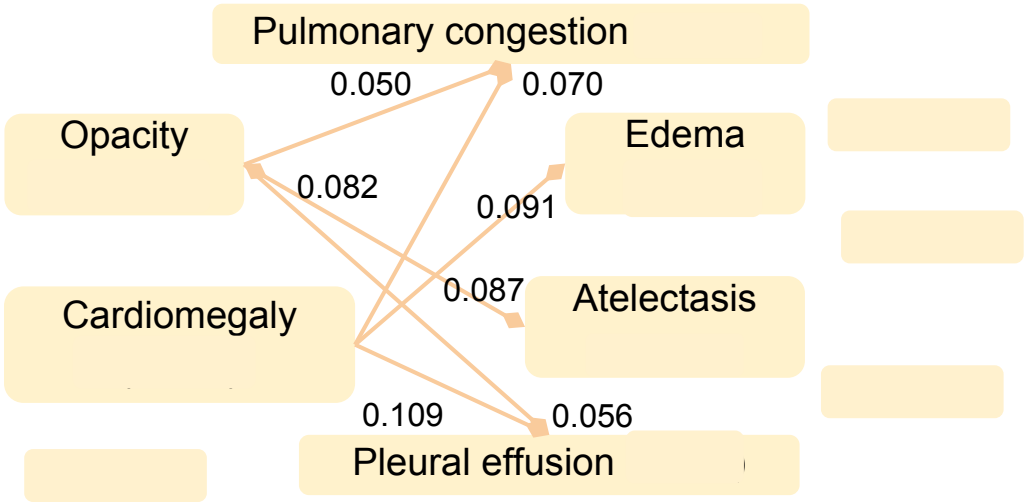
Improving Abnormality Description - II

- Structured medical knowledge
  - Common types of abnormality:
    - presence of abnormal attributes
    - absence of typical attributes
    - abnormal change of object shape or location
  - Different abnormality exhibits certain correlation patterns

Improving Abnormality Description - II

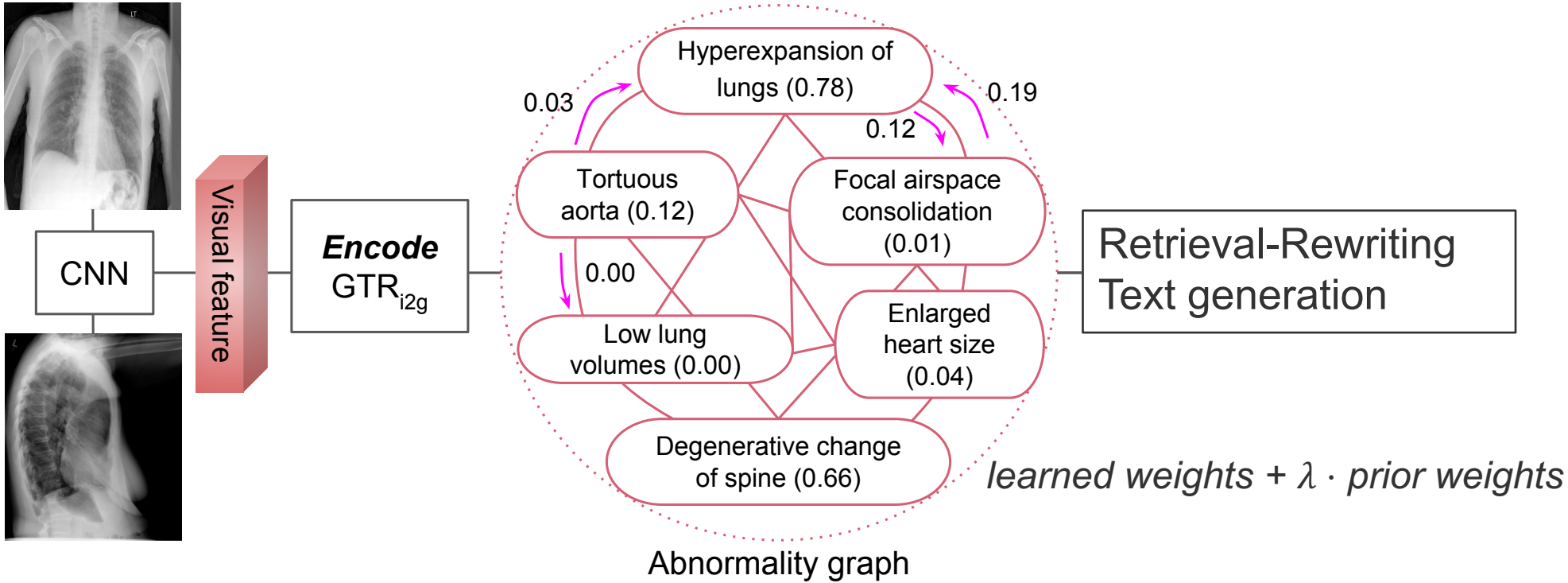
- Structured medical knowledge
  - Common types of abnormality:
    - presence of abnormal attributes
    - absence of typical attributes
    - abnormal change of object shape or location
  - Different abnormality exhibits certain correlation patterns

Construct a prior abnormality graph from data that captures abnormality co-occurrence patterns



Improving Abnormality Description - II: Incorporating structured knowledge

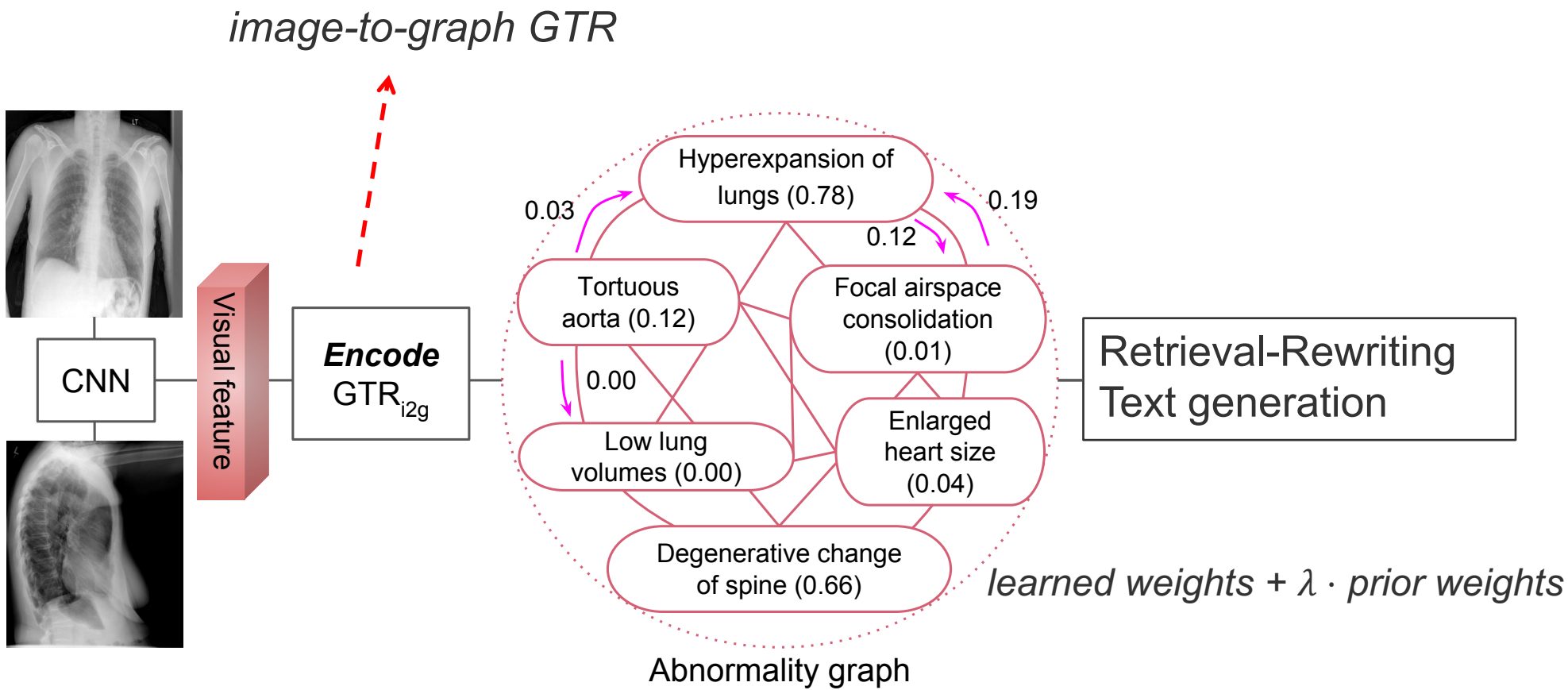
- Graph Transformer (GTR):
  - A universal transformation model between modalities



learned weights + λ · prior weights

Improving Abnormality Description - II: Incorporating structured knowledge

- Graph Transformer (GTR):
  - A universal transformation model between modalities



Improving Abnormality Description - II: Incorporating structured knowledge

- Graph Transformer (GTR):
  - A universal transformation model between modalities

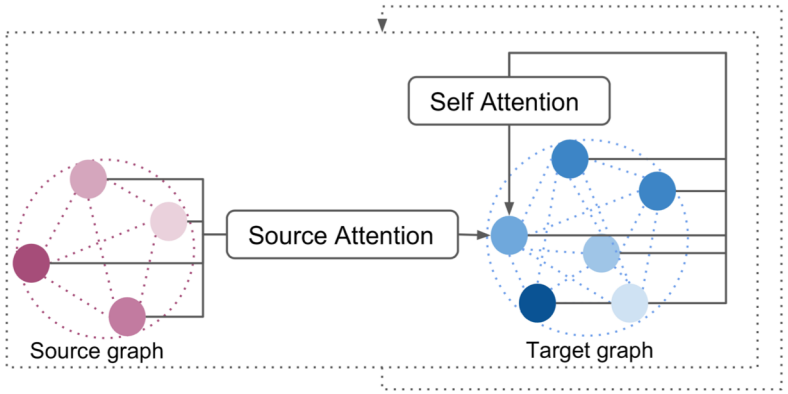
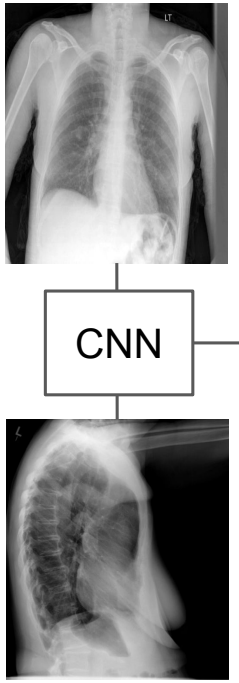


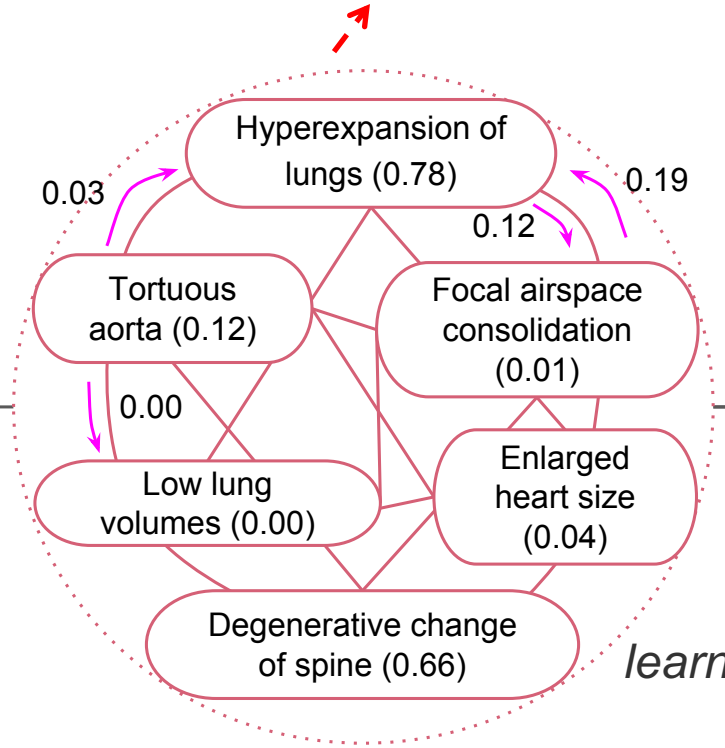
image-to-graph GTR *graph-to-graph GTR*



CNN

Visual feature

Encode GTR_{i2g}

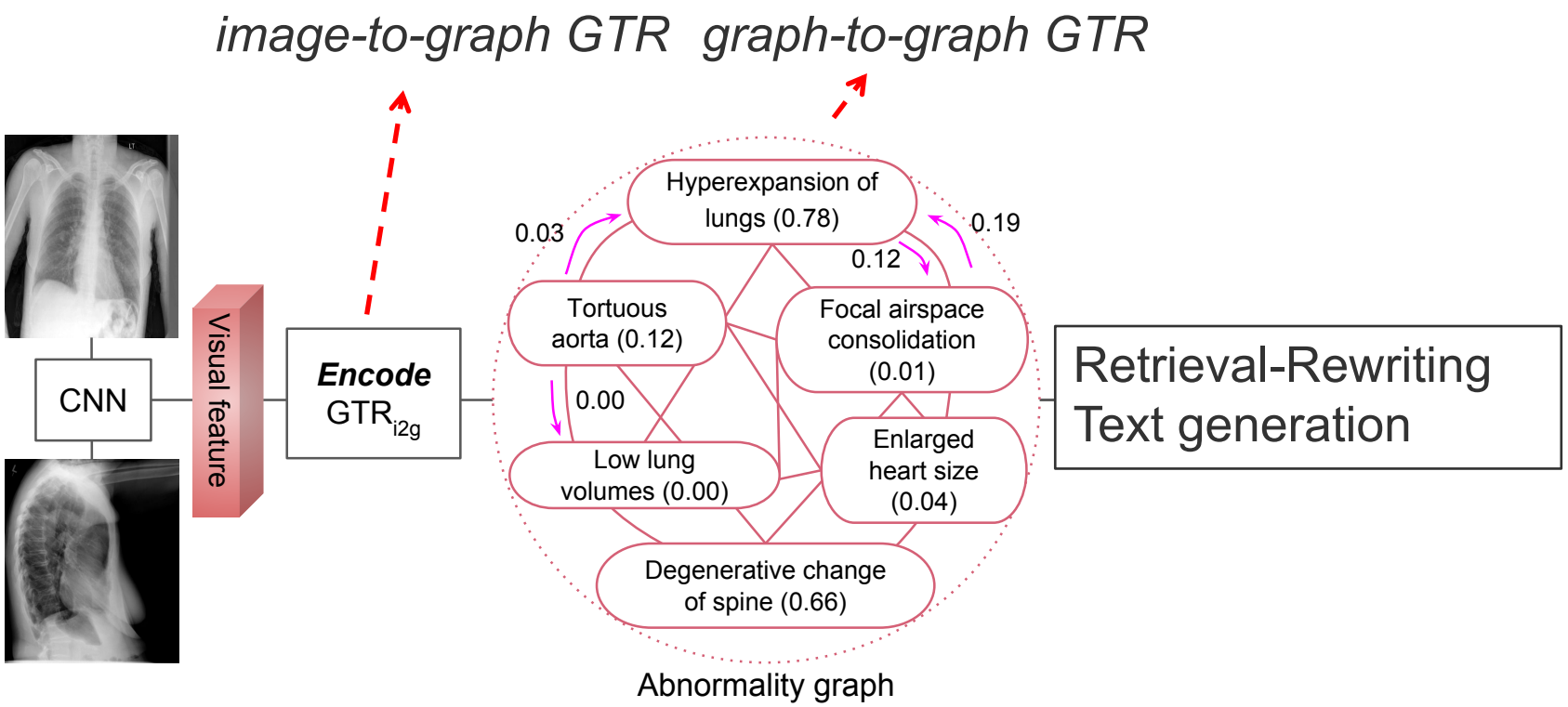


Abnormality graph

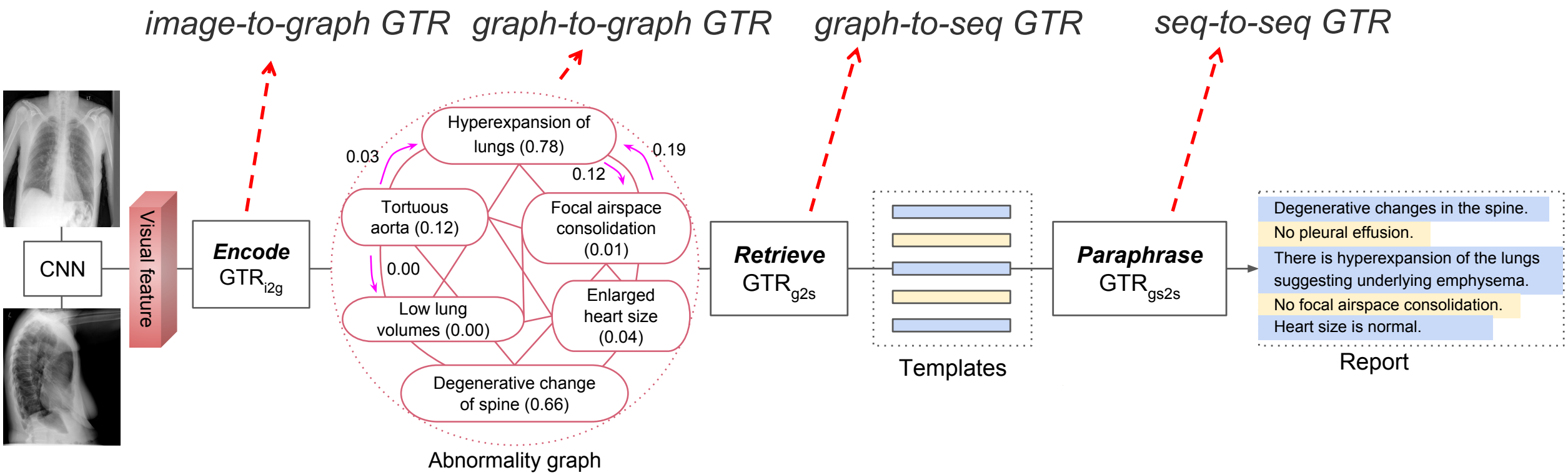
Retrieval-Rewriting Text generation

learned weights + $\lambda \cdot$ prior weights

Improving Abnormality Description



Improving Abnormality Description



Empirical Results

Dataset	Model	CIDEr	ROUGE-L	BLEU-1	BLEU-2	BLEU-3	BLEU-4	Hit (%)
IU X-Ray	CNN-RNN	0.294	0.307	0.216	0.124	0.087	0.066	–
	LRCN	0.285	0.307	0.223	0.128	0.089	0.068	–
	AdaAtt	0.296	0.308	0.220	0.127	0.089	0.069	–
	Att2in	0.297	0.307	0.224	0.129	0.089	0.068	–
	CoAtt*	0.277	**0.369**	0.455	0.288	0.205	0.154	24.100
	HRGR-Agent	**0.343**	0.322	0.438	0.298	0.208	0.151	
	KER	0.318	0.335	0.455	0.304	0.210	–	
	KERP	0.280	0.339	**0.482**	**0.325**	**0.226**	**0.162**	**57.425**
CX-CHR	CNN-RNN	1.580	0.578	0.592	0.506	0.450	0.411	–
	LRCN	1.589	0.577	0.593	0.508	0.459	0.413	–
	AdaAtt	1.568	0.576	0.588	0.505	0.446	0.409	–
	Att2in	1.564	0.576	0.587	0.503	0.447	0.403	25.937
	HRG	2.800	0.588	0.629	0.547	0.497	0.463	–
	HRGR-Agent	**2.895**	0.612	**0.673**	0.587	0.530	**0.486**	-
	KER	0.817	0.552	0.609	0.489	0.400	0.335	–
	KERP	2.850	**0.618**	**0.673**	**0.588**	**0.532**	0.473	**67.820**

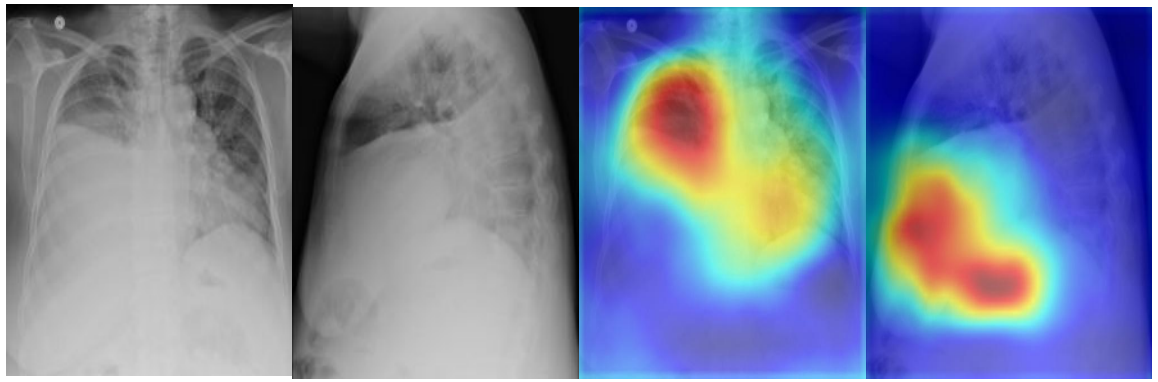
Table 1: Automatic and human evaluation on IU X-Ray (upper part) and CX-CHR dataset (lower part) compared with CNN-RNN (Vinyals et al. 2015), LRCN (Donahue et al. 2015), AdaAtt (Lu et al. 2017), Att2in (Rennie et al. 2017), CoAtt (Jing, Xie, and Xing 2018), and HRGR-Agent (Li et al. 2018). * indicates re-training and evaluation on our data split.

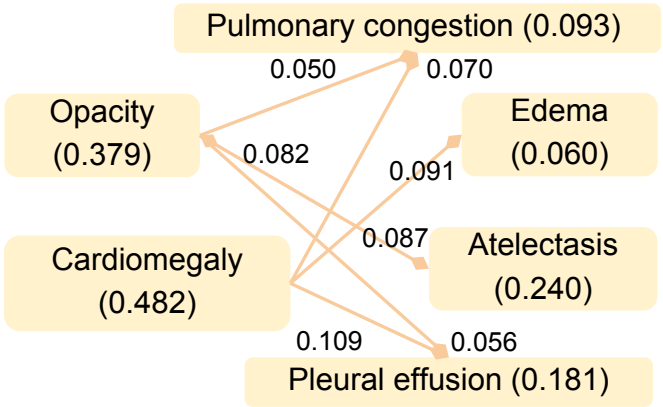
Empirical Examples

Images

Attention Maps

Abnormality graph





Ground truth report

The **cardiac silhouette is mildly enlarged.** Mediastinal contours are within normal limits. The **pulmonary vascularity is increased.** There is **large right - sided pleural effusion and probable underlying associated compressive atelectasis.** Mild **perihilar xxxx opacities , xxxx edema**. No pneumothorax is seen.

Retrieved templates

There is a **small left pleural effusion.** No pneumothorax.
Heart size normal
the lungs are clear.

Generated report

There are **bilateral pleural effusions with bibasilar airspace disease , <u>right greater than left.</u>** No pneumothorax. **<u>Cardiac silhouette is at the upper limits of normal.</u>** Clear lungs.

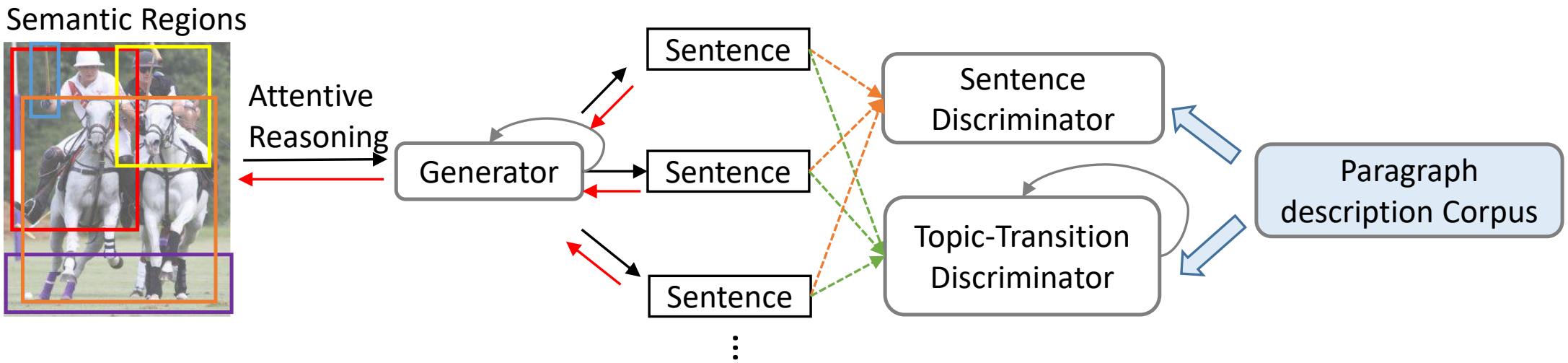
Summary: Medical Image Report Generation

- Challenges: (1) abnormality detection and description; (2) Long paragraph generation; (3) Accurate, visually-grounded description
- A set of techniques for solution
  - Cross modalities: images, text, graphs
    - *Graph Transformer*
  - Long text generation in a more structured way
    - *Hierarchical generation*
    - *Combination of retrieval + generation*
  - Integrating structured knowledge, visual grounding
    - *Structured medical knowledge*
    - *Visual-semantic co-attention*

Outline

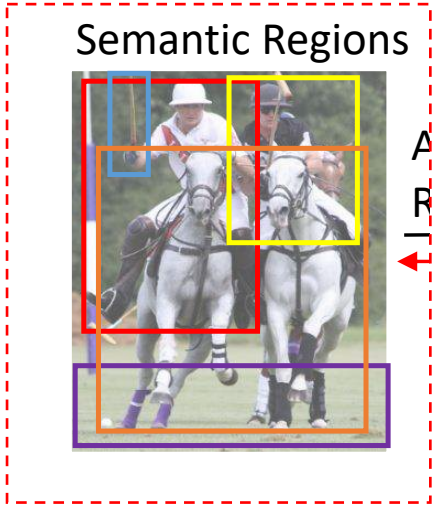
- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Paragraph description of natural images

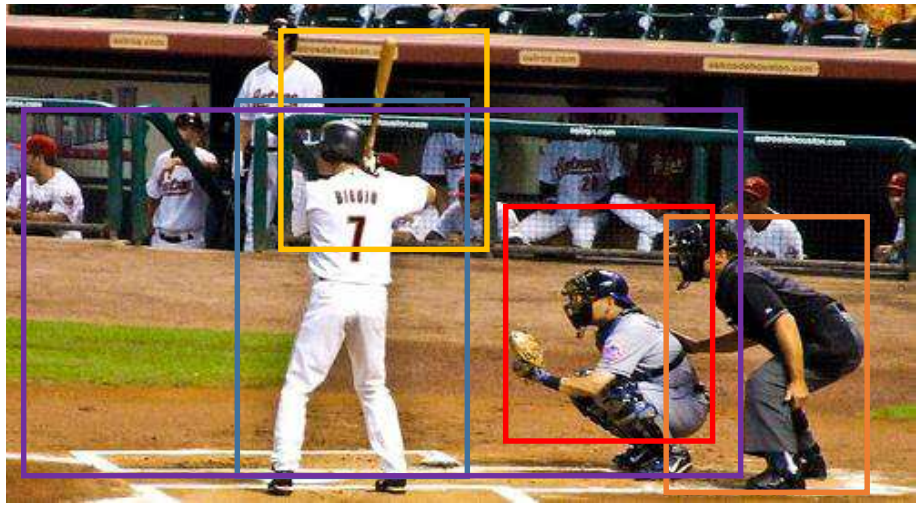


[Liang et al., 2017]

Paragraph description of natural images



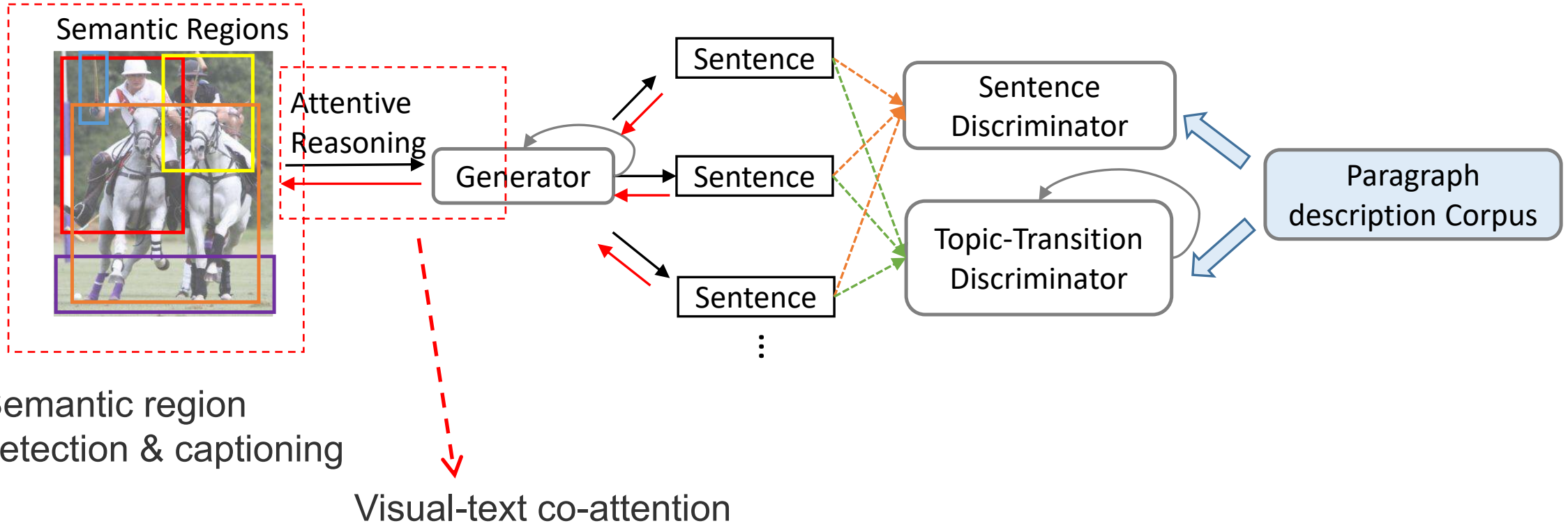
Semantic region detection & captioning



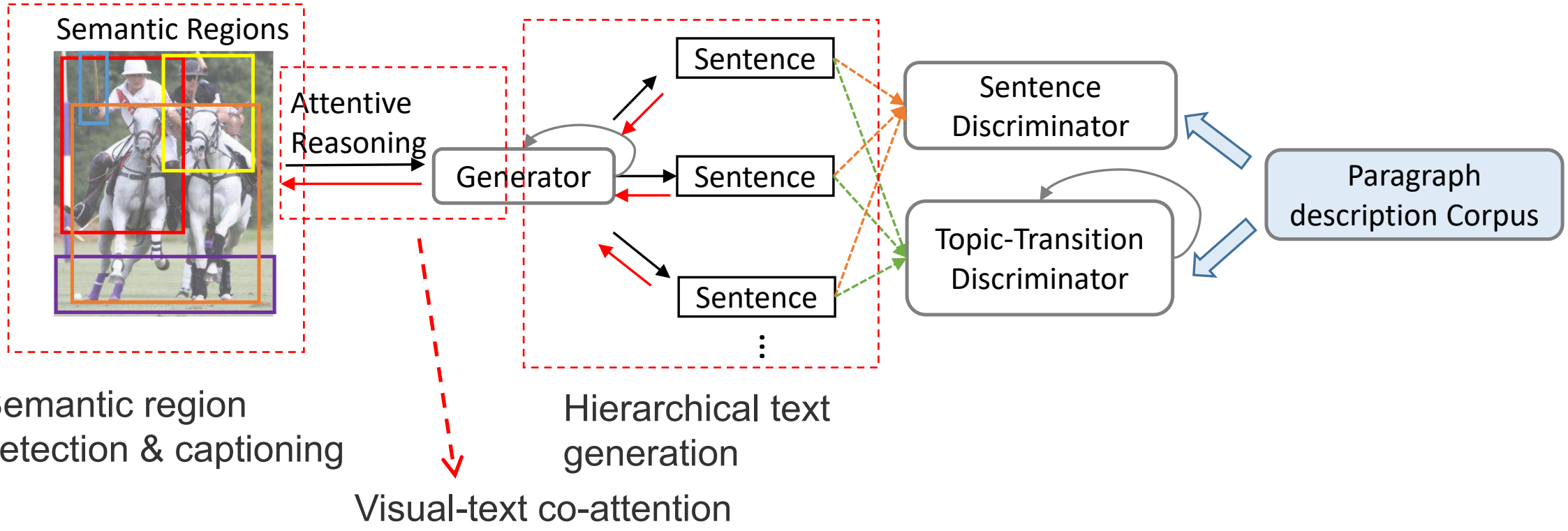
Local Phrases

- people playing baseball
- a man wearing white shirt and pants
- man holding a baseball bat
- person wearing a helmet in the field
- a man bending over

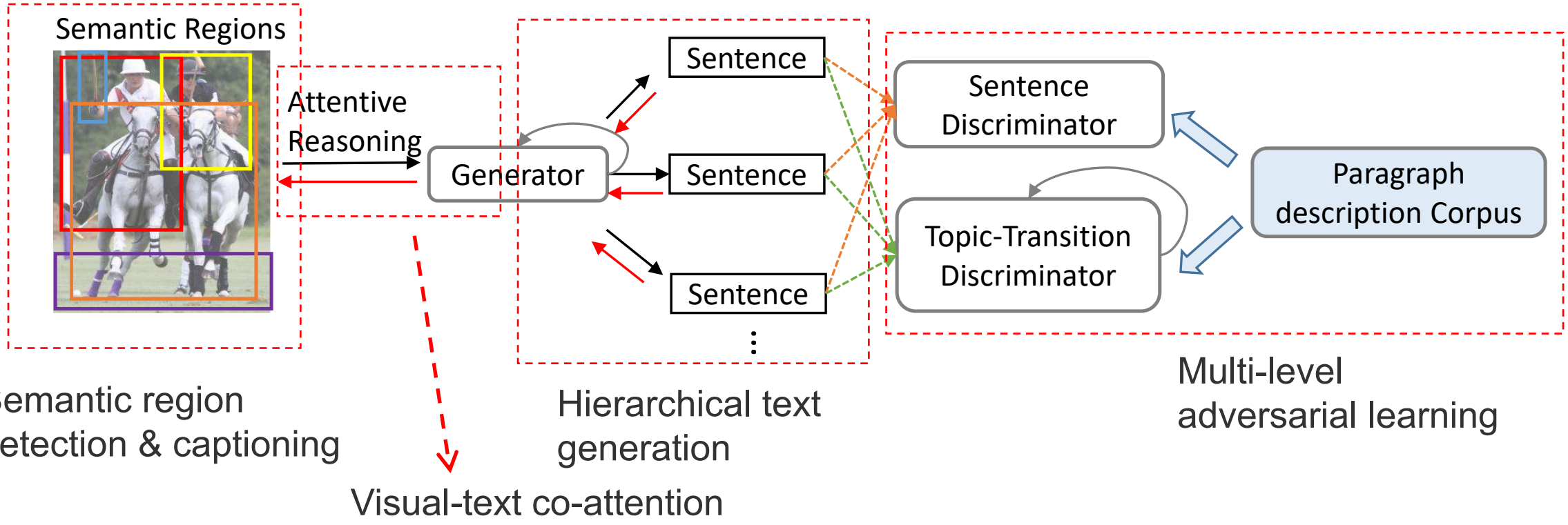
Paragraph description of natural images



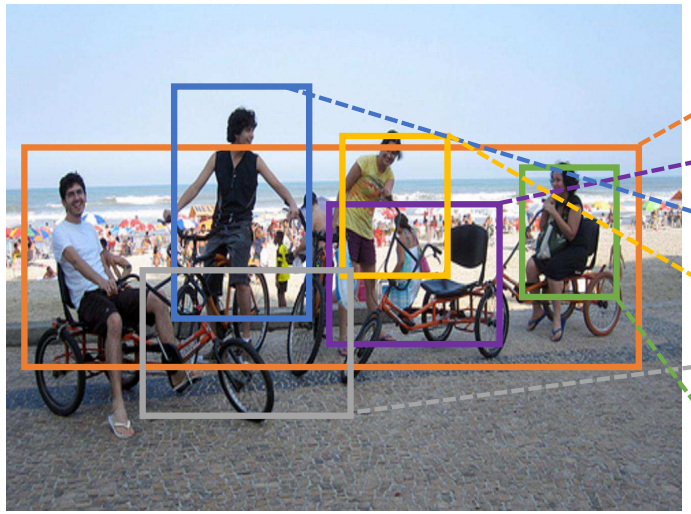
Paragraph description of natural images



Paragraph description of natural images



Paragraph description of natural images



1) people riding a bike
2) a bicycle parked on the sidewalk
3) man wearing a black shirt
4) a woman wearing a yellow shirt
5) a red and black bike
6) a woman wearing a shirt

Paragraph: *A group of people are riding bikes. There are two people riding bikes parked on the sidewalk. He is wearing a black shirt and jeans. A woman is wearing a short sleeve yellow shirt and shorts. There are many other people on the red and black bikes. A woman wearing a shirt is riding a bicycle.*

Paragraph description of natural images

Method	METEOR	CIDEr	BLEU-1	BLEU-2	BLEU-3	BLEU-4
Sentence-Concat	12.05	6.82	31.11	15.10	7.56	3.98
Template	14.31	12.15	37.47	21.02	12.03	7.38
Image-Flat [14]	12.82	11.06	34.04	19.95	12.20	7.71
Regions-Hierarchical [16]	15.95	13.52	41.90	24.11	14.23	8.69
RTT-GAN (Semi- w/o discriminator)	12.35	8.96	33.82	17.40	9.01	5.88
RTT-GAN (Semi- w/o sentence D)	11.22	10.04	35.29	19.13	11.55	6.02
RTT-GAN (Semi- w/o topic-transition D)	12.68	12.77	37.20	20.51	12.08	6.91
RTT-GAN (Semi-)	14.08	13.07	39.22	22.50	13.34	7.75
RTT-GAN (Fully- w/o discriminator)	16.57	15.07	41.86	24.33	14.56	8.99
RTT-GAN (Fully-)	17.12	16.87	41.99	24.86	14.89	9.03
RTT-GAN (Semi + Fully)	**18.39**	**20.36**	**42.06**	**25.35**	**14.92**	**9.21**
Human	19.22	28.55	42.88	25.68	15.55	9.66

Outline

- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Beyond Image-to-Text Generation

- Controlled Generation of Text [Hu et al., 2017]
  - With **control** over content, attributes, stylistic characteristics, …
  - At sentence / discourse level
  - In supervised / unsupervised settings

Beyond Image-to-Text Generation

- Controlled Generation of Text [Hu et al., 2017]
  - With **control** over content, attributes, stylistic characteristics, …
  - At sentence / discourse level
  - In supervised / unsupervised settings
  - E.g., **Text Style Transfer** [Hu et al., 2017; Yang et al., 2018]

Original: if i could give them a zero star review i would !
Output: if i can give them a five star review i would !

Original: the food is fresh and the environment is good
Output: the food is bland and the environment is bad

Original: i thought the movie was too bland and too much
Output-1: i guess the movie is too bland and too much
Output-2: i guess the film will have been too bland

Beyond Image-to-Text Generation

- Controlled Generation of Text [Hu et al., 2017]
  - With **control** over content, attributes, stylistic characteristics, …
  - At sentence / discourse level
  - In supervised / unsupervised settings

  - E.g., **Text Style Transfer**
    [Hu et al., 2017; Yang et al., 2018]
  - E.g., **Content Manipulation**
    [Wang, Hu, et al., 2019]

Content Record	**PLAYER**	**PT**	**RB**	**AS**	**PLAYER**	**PT**
	LeBron_James	32	4	7	Kyrie_Irving	20
Reference Sentence	Jrue_Holiday led the way with 26 points and 6 assists , while Goran_Dragic scored 23 points and pulled down 8 rebounds .					
Output	LeBron_James led the way with 32 points , 7 assists and 4 rebounds , while Kyrie_Irving scored 20 points .					

Beyond Image-to-Text Generation

- Controlled Generation of Text [Hu et al., 2017]
  - With **control** over content, attributes, stylistic characteristics, …
  - At sentence / discourse level
  - In supervised / unsupervised settings

  - E.g., **Text Style Transfer**
    [Hu et al., 2017; Yang et al., 2018]
  - E.g., **Content Manipulation**
    [Wang, Hu, et al., 2019]
  - E.g., **Conversation Strategy Control**
    [Tang et al., 2019]

Guide conversation to a target topic: e-books

Agent : hi how are you doing
Human : not so good . i am really tired
Agent : oh i'm sorry to hear
Human : i have a lot of work to do
Agent : what kind was it ?
Human : i am writing a chatbot program
Agent : interesting . i read about coding from a book
Human: really
Agent : i prefer e-books over paperback book

Outline

- Medical image report generation
  - Co-attention, hierarchical generation, multi-task
  - Further improvement: retrieval+generation, structured knowledge
- Paragraph description of natural images
- Text generation under control
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Text Generation Tasks

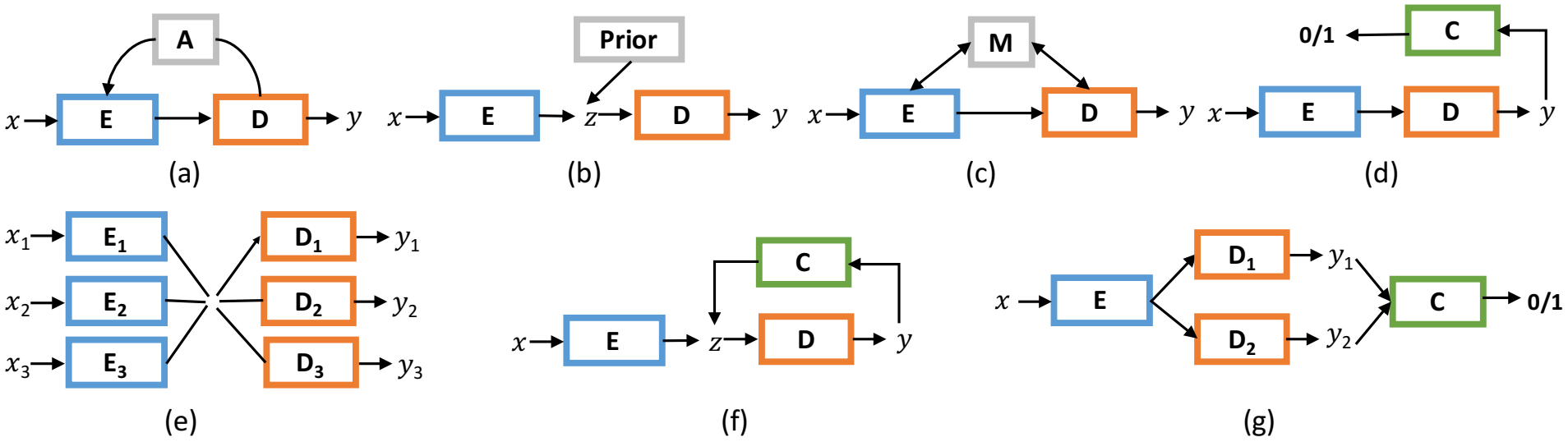
- Generates *natural language* from input *data or machine representations*
- Spans a broad set of natural language processing (NLP) tasks:

Task	Input *X*	Output *Y* (Text)
Chatbot / Dialog System	Utterance	Response
Machine Translation	English	Chinese
Summarization	Document	Short paragraph
Description Generation	Structured data	Description
Captioning	Image/video	Description
Speech Recognition	Speech	Transcript

Courtesy: Neubig, 2017

Various (Deep Learning) Techniques

- Various model architectures



E: encoder, D: decoder, C: Classifier, A: attention, Prior: prior distribution, M: memory

Texar

A General-Purpose Text Generation Toolkit

Texar

A General-Purpose Text Generation Toolkit

Hundreds of thousands of lines of code

Texar

A General-Purpose Text Generation Toolkit

Hundreds of thousands of lines of code

With the toolkit

Texar Overview

- A **unified platform** aiming to cover many text generation tasks
  - Provide the most comprehensive set of well-tailored and ready-to-use modules
  - Enable reuse of common components and functionalities
  - Standardize design, implementation, and experimentation
  - Encourage technique sharing among different tasks
- Based on TensorFlow
- Open-source under Apache License 2.0

Texar Highlights

Texar Highlights

Modularized

Assemble any complex model like playing building blocks

Texar Highlights

Modularized

Assemble any complex model like playing building blocks

Versatile

Supports a large variety of applications/models/algorithms

…

Texar Highlights

Modularized

Assemble any complex model like playing building blocks

Versatile

Supports a large variety of applications/models/algorithms

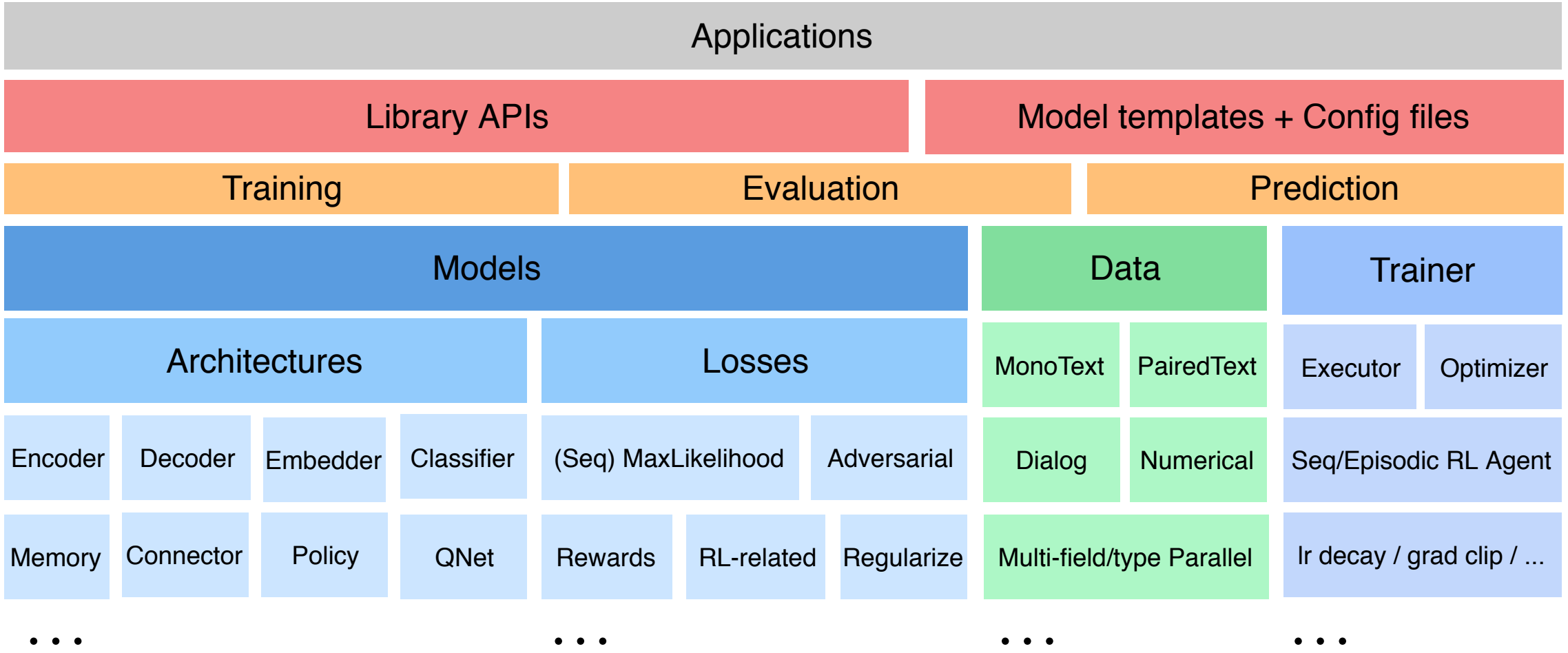
…

Extensible

Allows to plug in any customized or external modules

Texar Stack

***Texar** stack*



Module Catalog

Model architecture

- **Encoder**
  - UnidirectionalRNNEncoder
  - BidirectionalRNNEncoder
  - HierarchicalRNNEncoder
  - ConvEncoder
  - TransformerEncoder
  - ...
- **Decoder**
  - BasicRNNDecoder
  - AttentionRNNDecoder
    - BahdananAttn
    - LuongAttn
    - MonotonicAttn
  - TransformerDecoder
  - Greedy/Sample/BeamSearch/ GumbelSoftmax/... Decoding
  - ...
- **Embedder**
  - WordEmbedder (one-hot / soft)
  - PositionEmbedder
    - Parametrized
    - Sinusoids
  - ...
- **Classifier/ Discriminator**
  - RNNClassifier
  - ConvClassifier
  - HierarchicalClassifier
  - ...
- **Connector**
  - MLPTransformer
  - Stochastic
  - ReparameterizedStochastic
  - Concat
  - Forward
  - ...

. . .

Model loss

- **Loss**
  - MLE Loss
    - (Sequence) Cross-entropy
    - ...
  - Adversarial Loss
    - Binary Adversarial Loss
    - ...
  - Rewards
  - ...

Trainer

- **RL Agent**
  - Seq RL Agent
    - Seq Policy Gradient Agent
    - ...
  - Episodic RL Agent
    - Policy Gradient Agent
    - DQN Agent
    - Actor-critic Agent
    - ...
- **Optimization**
  - Optimizer
    - Adam/SGD/...
  - Learning Rate Decay
    - Piecewise/Exp/...
  - ...

Data

- **Data**
  - MonoText
  - PairedText
  - MultiAligned
  - Dialog
  - Numerical
  - ...

. . .

Example: Build a sequence-to-sequence model

```yaml
source_embedder: WordEmbedder
source_embedder_hparams:
  dim: 300
encoder: UnidirectionalRNNEncoder
encoder_hparams:
  rnn_cell:
    type: BasicLSTMCell
    kwargs:
      num_units: 300
    num_layers: 1
    dropout:
      output_dropout: 0.5
      variational_recurrent: True
embedder_share: True
decoder: AttentionRNNDecoder
decoder_hparams:
  attention:
    type: LuongAttention
beam_search_width: 5
optimization: …
```

(1) Customize model template via a YAML config file

```python
# Read data
dataset = PairedTextData(data_hparams)
batch = DataIterator(dataset).get_next()

# Encode
embedder = WordEmbedder(dataset.vocab.size, hparams=embedder_hparams)
encoder = TransformerEncoder(hparams=encoder_hparams)
enc_outputs = encoder(embedder(batch['source_text_ids']),
                      batch['source_length'])

# Decode
decoder = AttentionRNNDecoder(memory=enc_outputs,
                              hparams=decoder_hparams)
outputs, length, _ = decoder(inputs=embedder(batch['target_text_ids']),
                             seq_length=batch['target_length']-1)

# Loss
loss = sequence_sparse_softmax_cross_entropy(
  labels=batch['target_text_ids'][:,1:], logits=outputs.logits, seq_length=length)

```

(2) Program with Texar Python Library APIs

Resources

- Website: https://texar.io
- GitHub: https://github.com/asyml/texar
- Examples: https://github.com/asyml/texar/blob/master/examples
- Documentation: https://texar.readthedocs.io/
- Blog: https://medium.com/@texar
- Tech report: https://arxiv.org/pdf/1809.00794.pdf

The Petuum Vision

Industry Agnostic

Robo-radiologist

Insurance Auto-Report

Virtual EA

Smart Expense Reports

Smart Catalog

Robot Store Staff

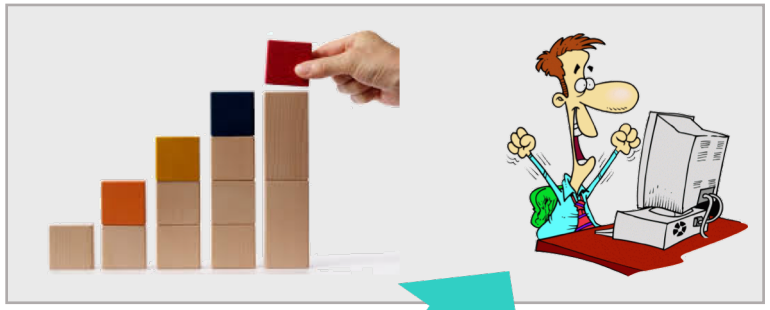
Building AI Like Lego

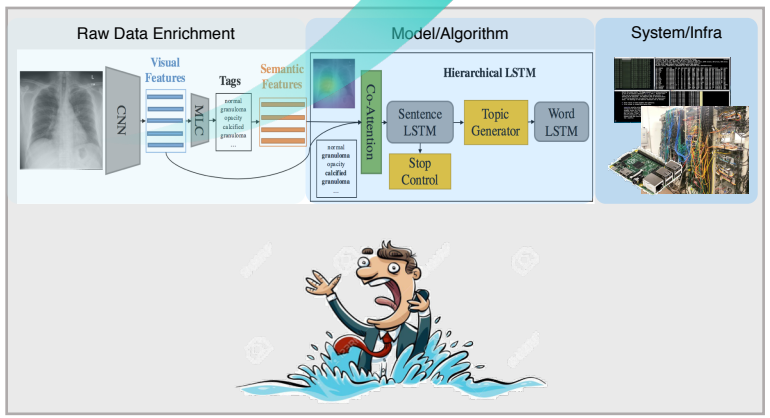
Completed software

10%
White-Glove Assembly

90%
Completed building blocks

AI With No Tears



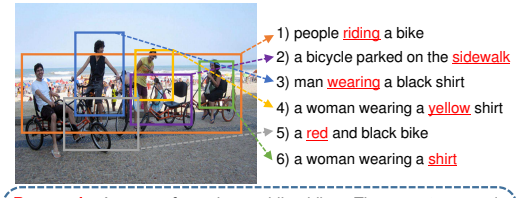


Petuum OS

Summary

- Medical image report generation
  - Cross modalities: images, text, graphs
  - Structured long text generation: retrieval/gen; hierarchical
  - Integrating medical knowledge
- Paragraph description of natural images
- Text generation under **control**
  - Various text properties, granularities, amount of supervision
- All in one toolkit: Texar

Findings:
There are no focal areas of consolidation.
No suspicious pulmonary opacities.
Heart size within normal limits.
No pleural effusions.
There is no evidence of pneumothorax.
Degenerative changes of the thoracic spine.



Paragraph: *A group of people are riding bikes. There are two people riding bikes parked on the sidewalk. He is wearing a black shirt and jeans. A woman is wearing a short sleeve yellow shirt and shorts. There are many other people on the red and black bikes. A woman wearing a shirt is riding a bicycle.*

Content Record	**PLAYER**	**PT**	**RB**	**AS**	**PLAYER**	**PT**
	LeBron_James	32	4	7	Kyrie_Irving	20
Reference Sentence	Jrue_Holiday led the way with 26 points and 6 assists , while Goran_Dragic scored 23 points and pulled down 8 rebounds .					
Desired Output	LeBron_James led the way with 32 points , 7 assists and 4 rebounds , while Kyrie_Irving scored 20 points .					