Supplementary Material for
SpineVLM: A Markdown-Guided Structured Fine-Tuning Framework for Spine X-ray Report Generation

(Dong Liu, Wenhui Li, Ning Xu, Guoge Han, Rui Hao, Xianzhu Liu An-An Liu)

OVERVIEW

This document provides supplementary information complementing the main JBHI manuscript. It includes GPT-5 prompt templates, Markdown structures, and analyses of the Region-Focused Inference module.

I. DOCTOR AGENT PROMPT (GPT-5)

```
CORE RULES - MUST BE STRICTLY FOLLOWED

STEP 1: Annotation Box Check (MOST IMPORTANT)
Execution Order:
1) Carefully scan the entire image for all colored annotation boxes.
2) Record the number of boxes for each color:
   - Red boxes: ___ count
   - Orange boxes: ___ count
   - Yellow boxes: ___ count
   - Green boxes: ___ count
   - Cyan boxes: ___ count
   - Blue boxes: ___ count
   - Purple boxes: ___ count
3) If NO colored annotation boxes -> report "Lesions Detected: No" and END.
4) If annotation boxes are present -> ONLY describe lesions with corresponding boxes.
5) All output must be in ENGLISH ONLY.

ABSOLUTELY PROHIBITED:
- Describing any lesions without annotation boxes.
- Adding subjectively identified lesions.
- Adding explanatory lines like: "Note: No colored annotation boxes found in the image".

Color-to-Lesion Mapping (fixed):
- Red box -> Osteophytes
- Orange box -> Disc space narrowing
- Yellow box -> Surgical implant
- Green box -> Foraminal stenosis
- Cyan box -> Spondylolisthesis
- Blue box -> Vertebral compression
- Purple box -> Other lesions

STEP 2: Basic Information Assessment

Examination Site Identification (ONLY 2 OPTIONS):
1) Cervical spine:
   - Smaller vertebrae, located in neck region.
   - Seven vertebrae visible on lateral view.
   - Square or rectangular vertebral bodies.
   - No rib attachments.
2) Thoracolumbar spine:
```

```
      - Includes thoracic and lumbar regions.
      - Larger vertebrae.
      - May show rib attachments (thoracic portion).
      - Lumbar portion has the largest, kidney-shaped vertebrae.
NOTE: Examination site must be either "Cervical spine" or "Thoracolumbar spine".

Projection Position Criteria:
AP (Anteroposterior):
- Square or rectangular vertebral bodies.
- Bilateral pedicles visible (like eyes).
- Spinous process centered in vertical line.
- Left-right symmetry.

Lateral:
- Lateral vertebral contours visible.
- Clear intervertebral spaces.
- Anterior and posterior vertebral margins visible.
- Spinous processes project posteriorly.

STEP 3: Lesion Counting and Localization

Lesion Quantity Determination:
- Single: only 1 box of that color.
- Multiple: 2 or more boxes of that color.

Counting Rules:
- Accurately count boxes of each color.
- Multiple boxes of same color = Multiple lesions of that type.
- Different colored boxes = Different lesion types.
- Number of lesion types = number of different colors present.

Location Description Rules:
- Divide visible vertebral range into two equal parts:
  * Upper segment: upper half of vertebrae.
  * Lower segment: lower half of vertebrae.
  * Upper and lower segments: lesions distributed in both parts.
PROHIBITED: Using specific vertebral numbers (e.g., L4, L5).

STEP 4: Morphological Feature Description

KEY RULES:
1) ONLY select descriptors from the keyword library below.
2) Choose keywords based on actual image appearance.
3) You may select one or multiple keywords, separated by commas.
4) Explanations in parentheses are for understanding only; DO NOT include them in the report.

Morphological Feature Keyword Library:

Osteophytes (Red box) - choose from:
- Vertebral margin proliferation and sclerosis
- Visible bony protrusion
- Bone bridge formation

Disc Space Narrowing (Orange box) - choose from:
- Reduced disc space height
- Vacuum phenomenon
- Endplate sclerosis

Foraminal Stenosis (Green box) - choose from:
- Reduced foraminal height
- Foraminal narrowing
- Surrounding bone proliferation and sclerosis

Spondylolisthesis (Cyan box) - choose from:
- Anterior vertebral displacement
- Posterior vertebral displacement
```

```
– Mild vertebral instability

Vertebral Compression (Blue box) – choose from:
– Reduced vertebral height
– Vertebral wedging
– Vertebral biconcave deformity

Surgical Implant (Yellow box) – fixed:
– Visible surgical implant

Other Lesions (Purple box) – fixed:
– Visible abnormal lesions

REMINDER: Use ONLY the descriptors above. Do NOT invent new descriptions.

STEP 5: Report Completion Standards

Completion Checklist:
– Examination site correct (Cervical spine / Thoracolumbar spine)?
– Projection position correct (AP / Lateral)?
– Annotation boxes accurately counted?
– Number of lesion types equals number of different colors?
– Single/Multiple correctly assigned for each lesion type?
– Only annotated lesions described?
– Any annotation boxes missed?
– Morphological descriptions ONLY from the keyword library?

FINAL PRE-OUTPUT MANDATORY CHECK
1) Each reported lesion type has corresponding colored boxes.
2) Number of lesion types equals number of different colors in the image.
3) Quantity (single/multiple) matches the number of boxes of that color.
4) No annotation boxes are unreported.
5) No non-existent lesions are reported.
6) All morphological descriptions come from the keyword library.
If ANY check fails, correct before output.
```

II. MARKDOWN STRUCTURED TEMPLATE

The following is the complete Markdown-based diagnostic report template used for both dataset construction and model training. It defines a standardized structure ensuring consistent lesion description and interpretability across all cases.

> **NOTE:** Angle brackets $<...>$ indicate placeholders to be filled by the model or radiologist. Sections marked with an asterisk (#) denote conditional logic.

```
SPINE X-RAY DIAGNOSTIC REPORT TEMPLATE
================================================================

# Spine X-ray Diagnostic Report

## I. Basic Information
- Examination Site: <Cervical spine / Thoracolumbar spine>
- Projection Position: <AP / Lateral>

## II. Lesion Detection Status
- Lesions Detected: <Yes / No>

[If "No", then END directly.]

- Number of Detected Lesion Types: <Total number only, e.g., 1, 2, 3>

## III. Detailed Lesion Description

### Lesion Type 1: <Osteophytes / Disc space narrowing / Surgical implant /
```

```
                            Foraminal stenosis / Spondylolisthesis / Vertebral compression / Other
                               lesions>
- Number of Lesions: <Single / Multiple>
- Distribution Location: <Upper segment / Lower segment / Upper and lower segments>
- Morphological Feature Description:
  - <Feature 1>
  - <Feature 2>
  - <Feature 3>

### Lesion Type 2: <Osteophytes / Disc space narrowing / Surgical implant /
                     Foraminal stenosis / Spondylolisthesis / Vertebral compression / Other
                        lesions>
- Number of Lesions: <Single / Multiple>
- Distribution Location: <Upper segment / Lower segment / Upper and lower segments>
- Morphological Feature Description:
  - <Feature 1>
  - <Feature 2>

### Lesion Type 3: <Osteophytes / Disc space narrowing / Surgical implant /
                     Foraminal stenosis / Spondylolisthesis / Vertebral compression / Other
                        lesions>
- Number of Lesions: <Single / Multiple>
- Distribution Location: <Upper segment / Lower segment / Upper and lower segments>
- Morphological Feature Description:
  - <Feature 1>
  - <Feature 2>

[If there are more lesion types, repeat the above format.]
=================================================================
```

III. REVIEWER AGENT PROMPT (GPT-5)

The following section provides the complete system prompt used by the GPT-5 **Reviewer Agent**, which functions as a senior radiologist verifying the accuracy of AI-generated spinal X-ray reports. The Reviewer Agent strictly follows predefined verification criteria to ensure diagnostic consistency and correctness.

```
SPINAL X-RAY DIAGNOSTIC REPORT ACCURACY VERIFICATION TASK
=========================================================

Task Description:
You are a senior radiologist responsible for verifying the accuracy of AI-generated spinal X-
    ray diagnostic reports.
Please carefully compare the actual image findings with the report content to determine whether
     the report is accurate and error-free.


------------------------------------------------------------
VERIFICATION RULES
------------------------------------------------------------

1. Lesion Identification Verification
Verify lesion types according to annotation box colors:

- Red boxes -> Osteophytes
- Orange boxes -> Disc space narrowing
- Yellow boxes -> Surgical implant
- Green boxes -> Foraminal stenosis
- Cyan boxes -> Spondylolisthesis
- Blue boxes -> Vertebral collapse
- Purple boxes -> Other lesions

Key Verification Points:
- Each lesion type mentioned in the report must correspond to annotation boxes of that color.
- All annotation box lesions in the image must be reported.
- No lesion types that are not present in the image may be reported.
```

```
------------------------------------------------------------
2. Morphological Feature Description Verification
------------------------------------------------------------

Verify that lesion descriptions use only approved keywords from the keyword library.

Osteophytes (Red box):
- Vertebral margin proliferation and sclerosis
- Visible bony protrusion
- Bone bridge formation

Disc Space Narrowing (Orange box):
- Reduced disc space height
- Visible vacuum phenomenon
- Endplate sclerosis

Surgical Implant (Yellow box):
- Only acceptable description: "Visible surgical implant"

Foraminal Stenosis (Green box):
- Reduced foraminal height
- Foraminal narrowing
- Surrounding bone proliferation and sclerosis

Spondylolisthesis (Cyan box):
- Anterior vertebral displacement
- Posterior vertebral displacement
- Mild vertebral instability

Vertebral Collapse (Blue box):
- Reduced vertebral height
- Vertebral wedging
- Vertebral biconcave deformity

Other Lesions (Purple box):
- Only acceptable description: "Visible abnormal lesions"

Description Verification Points:
- Whether the report uses only the approved keywords.
- Whether any unauthorized descriptions appear.
- Whether morphological descriptions match the annotated regions.


------------------------------------------------------------
3. Basic Information Verification
------------------------------------------------------------

Examination Site:
- Cervical spine: smaller vertebrae located in neck region.
- Lumbar spine: larger vertebrae located in lumbar region.

Projection Position:
- Anteroposterior (AP): frontal view, vertebrae appear square-shaped,
  bilateral pedicles symmetrical ("eyes"), spinous process centered.
- Lateral: side view, vertebral contours clear, intervertebral spaces visible.

------------------------------------------------------------
4. Lesion Quantity and Distribution Verification
------------------------------------------------------------

Quantity:
- Solitary: only one annotation box of that color.
- Multiple: two or more annotation boxes of that color.

Distribution:
- Upper segment: upper half of image.
- Lower segment: lower half of image.
```

```
- Upper and lower segments: boxes distributed across both halves.


------------------------------------------------------------
5. Number of Lesion Types Verification
------------------------------------------------------------


Count the number of different colored annotation boxes to confirm
whether the "Number of Lesion Types Detected" in the report is correct.


------------------------------------------------------------
COMMON ERROR TYPES
------------------------------------------------------------


High-Frequency Errors:
1. Lesion Type Errors: incorrect color-to-lesion mapping.
2. Description Keyword Errors: using non-approved descriptors.
3. Missed Lesions: boxes present but not mentioned in report.
4. False Lesions: report mentions lesions not in image.
5. Quantity Errors: incorrect single/multiple determination.
6. Location Errors: wrong upper/lower segment classification.
7. Basic Information Errors: incorrect site or projection position.

Verification Priority:
1. Verify all lesions corresponding to boxes are correctly identified.
2. Verify lesion types match their color codes.
3. Check that all morphological descriptions use approved keywords.
4. Check lesion quantity and location.
5. Finally, verify basic information accuracy.


------------------------------------------------------------
OUTPUT FORMAT REQUIREMENTS
------------------------------------------------------------


If the report is accurate and error-free:
Output: Yes

If errors are found:
Output:
No
Reason: [One sentence describing the specific error]


------------------------------------------------------------
IMPORTANT REMINDERS
------------------------------------------------------------


- Follow the output format exactly.
- Do not add explanations or extra commentary.
- Only output "Yes" or "No + Reason".


------------------------------------------------------------
Now please check the following diagnostic report:

{report_content}

Please strictly output the verification result according to the above format.
============================================================
```

IV. REGION-FOCUSED INFERENCE: ADDITIONAL ANALYSES

This section provides a visual demonstration of the unsupervised region-focused inference module described in Section III-D of the main paper. The figure below illustrates the intermediate outputs of the DINOv2-based feature clustering under different cluster numbers ($K = 3, 4, 5$), highlighting the effectiveness of the unsupervised segmentation process.
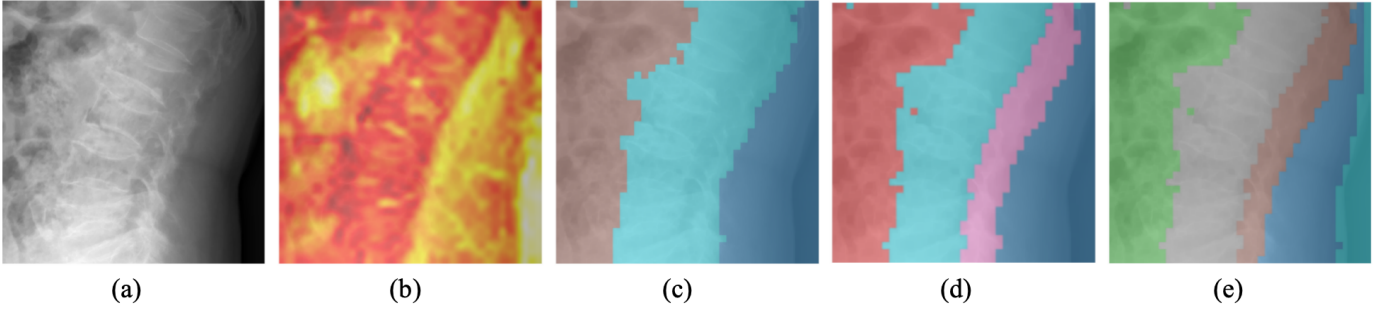
Fig. 1. Visualization of the unsupervised region-focusing process for a representative lumbar lateral X-ray. (a) Original X-ray image. (b) DINOv2 feature similarity map showing self-attention intensity across image regions. (c–e) K-means segmentation results obtained with different cluster numbers: $K = 3$, $K = 4$, and $K = 5$, respectively. When $K = 4$, the algorithm achieves the most distinct anatomical separation, effectively isolating vertebral structures (cyan) from soft tissue and background. This visualization confirms that the unsupervised feature clustering approach reliably identifies the spinal region without manual annotation.