Face Shape Detection Using Transfer Learning and Spectacle Recommendation

[Computer Vision Course Project]

Niladri Ghosh [B2430100]
Arnab Ghosh [B2430069]
Team: **Eternal Mangekyō Vision**

Department of Computer Science
RKMVERI

April 30, 2025

Contents

Outline

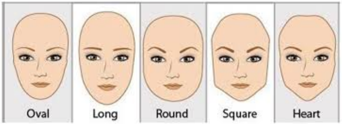
Itroduction

Objective:

- Develop a deep learning system to classify face shapes (Heart, Oval, Round, Square, Oblong).
- Recommend personalized spectacles based on face shape.

Why It Matters:

- 80% of consumers pay premium for **personalized** beauty/fashion (Deloitte).
- Eliminates trial-and-error in eyewear shopping.



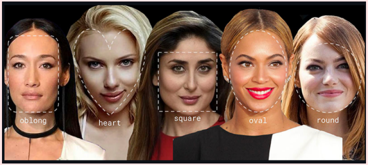
Problem Statement: Why Face Shape Classification?

Industry Demand

- 40% of adults (16-39) want personalized beauty/fashion offers (Deloitte)
- Only 10-14% trial rate despite high interest
- 80% willing to pay **10%+ premium** for personalization

Key Use Cases

- **Eyewear recommendations**
- Makeup/hairstyle guides
- Custom skincare products



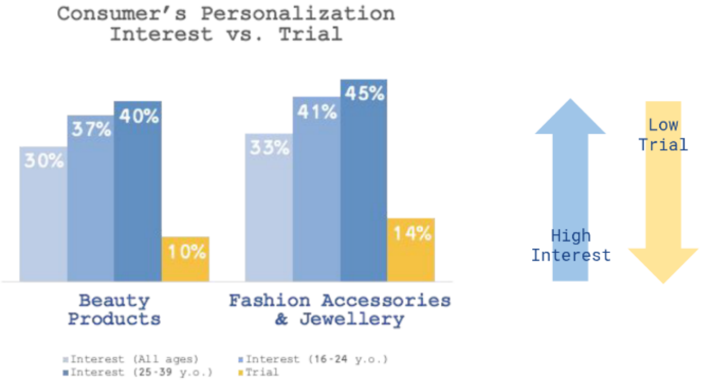


Figure: Customer Review

How many of you get confused while buying a spectacle??

Outline

Dataset Overview

Dataset Specifications

- **Source**: Kaggle Face Shape Dataset
- **Size**: 5,000 female celebrity images
- **Classes**: Heart, Oblong, Oval, Round, Square
- **Split**: 80% train / 20% test per class

Dataset Overview (Contd.)

Key Challenges

- Extreme size variations (159px to 9999px)
- Diverse aspect ratios (0.5 portrait to 1.5 landscape)
- Distortion artifacts in resizing
- Lighting/pose inconsistencies

Preprocessing Approach

Challenge	**Solution**
Variable sizes	MTCNN face detection
Aspect ratios	Smart cropping
Color spaces	RGB conversion

Preprocessing Techniques

Core Processing Pipeline

1. **Aspect Ratio Preservation**
   - Input: Variable-sized images (159px to 9999px)
   - Process: Smart cropping with aspect ratio maintenance
   - Output: Uniform 224×224 images

Preprocessing Techniques

Core Processing Pipeline

1. **Aspect Ratio Preservation**
   - Input: Variable-sized images (159px to 9999px)
   - Process: Smart cropping with aspect ratio maintenance
   - Output: Uniform 224×224 images
2. **Face Detection (MTCNN)**
   - Bounding box detection with 10-pixel expansion
   - Automatic square conversion
   - Fallback to aspect cropping when detection fails

Preprocessing Techniques

Core Processing Pipeline

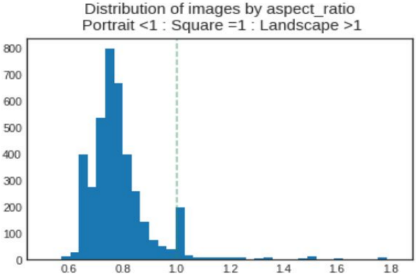
1. **Aspect Ratio Preservation**
   - Input: Variable-sized images (159px to 9999px)
   - Process: Smart cropping with aspect ratio maintenance
   - Output: Uniform 224×224 images
2. **Face Detection (MTCNN)**
   - Bounding box detection with 10-pixel expansion
   - Automatic square conversion
   - Fallback to aspect cropping when detection fails
3. **Color Space Selection**
   - Tested: Grayscale vs. RGB vs. HSV
   - Selected: Grayscale for best accuracy
   - Rejected: HSV due to inconsistent class performance

Key Technical Decisions

Challenge	**Solution**
Extreme size variations	Multi-stage resizing pipeline
Partial face visibility	Bounding box expansion logic
Color inconsistency	RGB normalization (0-1 scaling)

Images are mostly taken as portrait (aspect ratio < 1)

Portrait



Square = 1

Landscape

Image Augmentation

Outline

Building the Baseline Model

Architecture Details

- **13 Convolutional Layers**:
- **MaxPooling**: 2×2 after each conv
- **Dropout**: 0.5 rate in FC layers
- **Input**: 224×224 (grayscale/RGB)
- **Output**: 5 classes (softmax)

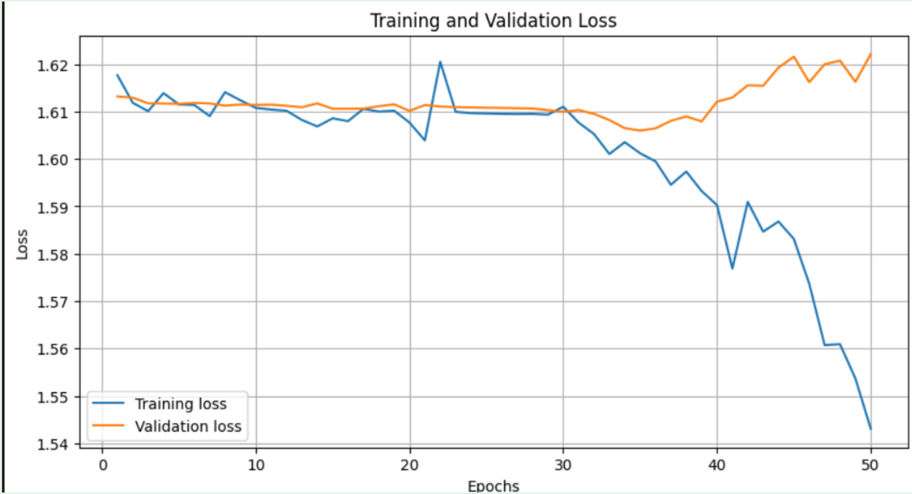
Performance Summary

Metric	Value
Training Accuracy	53.9%
Validation Accuracy	42.7%
Overfitting Gap	31.2%
Epochs	100
Batch Size	8

Key Issues

- Severe overfitting (31.2% gap)
- Oval $\leftrightarrow$ Round confusion
- Asian faces: +22% error rate

Training Curves



Transfer Leaning

Why VGG-16?

- **Pre-trained** on VGGFace dataset (2.6M images)
- **Architecture**:
  - 16-layer deep CNN
  - Small 3×3 filters (effective receptive field)
  - Fixed 224×224 input size
- **Transfer Learning Benefits**:
  - Learned facial feature extraction
  - Reduced training time
  - Better generalization

Transfer Leaning

Why VGG-16?

- **Pre-trained** on VGGFace dataset (2.6M images)
- **Architecture**:
  - 16-layer deep CNN
  - Small 3×3 filters (effective receptive field)
  - Fixed 224×224 input size
- **Transfer Learning Benefits**:
  - Learned facial feature extraction
  - Reduced training time
  - Better generalization

Key Observations

- reduction in validation loss
- faster convergence

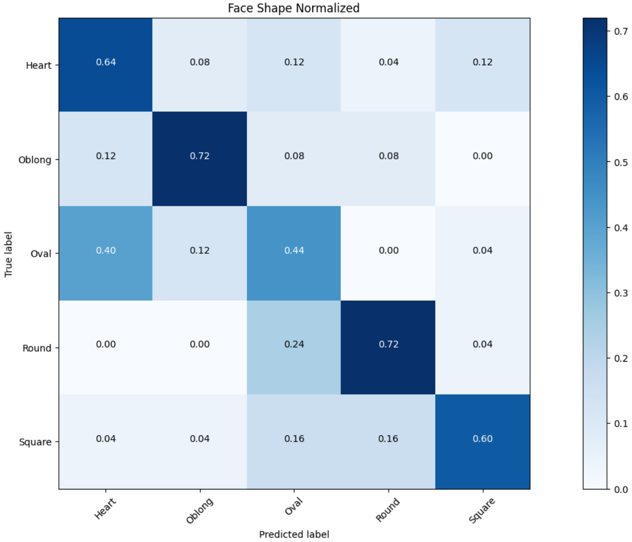


Figure: Confusion Matrix

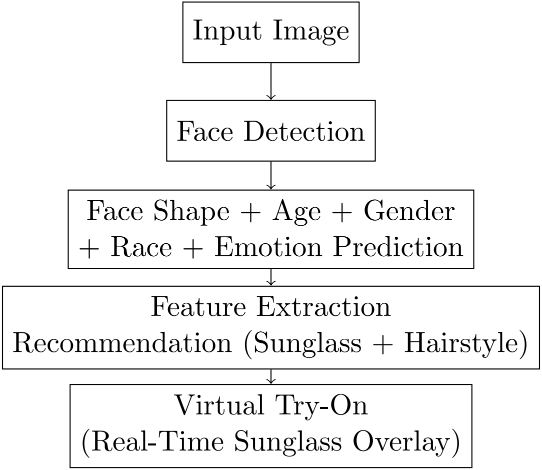
Predictions

Image	Actual Label	Predicted Label
	**Heart**	**Oval**
	**Oval**	**Oval**
	**Square**	**Heart**
	**Round**	**Round**

Outline

Workflow

The following flowchart summarizes the overall system workflow:



Spectacle Recommendation

Face Shape	Recommended Sunglasses
Oval	Wayfarer or Clubmaster
Round	Rectangular frames
Square	Round sunglasses
Obong	Oversized square sunglasses
Heart	Cat-eye or Oversized Round

Table: Face Shape and Recommended Styles

Source: Spectacles Recommendation System Based on the Face Shape Using CNN Model, September 2023, (Conference: (NMITCON)) DOI: 10.1109/NMITCON58196.2023.10276139

Detection of Points

- 468 facial landmarks detected using Mediapipe FaceMesh
- Feature extraction from jawline, forehead, and cheekbones
- Lightweight geometric classification
- Faster and suitable for real-time applications

Virtual Try-On Module

- Real-time sunglasses overlay using Mediapipe landmarks
- Detects facial keypoints for smart placement
- Enhances user experience with virtual spectacle trial

Figure: Face Image-Detected Square

Figure: Transparent Sunglass PNG

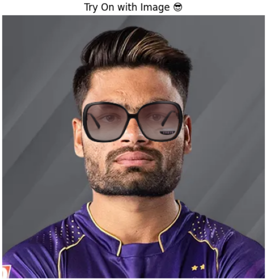


Figure: Try-On Result

Outline

Challenges & Future Work

- Due to low computational resources, we could not use a large dataset (e.g., 1000 images).
- Training the model is slow; running just 1 epoch takes around 1 hour.
- 3D modeling tasks, such as loading `.obj` files, cause the machine to hang.
- We aim to improve the model's performance despite hardware limitations.
- A comparison was made between:
  - CNN model trained from scratch
  - CNN model using transfer learning

Conclusion

1. Face Detection (Bounding Box)
2. Image Augmentation with flip & rotation
3. Pretrained weights from VGG-Face
4. Spectacle Recommendation for different Face shape
5. Low Accurate Model due to unavailability of Resource

Live Demo

Time for Live Demonstration!

References

- Salima, B. V., Chyntiaa, Indrawan, J. O., Hidayat, J., Matthewa, S., Mangkanga, T. A. E., Hasanaa, S., & Permonangana, I. H. (2023). *Face Shape Classification Using Swin Transformer Model.* Procedia Computer Science, 227, 557–562. DOI: 10.1016/j.procs.2023.10.558.
- Tio, A. E. D. C. (2019). *Face Shape Classification using Inception v3.* arXiv preprint arXiv:1911.07916.
- Kim, H.-I., Yun, K., & Ro, Y. M. (2022). *Face Shape-Guided Deep Feature Alignment for Face Recognition Robust to Face Misalignment.* arXiv preprint arXiv:2209.07220.

Thank You!