

BU AI Tour Assistant



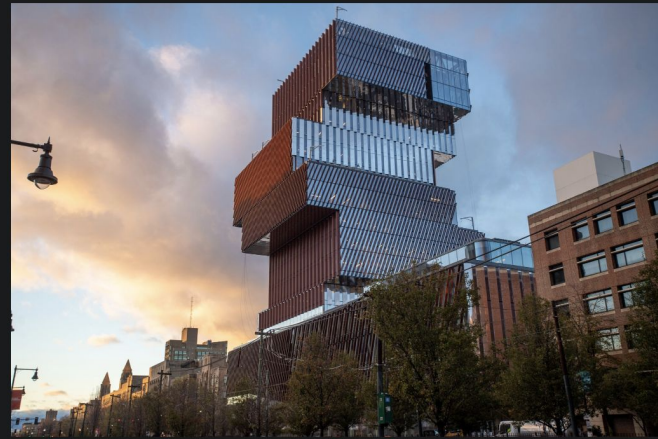
create an animed volcano and lava coming out



Motivation

Unhappy with chatgpt4 vision output on BU

Open-source



This is another building at BU, which is it?



ChatGPT

The building in the image is the Rajen Kilachand Center for Integrated Life Sciences & Engineering at Boston University. It's a state-of-the-art research facility that supports interdisciplinary research in life sciences and engineering. ↓



Tech stack + budget

Tensorflow

Huggingface

Runpod(GPU rented: A6000, A100)

Streamlit

Ollama

Colab pro(Local Gradio: A100)

Dataset

- 120 total images of CDS, CAS, and GSU (and some other locations)
 - Roughly 40 per class
 - Self-photographed, taken from Google Images/Streetview
- Annotations for images
 - E.g., “Provide a brief description of the given image”
 - "The image displays the College of Arts and Sciences (CAS) building with the Tsai Performance Center, part of Boston University's East Campus..."

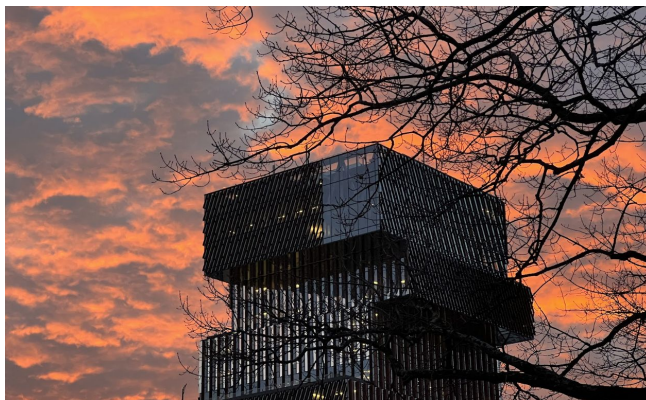


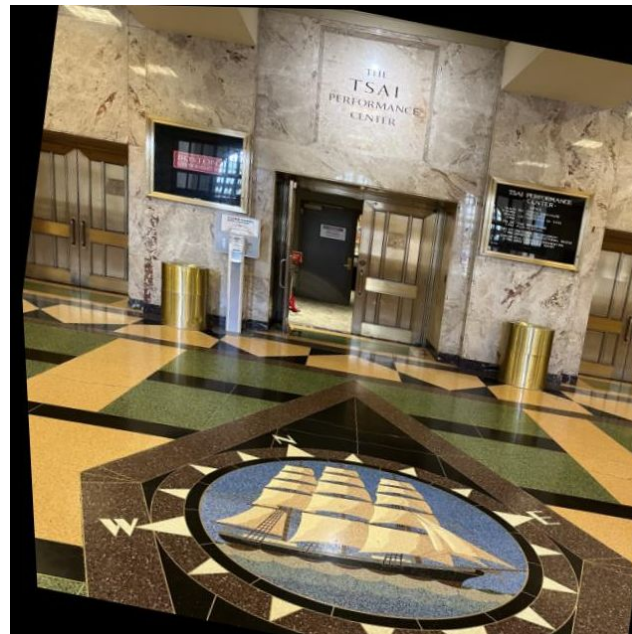


Image Classification

CNN, Transfer Learning, Fine Tuned Vision Transformer

Preprocessing - Roboflow

- Auto-Orient
 - Discards metadata rotations and standardize pixel ordering
- Resize to 640x640
- Augmentation
 - Crop
 - Rotation
 - Shear
- 120 → 281 images
 - 246 training, 23 validation, 12 test (88%, 8%, 4%)



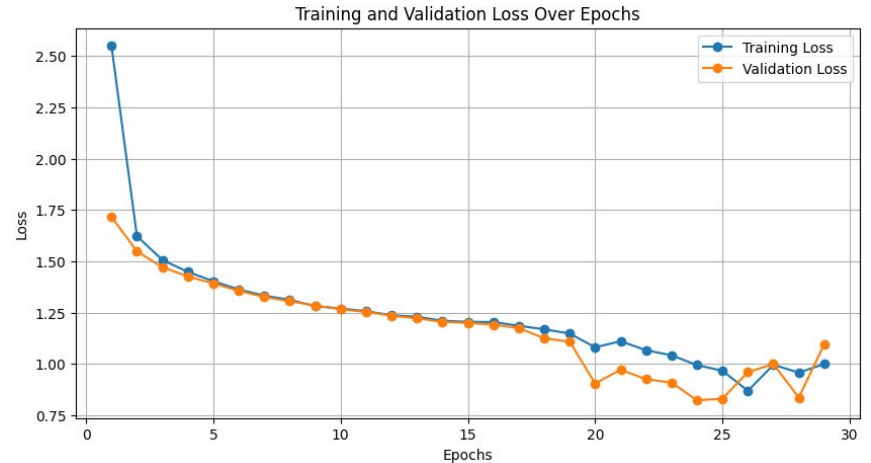
CNNs

1. Baseline model

- Two Conv2d + MaxPooling, 1 Dropout, 1 Dense layer
- Val_loss: 0.9431, val_accuracy: 0.6522

2. More conv layers, dropout, regularization, early stopping

- Four conv2d + MaxPooling, 3 Dropout, 2 Dense
- Val_loss: 1.0968, val_accuracy: 0.6087

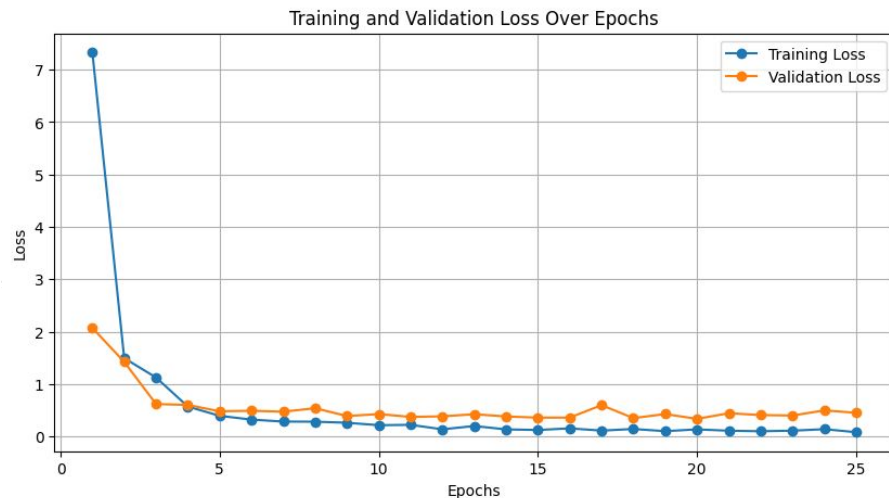


Transfer Learning

Base model: VGG16 w/ imagenet weights

Freeze layers in base model, add custom fully connected dropout layers

Val_loss: 0.4512, val_accuracy: 0.8696



Pred: CASJPG
True: CASJPG



Pred: CASJPG
True: CASJPG



Pred: CASJPG
True: CASJPG



Pred: CASJPG
True: CASJPG



Pred: CASJPG
True: CASJPG



Pred: CASJPG
True: CASJPG



Pred: CDSJPG
True: CDSJPG



Pred: CDSJPG
True: CDSJPG



Pred: CDSJPG
True: CDSJPG



Pred: CASJPG
True: CDSJPG



Pred: CASJPG
True: GSUJPG



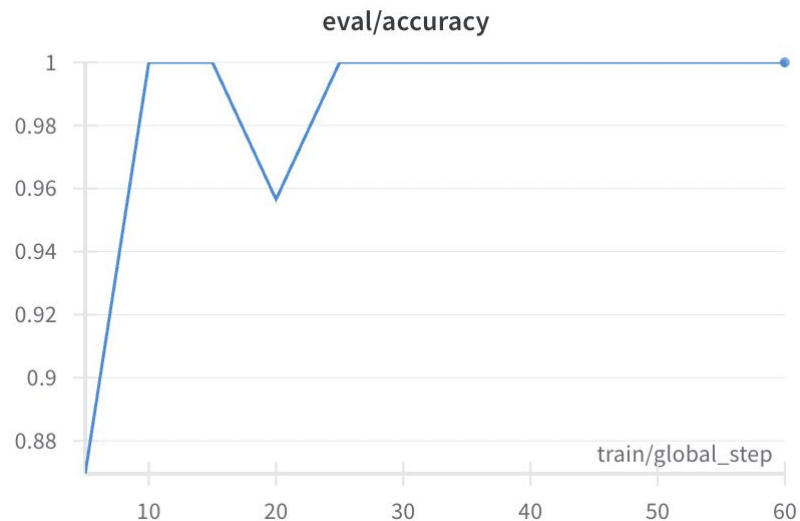
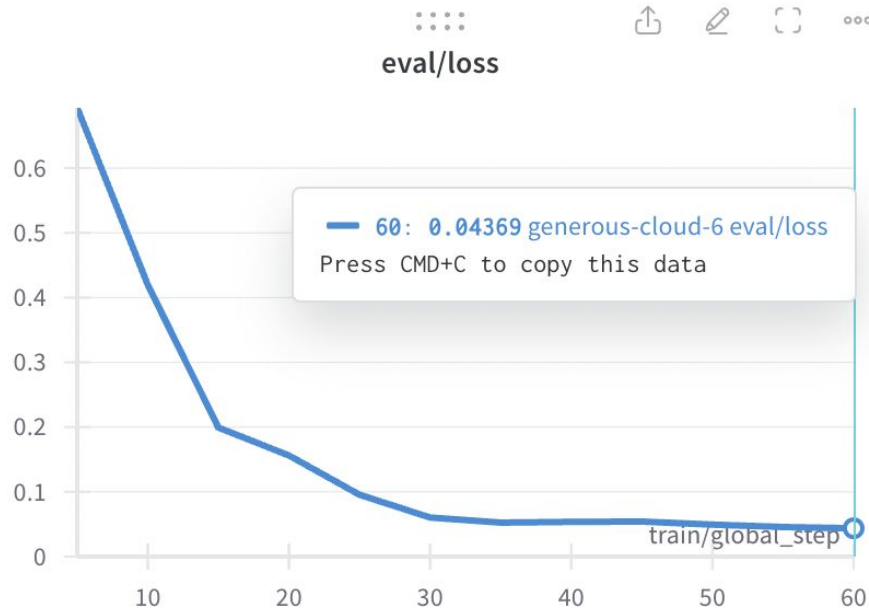
Pred: GSUJPG
True: GSUJPG



Vision Transformer Fine Tune

Model: google/vit-base-patch16-224-in21k,

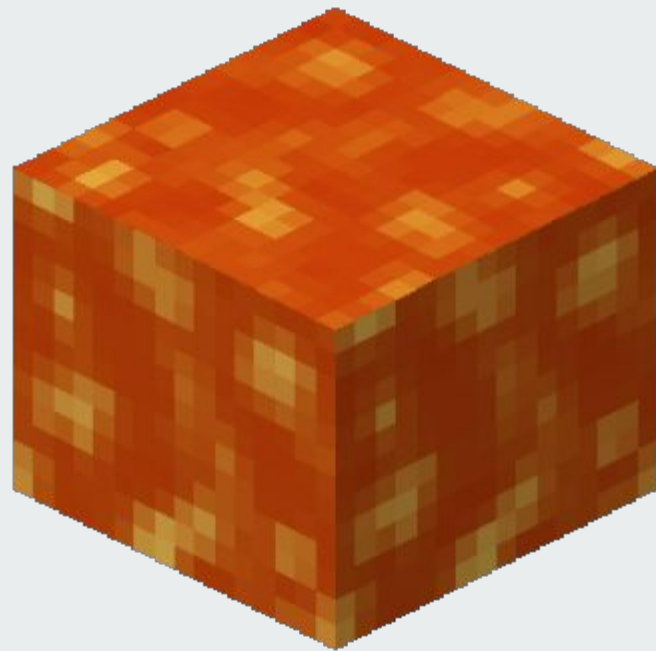
- Pre-trained on ImageNet-21k dataset





Llava

Out of the box Llava, Fine Tuned Llava





OOB Llava

Downloaded from Ollama -> streamlit vision app

- 4.7 GB
- 7 billion parameters

Does not perform very well 😞



Uploaded Image.

✓ Processing image file. DON'T LEAVE THIS PAGE WHILE IMAGE FILE IS BEING ANALYZED...

Analyzing Image File...

This image appears to show the "Fortified Campus" building at Boston University (BU), which is a student center and residence hall. It's not clear from this image alone, but it could be one of several buildings on the Fortified Campus that share a similar design aesthetic.

Done analyzing image file

LLava Image Analyzer

Select an image file to analyze.

Choose image file



Drag and drop file here

Limit 200MB per file • PNG, JPG, JPEG

Browse files



cas.jpeg 310.5KB



Uploaded Image.

✓ Processing image file. DON'T LEAVE THIS PAGE WHILE IMAGE FILE IS BEING ANALYZED...

Analyzing Image File...

The building in the image appears to be the BU Dining Center, which is a part of the Boston University campus.

Done analyzing image file



TinyLLaVA: A Framework of Small-scale Large Multimodal Models

Baichuan Zhou¹ Ying Hu² Xi Weng¹ Junlong Jia¹ Jie Luo¹ Xien Liu² Ji Wu² Lei Huang^{1*}

¹SKLCCSE, Institute of Artificial Intelligence, Beihang University, Beijing, China

²Department of Electronic Engineering, Tsinghua University, China

Updated: 2024-03-10

High level: Smaller parameters + phi-2 LLM

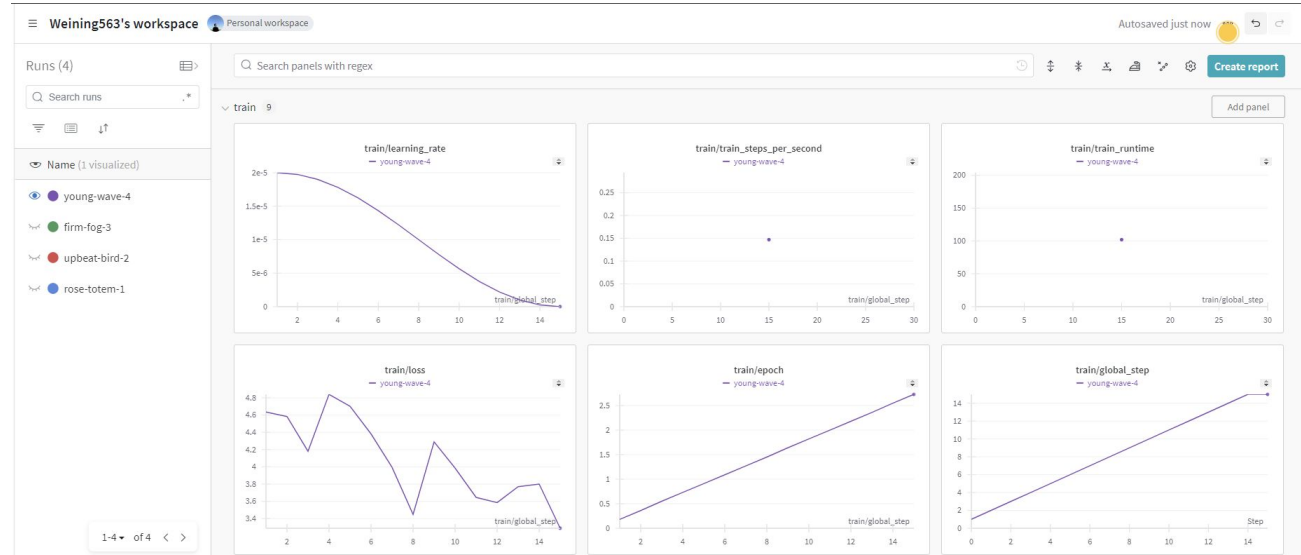
Demo



Weight and biases

Limitation

- Better quality data
- More images





Thank You