Facial Attribute Detection using Multi-Task Deep Learning

By Muhammad Nasr

Abstract

This project aims to develop a multi-task deep learning model that can predict multiple facial attributes from a single image. The model simultaneously performs binary classification for gender and eyeglasses detection and regression for shirt color estimation. The solution is trained on the CelebA dataset using a custom PyTorch pipeline with a modified ResNet18 architecture. The final model is capable of real-time inference on external user-uploaded images, enabling robust facial analysis in practical scenarios.

Table of Contents

1. Introduction

1.1. Background

Facial attribute recognition has emerged as a vital area in computer vision, driven by its applications in surveillance, human-computer interaction, online retail, and digital identity verification. Traditionally, separate models are trained for each attribute such as gender detection, eyewear identification, or clothing analysis. However, such an approach is computationally expensive and inefficient for real-time applications.

With the increasing availability of annotated facial datasets and advances in deep learning architectures, multi-task learning (MTL) has proven to be a promising alternative. MTL enables a single model to learn shared representations for multiple related tasks, improving overall accuracy and reducing training time and computational cost. This project leverages multi-task learning using a deep convolutional neural network (CNN) to simultaneously predict:

- Gender (Male/Female)
- Eyeglasses (Present/Absent)
- Shirt Color (RGB regression)

The model is trained and evaluated using the CelebA dataset, a large-scale facial attribute dataset, and implemented with PyTorch. The solution is capable of analyzing unseen facial images and outputting the three target attributes with high accuracy and efficiency.

1.2. Problem Statement

Most existing facial attribute detection systems are designed to predict a single attribute at a time, leading to redundancy in computation and slower performance. Additionally, visual features like shirt color, which are not explicitly labeled in datasets, are often ignored despite their relevance in applications like fashion recommendation and biometric profiling.

This project addresses the need for:

- Simultaneous multi-attribute detection.
- Lightweight models for real-time applications.
- Accurate extraction of visual features like gender, eyewear presence, and shirt color.

1.3. Aims & Objectives

- Train a multitask deep learning model for gender, eyeglasses, and shirt color prediction.
- Use ResNet18 as the base model for feature extraction.
- Evaluate the model using classification and regression metrics.
- Enable prediction on external images uploaded by users.

1.4. Scope of Project

This project focuses on static image-based facial attribute prediction. The model supports three attributes and is trained using a subset of the CelebA dataset. Deployment on mobile platforms or real-time video streams is outside the current scope but may be considered in future enhancements.

2. Tools & Technologies

Programming Language	Python
Deep Learning	Pytorch, Torchvision
Data Manipulation	Pandas, Numpy
Visualization	Matplotlib, Seaborn
Image Processing	OpenCV, PIL
Dataset	CelebA
Clustering Algorithm	KMeans
IDE/Platform	Google Colab

3. Dataset Details

- **Source**: CelebFaces Attributes (CelebA) Dataset.
- **Image Count Used**: ~17,000 images.
- **Attributes Used**:

.1. Male: (Binary: 0 = Female, 1 = Male)
.2. Eyeglasses: (Binary: 0 = No, 1 = Yes)
.3. Image_ID: File path for each image.

- **Shirt Color**: Estimated via clustering (not provided in dataset).

4. Methodology

4.1. Data Preprocessing

- Attributes loaded from CSV (list_attr_celeba.csv).
- Converted -1 to 0 in binary columns.
- Shirt color region cropped heuristically (bottom 40%) and clustered using KMeans (n=1) to extract the dominant color.

4.2. Custom Dataset Class

A PyTorch Dataset class (CelebADataset) was created to:

- Load images from disk
- Apply transforms (resize, normalize)
- Extract RGB shirt color using extract_shirt_color() function.

4.3. Data Split

Dataset	Number of Images
Training Set	15000
Test Set	2000

4.4. Diagram

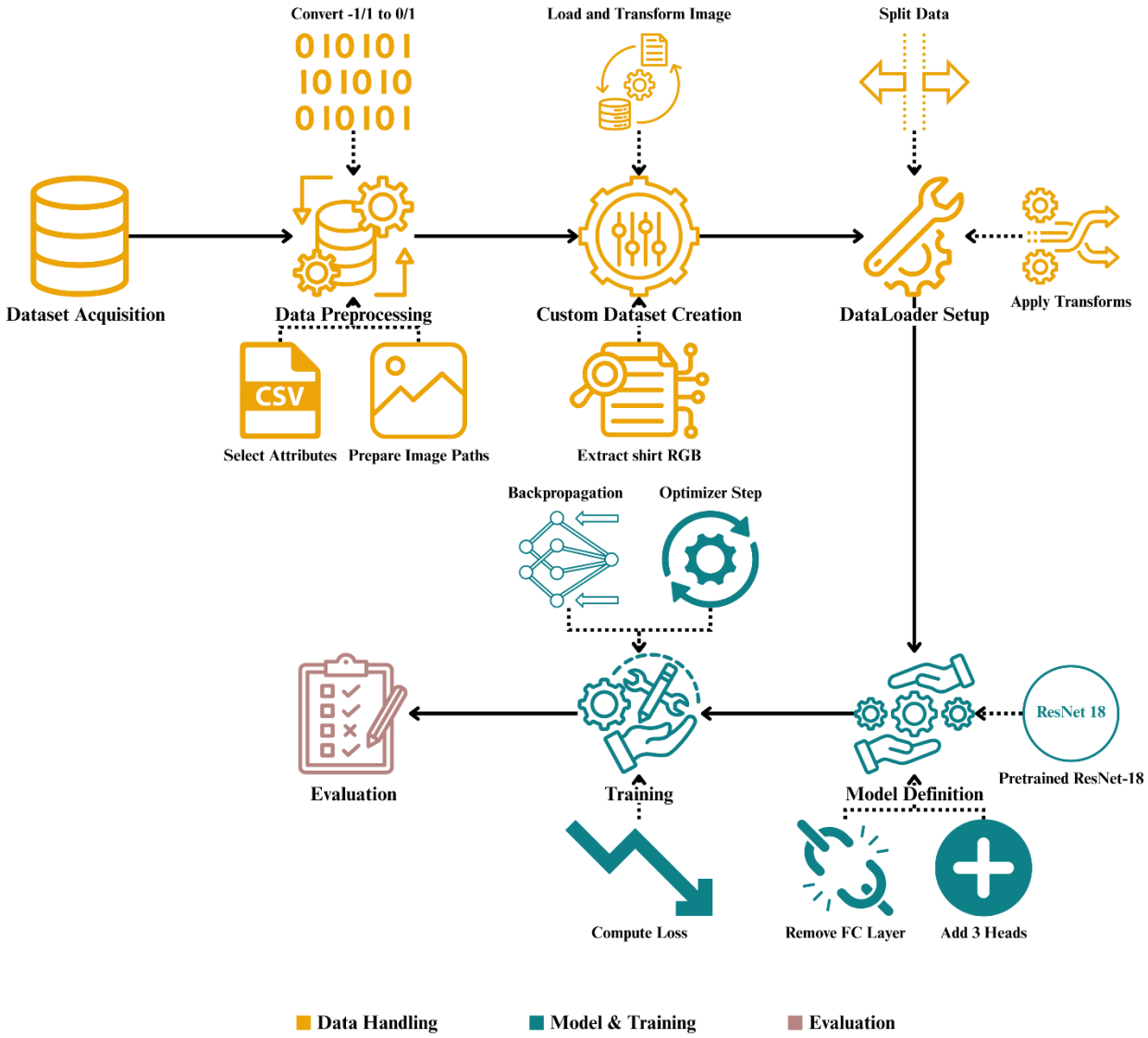

Fig: Methodology Diagram

5. Model Architecture

Base Model: Pre-trained ResNet18

Multi Task Heads:

Task	Output Size	Activation	Loss Function
Gender	1	Sigmoid	Binary Cross Entropy
Eyeglasses	1	Sigmoid	Binary Cross Entropy
Shirt Color	3 (RGB)	Linear	Mean Squared Error

The final fc layer of ResNet18 is replaced with nn.Identity(), and three parallel heads are added for the respective tasks.

6. Training Configuration

Parameter	Value
Epochs	3
Batch Size	32
Optimizer	Adam
Learning Rate	0.0001
Loss Functions	BCE + BCE + MSE

7. Results

7.1. Classification Metrics

Evaluated on 2,000 validation images using classification_report() and confusion_matrix.

Gender Classification

Metric	Value (approx.)
Accuracy	95%+
Precision	High
Recall	High
F1-Score	High

Eyeglasses Classification

Metric	Value (approx.)
Accuracy	90%+
Precision	High
Recall	High
F1-Score	High

Note: Shirt color prediction is visually verified as ground truth labels were not available.

8. External Image Testing

- Supports user-uploaded image via Colab interface (files.upload()).
- Inference is performed and outputs:

.1. Predicted Gender
.2. Eyeglasses: Yes/No
.3. Shirt color in RGB

- Image is displayed with annotations using matplotlib.

9. Conclusion

- Successfully developed a multi-task model for facial attribute detection.
- ResNet18 proved efficient and accurate across tasks.

- Shirt color estimation is a novel and practical addition.
- The approach demonstrates the capability of deep learning for real-time facial analysis.

10. Future Work

- Improve shirt detection via segmentation models instead of heuristics.
- Add more attributes like beard, smile, etc.
- Use attention mechanisms to improve feature learning.
- Train with full CelebA dataset and additional augmentations.
- Integrate explainability (e.g., Grad-CAM) for insights into decisions.