Multi-Modal Electronics Recommendation System

Dr. Ashirbad Pradhan

Experience

2012-16	**Undergraduate** Biomedical Engineering	....	Design a Biosignal Amplifier
2016-18	**Masters** Biomedical Engineering	....	Prosthetic Control Algorithms
2018-19	**Biomedical Scientist** Human Performance Lab	....	Robotic Hand Development, AI–ML
2020-24	**PhD** Systems Design Engineering	....	EMG–Biometrics
2023	**PhD Research Intern - Huawei** Human Machine Interaction Lab	....	Wearable sensors
2024-Now	**ML Research Scientist** Eaigle	....	Computer Vision, AI–ML

Table of contents

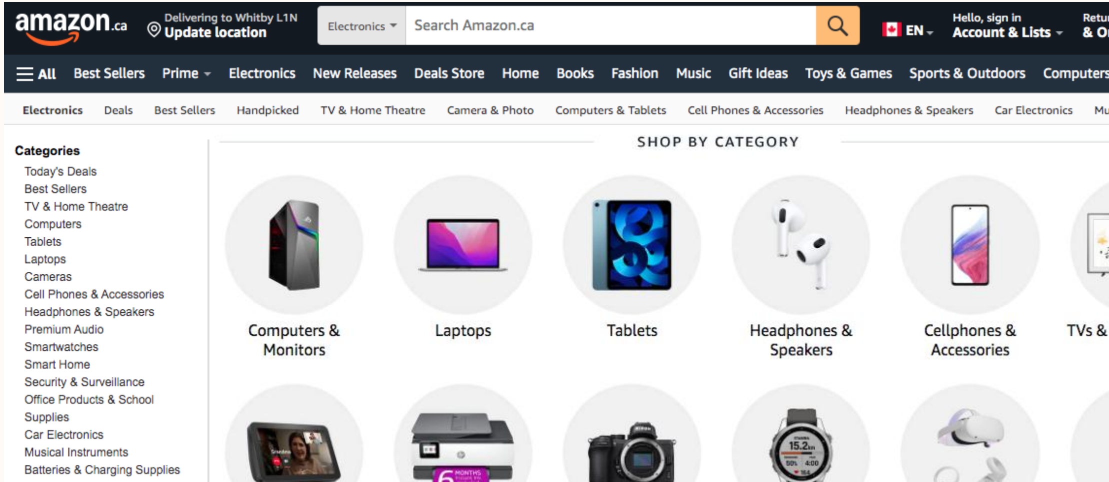
Introduction

- AI-Powered Recommendations – Multi-modal (text & image) approach
- **Use Case:** Amazon electronic products 2023 (10k, Kaggle)
- **LLMs:** CLIP (OpenAI, 2021) + Sentence-BERT (2019)

Combining computer vision and natural language processing to build a smarter product recommender, specifically showcasing it on a large data set of Amazon electronics

Problem Statement

- **Choice Overload:** Many products, hard to find the right one
- **Current Pain Point:** Users struggle to find similar electronics
- **Need:** Recommend products by content (image & text)
- **Further Need:** Can the customized chatbot learn our specific needs?



RAG and Fine-Tuning

- **RAG (Retrieval-Augmented Generation):** Combination of retrieval of relevant documents and images and language generation to produce more accurate and context-aware responses.

- However, what if the chatbot needs to find very specific details. For example, the different classes of cameras, or, finding defects on products

- Need more specialized retrieval of information based on image/text query

- **Fine-tuning:** Customizes a pre-trained model on domain-specific data to improve performance on targeted tasks.

Project Objective

Prototype a multi-modal recommender for Amazon electronic products

- **Input:** User query (text or image) → Output: similar electronics
- **Showcase:** Use AI embeddings to match products by content and generate the response (RAG)
- **Analysis:** Improve the performance of the chatbot using Fine-tuning

Project Breakdown

Prototype a multi-modal recommender for Amazon electronic products

- **Part1:** Building the RAG Chatbot
- **Part 2:** CLIP model performance analysis
- **Part 3:** CLIP model fine-tuning

Part 1: Building a RAG Chatbot

- Identifying an Image Text dataset *(Kaggle Amazon Electronics 2023 - 10k images-names)*
- Building an Image Vector database *(Chroma DB)*
- Building Text database *(Chroma DB)*
- Knowledge Retrieval, given a text query, get top 5 image and text matches, given an image query get top 5 image matches *(Langchain)*
- RAG Chatbot. Given the top matches provide a response

Amazon Electronics 2023

The dataset consists of the following columns:

1. **name:** The name of the electronic product.
2. **image:** The link to a high-quality image representing the product.
3. **link:** The reference link to the product on the Amazon website.
4. **ratings:** The average ratings given by Amazon customers for the product.
5. **no_of_ratings:** The number of ratings received by the product on Amazon.
6. **discount_price:** The discounted price of the product.
7. **actual_price:** The Manufacturer's Recommended Price (MRP) or original price of the product.

Amazon Electronics 2023

#	name	main_category	sub_category	image	link	ratings	no_of_ratings	discount_price	actual_price
	The name of the electronic product.	The main category to which the product belongs.	The sub-category specifying a more detailed classification of the product.	The link to a high-quality image representing the product.	The reference link to the product on the Amazon website.	Ratings of the Product [1-4]	Count of ratings	The discounted price of the product.	The Manufacturer's Recommended Price (MRP) or original price of the product.
0 9599	8800 unique values	1 unique value	1 unique value	8992 unique values	9600 unique values	4.3 14% 4.2 13% Other (7077) 74%	3456 unique values	[null] 5% ₹299 5% Other (8635) 90%	₹999 16% ₹1,999 5% Other (7554) 79%
0	Redmi 10 Power (Power Black, 8GB RAM, 128GB Storage)	tv, audio & cameras	All Electronics	https://m.media-amazon.com/images/I/81eM15lVcJL._AC_UL800_.jpg	https://www.amazon.in/Redmi-Power-Black-128GB-Storage/dp/B09Y64H8VS/ref=sr_1_4?qid=1679133649&s=elec...	4.0	965	₹10,999	₹18,999
1	OnePlus Nord CE 2 Lite 5G (Blue Tide, 6GB RAM, 128GB Storage)	tv, audio & cameras	All Electronics	https://m.media-amazon.com/images/I/71AvQd3VzqL._AC_UL800_.jpg	https://www.amazon.in/OnePlus-Nord-Lite-128GB-Storage/dp/B09WQYFLRX/ref=sr_1_5?qid=1679133649&s=elec...	4.3	113,956	₹18,999	₹19,999
2	OnePlus Bullets Z2 Bluetooth Wireless in Ear Earphones with Mic, Bombastic Bass - 12.4 Mm Drivers, 1...	tv, audio & cameras	All Electronics	https://m.media-amazon.com/images/I/51UhwaQXCpL._AC_UL800_.jpg	https://www.amazon.in/Oneplus-Bluetooth-Wireless-Earphones-Bombastic/dp/B09TVVGXWS/ref=sr_1_6?qid=16...	4.2	90,304	₹1,999	₹2,299

Knowledge Retrieval

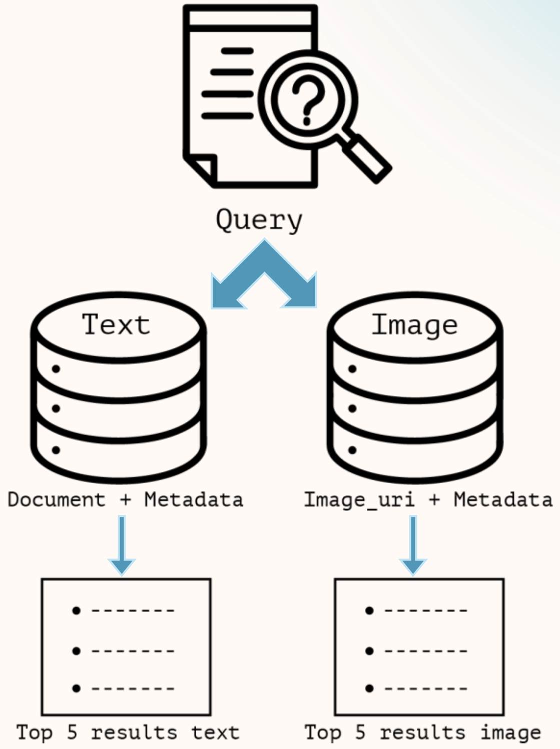
- Two *ChromaDB* databases are created: text_db and image_db

  Document = product name + rating + discount + price

  Image_uri = uri of the downloaded image

  Metadata = product_id, name, uri

- Given a query, get top 5 results from each database, based on distance



Prompt Engineering

System Template: Includes instructions on prioritization

- If brand name, discount, rating or price is mentioned in the query then prioritize the text query results
- If physical appearance (shape, color, size) is mentioned in the query then prioritize the Image query results
- If multiple descriptions are present then follow the sequence: Brand name > Physical appearance > Discount > Rating

```
# Create a custom prompt template
template = """
You are a helpful shopping assistant. Use the following product information to answer the question, while answering the question use metadata
If the query includes anything about the appearance (color size shape etc), then any other information such as rating, discount and price des
If the query is about the product/brand name,rating and discount, prioritize the text query results. Remeber the query results are sorted by
Give the answer in statements, not bullet points. make sure to include the product_id in the answer.
Question: {query}

Answer:
"""
```

Building QA chain

```
prompt = ChatPromptTemplate.from_messages([
    ("system", template),
    ("user",
     [
        {
            "type": "text",
            "text": "{texts}"
        },
        {
            "type": "image_url",
            "image_url": {'url': "data:image/jpeg;base64,{image_data_1}"}
        },
        {
            "type": "image_url",
            "image_url": {'url': "data:image/jpeg;base64,{image_data_2}"}
        },
        {
```

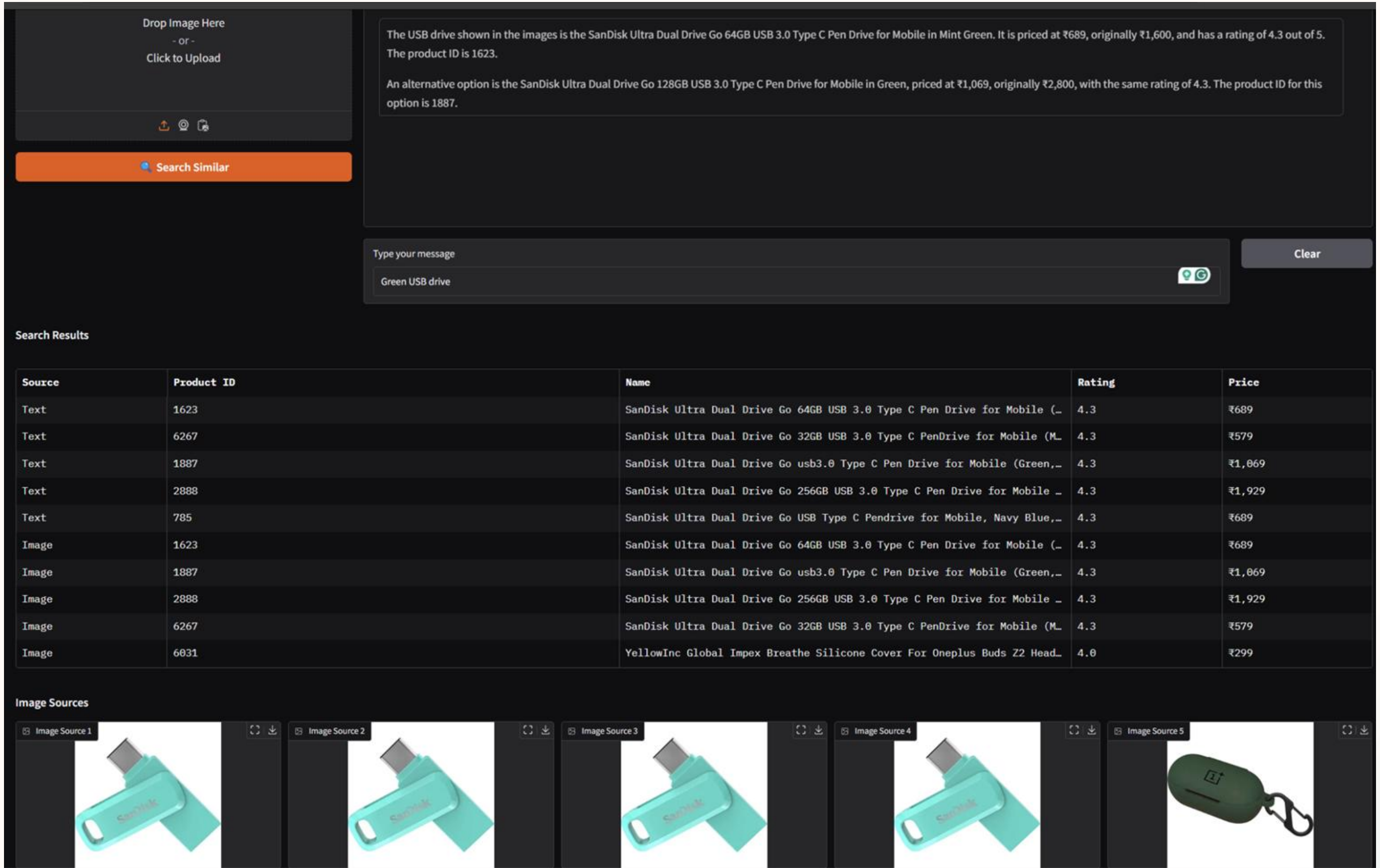
```
llm = ChatOpenAI(temperature=0.3, model="gpt-4o")
```

```
parser = StrOutputParser()
```

```
qa_chain = prompt | llm | parser
```

```
response= qa_chain.invoke(query)
```

UI Design


Drop Image Here
- or -
Click to Upload
Search Similar
The USB drive shown in the images is the SanDisk Ultra Dual Drive Go 64GB USB 3.0 Type C Pen Drive for Mobile in Mint Green. It is priced at ₹689, originally ₹1,600, and has a rating of 4.3 out of 5. The product ID is 1623.
An alternative option is the SanDisk Ultra Dual Drive Go 128GB USB 3.0 Type C Pen Drive for Mobile in Green, priced at ₹1,069, originally ₹2,800, with the same rating of 4.3. The product ID for this option is 1887.
Type your message
Green USB drive
Clear
Search Results

Source	Product ID	Name	Rating	Price
Text	1623	SanDisk Ultra Dual Drive Go 64GB USB 3.0 Type C Pen Drive for Mobile (…	4.3	₹689
Text	6267	SanDisk Ultra Dual Drive Go 32GB USB 3.0 Type C PenDrive for Mobile (M…	4.3	₹579
Text	1887	SanDisk Ultra Dual Drive Go usb3.0 Type C Pen Drive for Mobile (Green,…	4.3	₹1,069
Text	2888	SanDisk Ultra Dual Drive Go 256GB USB 3.0 Type C Pen Drive for Mobile …	4.3	₹1,929
Text	785	SanDisk Ultra Dual Drive Go USB Type C Pendrive for Mobile, Navy Blue,…	4.3	₹689
Image	1623	SanDisk Ultra Dual Drive Go 64GB USB 3.0 Type C Pen Drive for Mobile (…	4.3	₹689
Image	1887	SanDisk Ultra Dual Drive Go usb3.0 Type C Pen Drive for Mobile (Green,…	4.3	₹1,069
Image	2888	SanDisk Ultra Dual Drive Go 256GB USB 3.0 Type C Pen Drive for Mobile …	4.3	₹1,929
Image	6267	SanDisk Ultra Dual Drive Go 32GB USB 3.0 Type C PenDrive for Mobile (M…	4.3	₹579
Image	6031	YellowInc Global Impex Breathe Silicone Cover For Oneplus Buds Z2 Head…	4.0	₹299

Image Sources
Image Source 1
Image Source 2
Image Source 3
Image Source 4
Image Source 5

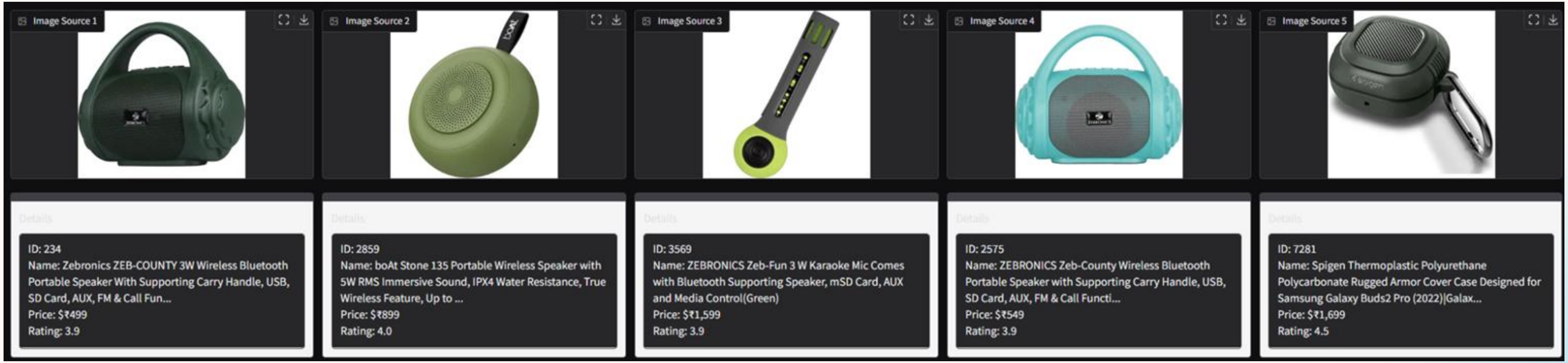
Text Query

Chat History

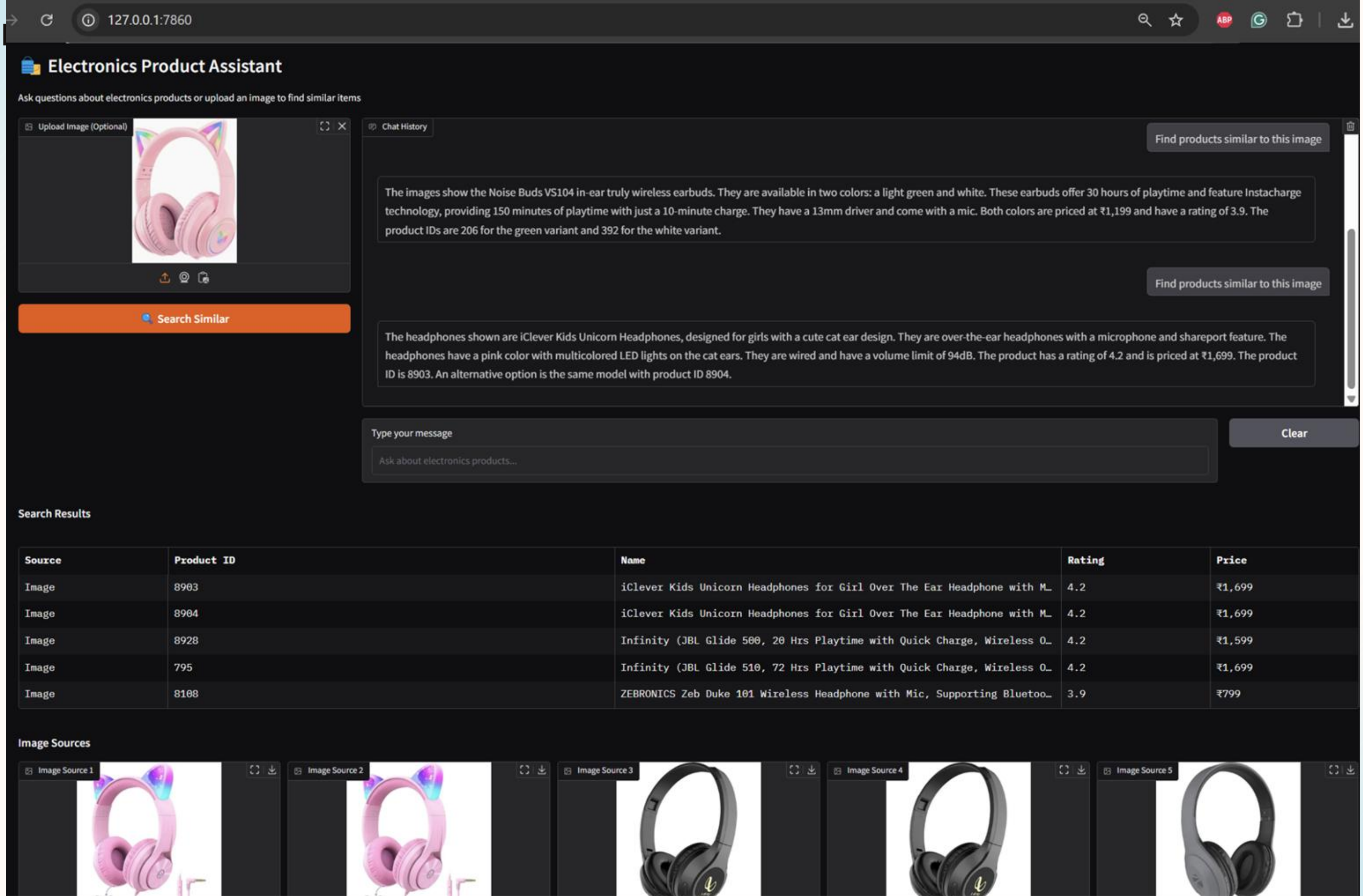
I am looking for a green speaker with a handle, with a high rating

The green speaker with a handle is the Zebronics ZEB-COUNTY 3W Wireless Bluetooth Portable Speaker. It features a carry handle and supports USB, SD Card, AUX, FM, and call functions. It has a rating of 3.9 and is priced at ₹499. The product ID is 234.

Alternatively, the boAt Stone 135 Portable Wireless Speaker is another green option, though it has a strap instead of a handle. It offers 5W RMS immersive sound, IPX4 water resistance, and a rating of 4.0. It is priced at ₹899, with the product ID 2859.



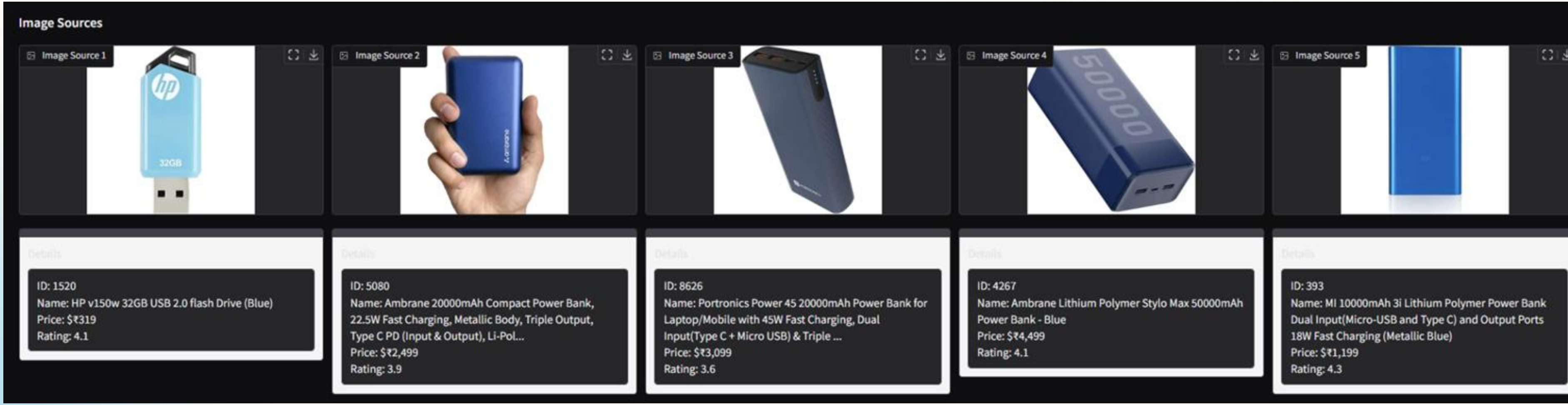
Part 1: I



127.0.0.1:7860
Electronics Product Assistant
Ask questions about electronics products or upload an image to find similar items
Upload Image (Optional)
Search Similar
Chat History
Find products similar to this image
The images show the Noise Buds VS104 in-ear truly wireless earbuds. They are available in two colors: a light green and white. These earbuds offer 30 hours of playtime and feature Instacharge technology, providing 150 minutes of playtime with just a 10-minute charge. They have a 13mm driver and come with a mic. Both colors are priced at ₹1,199 and have a rating of 3.9. The product IDs are 206 for the green variant and 392 for the white variant.
Find products similar to this image
The headphones shown are iClever Kids Unicorn Headphones, designed for girls with a cute cat ear design. They are over-the-ear headphones with a microphone and shareport feature. The headphones have a pink color with multicolored LED lights on the cat ears. They are wired and have a volume limit of 94dB. The product has a rating of 4.2 and is priced at ₹1,699. The product ID is 8903. An alternative option is the same model with product ID 8904.
Type your message
Ask about electronics products...
Clear
Search Results
Source | Product ID | Name | Rating | Price
Image | 8903 | iClever Kids Unicorn Headphones for Girl Over The Ear Headphone with M… | 4.2 | ₹1,699
Image | 8904 | iClever Kids Unicorn Headphones for Girl Over The Ear Headphone with M… | 4.2 | ₹1,699
Image | 8928 | Infinity (JBL Glide 500, 20 Hrs Playtime with Quick Charge, Wireless O… | 4.2 | ₹1,599
Image | 795 | Infinity (JBL Glide 510, 72 Hrs Playtime with Quick Charge, Wireless O… | 4.2 | ₹1,699
Image | 8108 | ZEBRONICS Zeb Duke 101 Wireless Headphone with Mic, Supporting Bluetoo… | 3.9 | ₹799
Image Sources
Image Source 1
Image Source 2
Image Source 3
Image Source 4
Image Source 5

Observations

- The results are accurate in terms of text sources
- The image-query-from-text is correct, however, image-query-from-image has some mistakes
- Image-text dataset has no distinct relationship



Part 2: CLIP Performance analysis

- **Pseudo labels** using OpenAI GPT-4o with the product name as the input
- Class_labels = {"camera", "headphones", "home appliance", "keyboard", "laptop", "monitor", "mouse", "phone case", "printer", "smartphone", "speaker", "tablet", "television", "tripod", "usb stick", "smartwatch", "earbuds", "phone charger", "power bank", "microSD card", "hard drive", "router", "wifi adapter", "calculator", "laptop charger", "stationery", "general accessory",, "computer accessory", "phone accessory"}
- Select class labels for CLIP classification analysis (~25)

Pseudo-Labels Class Distribution

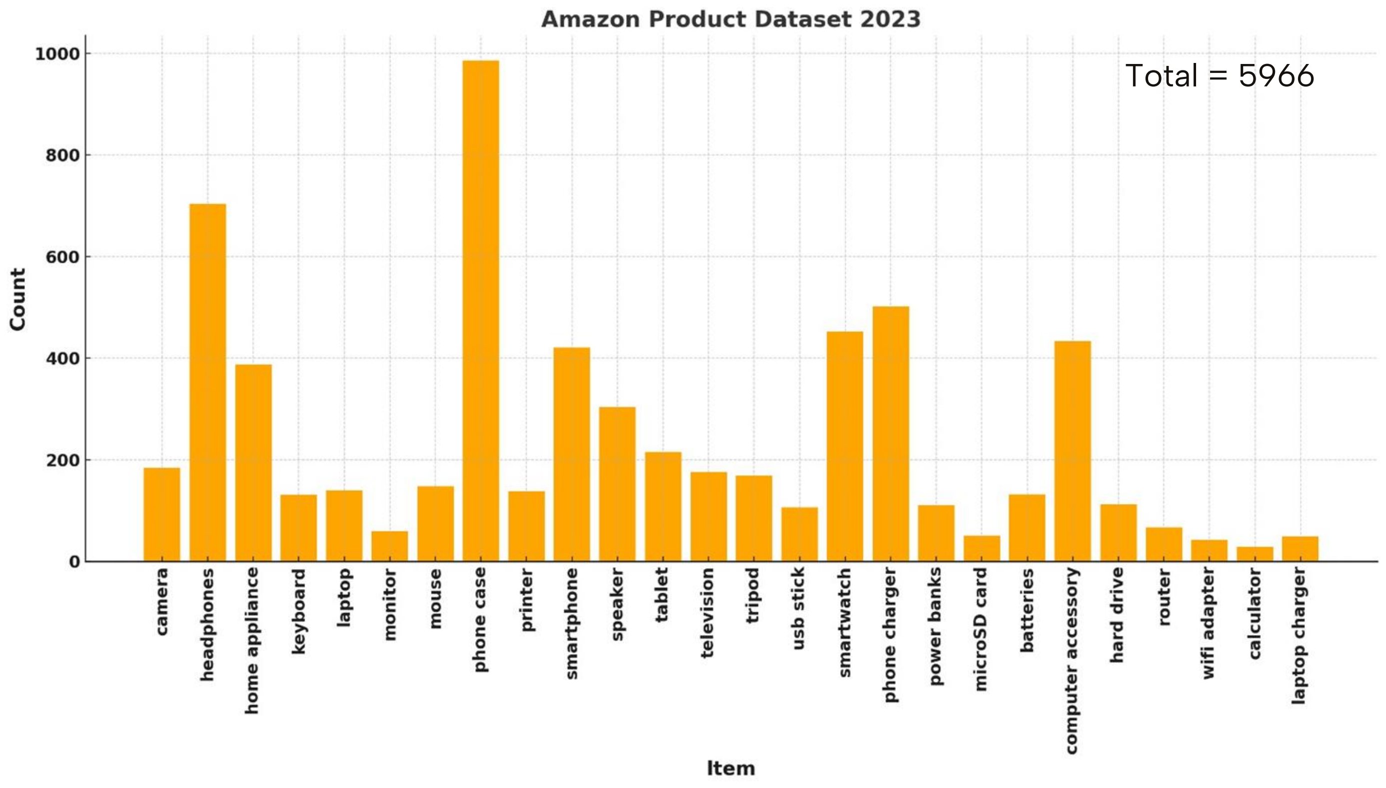
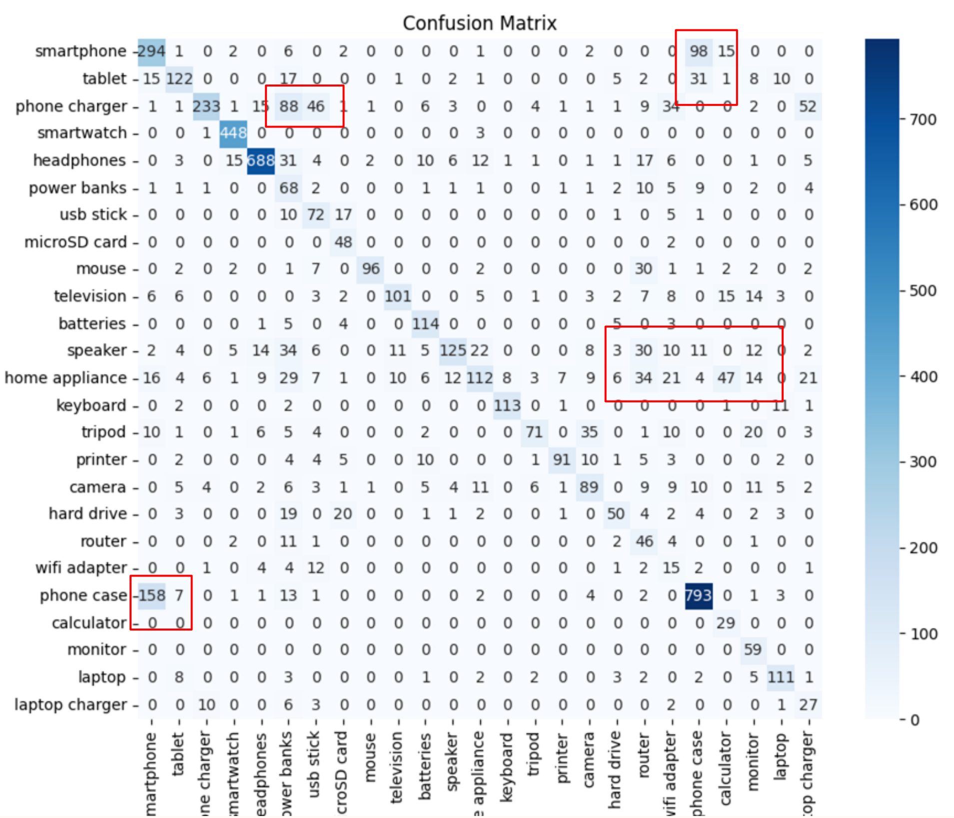


Image Classification Performance

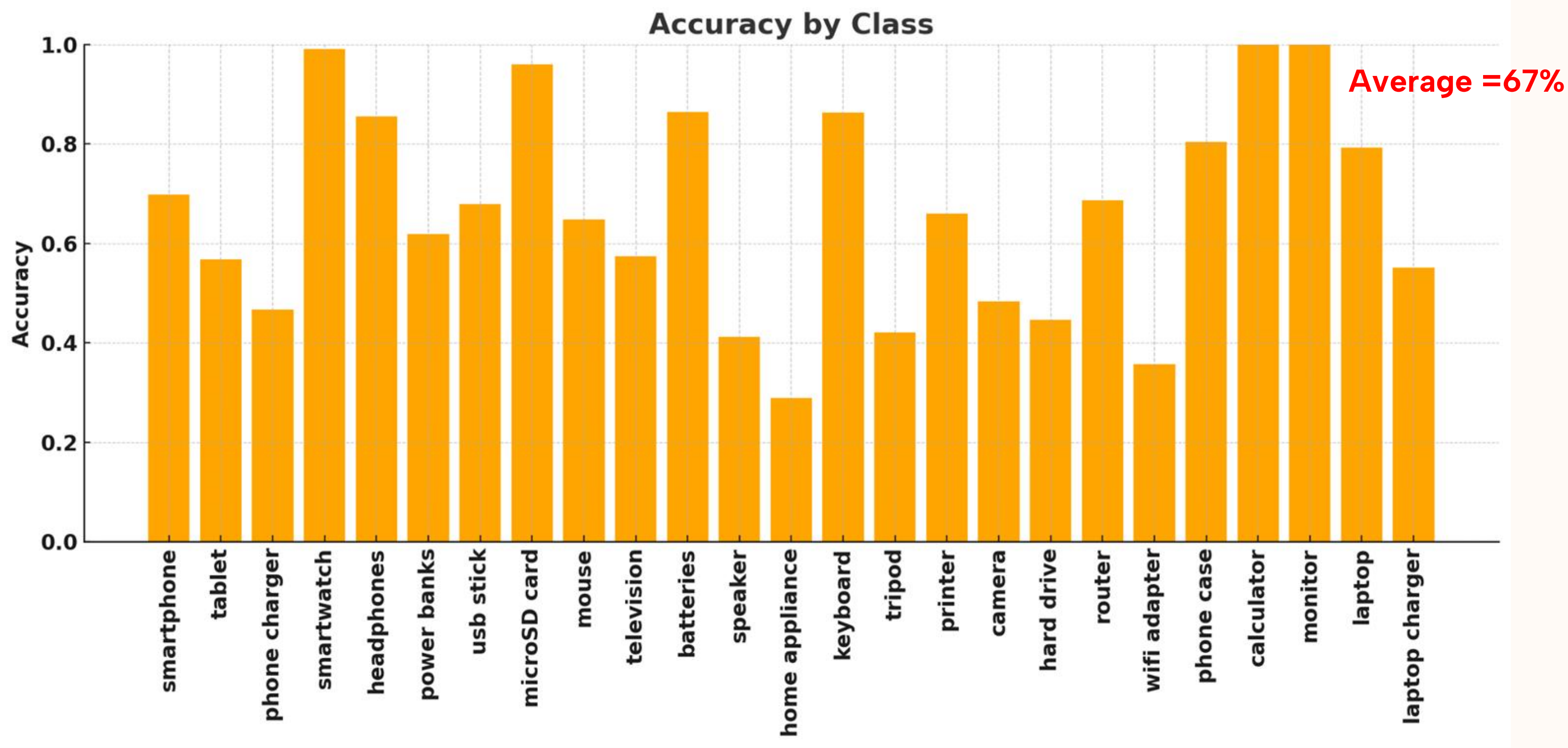
- `texts = [f“a photo of {`**`class_label`**`}” for class_label in class_label_list]`
- Given an image, find the text with top similarity score
- Predicted_label is class_label in the text
- `calculate_accuracy(predicted_labels, pseudo_labels)`

Baseline CLIP Performance



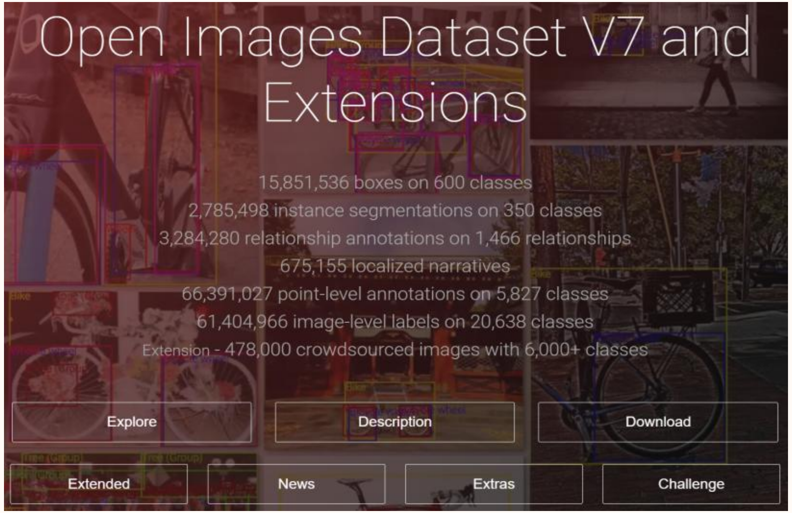
Average =67%

Baseline CLIP Performance

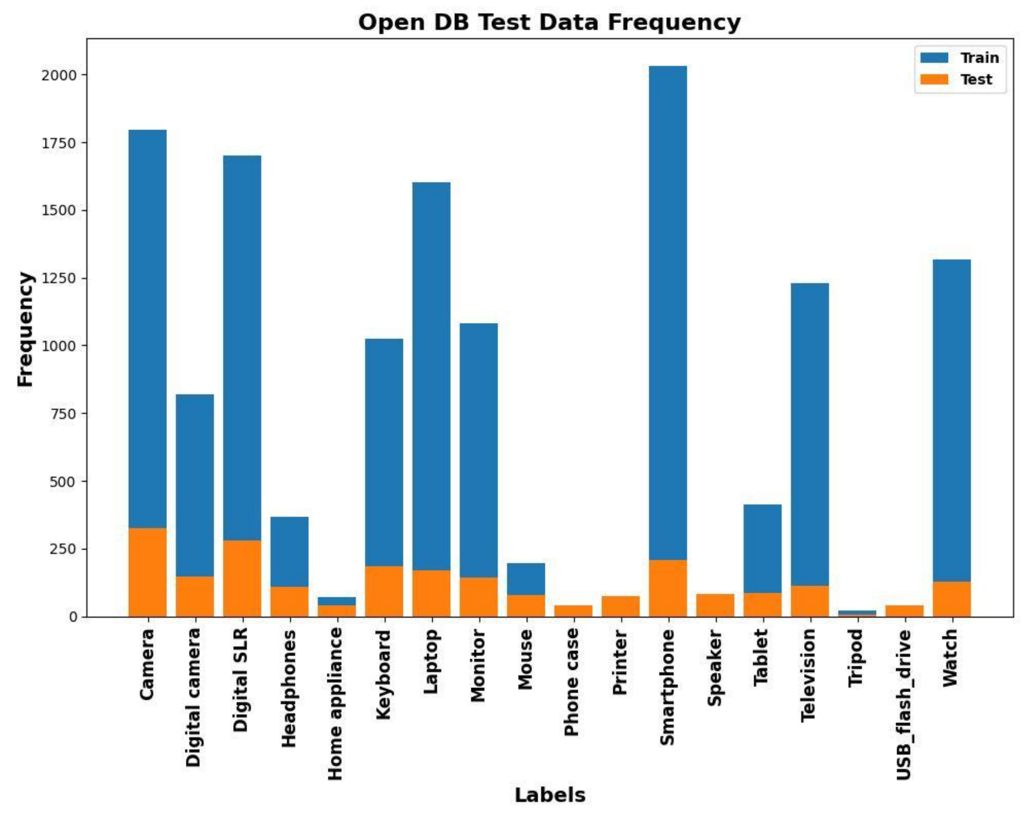


Part 3: Model Fine-tuning

- To improve the classification performance, we need to finetune the CLIP model on the select data classes
- Use of Open Images Dataset V7
- Transfer Learning
- Collected train–test data of selected classes



Open DB 7 Image Class Histogram



Num_classes = 16
Train Size = 13682
Test Size = 2088

Hyper Parameter Tuning

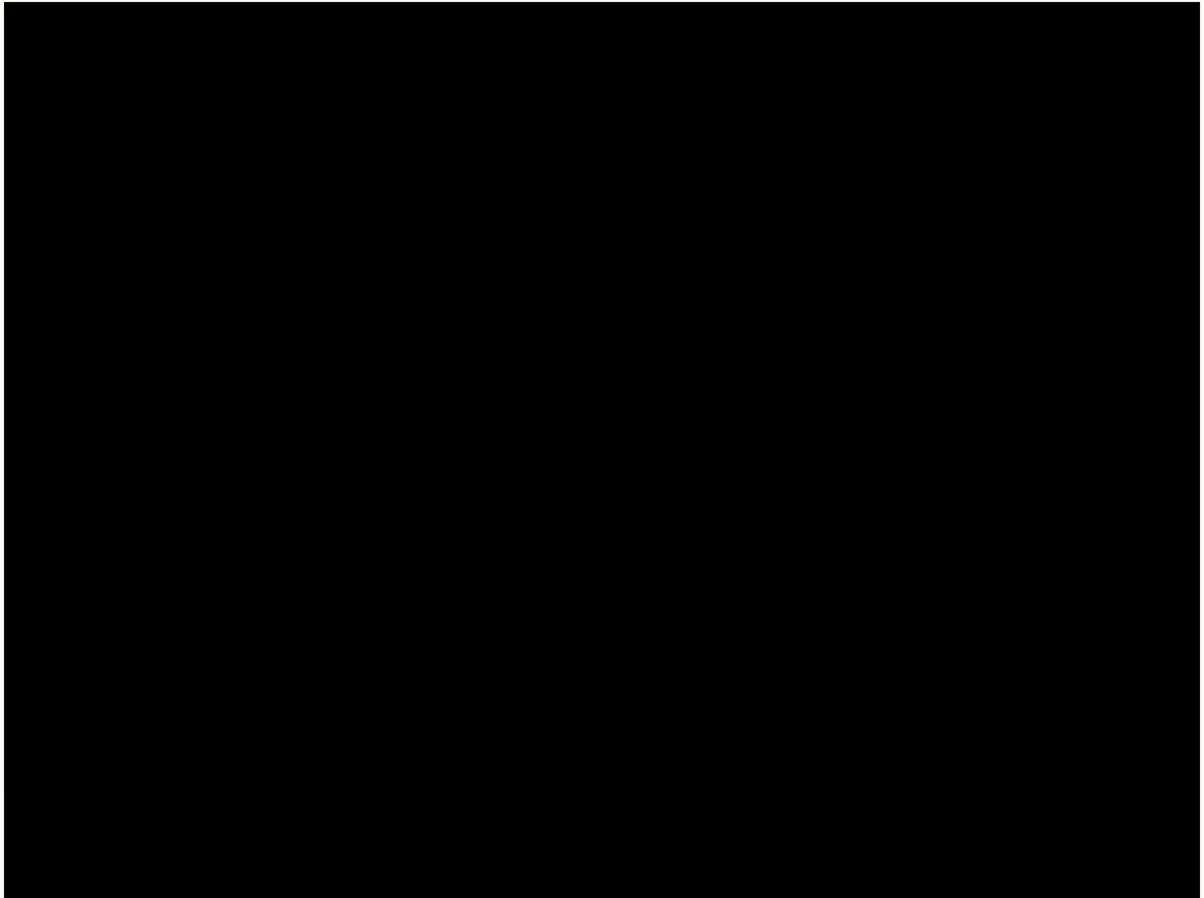
- Model training using MLFLow

- Base Model = CLIP ViT-B/32
- CrossEntropyLoss: (text_loss + image_loss)/2
- Adam Optimizer

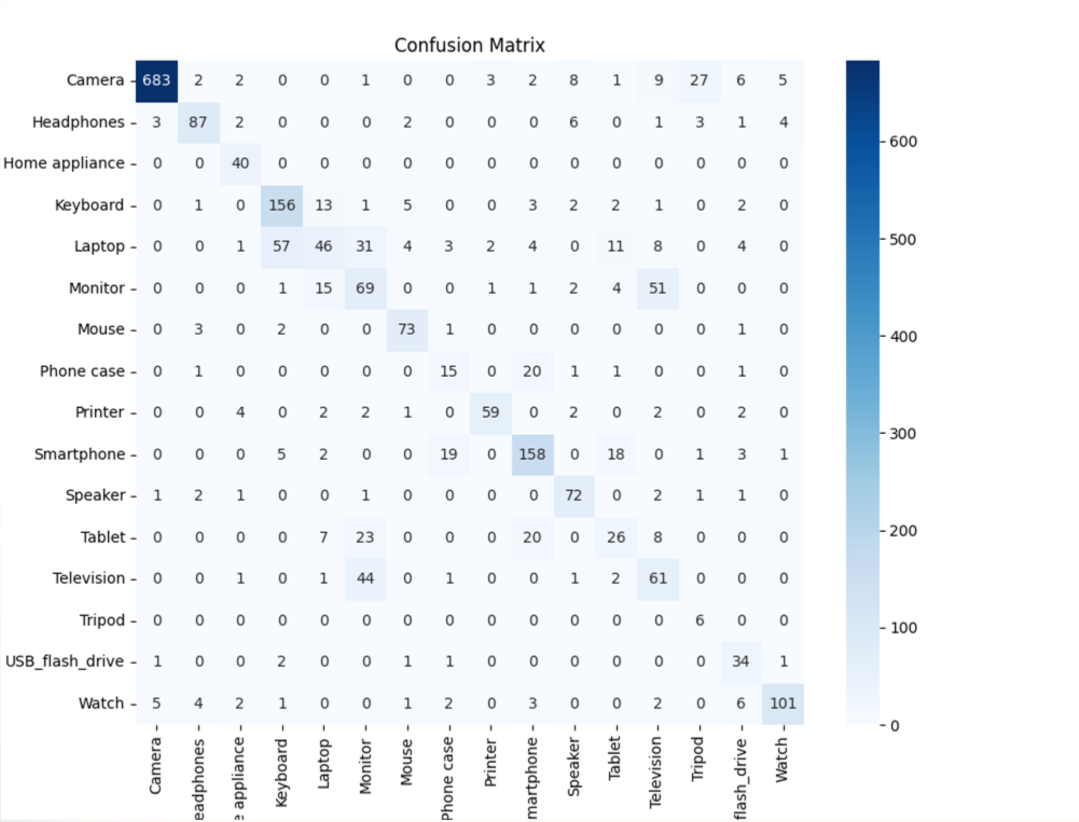
- Log the class_wise results for parameters:
  `{epochs-training, batch size, number of frozen layers, learning rate}`

epochs-training	2-10
batch size	16, 32, 64
number of frozen layers	None, 4, 6, 8, 10
Learning rate	1e-7, 5e-6, 1e-6

Model Tracking - MLFlow

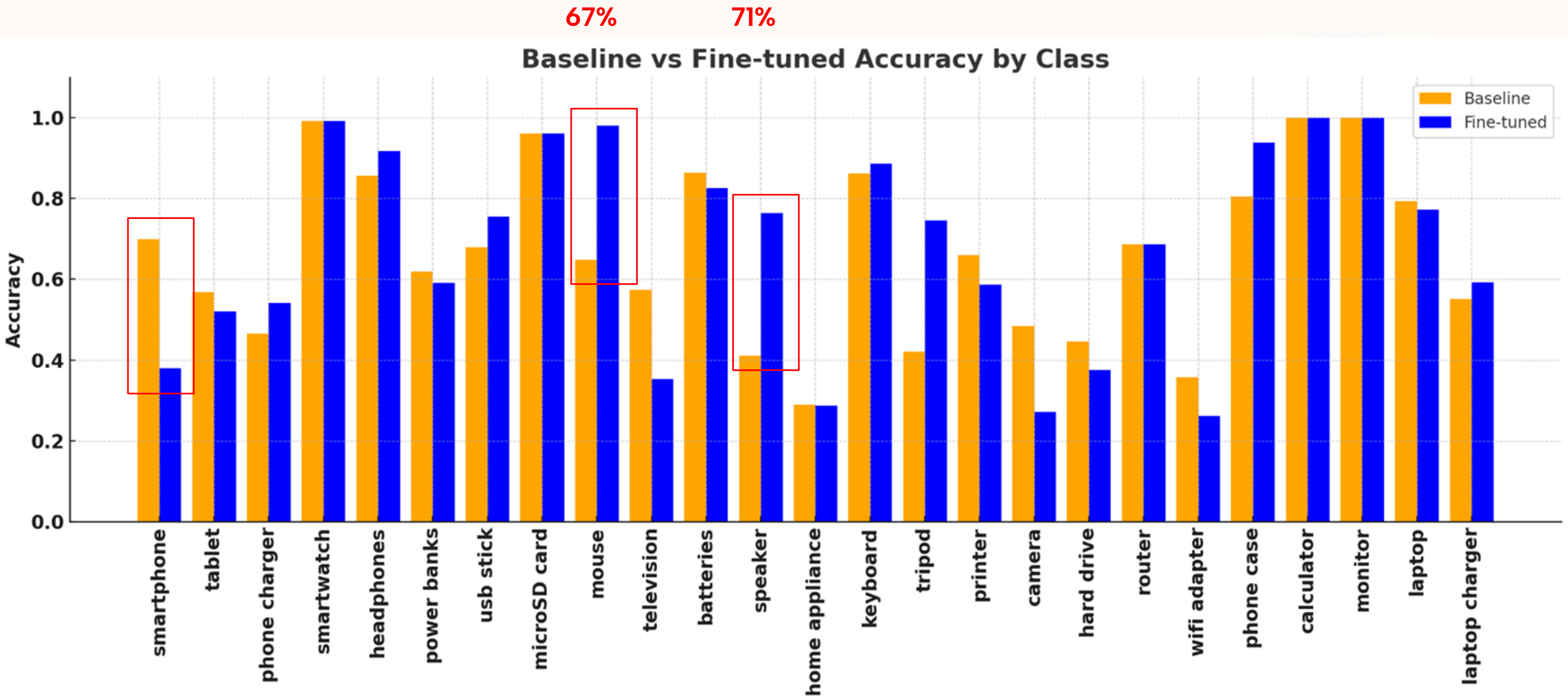


OpenDB 7 Classification Results



Overall Accuracy = 77%

Amazon Electronics Data Classification



Chatbot Observations

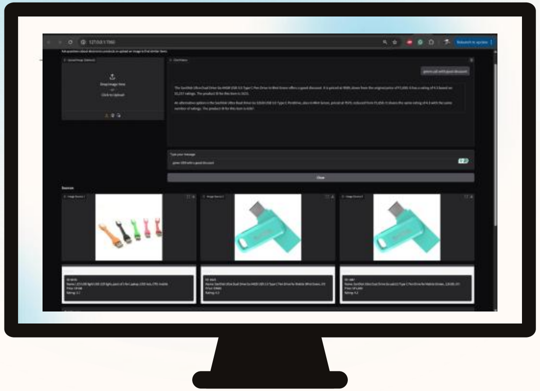
- Classification performance increased with Fine-tuning (4%)
- Update the fine-tuned clip model in the chatbot
- Little to no improvements in query results
- Results reduced on classes such as Smartphones

Discussion - Challenges

- The Image-Text pair is not a correct sample for CLIP Fine-tuning
  - A photo of {class label} ❌
- CLIP model might perform better in knowledge retrieval if trained on Amazon Electronics Dataset
  - The Amazon database has marketing-style images, while OpenDB is more real-life
- Fine-tuned with both losses (Image and Text)
- For downstream applications such as classification, fine-tuning is more relevant

Demo

Live walkthrough of the recommendation system



- **Scenario 1:** Text query → recommended products
- **Scenario 2:** Image query → recommended products
- **Visualize:** Show retrieved similar items and why they match

http://127.0.0.1:7860/

Future Work

01

Train on Amazon Dataset

02

Explore Other Vision Models for Embeddings

03

Data Annotation

04

Grounding Dino for Text based Bounding Box Detection

05

Grounded SAM for text based segmentation

Tools Used

- Python
- Pandas
- OpenAI CLIP
- Sentence Transformer, Huggingface
- GPT4o
- Pytorch
- Gradio
- Open Images Dataset V7
- MLFLow
- SkLearn, Seaborn

Contribution

01
Building RAG Chatbot

02
Prompt Engineering

03
Understanding the Embeddings of VLM

04
Model Tracking Platform

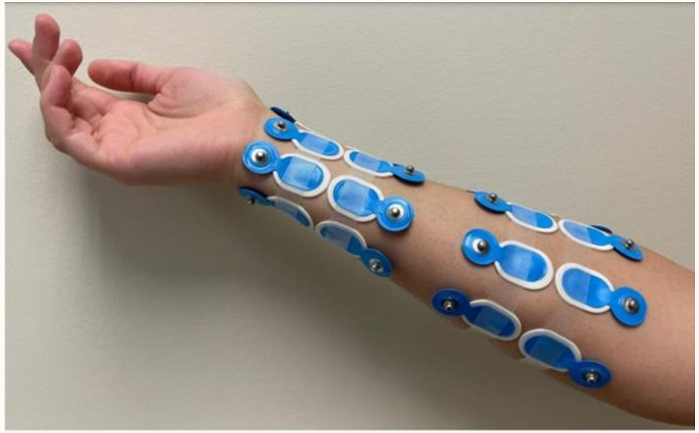
05
Model Fine Tuning/ Transfer Learning

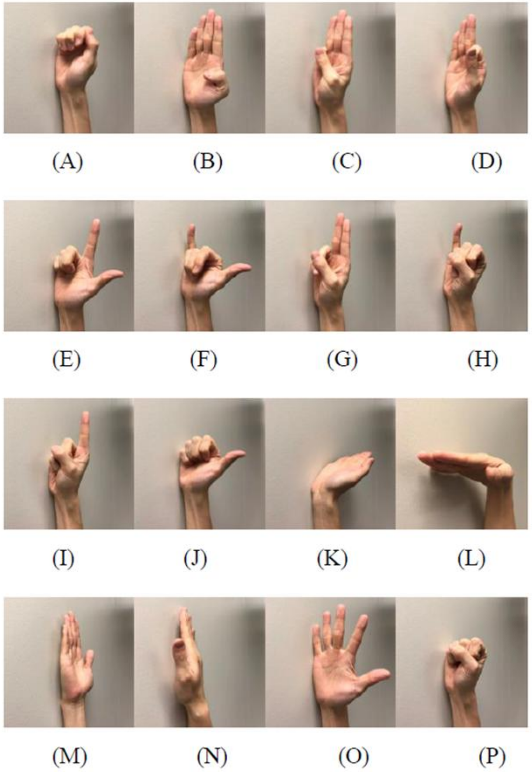
Any questions?

Thank you!

Master’s Research

PhD Research





List of Hand Gestures (Codes)
A) lateral prehension (LP)
B) thumb adduction (TA)
C) thumb and little finger opposition (TLFO)
D) thumb and index finger opposition (TIFO)
E) thumb and little finger extension (TLFE)
F) thumb and index finger extension (TIFE)
G) index and middle finger extension (IMFE)
H) little finger extension (LFE)
I) index finger extension (IFE),
J) thumb extension (TE)
K) wrist flexion (WF)
L) wrist extension (WE)
M) forearm supination (FS)
N) forearm pronation (FP)
O) hand open (HO)
P) hand close (HC)

EMG-BASED BIOMETRICS

Authentication & Identification

Contrastive Language-Image Pre-training

```
# Forward pass
  logits_per_image, logits_per_text = model(images, texts)

  # Compute loss
 ground_truth = torch.arange(len(images),dtype=torch.long,device=device)
 total_loss = (loss_img(logits_per_image,ground_truth) + l
oss_txt(logits_per_text,ground_truth))/2
```