SIIM CAIMI25
Conference on Artificial Intelligence in Medical Imaging
OCT 20-21 | UNIVERSITY OF CALIFORNIA, IRVINE

Vector Contours to the Rescue: How Basic Shape Outlines Keep Vision-LLM Generated Reports Clinically Grounded

Krish Malik, Parkland High School; Vedant Malik

Introduction/Background

Vision based Large Language Models focused on Radiology generate fluent reports but invent findings not present in the MRI/CT/X-Ray images. Current methods such as rule based engines, confidence scores and prompt based methods constrain the LLM to generating non speculative reports. However, vision based LLMs can invent nodules or laterality that do not exist. They may provide measurements that differ from actual readings. Because the vision model's reasoning is hidden and black box, it is hard for radiologists to trace why an error occurred, undermining clinician trust and slowing clinical adoption.

Methods/Intervention

The proposed two-stage method grounds radiology report generation in explicit anatomical evidence using two LoRA adapters on top of a fine-tuned vision Large Language Model. The first adapter (LoRA-Seg) is trained on image-mask pairs (mask converted to SVG vectors) to produce SVG contours (Scalable Vector Graphics) that outline key anatomical structures. The second adapter (LoRA-Report) is trained on image-report pairs to generate radiology text. During inference, the model first generates SVG contours and a corresponding radiology report. A test time validation function then checks each geometric finding in the report against SVG coordinates. Any mismatch is fed back to LoRA-Report as a feedback loop, ensuring the final report is fully grounded.

Results/Outcome

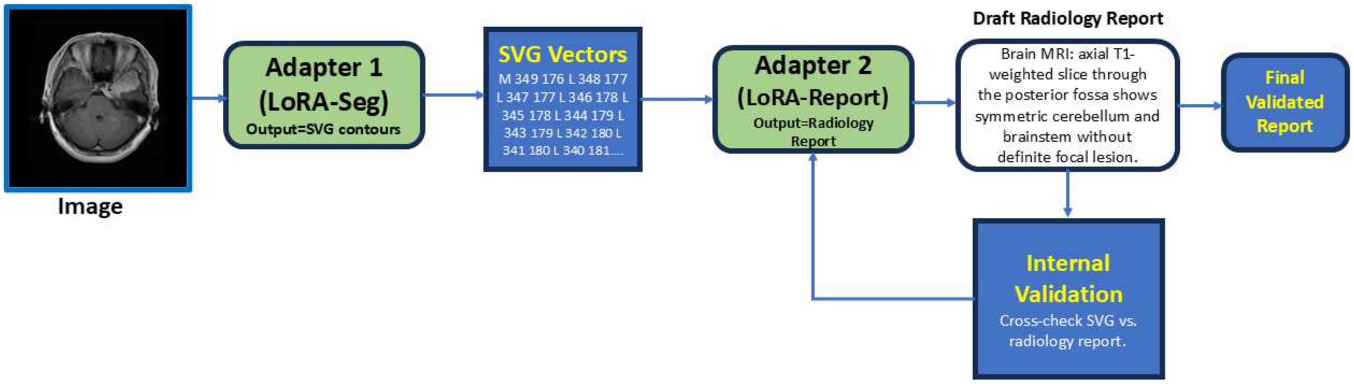
On a hold out set of 180 unseen BraTS brain tumor MRIs, the two-stage model achieved best Dice score of 0.93 and mean Dice score of 0.89, showcasing accurate tumor contour and spatial fidelity. With test time SVG-based feedback loop, the average F1 score increased from 0.78 to 0.91. This showed that radiology report when validated against SVG contours, grounds the report and avoids generation of new lesions.

Conclusion

Validating the generated radiology report against SVG contours predicted from the medical image significantly reduces speculative errors that many vision-based Large Language Models suffer from. The model first predicts SVG tumor contours, then cross checks every geometric reference in the report against SVG contours.

Statement of Impact

Grounding every geometric assertion in a radiology report to SVG contours predicted from the image transforms a vision-LLM from a black box generator to evidence linked assistant. It also provides an auditable path to trustworthy radiology report generation.



Model Inference at Test Time

Vision-Large Language Model Training Loss

Keywords

Medical Imaging; Vision-LLM; Large Language Models; CT/MRI/X-Ray; Image Segmentation; Radiology Reports