Project Report: Visual Search Engine Using Vision-Language Models(VLMs)

# 1. Problem Statement

The goal of this project is to develop a visual search engine that leverages Vision-Language Models (VLMs) to retrieve relevant images based on textual queries or sample images. The system should embed both text and images into a shared representation space, allowing users to search via: keywords, natural language descriptions, similar images.

# 2. Project Category

This project falls under:

* - Computer Vision (Image Embedding & Similarity Search)
* - Information Retrieval (Search Engine for Images)
* - Machine Learning (Deep Learning-based Feature Extraction)

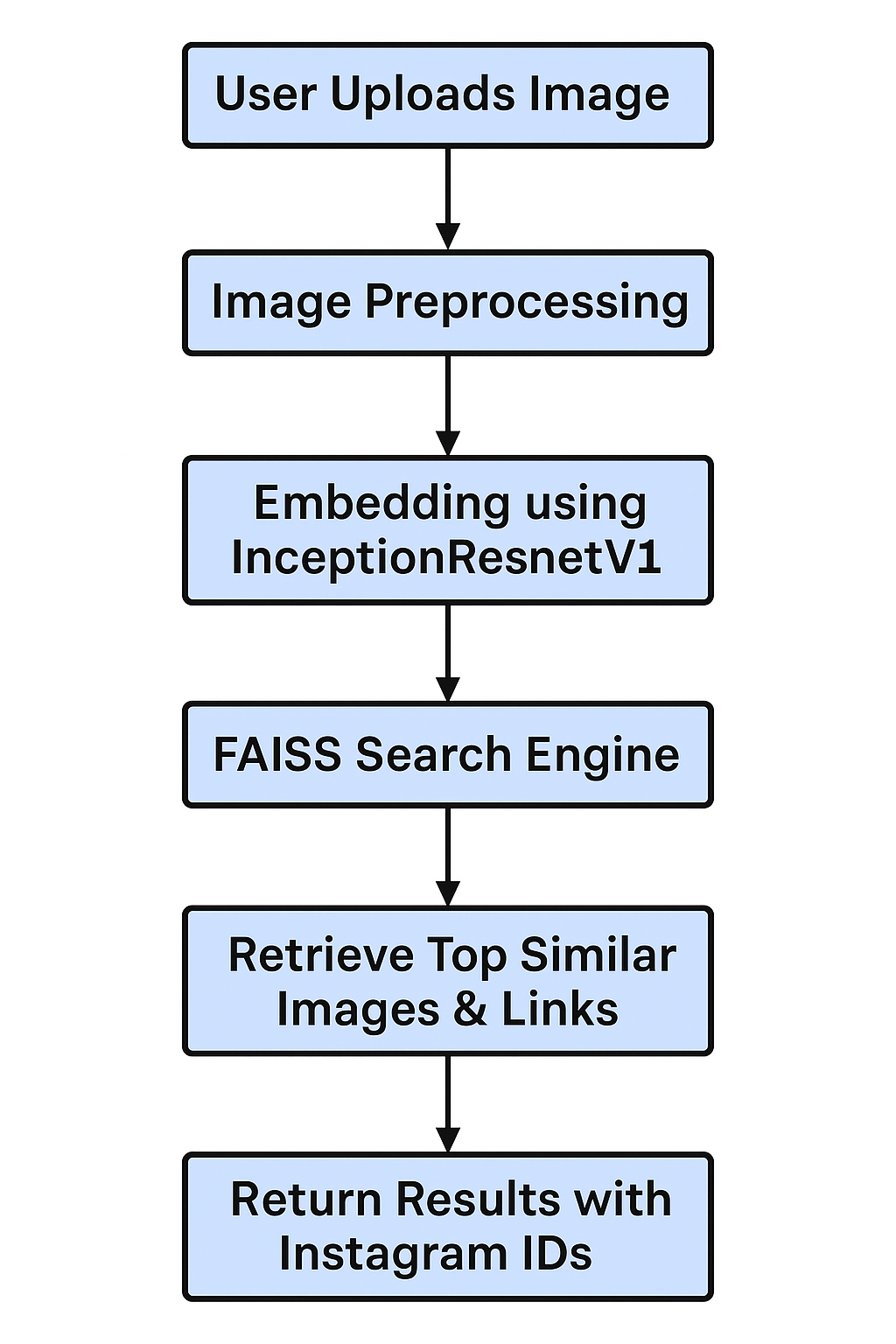
# 3. Solution Approach

To solve the problem, we built a visual search engine that:  
  
1. Preprocesses and embeds images from a dataset (Instagram images in this case) using a deep learning model.  
  
2. Stores embeddings in a vector database for efficient similarity search.  
  
3. Takes a user-uploaded image, converts it into an embedding, and searches for similar images in the dataset.  
  
4. Returns the most similar images along with their source links (Instagram profiles).

## Key Steps:

* - Dataset Preparation: Collected and preprocessed 3000 Instagram images.
* - Feature Extraction: Used InceptionResnetV1 (FaceNet) to generate image embeddings.
* - Vector Search: Used FAISS (Facebook AI Similarity Search) for fast nearest-neighbor search.
* - User Interface: Built a Gradio-based web interface for easy interaction.

## 4. Architecture Diagram



# 5. Flow of the Project:

1. Dataset Processing:  
- Images are preprocessed and embedded into vectors.  
- Vectors are stored in a FAISS index.  
  
2. User Interaction:  
- User uploads an image via Gradio interface.  
- The system computes its embedding.  
  
3. Search & Retrieval:  
- FAISS searches for the closest vectors.  
- The system retrieves and displays similar images with Instagram links.

# 6. Code Workflow (High-Level Explanation)

1. Image Embedding Generation:  
- Used InceptionResnetV1 to extract feature vectors from images.  
- Stored embeddings in a FAISS index for efficient search.  
  
2. Query Processing:  
- User uploads an image.  
- The system converts it into an embedding.  
- Searches the FAISS index for the nearest matches.  
  
3. Result Retrieval:  
- Returns the top similar images.  
- Displays Instagram profile links associated with the matched images.

# 7. Team Members & Contributions

|  |  |
| --- | --- |
| Team Member | Contribution |
| Arshdeep singh | Implemented image embedding, FAISS search, and Gradio UI. |
| seema | Assisted in dataset collection and preprocessing. |
| Rishabh | Worked on evaluation and performance testing. |

# 8. Future Enhancements

* - Support for Text Queries: Integrate a Vision-Language Model (e.g., CLIP) for text-based search.
* - Expand Dataset: Include more social media platforms (Facebook, LinkedIn, etc.).
* - Optimize Search: Improve speed and accuracy with better indexing techniques.

# 9. Conclusion

This project successfully demonstrates a visual search engine that retrieves similar images from an Instagram dataset. While the current implementation focuses on image-based search, future work will extend it to text-based queries using VLMs for a more comprehensive search experience.