

Capstone Proposal

Domain Background

The fashion industry has always been at the forefront of innovation and change. With the advent of digital technology, the domain has seen a significant shift towards online shopping and personalized experiences. In-store clothes retrieval is a growing field within this domain, driven by the need for efficient and accurate recommendation systems that can help customers find similar or complementary clothing items based on an uploaded image. This project aims to leverage the power of machine learning and computer vision to enhance the shopping experience by developing a system that allows users to upload a photo of a garment and retrieve similar items from an online store inventory.

Historically, the challenge of finding similar items based on visual attributes has been addressed using various techniques, from manual tagging to basic image processing methods. However, the rise of deep learning has revolutionized this field, enabling more accurate and scalable solutions.

Problem Statement

The problem to be solved is the inefficiency and inaccuracy of finding similar clothing items based on a user's uploaded photo. Traditional methods often rely on manual tagging or simple keyword searches, which may not capture the nuanced visual similarities between different garments. This problem is quantifiable, measurable, and replicable as it can be addressed by developing a system that processes images and retrieves visually similar items from an inventory.

Solution Statement

The proposed solution is to develop 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 solution will use a convolutional neural network (CNN) pre-trained to extract meaningful features from clothing images. These features will be indexed in a vector database (such as FAISS, Pinecone, or Milvus), allowing for efficient similarity searches.

The solution will be quantifiable through retrieval accuracy and response time and replicable using publicly available datasets and models.

Datasets and Inputs

The project will use the DeepFashion dataset, specifically the "In-shop Clothes Retrieval" subset, which contains over 52,000 images of various clothing items. The dataset is structured with images categorized into men and women, different clothing types, and sub-categories representing different garments of the

same type. 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.

Benchmark Model

A baseline model will be implemented using a pre-trained MobileNet network, known for its lightweight architecture and efficiency. This makes MobileNet an ideal choice for the initial implementation, as it allows for faster training and inference times without requiring extensive hardware resources.

To ensure a fair comparison, we will first use MobileNet as a baseline model and evaluate its performance under the same conditions as our final optimized solution. The results of MobileNet will serve as a benchmark to compare against our improved solution.

If MobileNet does not meet the desired accuracy and retrieval performance, we will experiment with a more powerful model, such as ResNet50. ResNet50 generally provides higher accuracy at the cost of increased computational requirements.

Evaluation Metrics

The primary evaluation metrics for this project will be precision and recall at k (e.g., precision@5, recall@5), which measure 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 appropriate given the context of image retrieval, where the goal is to find the most visually similar items from an inventory.

Project Design

The theoretical workflow for this project includes several key steps:

1. **Data Preparation:**
 - Organize and preprocess the DeepFashion dataset.
 - Extract features using a pre-trained model.
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 vector database to create a vector index of the extracted features.
 - Ensure the index is optimized 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 for user accessibility.

Small visualizations, such as diagrams of the workflow and examples of image retrieval results, will be included to aid understanding.