Supplementary for "GLACIA: Instance–Aware Positional Reasoning for Glacial Lake Segmentation via Multimodal Large Language Model"

Lalit Maurya[1,*]   Saurabh Kaushik[2,*]   Beth Tellman[2]

[1]Portsmouth AI and Data Science Centre (PAIDS), School of Computing, University of Portsmouth, Portsmouth, PO1 3HE, UK

[2]Center for Sustainability and the Global Environment (SAGE), University of Wisconsin–Madison, Madison, WI, 53726 USA

lalit.maurya@port.ac.uk, skaushik8@wisc.edu, beth.tellman@wisc.edu

A. Appendix

A.1. Generation of GLake-Pos: Question-Answer Pairs

To construct the reasoning-enhanced supervision used in GLACIA, we generate spatially grounded natural-language explanations directly from the segmentation masks. For each image, connected-component analysis is applied to the binary glacier-lake mask to identify individual lakes and extract their geometric attributes, including bounding box, centroid, and distance from the image center. Based on these measurements, each lake is assigned a positional label—such as top left, bottom right, or center—together with a proximity descriptor (near or far from the center). These spatial descriptors are then inserted into multiple linguistic templates to produce diverse reasoning statements.

For example, in the image shown in Fig. A.1, the mask contains two distinct glacial lakes. The first lake is located in the top right, far from the center of the image, while the second lake appears in the bottom left, near the center. Using our template-based generation system, GLACIA produces a corresponding reasoning-aware annotation such as: "There are two glacial lakes: the 1st is in the top right, far from the center, and the 2nd is in the bottom left, near the center." Such structured reasoning annotations are stored alongside each image and question, forming a comprehensive dataset that teaches the model not only to recognize glacial lakes but also to articulate their spatial relationships. The process consists of three major stages: (i) extraction of glacial lake positions from binary masks, (ii) creation of structured descriptive sentences, and (iii) filling question–answer templates to produce natural-language annotations. The overall pipeline is described below.

(1) Determining the Spatial Position of Each Glacial Lake
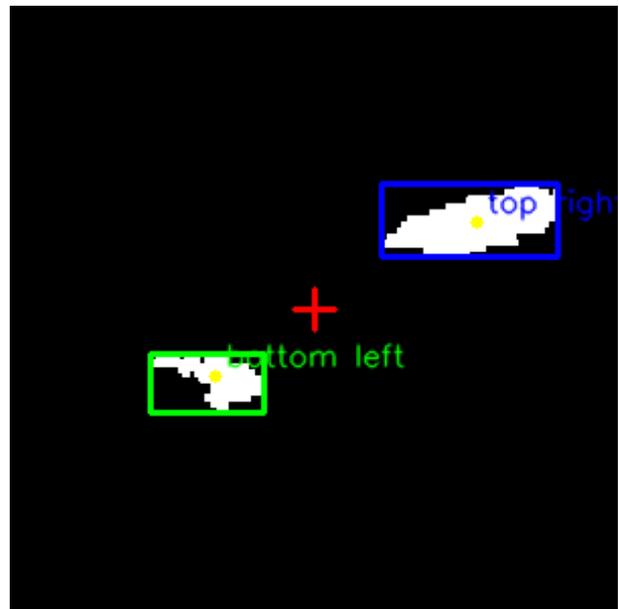
For every mask image, we per-formed connected-component analysis using cv2.connectedComponentsWithStats. Each connected component corresponds to a single glacial lake. For each detected lake, we computed:

- its bounding box,
- the centroid location (x_c, y_c),
- its relative position with respect to the image center,
- and proximity to the center ("near the center" or "far from the center").

Each lake is then assigned one of the following spatial categories: *top left*, *top right*, *bottom left*, *bottom right*, or *center*, with additional proximity information attached. For example, a resulting descriptor may be: *"bottom right, near*

Figure A.1. The position of Glacial Lake form the center

1

the center".

(2) Constructing Descriptive Answer Sentences

Let N denote the total number of detected glacial lakes. For each lake i, we generate an ordinal label (e.g., "1st", "2nd", "3rd") and pair it with its computed positional description.

Depending on N, three levels of answer construction are used:

- For $N = 1$, a single-sentence answer template is chosen from a set of 10 stylistic alternatives.
- For $N = 2$, a dual-lake template is used, explicitly describing the locations of the 1st and 2nd lakes.
- For $N \geq 3$, the answer consists of:
  1. a high-level count sentence (e.g., "There are 5 glacial lakes in the image."), and
  2. individual per-lake sentences generated randomly from a set of per-lake templates such as: "The {number_position} glacial lake lies in the {position_description}."

All per-lake sentences are concatenated to form the final answer string.

(3) Template-Based Question Generation

A large collection of question templates, written in JSON format, was used to generate natural-language questions. Each template requests the number of glacial lakes and their locations. Examples include:

- "In the provided remote sensing image, how many glacial lakes are visible and where are they located?"
- "Identify all glacial lakes in the image and give their respective positions."
- "Provide the glacial lake count and describe the location of each glacial lake."

For each image, one template is sampled at random. Placeholders such as {all_glacial_sentences} and {total_number} are dynamically replaced based on the detected lake count and generated descriptions.

(4) Final Q–A Output

Each image thus receives:

- a natural-language question asking for the number and location of the glacial lakes, and
- a systematically constructed answer describing each lake in ordinal order with precise spatial reasoning.

An example of final answer for three lakes might be:

> "There are 3 glacial lakes in the image. The 1st glacial lake lies in the top left, near the center. The 2nd glacial lake can be seen at bottom right, far from the center. The 3rd glacial lake appears in the top right, near the center."

A.2. Position-Based Question–Answer Templates.

The following templates are used for generating spatial reasoning questions and answers, where {position_description} is automatically derived from the mask-based positional analysis described above:

- **Q:** In the provided remote sensing image, where is the glacial lake located?
  A: The glacial lake can be observed in the {position_description} area of the image.
- **Q:** Specify the glacial lake location in this remote sensing view.
  A: The glacial lake is clearly seen in the {position_description} region.
- **Q:** Can you detect where the glacial lake is in this remote sensing image?
  A: The glacial lake is present in the {position_description} area.
- **Q:** Which area of the remote sensing image shows the glacial lake?
  A: The glacial lake is seen in the {position_description}.
- **Q:** From this remote sensing image, what is the glacial lake position?
  A: The glacial lake is located in the {position_description} part.
- **Q:** Could you point out where the glacial lake is located in this remote sensing scan?
  A: The glacial lake can be observed in the {position_description} area of the scan.
- **Q:** What part of the image contains the glacial lake in this remote sensing?
  A: In this remote sensing image, the glacial lake is found in the {position_description} section.
- **Q:** Identify the region in this remote sensing image that shows the glacial lake.
  A: The region showing the glacial lake is the {position_description}.
- **Q:** Can you specify the glacial lake location based on the remote sensing image provided?
  A: Based on the image, the glacial lake lies in the {position_description} region.
- **Q:** From this remote sensing image, where would you say the glacial lake is located?
  A: Judging from the image, the glacial lake is located at the {position_description}.
- **Q:** In the provided remote sensing image, can you locate the glacial lake?
  A: The glacial lake appears in the {position_description} area.
- **Q:** Which section of this remote sensing image contains the glacial lake?
  A: The glacial lake is present in the {position_description} section.
- **Q:** Where does the glacial lake appear in the remote sensing image?
  A: The glacial lake is situated in the {position_description} region.

- **Q:** Highlight the glacial lake location in this remote sensing scan.
  A: The glacial lake is located in the {position_description} part of the image.
- **Q:** Can the glacial lake be seen in the remote sensing image? If so, where?
  A: The glacial lake can be observed in the {position_description}.
- **Q:** In the remote sensing image provided, what is the glacial lake position?
  A: The glacial lake is found in the {position_description} region.
- **Q:** Could you identify the glacial lake in this remote sensing image?
  A: The glacial lake is located in the {position_description} area.
- **Q:** Where is the glacial lake visible in this remote sensing view?
  A: The glacial lake can be observed in the {position_description} region.
- **Q:** Which part of the remote sensing scan shows the glacial lake?
  A: The glacial lake is present in the {position_description} section.
- **Q:** Based on this remote sensing image, where is the glacial lake located?
  A: The glacial lake lies in the {position_description} area.
- **Q:** Identify where the glacial lake is in the remote sensing image.
  A: The glacial lake is found in the {position_description} section.
- **Q:** Can you determine the glacial lake location in this remote sensing image?
  A: The glacial lake is located in the {position_description} region.
- **Q:** Point out the glacial lake in the remote sensing image.
  A: The glacial lake can be observed in the {position_description} part.
- **Q:** In the provided remote sensing scan, which section shows the glacial lake?
  A: The glacial lake appears in the {position_description} area.
- **Q:** Where is the glacial lake situated in this remote sensing image?
  A: The glacial lake is seen in the {position_description} region.
- **Q:** Specify the area of this remote sensing image that contains the glacial lake.
  A: The glacial lake is located in the {position_description} section.
- **Q:** Which part of the remote sensing view shows the glacial lake?
  A: The glacial lake is present in the {position_description} area.
- **Q:** Can the glacial lake be identified in this remote sensing scan?
  A: Yes, the glacial lake is in the {position_description} region.
- **Q:** Highlight the glacial lake location in the provided remote sensing image.
  A: The glacial lake is situated in the {position_description} area.
- **Q:** In this remote sensing image, can you specify the glacial lake region?
  A: The glacial lake is found in the {position_description}.
- **Q:** Where does the glacial lake appear in the remote sensing scan?
  A: The glacial lake is observed in the {position_description} part.
- **Q:** Point out where the glacial lake is located in this remote sensing image.
  A: The glacial lake can be seen in the {position_description} section.
- **Q:** Which region of the remote sensing image shows the glacial lake?
  A: The glacial lake is visible in the {position_description} region.
- **Q:** Based on this remote sensing view, identify the glacial lake location.
  A: The glacial lake is in the {position_description} area.
- **Q:** Can you locate the glacial lake in the provided remote sensing scan?
  A: The glacial lake appears in the {position_description} section.

A.3. Instance-Aware Position Question–Answer Templates.

The Q&A generation for glacial lake segmentation is based on the number and positions of detected lakes in an image. We use different templates depending on whether there is a single lake, exactly two lakes, or more than two lakes.

Single Glacial Lake: If the image contains only one glacial lake, the question is paired with an answer template such as:

- **Question:** How many glacial lakes are present in this image? **Answer:** There is one glacial lake located in `position_description`.
- **Question:** Count the number of glacial lakes. **Answer:** A single glacial lake can be seen at `position_description`.
- **Question:** Identify the locations of glacial lakes. **Answer:** The only glacial lake in this image lies in the `position_description`.

Two Glacial Lakes: If the image contains exactly two glacial lakes, the templates include:

- **Question:** How many glacial lakes are present in this image? **Answer:** There are two glacial lakes: the 1st is in {position_description[0]}, and the 2nd is in {position_description[1]}.
- **Question:** Count and locate all glacial lakes. **Answer:** Two glacial lakes can be seen here: the 1st lies in {position_description[0]} and the 2nd is in {position_description[1]}.
- **Question:** How many glacial lakes and where are they positioned? **Answer:** There are exactly two glacial lakes: the first is at {position_description[0]} and the second at {position_description[1]}.

Multiple Glacial Lakes: If the image contains more than two glacial lakes, we use a combination of total count and per-lake sentences:

- **Question:** How many glacial lakes can be observed? **Answer:** There are {total_number} glacial lakes. {all_glacial_sentences}
- **Question:** Provide the count and locations of glacial lakes. **Answer:** glacial lake count in this scan is {total_number}. {all_glacial_sentences}
- **Question:** Describe the positions of glacial lakes in the image. **Answer:** Observed glacial lakes: {total_number}. {all_glacial_sentences}

A.3.1. Notes on Placeholders

- {position_description}: Position of a single glacial lake relative to the image center (e.g., *top left, near the center*).
- {position_description[0]}, {position_description[1]}: Positions of the first and second glacial lakes when exactly two are present.
- {total_number}: Total number of glacial lakes in the image.
- {all_glacial_sentences}: Dynamically generated per-lake description for images with more than two glacial lakes.