CHARTIST: A Model of Task-driven Eye Movement Control for Chart Reading

A EXAMPLE CASE

We demonstrate an example case (Fig. A1) to see the comparison among human scanpath and model predictions on a chart about the world's largest earthquakes [1]. Our approach and VQA approach are task-driven. UMSS and DeepGaze iii randomly sampled scanpaths three times.

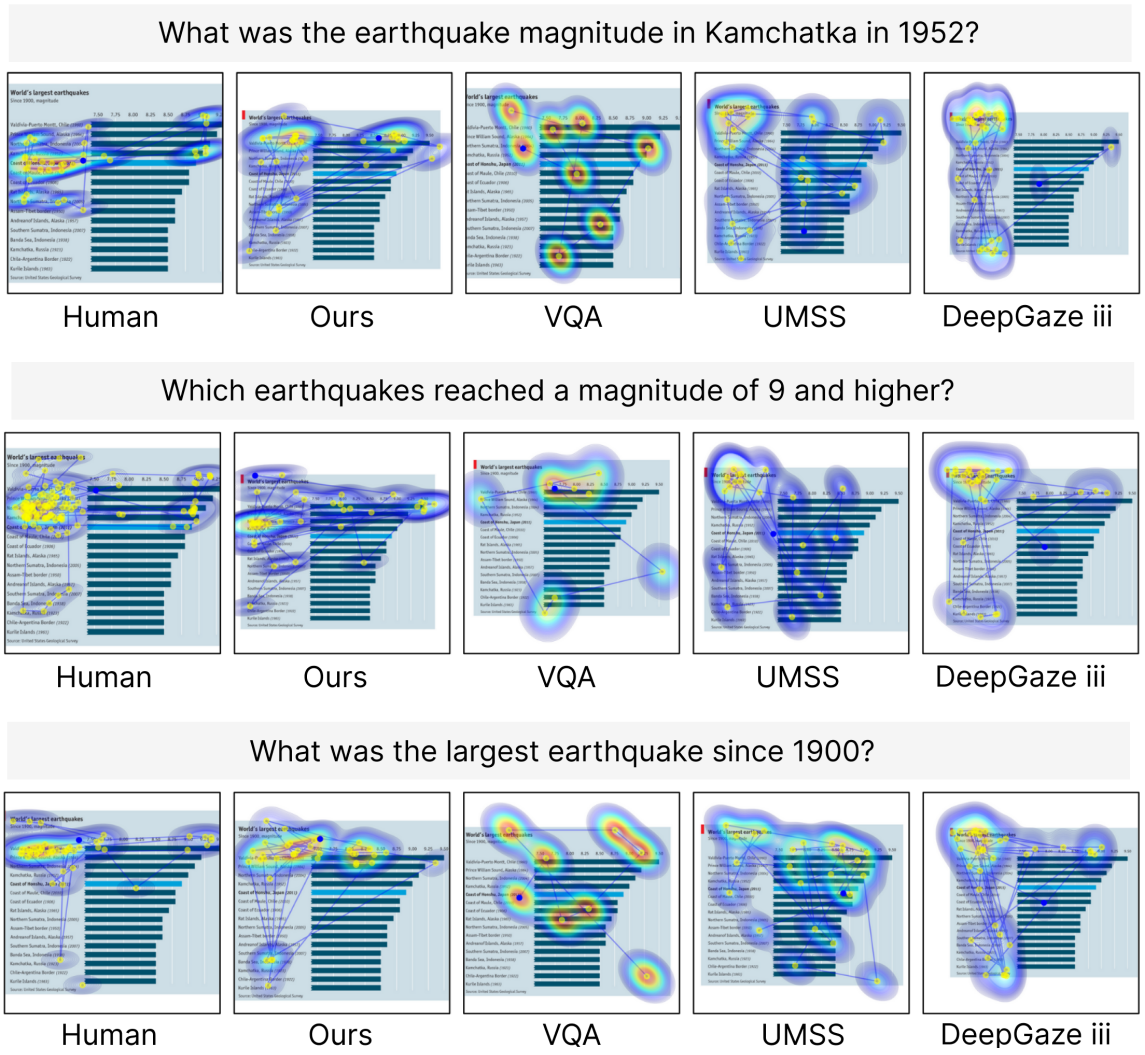

Fig. A1. A comparison of task-driven scanpaths across all model predictions and human data.

[1] https://www.economist.com/graphic-detail/2011/03/11/terrifying-tremors

B DETAILED PROMPTS

We list the detailed prompts for memory summary and LLM policy:

Memory Summary

```
[
    {
        "role": "system",
        "content": "You are looking at {chart}, and your task is {task}.
        Your goal is to read the chart and solve the task with as few actions
        as possible. You can move your gaze to gather information from the chart.
        After each step, you get information from the chart and it will also
        update your memory.
        The memory is a list of the information in patches you have gathered
        from the chart with position (x, y), action, and corresponding text.
        The chart image is divided into 20 x 20 grids.
        The position means where the gaze is from (0, 0) to (19, 19).
        From the chart image, it's from left-top to right-bottom.
        You can only solve the task by using the information in your memory."
    },
    {
        "role": "user",
        "content": "Your current memory in order is {memory}.
        Please summarize the memory into one paragraph."
    }
]
```

LLM Policy

```
[
    {
        "role": "system",
        "content": "You are looking at {chart}, and your task is {task}".
        Your goal is to read the chart and solve the task
        with as few actions as possible.
        You can move your gaze to gather information from the chart.
        After each step, you get information from the chart and it will also
        update your memory.
        The memory is a list of the information you have gathered from
        the chart with position (x, y), action, and corresponding text.
        The chart image is divided into 20 x 20 grids. The position means
        the patch where the gaze is from (0, 0) to (19, 19).
        From the chart image, it's from left-top to right-bottom.
        You can only solve the task by using the information in your memory.
```

```
        Initially, you have no idea about any information in the chart,
        so the memory will be an empty list
        There is a list of actions, you can use any of them in any order
        you want:
        search_text(task): visually searching for an item label or a value
        label related to the task. This is highly focused on the task itself;
        find_mark(ref_pos): visually searching for a graphic mark of the data
        point given a reference label.
        This approach is used to look at data points in the chart;
        read_axis(ref_pos): visually searching to read the associated value
        or text label of a given mark. This is used to read the value of a
        related bar;
        answer(task): stop gaze movement, and answer the task question.
        You can use these actions to solve the task. It is okay to reuse each
        of the actions.
        If the action is answer, it means you can solve the question with
        the information you have."
    }
    {
        "role": "user",
        "content": "Your current memory in order is {memory}. In summary,
        the information in your memory: {memory_summary}. Please decide
        your next action just based on your memory.
        Your goal is to read the chart and solve the task with as few
        actions as possible.
        Ensure the action is in a JSON format that can be directly parsed
        with the function name and its parameter.
        The parameter is a coordinate (x,y).
        If there is no parameter, you can give an empty input.
        Please also give a reason for your action.
        Please ensure the action has corresponding parameters."
    }
]
```