SIIM CAIMI26 Conference on Artificial Intelligence in Medical Imaging
OCT 26-27 | University of Pennsylvania, Philadelphia, PA

Anatomy-Grounded CT Report Generation Guided by Organ-Hierarchical Clinical Knowledge

Guoliang You, PhD, University of Pennsylvania; Hongming Li, PhD; Yuanwang Zhang, PhD Student; Yong Fan, PhD

Introduction/Background
AI-based report-generation methods for three-dimensional computed tomography (CT) scans commonly rely on a single global representation of the entire image volume. Such representations may dilute sparse, organ-specific abnormalities and provide insufficient anatomical grounding for generated findings. We hypothesized that organ-hierarchical clinical knowledge could organize visual evidence by anatomy and thereby improve CT report generation.

Methods/Intervention
We developed an organ-hierarchical knowledge-guided framework for three-dimensional chest CT report generation, termed as OKA-RG. Free-text reports were structured into organ-conditioned clinical knowledge representing abnormality status, disease or imaging findings, anatomical location, and severity-related attributes. A three-dimensional encoder was pretrained using whole-volume disease supervision, organ-hierarchical supervision, and structured report-image alignment. This design preserved global disease context while learning anatomy-aligned representations of the lungs, heart, esophagus, and aorta. One global token and four organ-conditioned tokens were projected into the language model's embedding space and prepended to the report-generation instruction. During report-generation training, the visual encoder was frozen, while the projection layers and low-rank adaptation parameters were optimized autoregressively. Models with 0.5-billion and 7-billion language-model parameters were evaluated on the same CT-RATE subset and compared with the 8-billion-parameters CT-CHAT baseline.

Results/Outcome
The organ-hierarchical knowledge-guided encoder achieved a macro-AUROC of 84.9 on CT-RATE, compared with 73.1 for CT-CLIP, 72.8 for Merlin, and 77.8 for fVLM. Despite using substantially fewer language-model parameters, OKA-RG with 0.5 billion parameters outperformed CT-CHAT on BLEU-1/2/3/4, achieving 0.3569/0.2564/0.1938/0.1515 versus 0.3479/0.2486/0.1850/0.1422. Scaling OKA-RG to 7 billion parameters further improved BLEU-4 to 0.2048, ROUGE-L to 0.3631, and token-F1 to 0.4811, yielding the strongest overall performance.

Conclusion
Organ-hierarchical clinical knowledge provides an effective bridge between anatomy-grounded visual representation learning and efficient CT report generation. Integrating complementary global and organ-conditioned evidence improved report quality, with further gains achieved by increasing language-model capacity.

Statement of Impact

Structured global and organ-level visual evidence may reduce dependence on very large language models while improving anatomical grounding and coverage of multi-organ finding in automatically generated chest CT reports.

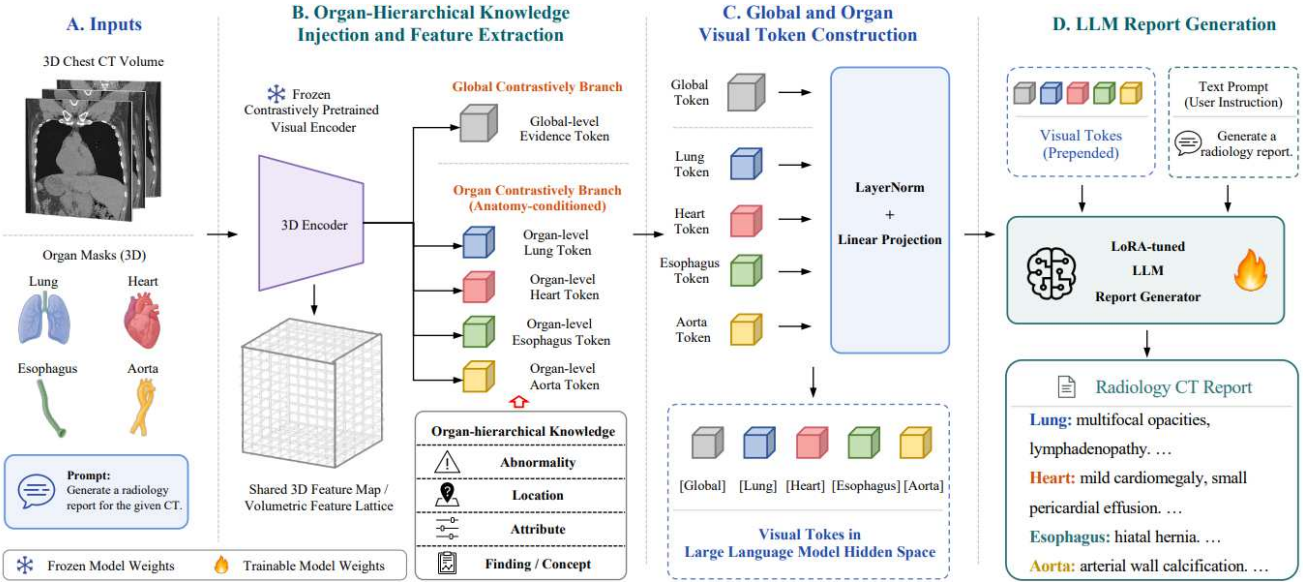

Figure 1. Organ-hierarchical visual evidence for anatomy-grounded CT report generation. A contrastively pretrained three-dimensional encoder extracts a shared volumetric feature map from the CT volume. Whole-volume and anatomy-conditioned branches generate one global evidence token and organ-level tokens representing the lungs, heart, esophagus, and aorta. Report-derived knowledge describing abnormality status, imaging finding, anatomical location, and severity-related attributes shapes the organ-conditioned visual evidence space. The global and organ-level tokens are projected into the language-model's embedding space, prepended to the report-generation instruction, and decoded by a low-rank-adapted causal language model to generate a radiology report.

Table 1. Diagnostic performance of anatomy-grounded visual representations on CT-RATE

Method	AUROC
CT-CLIP	73.1
Merlin	72.8
fVLM	77.8
Organ-hierarchical knowledge-guided encoder (Ours)	84.9

Table 2. Report-generation performance on the CT-RATE evaluation subset

Method	BLEU-1	BLEU-2	BLEU-3	BLEU-4	ROUGE-L	token-F1
CT-CHAT (8B)	0.3479	0.2486	0.1850	0.1422	0.2845	0.4547
OKA-RG (0.5B)	0.3569	0.2564	0.1938	0.1515	0.2828	0.4442
OKA-RG (7B)	0.3590	0.2794	0.2347	0.2048	0.3631	0.4811

Keywords

computed tomography; radiology report generation; organ-hierarchical knowledge; anatomical grounding; vision-language models; large language models