Visual Search Engine Using Visual Language Model

AI-Powered Image Retrieval Using Vision-Language Models

Krishnandu Halder
Mentor: Dr Shivnath Ghosh
2025

Visual Search Engine using VLMs

📌 Project Report

🔍 Introduction

The Visual Search Engine utilizes Vision-Language Models (VLMs) to retrieve visually similar images based on textual queries or example images. This system embeds both text and images into a shared representation space, enabling advanced AI-driven search capabilities.

🏗️ Project Overview

📌 Problem Statement

Develop a visual search engine that can efficiently find relevant images using image-based and text-based queries. The system should leverage deep learning models to generate robust feature embeddings and provide an intuitive interface for users.

🛠️ Technologies Used

- **Programming Language**: Python
- **Deep Learning Framework**: PyTorch & Torchvision
- **Feature Extraction Model**: ResNet-18
- **Web Interface**: Streamlit
- **Data Processing**: NumPy, SciPy, PIL
- **Version Control**: Git & GitHub

🚀 Implementation

1️⃣ Feature Extraction

- Used **ResNet-18** to extract feature vectors from images.
- Applied preprocessing: resizing, normalization, and tensor conversion.
- Stored extracted embeddings for efficient similarity search.

2️⃣ Search Mechanism

- **Image-based Search**: Finds similar images using vector distances (cosine similarity / Euclidean distance).

3️⃣ User Interface

- Implemented a **Streamlit** web UI for interactive image selection and similarity search.

📂 Project Structure

```
📂 Visual_Search_Engine_using_VLM
├── 📂 images/                # Dataset of images
├── 📂 models/                # Pre-trained model storage
├── 📜 app.py                 # Streamlit UI application
├── 📜 feature_extraction.py  # Extracts image feature vectors
├── 📜 requirements.txt       # Required dependencies
├── 📜 README.md              # Project documentation
```

🔧 Setup & Installation

1️⃣ Clone the Repository

```
git clone https://github.com/Krishnandu-Halder/Visual_Search_Engine_using_VLM.git
cd Visual_Search_Engine_using_VLM
```

2️⃣ Create a Virtual Environment

- **Windows**
  ```
  python -m venv venv
  ```
- ```
  venv\Scripts\activate
  ```
-
- **Linux & macOS**
  ```
  python3 -m venv venv
  ```
- ```
  source venv/bin/activate
  ```
-

3️⃣ Install Dependencies

```
pip install -r requirements.txt
```

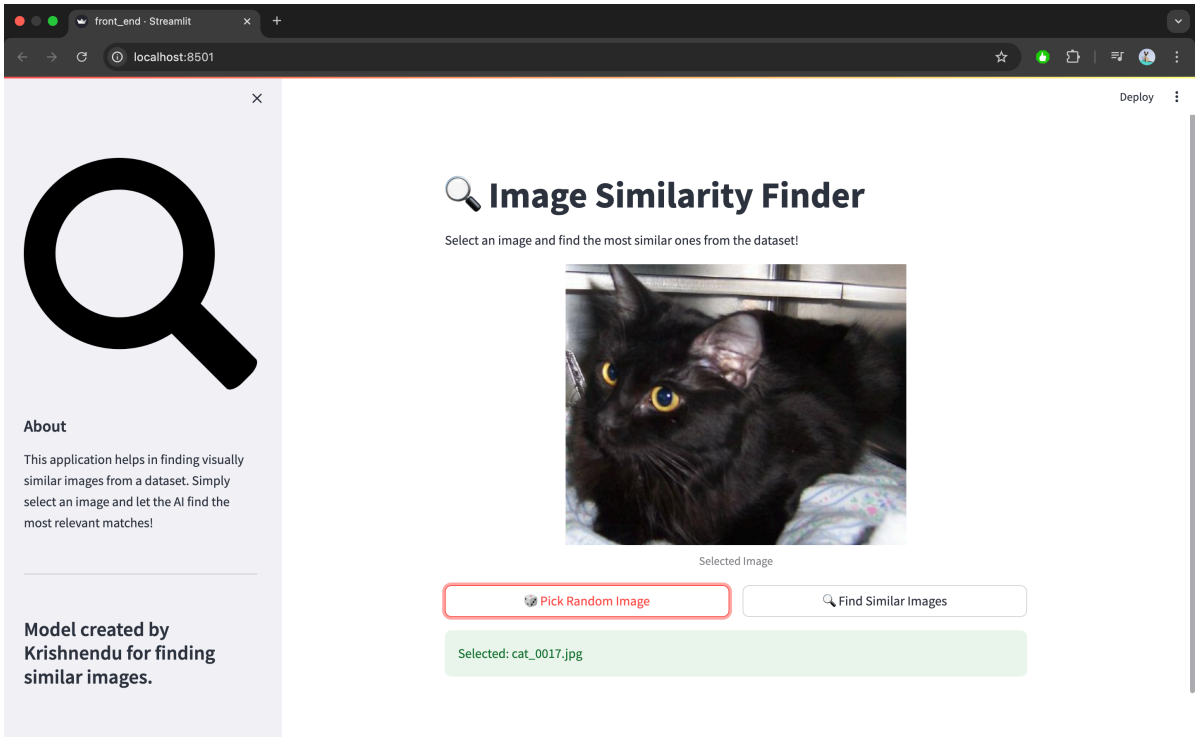
4️⃣ Running the Application

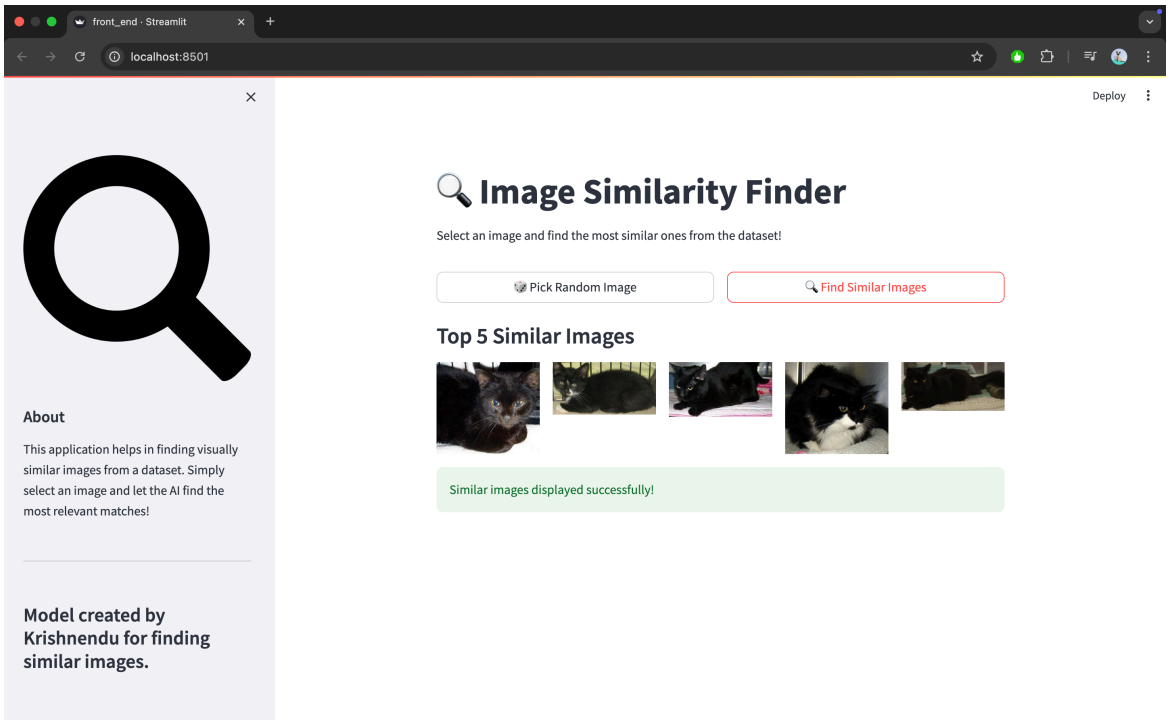
```
streamlit run app.py
```

5️⃣ Deactivating Virtual Environment

```
deactivate
```

🖼️ Preview





📌 Future Improvements

- 🔷 **Integrate CLIP/DINO models** for enhanced accuracy.
- 🔷 **Deploy to Cloud** (AWS/GCP) for scalability.
- 🔷 **Optimize database storage** for large-scale image retrieval.

📜 License

This project is licensed under the **MIT License**.

👤 Author

Krishnandu Halder - Frontend & AI Developer

For any queries, contact: Krishnandu.work@gmail.com

🚀 Thank you for exploring the Visual Search Engine!