Rethinking Whole-Body CT Image Interpretation: An Abnormality-Centric Approach

Appendix

Contents

A. Qualitative Experiment Results

Fig. 4 presents several cases from the **grounded report generation** task, comparing the segmentation and report generation results between OmniAbnorm-CT and BiomedParse+LLaVA-Med. It clearly shows that BiomedParse fails to detect any abnormalities in the latter two cases, and consequently, LLaVA-Med generates reports irrelevant to the abnormalities. In contrast, OmniAbnorm-CT successfully localizes the abnormalities across all cases and produces more accurate reports.

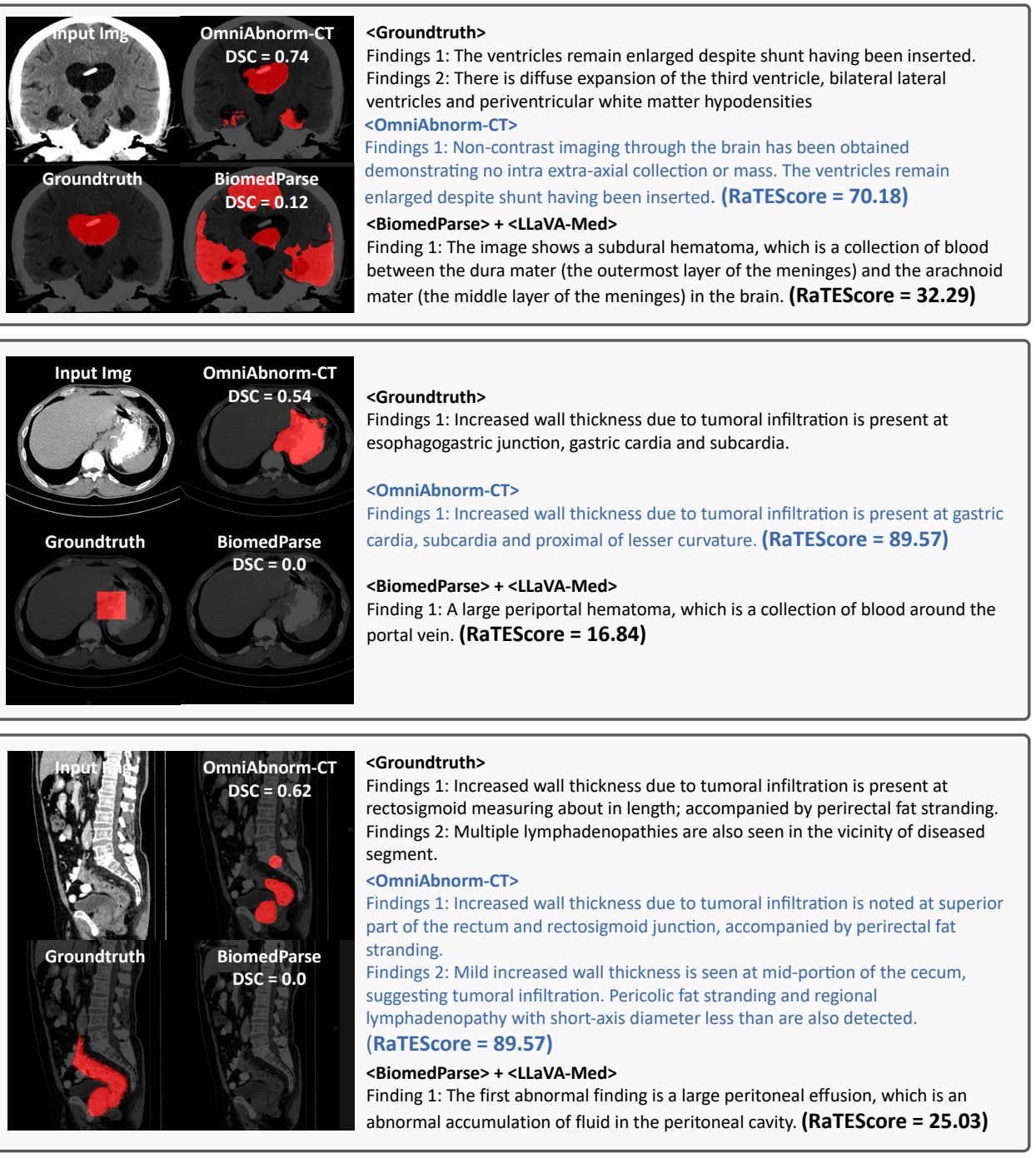


Figure 4. Qualitative comparison on the grounded report generation task.

In Fig. 5, we compare OmniAbnorm-CT with BiomedParse+LLaVA-Med on the **text-guided grounded report generation**

task. BiomedParse mislocalizes the queried findings in first two cases, while LLaVA-Med fails to correctly interpret the segmentation results in last two cases. In contrast, OmniAbnorm-CT consistently localizes the queried abnormalities and produces more precise, clinically aligned reports.

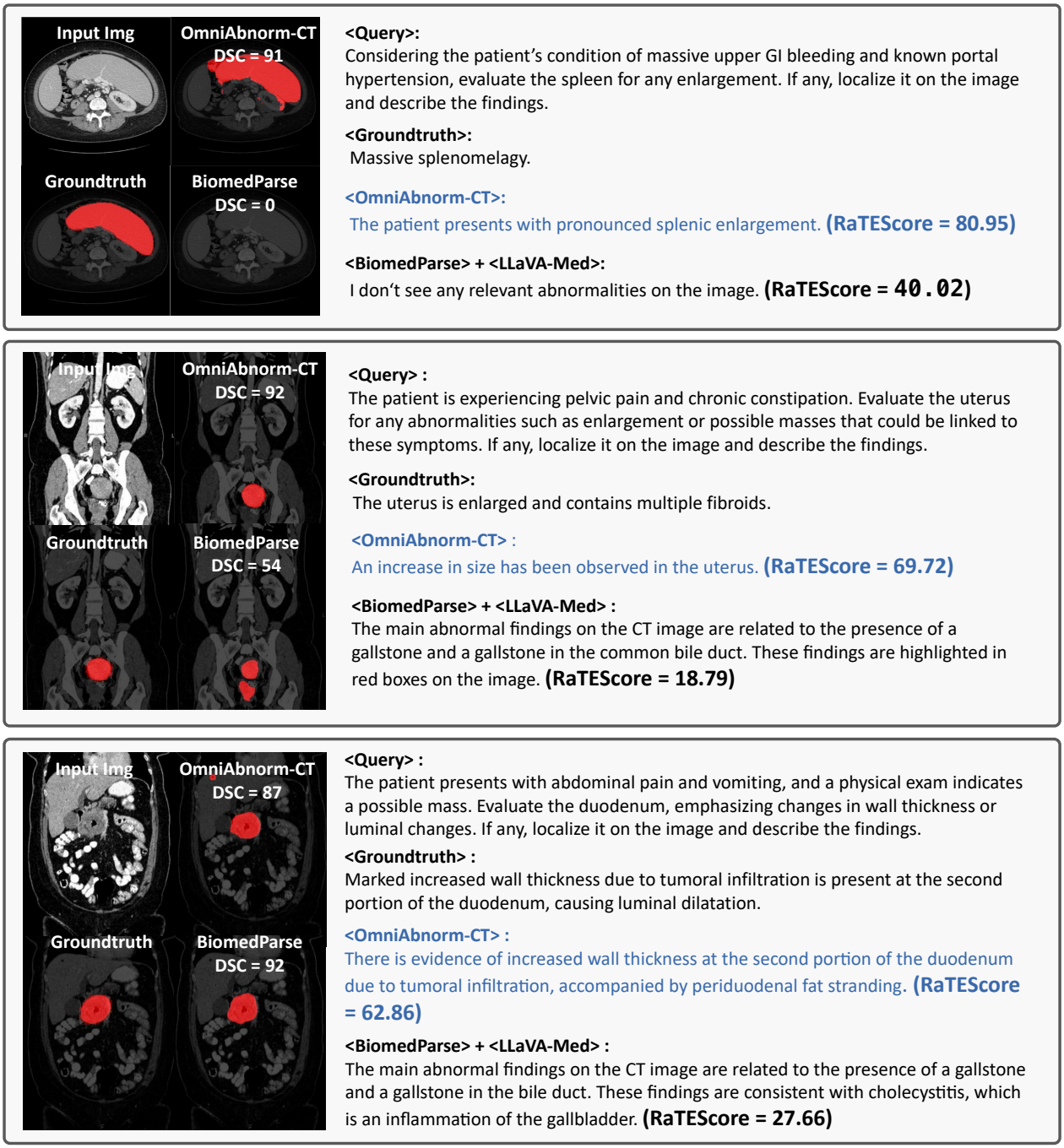
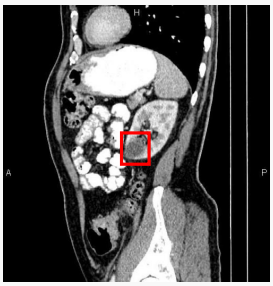


Figure 5. Qualitative comparison on the text-guided grounded report generation task.

In Fig. 6, we compare OmniAbnorm-CT with LLaVA-Med and QWen2.5-VL-7B on visual prompted report generation. The baselines show some fundamental mistakes, *e.g.*, LLaVA-Med mislabels the kidney as the gallbladder in the first case, and QWen2.5-VL-7B confuses the liver and gallbladder in the last case. In contrast, OmniAbnorm-CT accurately identifies the marked abnormalities and produces higher-quality, clinically consistent reports.

Input Image:



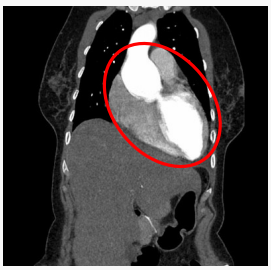
<Groundtruth> : A simple cortical cyst is noted in the left kidney.

<OmniAbnorm-CT> : The kidneys exhibit several uncomplicated cortical cysts that do not show enhancement. **(RaTEScore = 81.03)**

<LLaVA-Med> : The main findings indicated by the red box in this CT image are a dilated gallbladder and a dilated common bile duct. **(RaTEScore = 23.09)**

<QWen2.5-VL-7B> : The red box highlights an area of increased density within the mesenteric fat, suggesting the presence of a mesenteric lymph node or a similar soft tissue mass. Further evaluation is recommended for definitive diagnosis. **(RaTEScore = 25.01)**

Input Image:



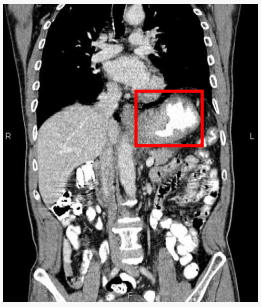
<Groundtruth> : Gross cardiomegaly

<OmniAbnorm-CT> : The patient presents with an enlargement of the heart, indicative of cardiomegaly. **(RaTEScore = 81.88)**

<LLaVA-Med> : The main findings indicated by the red circular/elliptical demarcation in the image are a large pericardial effusion and a large pericardial mass. **(RaTEScore = 42.35)**

<QWen2.5-VL-7B> : The red circular/elliptical demarcation highlights an area of increased density within the mediastinum, which appears to be consistent with a mass or lesion. Further evaluation is required for definitive diagnosis. **(RaTEScore = 29.38)**

Input Image:



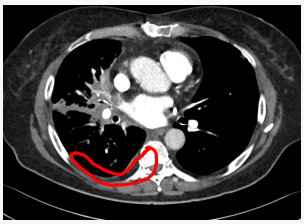
<Groundtruth> : Marked increased wall thickness due to tumoral infiltration is present at gastric body and antrum accompanied by several regional lymphadenopathies.

<OmniAbnorm-CT> : Increased wall thickness due to tumoral infiltration is present at gastric cardia and body accompanied by perigastric fat stranding. **(RaTEScore = 83.88)**

<LLaVA-Med> : The main findings indicated by the red box in this CT image are a dilated stomach and a dilated duodenum. **(RaTEScore = 42.81)**

<QWen2.5-VL-7B> : The red box highlights an area of increased density within the liver parenchyma, suggesting a possible focal lesion. Further characterization is required for definitive diagnosis. **(RaTEScore = 39.78)**

Input Image:



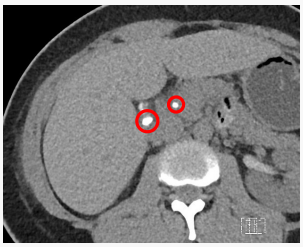
<Groundtruth> : Small right pleural effusion.

<OmniAbnorm-CT> : Minimal pleural effusion mainly on the right. **(RaTEScore = 85.64)**

<LLaVA-Med> : The main findings indicated by the red-outlined area in the image are a large pericardial effusion and a moderate-sized pleural effusion. **(RaTEScore = 34.89)**

<QWen2.5-VL-7B> : The red-outlined area appears to be a normal lung. **(RaTEScore = 20.80)**

Input Image:



<Groundtruth> : The gallbladder is not abnormally distended and shows multiple gallstones.

<OmniAbnorm-CT> : Numerous gallstones identified within the gallbladder. **(RaTEScore = 88.13)**

<LLaVA-Med> : The main findings indicated by the red-outlined area in the image are a dilated common bile duct (CBD) and a gallstone. **(RaTEScore = 59.10)**

<QWen2.5-VL-7B> : The red-outlined area indicates a small, well-defined, hypodense lesion within the liver parenchyma. The lesion appears to be homogenous in density and does not show significant enhancement on this particular phase of contrast administration. Further imaging or clinical correlation is recommended for definitive diagnosis**. (RaTEScore = 23.78)**

Figure 6. Qualitative comparison on the visual prompted report generation task.

B. External Validation

To assess the generalizability of our approach, we construct an external validation set by applying the annotation pipeline described in Section 4.2 to CT-RATE [21], a large-scale non-contrast chest CT dataset from Turkey clinical centers. The external validation follows the three tasks defined in Section 6.1. In this section, we first detail the construction of the external validation set in Section B.1, then demonstrate the validation results in Section B.2.

B.1. Data Construction

We randomly sample 65 CT volumes for annotation and follow the pipeline introduced in Section 4.2. Since we can hardly identify the representative slices of the volume, we require the annotator to annotate all the abnormalities visible in the full-volume. As a result, the external validation set comprises 271 abnormality annotations, each with a segmentation mask, linked to the corresponding text description in the paired report and categorized to the proposed taxonomy. Since OmniAbnorm-CT requires up to 9 consecutive slices, we select the most prominent slice (with the largest segmentation area) for each abnormality for evaluation.

B.2. Results

As demonstrated in Table 4, on the external validation set, OmniAbnorm-CT maintains strong performance across all three tasks. In visual prompted report generation, OmniAbnorm-CT achieves the highest AbnormRubric score (+4.94 over the best baseline), indicating that its generated reports align more closely with clinical criteria. Moreover, in both grounded report generation and text-guided grounded report generation, OmniAbnorm-CT surpasses all baselines on report generation metrics and shows superior grounding ability (DSC improvements of +10.57 for grounded report generation and +2.44 for text-guided grounded report generation). These results suggest that OmniAbnorm-CT generalizes well to external datasets.

Model	DSC	Rubric	B-1	B-2	B-3	RaTESc	BERTSc	MTR	R-1	R-L	RadG
Visual prompted report generation (n=271)											
GPT4o	-	19.15	17.37	7.59	**1.81**	37.86	85.52	**20.46**	19.16	15.18	3.73
QWen2.5-VL-7B	-	15.15	6.41	2.57	0.42	37.04	82.84	14.62	9.52	7.37	2.51
ViP-LLaVA	-	10.31	17.59	6.20	0.55	32.88	85.46	15.33	19.18	16.36	1.95
MedDr ✓	-	13.52	3.24	1.13	0.17	20.36	54.28	3.58	4.94	4.29	0.74
LLaVA-Med ✓	-	12.22	17.63	7.40	1.20	34.96	85.52	19.81	19.01	15.64	2.92
BiomedGPT ✓	-	12.07	10.99	4.57	0.60	35.14	82.62	17.30	13.92	11.37	2.67
OmniAbnorm-CT ✓	-	**24.09**	**19.09**	**7.71**	1.67	**39.87**	**86.64**	19.46	**21.02**	**17.34**	**4.27**
Grounded report generation (n=271)											
MedULS + LLaVA-Med	11.63	3.96	12.86	7.09	1.32	33.53	81.88	17.31	15.88	12.87	2.50
LiSA + LLaVA-Med	18.34	4.37	12.33	6.54	0.62	**35.43**	83.57	17.03	15.87	12.71	2.69
BiomedParse + LLaVA-Med	11.78	5.32	11.98	6.51	0.74	34.76	83.47	16.59	16.10	13.08	**2.87**
OmniAbnorm-CT ✓	**22.20**	**8.62**	**18.55**	**12.10**	**8.14**	34.93	**86.09**	**21.61**	**22.36**	**18.32**	2.57
Text-guided grounded report generation (n=271)											
MedULS + LLaVA-Med	11.69	3.23	9.47	2.86	0.33	25.57	74.55	11.57	10.81	8.91	1.32
LiSA + LLaVA-Med	9.72	0.17	2.85	1.43	0.89	4.68	20.26	3.08	3.32	2.90	1.40
BiomedParse + LLaVA-Med	10.50	0.33	4.33	2.06	**1.44**	6.65	25.52	4.50	4.77	4.48	1.63
OmniAbnorm-CT ✓	**14.13**	**15.12**	**12.38**	**3.76**	1.18	**32.25**	**76.34**	**13.55**	**14.85**	**11.97**	**3.55**

Table 4. **External validation results across three tasks.** Metrics are averaged within each abnormality category and then across all categories. Within each block, the best result is **bolded** and the second best is underlined. For metrics that are not applicable to a given task, we report "-". Models optimized with medical data are marked with ✓.

C. Ablation Studies

We conduct a series of ablation studies on the key factors of OmniAbnorm-CT, from the segmentation module (Section C.1), the context range of input image (Section C.2), to the form of visual prompts (Section C.3). This section introduces the settings and results of these experiments sequentially.

C.1. Ablation Study on Segmentation Module

The integration of the segmentation module allows OmniAbnorm-CT to acquire grounding evidence for abnormal findings and generate descriptions accordingly. To validate its necessity, we conduct an ablation study comparing the full model against a variant without the segmentation module (OmniAbnorm-CT w/o Seg), which directly generates findings.

Text-guided grounded report generation. As shown in Table 5, removing the segmentation module significantly degrades performance on most metrics (23/27 across planes). Furthermore, our manual analysis reveals that, without the grounding evidence, the model struggles to accurately detect the queried abnormalities in CT images.

Grounded report generation. Similarly, Table 6 shows that OmniAbnorm-CT w/o Seg consistently underperforms the full model on all metrics, reaffirming that segmentation-driven evidence is essential for accurate grounding and generating clinically faithful reports.

Model	BLEU-1	BLEU-2	BLEU-3	RaTEScc	BERTScc	METEOR	ROUGE-1	ROUGE-L	RadG
Axial (n=2193)									
OmniAbnorm-CT w/o Seg	4.50	2.33	**1.21**	20.53	31.06	7.04	8.50	7.02	**7.25**
OmniAbnorm-CT	**11.94**	**3.33**	0.74	**34.80**	**83.85**	**12.94**	**14.65**	**11.94**	4.32
Coronal (n=750)									
OmniAbnorm-CT w/o Seg	4.50	2.47	1.27	19.07	28.34	6.27	7.96	6.43	**6.17**
OmniAbnorm-CT	**12.51**	**4.10**	**1.22**	**36.56**	82.98	**13.61**	**15.83**	**12.36**	5.72
Sagittal (n=591)									
OmniAbnorm-CT w/o Seg	4.17	2.16	1.12	16.62	25.43	5.84	7.06	5.53	**5.67**
OmniAbnorm-CT	**12.80**	**4.51**	**1.42**	**36.04**	**83.97**	**14.02**	**16.20**	**12.79**	4.85

Table 5. Ablation study of the segmentation module for **text-guided grounded report generation** task.

Model	BLEU-1	BLEU-2	BLEU-3	RaTEScc	BERTScc	METEOR	ROUGE-1	ROUGE-L	RadG
Axial (n=2193)									
OmniAbnorm-CT w/o Seg	7.73	4.74	3.16	27.87	73.14	17.65	15.18	13.58	3.42
OmniAbnorm-CT	**19.00**	**12.26**	**8.40**	**33.85**	**86.24**	**22.45**	**22.27**	**19.43**	**4.60**
Coronal (n=750)									
OmniAbnorm-CT w/o Seg	8.12	4.81	3.12	31.03	75.55	15.64	15.27	13.23	3.84
OmniAbnorm-CT	**19.28**	**12.03**	**7.96**	**35.40**	**85.56**	**20.81**	**22.35**	**18.47**	**4.65**
Sagittal (n=591)									
OmniAbnorm-CT w/o Seg	8.79	5.21	3.33	31.21	77.41	16.18	15.86	13.87	**3.91**
OmniAbnorm-CT	**16.69**	**10.41**	**6.98**	**33.45**	**84.94**	**18.77**	**20.67**	**17.61**	3.83

Table 6. Ablation study of the segmentation module in **grounded report generation** task.

C.2. Ablation Study on Context Range

In clinical practice, adjacent-slice context is important for CT interpretation, including segmentation and diagnosis [28, 67]. To assess its impact on report generation, we vary the number of input slices in visual prompted report generation, using a base QWen2.5-VL-7B. Surprisingly, Table 7 shows that adding slices yields no consistent gains across metrics. We hypothesize that current LVLMs, which encode the slices independently into tokens, struggle to model subtle inter-slice spatial relations. Accordingly, we use only the center slice as input to the LVLM in subsequent training and evaluation of OmniAbnorm-CT.

C.3. Ablation Study on Different Visual Prompts

Recent work shows VLMs respond exhibit varying perceptual capabilities for different prompt formats [10]. We therefore study visual prompt design for visual prompted report generation. For each abnormality, we simulate four prompts, *i.e.*, center crop, ellipse, contour, and box, and evaluate OmniAbnorm-CT separately. We also consider a per-case oracle that

Input Slices	BLEU-1	BLEU-2	BLEU-3	RaTEScore	BERTScore	METEOR	ROUGE-1	ROUGE-L	RadGraph
Axial (n=2193)									
4 Adjacent Slices	9.39	2.98	0.45	31.12	83.52	14.90	11.51	9.01	2.46
2 Adjacent Slices	9.24	3.03	0.46	31.47	83.49	15.05	11.41	8.99	2.49
No Adjacent Slices	**11.57**	**4.20**	**0.76**	**38.35**	**84.10**	**16.68**	**14.26**	**11.09**	**3.85**
Coronal (n=750)									
4 Adjacent Slices	10.93	**3.93**	**0.85**	32.95	83.49	15.82	13.29	9.79	3.62
2 Adjacent Slices	10.89	3.85	0.73	33.06	83.48	**15.87**	13.34	9.85	**3.71**
No Adjacent Slices	**13.62**	3.54	0.66	**37.26**	**84.19**	13.49	**14.97**	**11.47**	3.37
Sagittal (n=591)									
4 Adjacent Slices	9.39	2.98	0.45	31.12	83.52	14.90	11.51	9.01	2.46
2 Adjacent Slices	11.48	**4.16**	**0.94**	33.86	83.32	**15.70**	13.64	9.97	**3.75**
No Adjacent Slices	**13.28**	3.54	0.63	**35.67**	**83.98**	13.01	**14.75**	**11.34**	2.91

Table 7. Impact of multi-slice context range on visual prompted report generation, with the base Qwen2.5-VL-7B model as baseline.

takes the best score among the four, representing the theoretical best performance via selecting the optimal prompt for each specific abnormality, depending on its shape, location, and etc. As shown in Table 8, cropping performs worst on most metrics (24/27), likely due to lost context. Ellipse, contour, and box yield similar performance. The oracle notably improves all metrics over any single prompt, revealing large per-case variance. This suggests that adaptively choosing the visual prompt by morphology and location is beneficial, whereas cropping away context is harmful.

Prompt Type	BLEU-1	BLEU-2	BLEU-3	RaTEScore	BERTScore	METEOR	ROUGE-1	ROUGE-L	RadGraph
Axial (n=2193)									
Center Cropping	11.70	2.87	0.68	33.03	84.84	12.77	14.58	11.75	3.92
Ellipse	12.43	3.60	1.15	35.14	84.87	13.67	15.55	12.63	5.27
Contour	12.60	3.79	1.22	35.74	85.09	14.12	15.97	12.92	5.81
Bounding Box	12.35	3.69	1.26	35.46	84.92	13.65	15.76	12.80	5.63
Max	**18.03**	**6.65**	**1.80**	**42.81**	**86.35**	**19.40**	**21.66**	**17.61**	**9.98**
Coronal (n=750)									
Center Cropping	10.94	2.96	0.80	33.57	84.48	11.92	14.23	11.36	3.87
Ellipse	12.49	3.79	1.13	36.09	84.56	13.44	15.94	12.43	5.16
Contour	12.61	3.73	1.02	35.50	84.71	13.45	15.75	12.25	4.87
Bounding Box	12.15	3.70	1.12	35.12	84.38	13.27	15.68	12.13	4.68
Max	**18.38**	**7.19**	**2.39**	**42.88**	**86.00**	**19.00**	**21.30**	**16.87**	**9.44**
Sagittal (n=591)									
Center Cropping	10.98	2.73	0.39	32.23	84.20	11.98	14.45	11.01	3.05
Ellipse	11.37	3.10	0.65	34.36	84.01	12.19	14.45	10.95	3.46
Contour	11.08	2.95	0.57	34.19	84.22	12.01	14.38	11.10	3.22
Bounding Box	11.29	2.98	0.67	34.07	84.13	12.16	14.84	11.24	3.45
Max	**17.29**	**6.48**	**1.82**	**41.77**	**85.62**	**17.45**	**20.43**	**15.87**	**7.51**

Table 8. Comparison of different visual prompts in visual prompted report generation task, with OmniAbnorm-CT fixed as baseline.

D. Details of AbnormRubric

As introduced in Section 6.2, the computation of **AbnormRubric** is decomposed into three stages that explicitly mirror the way radiologists assess a report: (i) abnormality detection (recall), (ii) hallucination checking (precision), and (iii) description accuracy of radiological attributions. All three stages are implemented using LLM-based judgments given the ground-truth report and the generated report.

Detection of ground-truth abnormalities (Recall). Given the set of abnormalities in the ground-truth report, we iterate over each and query an LLM to determine whether this abnormality is explicitly present in the generated report. Abnormalities that are judged as present are counted as true positives (TP), while the remaining ground-truth abnormalities are counted as false negatives (FN). The detailed prompt is:

You are a medical imaging report analyst. Your task is to evaluate an AI-generated radiology report (or only a portion/fragment of a report) from two aspects:

1. Whether the report is a VALID radiology report (it makes a clear statement about what abnormalities are present OR absent on the current image).
2. Whether the report identifies a specific abnormality.

ABNORMALITY TO CHECK:

- Target Abnormality: `{abnormality_category}`
- Relevant Description in Detailed: `{groundtruth_description}`

AI-GENERATED REPORT TO ANALYZE: `{generated_report}`

EVALUATION CRITERIA:

1. **Validity Check:**
   - The generated report may contain medically irrelevant content, nonsensical phrases or garbled text. Completely ignore such content and focus only on medically relevant content.
   - The generated report may be blank, or entirely composed of non-medical gibberish, irrelevant text, or insufficient to determine presence/absence of abnormalities. Then ignore the following procedure and your evaluation is complete: `is_invalid_report=1`, `detected=0`, `description=None`.
2. **Detection:**
   - Report as `detected=1` if the target abnormality (`{abnormality_category}`) is reported in the generated report.
   - Consider synonyms and clinically equivalent terms as matches.
   - Regardless of the description accuracy, such as incorrect location, size measurements, morphological characteristics, density/attenuation, and so on.
   - The generated report may indicate that no abnormalities were detected on the image, for example, `I don't see any relevant abnormalities on the image.` Then ignore the following procedure and your evaluation is complete: `detected=0`.
3. **Description Extraction:**
   - If `detected=1`, collect ALL relevant description from the report.
   - Preserve original wording without modifications.
   - If multiple descriptions exist, concatenate them.
   - May include adjacent text if the description cannot be grammatically isolated.

REQUIRED OUTPUT FORMAT:

```
{
  "is_invalid_report": 0, // 1 or 0
  "detected": 1,      // 1 if detected, 0 if not
  "description": "exact text" // string if detected, null if not
}
```

STRICT PROHIBITIONS:

- NEVER add information not present in the generated report.
- NEVER paraphrase or summarize original text.
- DO NOT output any additional text except for a valid json.

Identification of hallucinated abnormalities (Precision). To assess hallucinations, we further instruct the LLM to enumerate any abnormalities that are described in the generated report but do not exist in the ground-truth report. Abnormalities that cannot be matched are counted as false positives (FP). The detailed prompt is:

Your task is to evaluate the hallucinations in an AI-generated radiology report (or only a portion/fragment of a report):
1. Suppose some ground truth abnormalities have been successfully detected by the AI-generated radiology report.
2. Identify any **ADDITIONAL** abnormalities mentioned in the generated report that are **NOT** covered by the ground truth list.

THE AI-GENERATED REPORT TO ANALYZE: `generated_report`
GROUND TRUTH ABNORMALITIES AND MATCHING RESULTS: `groundtruth_abnormalities`
INSTRUCTIONS:
1. Review matched content: The above shows which ground truth abnormalities were detected and their corresponding text in the generated report.
2. Find additional abnormalities: Identify ALL ADDITIONAL abnormal findings mentioned in the generated report that are NOT already matched to ground truth abnormalities.
3. Exclude already matched content: Do NOT include any abnormalities that are already matched to ground truth abnormalities.
4. Ignore normal findings: Only report abnormalities, not normal findings.
5. Output: ONLY valid JSON with no additional text.

NOTE:
1. The generated report may contain medically irrelevant content, nonsensical phrases or garbled text. Completely ignore such content and focus only on medically relevant content.
2. If the generated report is entirely composed of non-medical gibberish, irrelevant text, or is blank, then your evaluation is complete: `count = 0` and `hallucinated_abnormalities = []`.

REQUIRED OUTPUT FORMAT:

```
{
  "hallucinated_abnormalities": [...],  // list of hallucinated abnormalities
  "count": x  // number of hallucinated abnormalities
}
```

If no additional abnormalities found:

```
{
  "hallucinated_abnormalities": [],
  "count": 0
}
```

Accuracy of radiological attributions. For each *detected* ground-truth abnormality (TP), we assess how accurately its key radiological attributions are described in the generated report. Based on our abnormality taxonomy, radiologists define a set of clinically important attributions for each abnormality type (e.g., location, morphology, density, size). For a given abnormality, we consider only those attributions that are explicitly mentioned in the ground-truth report and ask the LLM to compare them with the corresponding text in the generated report. For each attribution, the LLM returns a binary judgment of whether the generated description is clinically consistent with the ground-truth description. The detailed prompt is:

You are a medical imaging report analyst. For a single abnormality in the image, the AI model has generated a description. Your task is to compare the AI-generated description with the ground-truth description and evaluate the AI's accuracy across specific attributes that will be provided. **EVALUATION INSTRUCTIONS: INPUT DATA:**
- Abnormality: `{abnormality_category}`
- Ground-Truth (GT) Description: `{gt_description}`
- AI-Generated (Pred) Description: `{pred_description}`
- Attributes: `{attribution_ls}`

1. **Detection of Attribute Mention:**
   - Evaluate ONLY the pre-defined attributes provided in the list.
   - For each attribute, check if it is mentioned in the GT description.
   - For attributes present in the GT description, check if they are also mentioned in the Pred description.
2. **Comparison:**
   - For attributes present in BOTH GT and Pred:
     - Extract the exact phrasing from both descriptions.
     - Determine if the descriptions are clinically equivalent (1) or not (0).
     - Consider synonyms and clinically equivalent terms as matches.
     - Provide brief reasoning for your judgment.
   - For attributes ONLY present in GT:
     - Mark as not clinically equivalent (0).
3. **Quantitative Data Handling:**
   - If GT contains quantitative data (e.g., measurements with units):
     - Extract the numerical value and unit from GT.
     - Check if Pred provides quantitative data for the same attribute.
     - If Pred lacks quantitative data: mark as non-equivalent (0).
     - If Pred provides data: calculate percentage error after unit conversion.
     - Error $\leq 20\%$: equivalent (1); Error $> 20\%$: non-equivalent (0).
     - In the reasoning field, include the brief calculation process.
4. **Special Cases:**
   - If Pred provides quantitative data but GT does not: mark as non-equivalent (0).
   - For ambiguous cases where clinical equivalence cannot be determined: mark as uncertain (-1).
5. **Strict Constraints:**
   - Base ALL judgments strictly on the provided GT and Pred texts.
   - NO additions, deletions, modifications, or external inferences.
   - Preserve exact wording when extracting descriptions.
   - Clinical relevance overrides literal text matching:
     - Ignore differences in writing style, terminology preferences, and phrasing variations.
     - Consider clinically synonymous terms as equivalent.
     - Focus on clinical semantic meaning rather than exact word matching.
     - Account for standard medical abbreviations and their equivalent expressions.

REQUIRED OUTPUT FORMAT (JSON only):

```
{
  "attributes": {
    "attribute_name_1": {
      "present_in_gt": 1, // 0 if not present, and the followings can be empty
      "gt_description": "exact text from GT",  // null if not present
      "pred_description": "exact text from Pred",  // null if not present
      "equivalent": 1,  // 0 if not equivalent, -1 if can not be determined
      "reasoning": "brief clinical justification",  // null if not present
      "gt_quantified": 1,  // 0 if not quantified
      "pred_quantified": 1,  // 0 if not quantified
      "error_percentage": 5.2 // null if not both quantified
      or can not be calculated
    },
    ...
  }
}
```

E. Supplementary Implementation Details

E.1. Supplementary Implementation Details of OmniAbnorm-CT

Training Hyperparameters. We adopt 8 NVIDIA A100-80G GPUs for training, with 1 sample per device and gradient accumulation set to 4. We train OmniAbnorm-CT for 250K iterations in total, and use 5% steps for warming up. We take AdamW [37] as optimizer with a learning rate of 2e-4. When optimizing the multi-modal language model and the segmentation module jointly, we assign equal weights to the text generation loss and segmentation loss, as defined in Equation 8. We set the rank of LORA to 32.

Training Data. We pad and rescale all images to 512×512. For training data sampling from four tasks: visual prompted report generation, grounded report generation, text-guided grounded report generation, and general VQA, we control the ratio to be 1:1:1:1. Additionally, in the abnormality grounding task, we maintain a ratio of 6:4 between OmniAbnorm-CT-14K and public lesion segmentation data. For each abnormality description, we use the following prompt to have GPT rewrite three different versions, then randomly select one as the ground truth:

> Your task is to produce THREE different rewrites of medical reports while preserving all original medical content and diagnostic information, changing only the expression style and sentence structure.
>
> Rules:
> 1. Strictly maintain all medical information, diagnostic results, numerical values, and key findings.
> 2. Do not add any new medical content or conclusions.
> 3. Do not remove any medical information from the original report.
> 4. Only change the wording, sentence structure, word order, and vocabulary choices.
> 5. Maintain professional medical language and terminology.
> 6. Preserve the overall structure of the original report (such as section divisions).
> 7. Ensure all rewritten reports remain medically rigorous and accurate.
> 8. Create THREE distinct rewrites with different phrasing and structure.
>
> Input: Original medical report written by a doctor
> Output: THREE rewritten versions with identical content but different expression
>
> Please rewrite the following medical report in three different ways:
> $Findings Description$
> Your response must follow this exact format:
> $Output Template$

E.2. Supplementary Implementation Details of Baselines

Visual prompted generation. we prompt each abnormality with all types of visual prompts, and take the maximum score. All the methods are prompted with the following template:

> You are a helpful medical assistant. Describe the abnormal findings indicated by the $Visual Prompt$.
> Please use precise medical terminology, maintain the concise reporting style used in formal radiology reports and provide only the specific radiological findings. Do not list general possibilities, explanations, or recommendations.

Grounded report generation. We integrate LLaVA-Med with 2D segmentation models as baselines. Specifically, we re-implement the MedULS [15] with a 2D nnU-Net [28] and the public lesion segmentation datasets in Table 9, covering all the datasets in the official ULS-23 challenge and 10 additional ones. To our knowledge, this represents the segmentation model with the broadest capability range (in terms of abnormality variety) that can be constructed from currently available public datasets. For BiomedParse [64] and LiSA [32], we prompt them with 'Abnormal findings on the CT image' to derive segmentation results for all the abnormalities on the input CT image. Then, we combine each segmentation model with LLaVA-Med by converting the segmentation results into bounding boxes overlaid on the CT image, and prompt LLaVA-Med to generate the report based on these visual cues:

> You are a helpful medical assistant. The abnormal findings are highlighted in red boxes on this CT image, if present. Please describe each abnormal finding indicated by the red boxes using the format 'Finding 1: [description]', 'Finding 2: [description]', etc.
> Use precise medical terminology, maintain the concise reporting style used in formal radiology reports and provide only the specific radiological findings. Do not list general possibilities, explanations, or recommendations. Respond with 'I don't see any abnormalities on the image.' if no abnormalities are present.

Text-guided grounded report generation. We use the simulated text queries as prompts for BiomedParse and LiSA to derive segmentation results for the queried abnormalities, which are detailed in Section 4.3. For MedULS, since it doesn't support text-prompted segmentation, we simply use its unconditioned segmentation results. Similarly, we convert their segmentation results into bounding boxes overlaid on the CT image, and prompt LLaVA-Med to generate the report based on them:

> You are a helpful medical assistant. The abnormal findings are highlighted in red boxes on this CT image, if present. Please describe the abnormal findings indicated by the red boxes.
> Use precise medical terminology, maintain the concise reporting style used in formal radiology reports and provide only the specific radiological findings. Do not list general possibilities, explanations, or recommendations. Respond with 'I don't see any relevant abnormalities on the image.' if no abnormalities are present.

E.3. Runtime and Computation Analysis

Training. We use 4 A100-80G GPUs for training OmniAbnorm-CT. The segmentation module is first pre-trained for 50,000 steps with batch size 16, which takes around 10 hours. Then the Qwen2.5-VL is fine-tuned on the visual-prompted generation task for 20,000 steps, with batch size of 1 and gradient accumulation every 8 steps, which takes around 8 hours. Finally, the VLM and segmentation modules are jointly trained for grounded report generation, with batch size of 1 and gradient accumulation every 8 steps. This final stage takes approximately 48 hours for 100,000 steps.
Inference. All inferences are conducted on 1 A100-80G GPU. We use mixed-precision without quantization. The average inference latency is 0.49 second per sample for visual-prompted generation, 0.98 second per sample for grounded report generation, and 1.01 second per sample for text-guided grounded report generation.

F. Supplementary Details of OmniAbnorm-CT-14K

F.1. Quality Verification

Annotation Quality. We conducted rigorous and comprehensive quality verification on the annotations in OmniAbnorm-CT-14K. For each annotator, we randomly sampled 100 annotated images on the axial plane, 50 on the coronal, and 50 on the sagittal plane. Then a senior radiologist with 12 years of experience assesses the annotation quality on four key metrics: (i) **Detection rate** measures the percentage of abnormalities successfully identified and annotated by the annotator, without any omission; (ii) **Grounding precision** evaluates the percentage of grounding annotations that properly encompass the primary regions of the abnormality, while minimizing false positive areas; (iii) **Report concordance** quantifies the percentage of description annotations that faithfully reflect the linked abnormal findings on the image; (iv) **Classification accuracy** measures the percentage of abnormalities that are correctly categorized. The verification results demonstrated that: (i) 3 out of 4 annotators achieved perfect scores across all four metrics (100%). (ii) Only one annotator had 10 abnormality annotations where the descriptions were not sufficiently accurate and the category labels were also incorrect, resulting in Report concordance and Classification accuracy of 95%, while maintaining 100% Detection rate and Grounding precision. These results confirm the high quality of annotations in our OmniAbnorm-CT-14K.
Cross-annotator Consistency. We further conducted inter-annotator consistency checks across 4 annotators on 40 random samples. For each sample, we enumerated all annotator pairs and, for each pair, computed the DSC between their segmentation masks and the BLEU-1 score between their abnormality-level report descriptions. Averaging these metrics across all annotator pairs yielded mean pairwise scores of DSC=84.5 for segmentation and BLEU-1=72.2 for finding descriptions, indicating reasonably good agreement among the annotators.

F.2. Mitigation of the Long-Tail Distribution

To mitigate the long-tail distribution in OmniAbnorm-CT-14K, we identified underrepresented organs in our annotated corpus and strategically employed GPT-4o to analyze unlabeled reports. Using the following prompt, we efficiently filtered cases

containing abnormal findings related to these underrepresented organs:

> This is a report of a CT scan: $Report$
> Please help me carefully check if the report mentions any abnormal findings that belong to the following anatomical areas: $List of Underrepresented Organs$.
> If there are, please output 'YES', otherwise output 'NO'. Do not output any other information.

These identified cases were then prioritized in our annotation pipeline. To evaluate the effectiveness of this strategy, we randomly sampled 1,000 images before and after implementation. As shown in Fig. 7, prior to our intervention, the top 20 organs (out of 82 total) accounted for 79.8% of all annotations, while the top 85 abnormality categories (out of 340 total) constituted 80.1%. Following our strategy, the underrepresented organs and rare abnormality categories received notably increased annotation coverage: the representation of underrepresented organs increased by 19.6%, while rare abnormality categories increased by 11.6%, significantly enhancing the diversity of our dataset.

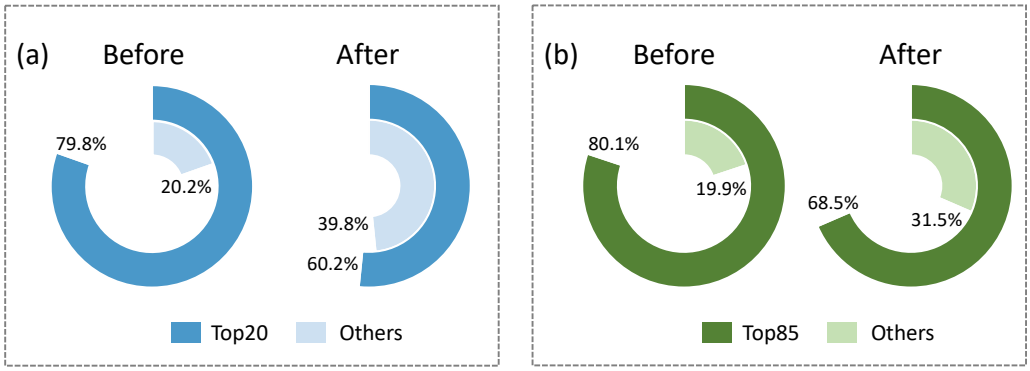


Figure 7. The distribution before and after prioritizing annotation for underrepresented organs. (a) Comparison of organ distribution in annotations; (b) Comparison of abnormality category distribution in annotations.

F.3. Comparison against Existing Datasets

We compare OmniAbnorm-CT-14K with existing public CT image datasets in Table 9, including those widely used for lesion segmentation, lesion detection, organ segmentation, or report generation. In contrast to the substantial limitations exhibited by these datasets, as detailed the main paper, OmniAbnorm-CT-14Kepresents the first large-scale dataset designed for abnormality grounding and description across CT images from different views and body regions.

Dataset	Task	Anatomy	Plane	Report	#Category	#Image
ULS Bone [15]	Lesion Seg.	Bone	Axial	✗	1	151
ULS Pancreas [15]	Lesion Seg.	Pancreas	Axial	✗	1	119
MSD Liver [5]	Lesion Seg.	Liver	Axial	✗	1	131
MSD Lung [5]	Lesion Seg.	Lung	Axial	✗	1	63
MSD Colon [5]	Lesion Seg.	Colon	Axial	✗	1	126
MSD Pancreas [5]	Lesion Seg.	Pancreas	Axial	✗	1	281
COVID19 [39]	Lesion Seg.	Lung	Axial	✗	1	20
KiTS23 [25]	Lesion Seg.	Kidney	Axial	✗	2	489
KiPA22 [23]	Lesion Seg.	Kidney	Axial	✗	2	70
NSCLC [8]	Lesion Seg.	Lung	Axial	✗	1	85
LIDC IDRI [6]	Lesion Seg.	Lung	Axial	✗	1	750
LNDb [47]	Lesion Seg.	Lung	Axial	✗	1	236
INSTANCE22 [35]	Lesion Seg.	Brain	Axial	✗	1	100
Seg.Rap2023 [38]	Lesion Seg.	Head & Neck	Axial	✗	2	120
FUMPE [43]	Lesion Seg.	Lung	Axial	✗	1	35
RibFrac [60]	Lesion Seg.	Rib	Axial	✗	1	420
Adrenal ACC Ki67 [4]	Lesion Seg.	Adrenal Gland	Axial	✗	1	29
LNQ2023 [17]	Lesion Seg.	Lung	Axial	✗	1	393
NIH-LN [52]	Lesion Seg.	Lung	Axial	✗	1	175
CCC-18 [54]	Lesion Seg.	Chest & Abdomen	Axial	✗	✗	404
DeepLesion [58]	Lesion Det.	Whole Body	Axial	✗	✗	33K
TotalSegmentator [55]	Organ Seg.	Whole Body	Axial	✗	✗	1.2K
AbdomenAtlas [48]	Organ Seg.	Chest & Abdomen	Axial	✗	✗	8K
CTRATE [20]	Report Gen.	Chest	Axial	✓	✗	26K
BIMCV-R [12]	Report Gen.	Chest	Axial	✓	✗	8K
ReXGroundingCT [7]	Lesion Seg.	Chest	Axial	✓	14	3K
	Lesion Seg.		**Axial**	✓	**340**	**10K**
OmniAbnorm-CT-14K	**Lesion Det.**	**Whole Body**	**Coronal**	✓	**255**	**2K**
	Report Gen.		**Sagittal**	✓	**223**	**1.6K**

Table 9. Comparison of key characteristics between OmniAbnorm-CT-14K and widely-used public CT imaging datasets. Note that CCC18 and DeepLesion has no category label for each annotated lesion. Even though some scans in these datasets are isotropic, they are acquired as axial-plane. Abbreviations: Category = Abnormality Category; Seg. = Segmentation; Det. = Detection; Gen = Generation.

F.4. Generation of Text Queries and Visual Prompts

We simulate four visual prompts for each annotated abnormality, mimicking how clinicians would select the most appropriate highlighting method based on the abnormality's shape and location: (i) **cropped region.** We extract the minimum bounding box that completely contains the annotated lesion region, add a 50-pixel padding around this region, and crop the original image to center the abnormality; (ii) **ellipse.** We fit an optimal ellipse to the largest contour extracted from the lesion mask. (iii) **contour.** We smooth the lesion mask using Gaussian blur, detect its contours, and refine them with polygon approximation. (iv) **bounding box.** We identify the minimum bounding box that completely contains the annotated lesion region, and add a 10-pixel padding around.

To simulate radiologists approaching CT images with prior knowledge, we provide GPT-4o with patient information (complaints, medical history, etc) to generate text queries that inquire specific abnormality based on such preliminary information. We use the following prompt template:

Assuming a CT image has one or more abnormal findings, I will provide detailed information about them. Please help me generate prompts to test a VLM's ability to localize and analyze specific abnormalities.

Requirements:
1. Generate prompts in English.
2. Ensure accurate information. The prompt content must stem from the real abnormality information, CT image reports, and doctor's discussion results I provided. You may use equivalents expressions in medical terminology, but do not introduce any content beyond these provided information.
3. Clear indication. Ensure each prompt refers to the corresponding abnormality without confusion with other abnormalities in the image.
4. Clear task. Each prompt should end with a clear request for the VLM to perform localization and analysis tasks.
5. Simulate realistic pre-examination clinical queries. Create prompts that reflect how clinicians approach CT images with preliminary information (such as the patient's complaints, medical history, other test results, etc., if available) and medical knowledge (such as common abnormalities associated with the patient's information or typically found in this imaged region), without being overly specific. Importantly, prompts should be broad enough to guide examination of suspicious areas and must not include detailed descriptions or conclusions from findings, reports and discussion results that would only be available after examining the CT image, such as specific abnormality details, exact measurements, precise locations, or definitive characteristics that could only be determined after image interpretation.
Some appropriate examples:
$Some Example Queries$
6. Avoid data leakage. To evaluate the VLM's ability to localize and analyze abnormalities independently, do not provide complete findings or diagnostic conclusions in the query. This prevents the VLM from bypassing the analytical process by retrieving answers directly from the prompt, while maintaining challenge authenticity.
7. Diversity. Generate at least 1 and at most 5 prompts with different perspectives for each abnormality. Each prompt should have clearly different focuses, avoiding content redundancy.

The information about all the abnormalities on the CT images:
$Abnormal Findings$
The clinical presentation of the patient corresponding to this CT image is:
$Presentation$
The overall report for the CT image containing these abnormalities is:
$Whole Report$
The doctors' discussion results for the patient corresponding to this CT image are:
$Impression$
Your response must follow this exact format:
$Output Template$

F.5. Detailed Distribution of OmniAbnorm-CT-14K

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Brain	Cerebral parenchyma	Brain parenchymal atrophy	31	2	4
		Brain parenchymal edema	49	13	4
		Brain parenchymal soft tissue mass	280	57	57
		Brain parenchymal thin-walled cystic mass	52	1	3
		Brain parenchymal thick-walled cystic mass	17	0	1
		Brain parenchymal hemorrhage or contusion	198	42	31
		Acute infarct	88	1	0
		Lacunar infarct	54	11	8
		Encephalomalacia	27	6	3
		Brain parenchymal morphological altrenation	102	31	23
		Other non-mass effect lesions	44	5	5
		Intracranial air	14	8	10
		Hyperdense lesions in brain parenchyma	72	24	44
		Others	35	13	6
	Ventricles and cisterns	Ventricular or cisternal enlargement	122	25	25
		Ventricular or cisternal soft tissue mass	49	8	6
		Ventricular or cisternal cystic mass	13	1	0
		Ventricular or cisternal hemorrhage	60	4	3
		Others	11	2	3
	Meninges (including dura mater, pia mater and arachnoid mater)	Meningeal cystic mass	4	0	0
		Meningeal soft tissue mass	8	9	6
		Meningeal hemorrhage	137	27	10
		Meningeal effusion	32	9	3
		Meningeal thickening	6	0	0
		Others	4	0	2
	Pituitary and Sellar Region	Pituitary stalk thickening	0	0	0
		Pituitary stalk lateral displacement	0	0	0
		Pituitary enlargement	0	0	2
		Pituitary atrophy	0	2	4
		Pituitary calcification	0	0	3
		Pituitary or sella region soft tissue mass	20	16	27
		Pituitary or sella region cysitc mass	3	5	6
		Others	1	0	0

Table 10. Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Spine	Spinal Cord	Spinal cord compression	3	0	2
		Spinal cord soft tissue mass	12	1	6
		Spinal cord hemorrhage or contusion	0	0	0
		Spinal morphological alteration	0	0	0
		Syringomyelia	0	0	0
		Others	4	0	1
	Intervertebral disc	Intervertebral disc morphological alteration	4	2	1
		Intervertebral disc ossification or calcification	0	0	0
		Intervertebral disc gas	4	3	2
		Intervertebral disc extrusion	0	0	0
		Intervertebral disc soft tissue mass	5	1	1
		Others	0	0	1
	Spinal meninge (including dura mater, arachnoid mater, and pia mater)	Spinal meningeal hemorrhage	0	0	0
		Spinal meningeal effusion	0	0	1
		Spinal meningeal thickening	0	0	0
		Meningeal soft tissue mass	7	2	6
		Meningeal cystic mass	0	2	0
		Others	2	1	5
Eye	Eyeball	Eyeball atrophy	4	1	2
		Eyeball positioning or morphological alteration	44	7	6
		Eyeball density alteration	45	20	14
		Eyeball soft tissue mass	19	3	1
		Eyeball wall thickening morphological alteration	5	3	0
		Complete or partial absence of eyeball structure	2	0	0
		Others	2	2	0
	Ocular Adnexa	Orbital density changes	54	36	15
		Intraorbital gas (e.g., soft tissue mass, fluid accumulation)	4	4	3
		Extraocular muscle hypertrophy or atrophy	16	43	11
		Optic nerve thickening or soft tissue mass	7	3	7
		Optic nerve atrophy	0	0	0
		Others	3	2	3
	Lacrimal gland and lacrimal sac	Lacrimal gland and lacrimal sac enlargement or mass	19	12	5
		Lacrimal gland and lacrimal sac calcification or fluid	4	2	0
		Others	1	0	1
Ear	External Ear	External auditory canal stenosis or atresia	5	1	2
		External auditory canal soft tissue mass	11	5	1
		Others	1	1	0

Table 11. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Ear	Middle Ear	Middle ear fluid or hemorrhage	5	0	0
		Middle ear soft tissue mass	23	23	3
		Ossicular chain destruction or deformity	7	2	0
		Tympanic membrane thickening or calcification	2	2	0
		Middle ear gas density change	6	0	1
		Others	1	2	0
	Inner Ear	Inner ear congenital structural alteration	2	1	0
		Inner ear bone destruction or sclerosis	11	4	0
		Internal auditory canal enlargement or narrowing	2	0	0
		Labyrinthine structural alteration	4	2	0
		Others	1	1	0
Sinus	Sinus cavity	Sinus effusion	18	8	0
		Sinus hemorrhage	6	3	3
		Sinus soft tissue mass	98	85	33
		Sinus cystic mass	8	5	1
		Sinus mucosal thickening	24	25	7
		Others	9	3	2
	Sinus ostium	Sinus obstruction or stenosis	1	4	0
		Sinus widening	0	0	1
		Others	1	1	0
	Nasal septum	Nasal septum deviation or thickening	7	6	0
		Nasal septum perforation or defect	1	1	0
		Others	6	2	0
Pharynx	Pharynx (including nasopharynx, oropharynx, and hypopharynx)	Pharyngeal narrowing or obstruction	4	2	1
		Pharyngeal soft tissue mass	46	14	10
		Pharyngeal cystic mass	10	2	4
		Pharyngeal wall thickening	4	3	0
		Pharyngeal foreign body	0	0	0
		Others	3	4	1
	Larynx	Laryngeal narrowing or obstruction	3	2	0
		Vocal cord asymmetry	2	0	0
		Vocal cord soft tissue mass	6	1	1
		Laryngeal cartilage calcification	0	0	0
		Laryngeal cystic mass	1	1	0
		Laryngeal foreign body	1	0	0
		Others	1	0	0
	Pharyngeal space	Pharyngeal space soft tissue mass	46	23	16
		Pharyngeal space cystic mass	13	6	2
		Pharyngeal emphysema	7	0	0
		Others	5	7	1

Table 12. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Pharynx	Pharyngeal space	Pharyngeal space soft tissue mass	46	23	16
		Pharyngeal space cystic mass	13	6	2
		Pharyngeal emphysema	7	0	0
		Others	5	7	1
Parotid gland	Parotid gland	Parotid gland enlargement	20	2	0
		Parotid gland atrophy	4	0	0
		Parotid gland soft tissue mass	33	2	2
		Parotid gland cystic mass	16	0	0
		Parotid gland calcification or stone	15	0	2
		Others	10	0	0
Thyroid gland	Thyroid gland	Thyroid enlargement	12	2	2
		Thyroid atrophy	1	0	0
		Thyroid soft tissue mass	41	4	0
		Thyroid cystic mass	3	1	0
		Thyroid calcification	1	0	0
		Ectopic thyroid gland	5	1	0
		Others	6	0	1
Trachea	Tracheal lumen	Tracheal stenosis or obstruction	12	4	7
		Tracheal dilatation	11	5	2
		Tracheal soft tissue mass	12	6	1
		Tracheal wall thickening	9	0	0
		Tracheal wall calcification	3	1	4
		Tracheal wall defect	3	0	2
		Others	1	0	2
Lung	pulmonary parenchyma	Atelectasis	106	10	3
		Incomplete lung expansion	52	3	1
		Pulmonary consolidation	203	24	2
		Pulmonary ground-glass opacities	343	52	5
		Pulmonary emphysema	45	10	2
		Pulmonary solitary nodule or mass	226	46	10
		Pulmonary diffusely distributed multiple nodules	285	76	3
		Pulmonary parenchymal fibrosis	42	12	0
		Pulmonary thin-walled cavitation	48	6	1
		Pulmonary thick-walled cavities	44	1	0
		Pulmonary cystic mass	53	14	1
		Others	14	6	1
	Bronchi	Bronchiectasis	97	19	6
		Bronchial wall thickening, stenosis, or occlusion	15	11	5
		Bronchial foreign body	4	1	0
		Others	3	0	1

Table 13. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Lung	Lung interstitial	Lung interstitial fibrosis and thickening	78	10	0
		Honeycomb lung	34	0	0
		Others	2	0	0
	Pleura	Pleural thickening	39	10	1
		Pleural effusion	268	27	3
		Pneumothorax	36	2	0
		Pleural calcification	24	0	1
		Pleural soft tissue mass	50	8	0
		Others	8	1	1
Mediastinum	Mediastinal soft tissue	Mediastinal shift	29	4	0
		Mediastinal soft tissue mass	200	31	15
		Mediastinal cystic mass	19	0	2
		Mediastinal hemorrhage	15	0	0
		Mediastinal emphysema	58	7	1
		Others	10	0	0
Diaphragm	Diaphragm	Diaphragmatic hernia	19	16	12
		Diaphragmatic elevation	5	4	0
		Diaphragmatic soft tissue mass	7	2	0
		Others	0	2	0
Thymus	Thymic parenchyma	Thymic enlargement	1	0	0
		Thymic atrophy	0	0	0
		Thymic soft tissue mass	6	0	0
		Thymic cystic mass	0	0	0
		Thymic calcification	4	0	0
		Others	1	0	0
Oral cavity	Oral Soft Tissue	Oral soft tissue mass	8	0	0
		Oral cystic mass	3	0	0
		Oral soft tissue calcification	0	0	0
		Others	0	0	0
	Teeth and alveolar bone	Dental developmental anomalies	1	0	0
		Dental positional anomalies	3	4	5
		Dental calcification or caries	3	0	0
		Alveolar bone resorption or hyperplasia	3	0	0
		Alveolar soft tissue mass	8	2	0
		Alveolar cystic mass	8	1	1
		Others	0	1	0
Heart	Cardiac chambers (atrium or ventricle)	Cardiac chamber enlargement	41	7	3
		Cardiac chamber mass	28	7	2
		Others	17	1	0

Table 14. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Heart	Myocardium	Myocardial hypertrophy	4	1	1
		Myocardial thinning	3	0	0
		Myocardial calcification	4	1	0
		Myocardial density alteration	6	0	0
		Others	3	3	0
	Heart valve	Valvular calcification	3	0	0
		Others	3	0	0
	Pericardium	Pericardial thickening	9	0	2
		Pericardial calcification	6	1	0
		Pericardial effusion	54	7	2
		Pericardial hemorrhage	2	0	0
		Pericardial emphysema	6	1	0
		Others	4	0	0
	Coronary arteries	Coronary artery myocardial bridge	1	0	0
		Coronary artery dilation	0	0	0
		Others	3	1	0
Vascular structure	Artery	Arterial widening	66	24	9
		Aneurysm	138	44	39
		Atherosclerotic plaque of arterial wall	39	13	4
		Arterial wall ulcer	8	2	2
		Arterial dissection or intramural hematoma	22	15	8
		Arterial stenosis	24	8	10
		Arterial occlusion	55	9	0
		Arterial filling defect	80	23	8
		Arterial contour abnormality	51	23	23
		Arteriovenous fistula	16	7	1
		Arterial wall inflammatory exudate	4	0	0
		Others	39	9	7
	Vein	Venous dilation	41	4	6
		Varicosity	23	11	9
		Venous wall inflammation	0	0	0
		Venous stenosis	7	1	1
		Venous occlusion	5	2	1
		Venous filling defect	48	12	3
		Venous morphological abnormality	29	6	3
		Venous wall inflammatory exudate	2	1	0
		Others	13	8	0
	Capillary	Capillary dilation	0	0	0
		Capillary malformation proliferation	1	0	0
		Others	0	0	0

Table 15. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Gastrointestinal tract (including esophagus, stomach, and intestines)	Gastrointestinal lumen	Gastrointestinal dilatation	265	212	112
		Gastrointestinal narrowing	20	5	6
		Gastrointestinal foreign body	2	6	0
		Gastrointestinal air-fluid level	31	8	5
		Gastrointestinal luminal contents abnormal density	17	6	4
		Others	8	3	1
	Gastrointestinal tract wall	Gastrointestinal wall thickening	398	212	120
		Gastrointestinal wall mass	103	56	26
		Gastrointestinal wall rupture and perforation	43	12	4
		Gastrointestinal wall ulceration	0	0	0
		Others	7	9	5
	Gastrointestinal positioning abnormality	Gastrointestinal herniation or deformity	117	67	30
		Others	5	1	1
	Gastrointestinal morphological abnormalities	Gastrointestinal diverticulum	53	35	15
		Gastrointestinal malrotation and volvulus	17	17	3
		Gastrointestinal annular or concentric abnormality	28	22	10
		Others	2	1	2
	Mesentery	Mesenteric volvulus	3	4	2
		Mesenteric edema	10	3	3
		Mesenteric panniculitis	26	6	6
		Mesenteric soft tissue mass	28	6	9
		Others	0	4	2
Liver	Hepatic parenchyma	Hepatic parenchymal morphological alteration	99	34	11
		Hepatic parenchymal hyperdensity	18	2	0
		Hepatic parenchymal hypodensity	71	25	9
		Hepatic parenchymal soft tissue mass	439	58	20
		Hepatic parenchymal thin-walled cystic mass	124	23	10
		Hepatic parenchymal thick-walled cystic mass	21	6	4
		Liver contusion or hemorrhage	10	3	0
		Intrahepatic bile duct emphysema	13	1	0
		Intrahepatic bile duct fluid accumulation	0	0	0
		Intrahepatic biliary stones or calcification	7	0	0
		Others	16	4	1
Gallbladder	Gallbladder lumen	Gallbladder distension	39	10	1
		Gallbladder atrophy or shrinkage	4	1	0
		Gallbladder stone	87	22	2
		Gallbladder contents density change	32	5	0
		Others	3	1	0

Table 16. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Gallbladder	Gallbladder wall	Gallbladder wall thickening	53	11	1
		Gallbladder wall calcification	9	1	0
		Gallbladder wall mass	16	2	1
		Gallbladder wall rupture	11	1	0
		Others	3	1	1
	Gallbladder morphology	Gallbladder position alteration	1	0	0
		Gallbladder congenital morphological variation	0	0	0
		Others	1	1	0
	Extrahepatic bile ducts	Extrahepatic bile duct dilation	26	8	0
		Extrahepatic bile duct wall thickening	2	0	0
		Extrahepatic bile duct soft tissue mass	1	0	1
		Extrahepatic bile duct cystic mass	1	1	0
		Extrahepatic bile duct stenosis or obstruction	2	0	0
		Extrahepatic bile duct content density alteration	6	1	0
		Extrahepatic bile duct stone	15	5	0
		Extrahepatic bile duct injury or rupture	0	0	0
		Others	1	0	0
Pancreas	Pancreatic parenchyma	Pancreatic parenchymal soft tissue mass	137	8	4
		Pancreatic cystic mass	69	7	4
		Pancreatic calcification	20	1	0
		Pancreatic enlargement	60	1	1
		Pancreatic atrophy	8	1	1
		Others	16	0	0
	Pancreatic duct	Pancreatic ductal dilatation	13	3	0
		Pancreatic ductal stone	3	0	0
		Others	0	0	0
	Pancreatic morphology	Pancreatic congenital anomaly	5	0	0
		Pancreatic positional displacement	1	0	0
		Others	1	0	0
Spleen	Splenic parenchyma	Splenic parenchymal calcification	15	0	1
		Splenic parenchymal soft tissue mass	52	11	3
		Splenic parenchymal cystic mass	27	10	0
		Splenic parenchymal infarct	19	1	0
		Splenic parenchymal rupture	12	1	0
		Others	18	2	0
	Spleen morphology	Spleen enlargement	102	22	5
		Others	17	1	0

Table 17. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Abdominopelvic peritoneum	Abdominopelvic peritoneum	Peritoneal inflammatory exudate	75	19	10
		Peritoneal thickening	19	3	1
		Peritoneal calcification	5	1	0
		Peritoneal soft tissue mass	364	135	90
		Peritoneal cystic mass	102	34	22
		Abdominopelvic fluid	280	96	43
		Abdominopelvic hemorrhage	68	3	1
		Abdominopelvic free air	61	17	10
		Retroperitoneal fibrosis	1	0	0
		Peritoneal or retroperitoneal lymph enlargement	17	18	4
		Extravasation of gastrointestinal content	0	1	0
		Abdominopelvic contrast agent leakage	22	13	4
		Others	12	12	2
Kidney	Renal Parenchyma	Renal parenchymal soft tissue mass	150	81	29
		Renal parenchymal cystic mass	239	68	21
		Renal parenchymal calcification	9	9	2
		Renal parenchymal or subcapsular hemorrhage	11	3	2
		Renal infarct	8	3	0
		Others	27	8	1
	Renal pelvis and ureter	Hydronephrosis	81	63	3
		Ureteral dilatation	35	29	4
		Ureteral stricture or obstruction	1	1	0
		Ureteric stone	59	60	19
		Double renal pelvis and/or double ureter anomaly	5	3	1
		Renal pelvis soft tissue mass	13	4	0
		Renal pelvic cystic mass	1	3	0
		Others	4	5	2
	Renal morphology and position abnormalities	Renal morphological anomaly	33	20	1
		Renal enlargement	56	16	1
		Renal atrophy	23	6	2
		Ectopic or transplanted kidney	17	13	5
		Others	3	2	0
Bladder	Bladder cavity	Bladder distention	4	0	1
		Bladder stone	24	5	3
		Bladder content density change	13	2	1
		Others	4	3	0

Table 18. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Bladder	Bladder cavity	Bladder distention	4	0	1
		Bladder stone	24	5	3
		Bladder content density change	13	2	1
		Others	4	3	0
	Bladder wall	Bladder wall diffuse thickening	17	5	6
		Bladder wall calcification	1	3	1
		Bladder wall focal thickening or soft tissue mass	69	7	3
		Bladder wall defect or fistula	4	0	1
		Bladder wall emphysema	9	0	1
		Bladder wall diverticulum	31	2	2
		Others	1	0	0
	Bladder morphology	Bladder position displacement	9	2	2
		Bladder morphological anomaly	2	0	0
		Others	1	0	2
Adrenal gland	Adrenal gland	Adrenal soft tissue mass	122	17	1
		Adrenal cystic mass	32	0	0
		Adrenal calcification	0	3	0
		Adrenal gland thickening	16	0	0
		Adrenal atrophy	0	0	0
		Others	15	0	0
Prostate	Prostate	Prostate enlargement	18	12	14
		Prostate atrophy	0	0	0
		Prostatic cystic mass	2	0	0
		Prostatic soft tissue density anomaly	2	0	1
		Prostatic calcification	2	1	2
		Prostatic hemarrhage	0	0	0
		Others	0	0	0
Seminal vesicle	Seminal vesicle	Seminal vesicle soft tissue mass	1	0	0
		Seminal vesicle calcification	0	0	1
		Seminal vesicle cystic mass	3	0	0
		Others	0	0	0
Testes, epididymis, and scrotum	Testis	Testicular enlargement	1	0	0
		Testicular atrophy	0	0	0
		Testicular soft tissue mass	1	0	0
		Testicular calcification	0	0	0
		Testicular cystic mass	0	0	0
		Testicular torsion	0	0	0
		Testicular hemorrhage and rupture	0	0	0
		Others	0	0	0

Table 19. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Testes, epididymis, and scrotum	Testis	Testicular enlargement	1	0	0
		Testicular atrophy	0	0	0
		Testicular soft tissue mass	1	0	0
		Testicular calcification	0	0	0
		Testicular cystic mass	0	0	0
		Testicular torsion	0	0	0
		Testicular hemorrhage and rupture	0	0	0
		Others	0	0	0
	Scrotum	Scrotal effusion	1	0	0
		Scrotal hematoma	0	0	0
		Scrotal soft tissue mass	1	0	0
		Scrotal wall thickening	0	0	0
		Others	2	0	0
	Epididymis	Epididymis enlargement	0	0	0
		Epididymal soft tissue mass	0	0	0
		Epididymal calcification	0	0	0
		Epididymal cystic mass	0	0	0
		Epididymal thickening	0	0	0
		Others	0	0	0
Penis	Penis	Penile morphological anomaly	0	0	0
		Penile soft tissue mass	0	0	0
		Penile calcification	0	0	0
		Urethral calculi or foreign body	6	2	1
		Urethral Stricture	0	0	0
		Urethral dilation	1	0	0
		Others	4	0	0
Uterus	Uterus	Uterine morphological anomaly	7	1	2
		Uterine enlargement	15	3	6
		Uterine soft tissue mass	50	4	7
		Uterine calcification	7	0	2
		Uterine cavity effusion	2	0	1
		Uterine cavity hemorrhage	2	0	0
		Uterine cystic mass	4	0	0
		Others	4	1	1
Fallopian tube	Fallopian tube	Fallopian tube thickening	3	0	0
		Fallopian tube cystic mass	2	0	1
		Fallopian tube soft tissue mass	12	4	2
		Fallopian tube effusion	2	0	0
		Fallopian tube hemorrhage	0	0	0
		Fallopian tube calcification	0	0	0
		Others	0	0	0

Table 20. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Ovary	Ovary	Ovarian enlargement	9	3	0
		Ovarian Atrophy	0	0	0
		Ovarian cystic mass	85	12	6
		Ovarian soft tissue mass	96	6	3
		Ovarian calcification	3	1	0
		Ovarian torsion	9	0	0
		Others	1	1	0
Vagina and vulva	Vagina and vulva	Vaginal soft tissue mass	0	0	0
		Vaginal cystic mass	9	0	0
		Vaginal hemorrhage	1	0	0
		Vaginal emphysema	3	0	4
		Vaginal anatomical anomaly	3	0	1
		Others	6	0	0
Breast	Breast gland	Breast gland enlargement	2	0	0
		Breast gland atrophy	0	0	0
		Breast gland soft tissue mass	36	9	2
		Breast gland calcification	0	0	0
		Breast gland cystic mass	0	0	0
		Others	6	0	0
	Breast duct	Breast duct dilation	0	0	0
		Breast duct calcification	0	0	0
		Others	0	0	0
	Nipple	Nipple retraction	0	0	0
		Nipple calcification	0	0	0
		Others	0	0	0
	Areola	Areola thickening	1	0	0
		Others	0	0	0
Skeletal system	Skeletal system	Osteoporosis	9	1	8
		Osteomalacia	2	0	1
		Bone destruction or soft tissue mass	219	124	117
		Bone cystic mass	19	32	17
		Osseous sclerosis	179	80	121
		Osteonecrosis	64	13	24
		Bone fracture	308	172	247
		Periosteal reaction	23	9	5
		Periosteal thickening	8	5	1
		Bone callus and post-fracture healing	10	8	2
		Scar of fracture fixation removal	0	0	0
		Bone deformation	67	44	73
		Skeletal asymmetry	10	2	1
		Cartilage calcification	4	4	8
		Chondral calcification	5	0	1
		Others	30	50	26

Table 21. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.

Organ	Anatomical Structure	Category	Axial	Coronal	Sagittal
Joint	Joint	Joint space narrowing	6	12	3
		Joint space widening	3	0	3
		Joint cartilage degradation	1	0	2
		Joint cartilage calcification	3	1	1
		Joint capsule thickening	0	0	0
		Intra-articular effusion	13	3	2
		Intra-articular hemorrhage	4	0	0
		Intra-articular gas	1	0	4
		Joint periarticular soft tissue swelling	2	0	0
		Irregular articular surfaces	12	5	2
		Joint subluxation or dislocation	30	14	22
		Intra-articular loose body	8	1	2
		Others	5	12	2
Muscle	Muscle	Muscle swelling	8	5	1
		Muscle atrophy	13	1	0
		Muscular soft tissue mass	40	9	2
		Muscular cystic mass	11	13	3
		Muscular hemorrhage	7	0	1
		Muscle calcification	3	0	0
		Muscle open injury and tear	0	0	0
		Tendon calcification	19	3	13
		Tendon tear or rupture	1	0	1
		Others	7	4	1
Skin and subcutaneous fat	Skin and subcutaneous fat	Subcutaneous soft tissue mass	141	47	28
		Subcutaneous edema	18	6	2
		Subcutaneous effusion	26	3	1
		Subcutaneous inflammatory exudate	10	2	1
		Subcutaneous swelling	60	10	15
		Subcutaneous open wound and laceration	15	5	2
		Subcutaneous calcification	12	0	0
		Subcutaneous fat necrosis	2	0	0
		Abdominal wall hernia	48	3	6
		Others	33	5	9
Agenesis or ectopia	Agenesis or ectopia	Congenital developmental anomaly	78	35	14
		Postoperative changes	42	16	5
		Situs inversus	10	7	1
		Others	2	0	0
Implantation of artificial object	Implantation of artificial object	Implantation of artificial object	179	91	53
		Others	17	3	8

Table 22. (Continued) Detailed distribution of abnormality categories in OmniAbnorm-CT-14K.