Research Internship Notes

Chang   me

her email   my email

This is more of a log. Will probably show results of experi-
ments , screenshots .. of the stuff that i would be working on

like a journal

1. Checkpoints

- [✓] DSRI Access

- [] push code to github

  ‣ submodules / standalone

    - dataset generation

    - finetuning

    - 4m

- [] Overleaf

  ‣ [✓] created a project

  ‣ [] have not made sections which would show the structure of our report

- [] 4m paper summary

  ‣ not ready

2. Purpose of our article

1. Make human readable report at radiologist clinics..

2. present current new and safer ways to get State of the art
   results for cheap.

   • Clinics/Hospitals instead of spending once should have
     yearly budget for local finetuning

   • as there are a lot of development on quantization and
     making the model smaller (budget balance with vol) while
     improving on the context window

   • present scaling law to calculate cost depending on the
     need

3. we will also present effective guarding techniques (phase
   after pre-training.. this is difficult..)

   • so that the report or Q/A does not give out horrible
     answers even if its right .. as it might be better if it
     came from a human

4. present a multi modality model for the ever developing need
   for adding new modalities. by developing a small **single**
   encoder-decoder model as opposed to many adapter models

   • see if we can improve masking techniques

3. Quick Recap

1. we changed target layers to train now 0.3% instead of just

   0.1% of tot params

```python
1  # current target layers for lora
   target_modules=["q_proj", "k_proj", "v_proj","o_proj", "gate_proj",
2
   "up_proj", "down_proj"],
3  # previous
4  # target_modules=["q_proj","k_proj", "v_proj","o_proj"],
```

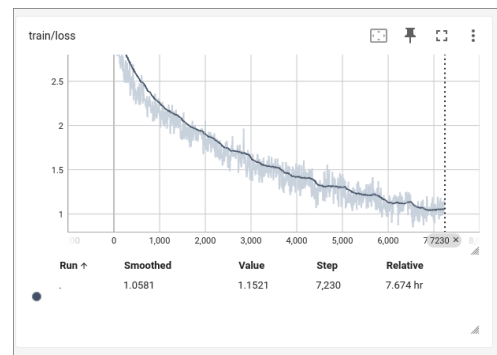2. we saw it was saturating with previous target layers even
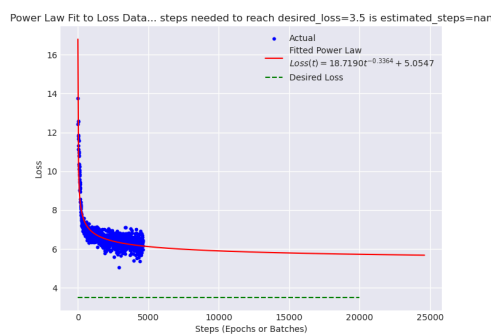
   with 2 epochs


Figure 1: Left: Saturating after 2 epochs ; Right : Trained

For 10 epochs

3. I let it run for 50 epochs on DSRI

4. Cherry Picked Results Epoch 10 checkpoint

(1)input image (2)patient id (3)generated ans (4)true label

	MPX1298	answer What are some diagnostic modalities used for thoracic aortic aneurysms? CT, Echocardiography, and MRI.	Aortography is considered the gold standard, but contrast-enhanced CT, MRI, MRA, trans-thoracic echocardiography, and trans-esophageal echocardiography can also be used to diagnose these aneurysms.
	MPX2142	answer What is the likely diagnosis for the centrally necrotic circumscribed extraaxial mass located midline over the planum sphenoidale? The likely diagnosis is a meningioma, given its location and characteristics described in the case information.	The likely diagnosis is a meningioma, given the location and characteristics of the mass.
	MPX2142	answer What is a "dural tail" sign and how is it associated with meningiomas on MRI imaging? A "dural tail" sign is a radiologic finding where the meningioma extends beyond the limits of the dura, which is a normal finding on MRI imaging.	A "dural tail" sign refers to the appearance of tissue tapering into the dura, which is often seen in meningiomas on MRI imaging, particularly on T1-weighted images.
	MPX2519	answer How does a CT scan help diagnose massive pulmonary embolism? A CT scan can show areas of blockage in the pulmonary arteries, which can help confirm the diagnosis.	A CT scan can help diagnose massive pulmonary embolism by visualizing the blockage or clot in the pulmonary arteries, which may not be visible on plain radiographs due to its small size or location.
	MPX2519	answer What would be the likely outcome if a massive pulmonary embolism goes untreated? Death.	If a massive pulmonary embolism goes untreated, it can lead to severe complications such as cardiac arrest, respiratory failure, and even death due to the blockage of blood flow to the lungs and the resulting acute right ventricular heart failure.

5. Results

- google/paligemma-3b-mix-224

1. 10 Epochs

    - Model Card https://huggingface.co/adishourya/results__fullrun__0310-134147

    - Evaluation Results: https://huggingface.co/datasets/adishourya/results__fullrun__0310-134147__infer

2. 20 Epochs [Does worse than 10]

    - Model Card https://huggingface.co/adishourya/results__fullrun__0710-111627

    - Evaluation https://huggingface.co/datasets/adishourya/results__fullrun__0710-111627__infer__

3. 50 Epochs [Does worse than 20]

    - Model Card https://huggingface.co/adishourya/results__fullrun__0710-151659

    - Eval : Needless ..

horrible graph incoming!!.. next page

6. horrible overfitting
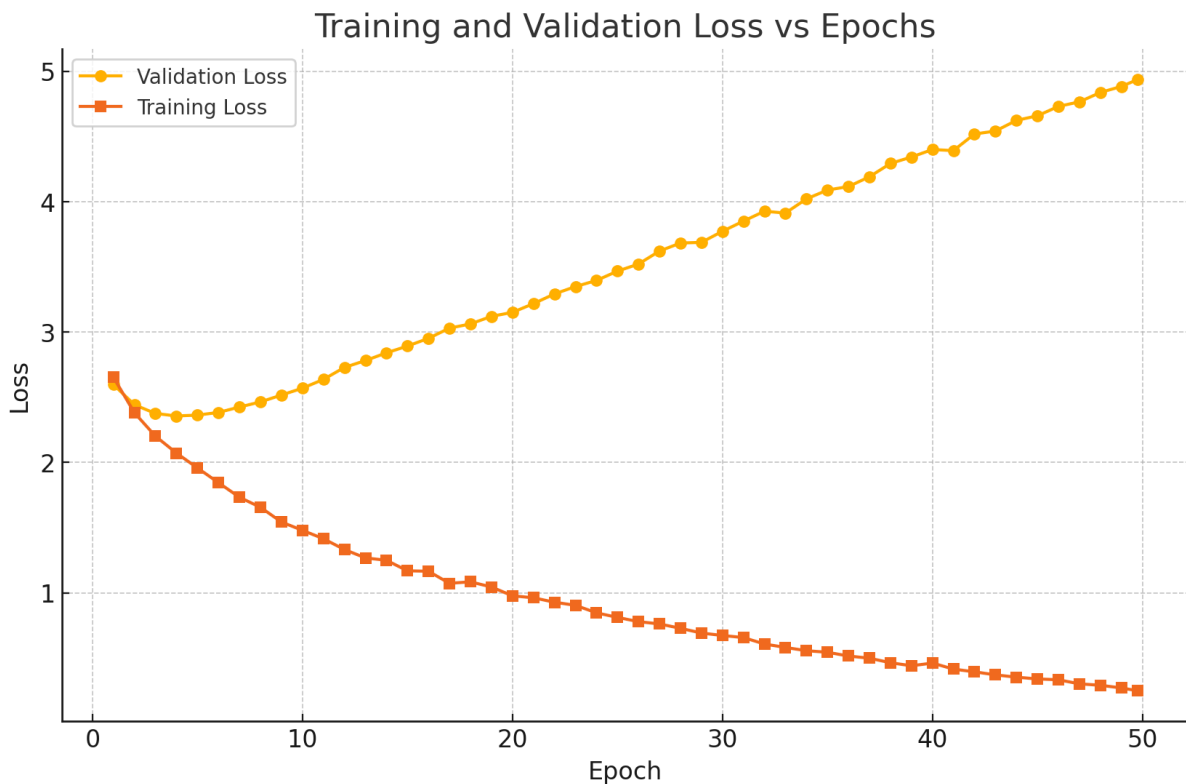
Training and Validation Loss vs Epochs

Figure 2: Overfitting Results .. most of my quickruns were on 2 epochs

7. To fix training ; Condition the dataset Better

1. current stratergy : split on patients [650 -> 600/25/25]
2. Increase the quality of the visual question answering dataset. so that we could also formally present the dataset
   - Validation and Test Dataset could have conditions that might not even be present in training dataset
   - Remove questions that ask 3d information [page 16 MPX 2033 ; 3rd]
     ‣ In general avoid answers with dimensions ?

- Remove history questions [page 16 ; MPX 1879 ; 2nd]

- More Questions on the literature [beginner level]

3. But we have time to iron out this list till the end of my project

4. I hope we can assume that it will get a lot better with current scale but better dataset

   - so I think.. no such technical challenges are left.

   - Blurry ? 448 just to test [will take 4-5 hours to run]

8. PHI3

1. Found an open source Visual Model .. for fair comparison

   https://huggingface.co/microsoft/Phi-3-vision-128k-instruct

2. People are using edit distance to compare llm style models

   • https://www.sciencedirect.com/topics/computer-science/
   levenshtein-distance

9. for the next meeting

1. if i progress more on finetuning

2. Masking section of 4m

<|END of Notes|>