

LlaVa Fine-tuning

Date : 05/09/24

LlaVa - (Large Language and Vision Assistant) Fine-tuning LlaVa with our custom own dataset..

LlaVa - Quantized 7b model has been used for fine-tuning process. ([llava-hf/llava-1.5-7b-hf](#))

Dataset:

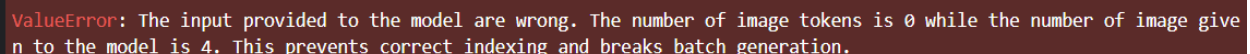
To fine-tune the LlaVa model, we annotated the [image](#) using the VIA (Visual Image Annotation) tool, resulting in a JSON-formatted dataset. For fine-tuning, the dataset must be structured in a specific format required by the LlaVa model. After converting our JSON file to the pre-defined structure for the LlaVa model, it will be in the format of [Llava_data](#).

Deepspeed Approach:

To fine-tune the LlaVa model, we researched different methods to optimize the fine-tuning process. We started with the Deepspeed approach to train the dataset, following the guidelines from a specific [blog](#). However, during this approach, we encountered a "module not found" error in the cloned repository of llava/train/train. Despite this issue, the Deepspeed script simplifies the fine-tuning process for the dataset.

Data Collator Approach:

We also explored a different approach using LlavaForConditionalGeneration to access the quantization model, following a [Colab](#) notebook to implement the LlaVa fine-tuning structure. During this process, we faced challenges with structuring the LlaVa Data Collator, which is needed to combine image and text pairs. We developed a logic to handle our specific dataset, but we still encountered issues making the model trainable. The problem occurred while training, as shown in the screenshot below.



```
ValueError: The input provided to the model are wrong. The number of image tokens is 0 while the number of image given to the model is 4. This prevents correct indexing and breaks batch generation.
```

To assist further, I am providing my Colab notebooks for reference on where we faced difficulties in fine-tuning our custom model.

Existing dataset Fine-tune:

To address the problems we encountered with our custom dataset, we decided to fine-tune the LlaVa model using an existing online dataset. However, during training, we encountered a "CUDA out of memory" error due to insufficient storage.

References:

→ [Colab Notebook-1](#)

→ [Concept of LLaVa](#)

Debugging Made through

→ [debug-DataCollator](#)

→ [debug-wrong model index](#)

Researched Different approaches (Yet to done):

→ [LLaVa Fine-tuning with Lightning PyTorch](#)