Capstone Proposal

Domain Background

The fashion industry has consistently been at the forefront of innovation, with rapid advancements in digital technology transforming the way consumers interact with brands and products. The shift toward online shopping has opened new avenues for personalized experiences, especially in the context of in-store clothes retrieval systems. These systems are driven by the increasing demand for efficient and accurate recommendation tools that help customers find similar or complementary clothing items based on an uploaded image.

In recent years, the fashion domain has leveraged machine learning and computer vision to address challenges in this space. Historically, finding visually similar items was achieved through manual tagging or basic image processing techniques, which often lacked precision. The advent of deep learning has revolutionized this field, providing more sophisticated and scalable solutions. For instance, convolutional neural networks (CNNs) have been particularly effective in feature extraction from images, enabling enhanced retrieval accuracy.

Relevant Studies and Applications:

- **Fashion Image Retrieval Systems** focus on matching user-uploaded images with similar items in a database, using advanced techniques like CNNs for clustering fashion products based on visual features. More details can be found in this review on fashion recommendation systems.
- **AI-Driven Fashion Recommender Systems** explore the use of machine learning and deep learning to create personalized recommendations and retrieval systems in the fashion industry. Read more about these approaches here.
- **Multi-Turn Fashion Image Retrieval** introduces an iterative approach to refining image retrieval based on continuous user input, which is particularly relevant for complex search queries in fashion. More information is available here.

Problem Statement

The inefficiency and inaccuracy of finding similar clothing items based on a user's uploaded photo is a significant challenge in the fashion industry. Traditional methods relying on manual tagging or keyword searches fail to capture the nuanced visual similarities between garments. This problem is quantifiable, measurable, and replicable. The goal is to develop a system that processes images and retrieves visually similar items from an inventory, thereby enhancing the shopping experience.

Solution Statement

The proposed solution involves creating a machine learning-based system that extracts features from an uploaded image of a garment and searches for similar items in a pre-indexed inventory. This system will utilize a convolutional neural network (CNN), specifically a pre-trained MobileNet model, for feature extraction. These features will be indexed using a vector database, such as FAISS or Pinecone enabling efficient similarity searches.

The solution will be developed on the AWS platform, leveraging its scalable infrastructure for both model training and deployment. AWS SageMaker will be used for training the model, given its robust environment for handling large-scale machine learning tasks. The performance of this system will be evaluated based on retrieval accuracy and response time, ensuring that it meets the requirements of an effective in-store clothes retrieval system.

Datasets and Inputs

The project will use the DeepFashion dataset, specifically the “In-shop Clothes Retrieval” subset. This dataset was obtained from the Multimedia Laboratory at The Chinese University of Hong Kong and is well-suited for training and evaluating image retrieval models. More information about the dataset can be found here.

The images are preprocessed to a uniform size and normalized to ensure consistency. Bounding box and landmark annotations provided with the dataset can be optionally used to improve the focus on the clothing items and enhance feature extraction.

Dataset Characteristics: - **Total Images**: 52,712

- **Categories**: Men’s and Women’s clothing
- **Sub-categories**:
  - **Men**: Denim, Jackets_Vests, Pants, Shirts_Polos, Shorts, Suiting, Sweaters, Sweatshirts_Hoodies, Tees_Tanks
  - **Women**: Blouses_Shirts, Cardigans, Denim, Dresses, Graphic_Tees, Jackets_Coats, Leggings, Pants, Rompers_Jumpsuits, Shorts, Skirts, Sweaters, Sweatshirts_Hoodies, Tees_Tanks
- **Image Dimensions**: 256 x 256 pixels
- **Total Size**: 900 MB
- **Annotations**: Bounding boxes, landmarks

Benchmark Model

A baseline model will be implemented using the pre-trained MobileNet network, chosen for its lightweight architecture and efficiency. MobileNet’s

fast training and inference times make it an ideal candidate for initial implementation, particularly when hardware resources are limited. This model will serve as a benchmark for comparing the performance of more advanced models.

If MobileNet does not achieve the desired accuracy and retrieval performance, a more powerful model like ResNet50 will be considered. ResNet50, known for its higher accuracy, may provide better results at the cost of increased computational requirements.

Evaluation Metrics

The primary evaluation metrics for this project will include **precision and recall at k** (e.g., precision@5, recall@5), which assess the accuracy of the top k retrieved items. **Mean Average Precision (mAP)** will also be used to evaluate the overall retrieval performance. These metrics are crucial for measuring the effectiveness of image retrieval systems, where the goal is to find the most visually similar items in an inventory.

Project Design

The workflow for this project involves several key steps:

1. **Data Preparation**:
    - Organize and preprocess the DeepFashion dataset.
    - Extract features using a pre-trained model.
    - Visualize data distribution and characteristics.
2. **Feature Extraction**:
    - Fine-tune the pre-trained model if necessary.
    - Extract and store feature vectors for all images in the inventory.
3. **Indexing**:
    - Use a vector database to create a vector index of the extracted features.
    - Optimize the index for fast retrieval.
4. **Query Processing**:
    - Develop an interface for users to upload images.
    - Extract features from the uploaded image using the trained model.
    - Query the vector index to find similar items.
5. **Evaluation**:
    - Evaluate the system using precision@k, recall@k, and mAP metrics.
    - Compare the results with the benchmark model.
6. **Deployment**:
    - Implement the system as a web application on AWS for user accessibility.