Face Recognition — Project Plan (v2)

Objective. Implement one classical and one deep face-recognition pipeline on Indian Celeb or CelebA, and evaluate beyond accuracy: robustness (lighting/quality/occlusions), explainability, fairness, and a crowd-recognition test. Deliver a clean demo and a 4–6 page report.

Deliverables (what we'll ship)

- **Two recognizers:** (i) Classical: Eigenfaces or LBP+SVM; (ii) Deep: ArcFace/FaceNet embeddings + classifier (prototype/kNN/SVM).
- **Eval suite:** closed-set ID metrics; (optional) verification (ROC/EER); five robustness buckets with $\Delta\%$; explainability figures; bias gaps; crowd test metrics.
- **Crowd dataset:** 50–100 group images, GT boxes + identity/Unknown, val/test split.
- **Demo:** CLI + notebook that runs on unseen images; "Unknown" thresholding.
- **Report (4–6 pages):** methods, metrics, plots/tables, robustness, explainability, bias, ethics.
- **Reproducibility:** YAML configs, seeds, cached embeddings, experiment logs.

Tech Stack

- **Language/Runtime:** Python 3.10+, PyTorch.
- **Detection/Alignment:** RetinaFace (preferred) or MTCNN; 5-point alignment; 112/160 px crops.
- **Deep Embeddings:** ArcFace (InsightFace) or InceptionResnetV1 (FaceNet). L2-normalized embeddings.
- **Classifiers on embeddings:** cosine-to-class centroid (prototype), kNN, Linear/SVM.
- **Classical:** PCA (Eigenfaces) &/or LBP histograms + χ^2/SVM.
- **Augs/Robustness:** Albumentations (brightness/contrast, noise/blur, JPEG), custom occlusion overlays (eyes/mouth/mask).
- **Analysis:** numpy/pandas, matplotlib; Grad-CAM (torch_cam/captum), LBP heatmaps, eigenfaces.
- **MLOps niceties:** Hydra/YAML configs, deterministic seeds, cached tensors; (optional) Weights & Biases.

Datasets and Splits

- **Identity subset:** 20–40 identities, 20–40 images/ID (balanced where possible).
- **Splits (closed-set identification):** train/val/test per identity; no image overlap.
- **Verification (optional):** generate matched/mismatched pairs from val/test.
- **Bias labels:** CelebA attributes (gender, age band, skin-tone proxy via luminance binning, hair color, makeup). For Indian Celeb, annotate a small subset or use a pretrained attribute tagger.
- **Crowd set:** 50–100 group photos (events/premieres); mix of 0/1/2/3+ known IDs; 20% images with no known ID. Provide `gt.json` with `[x,y,w,h]` and `id/Unknown`.

Pipelines (concise specs)

Classical. Grayscale → face align → (a) PCA: keep k components, cosine/kNN/SVM on projected space; or (b) LBP over 8×8 grid, histogram concat, χ^2 distance or SVM.

Deep. Detect+align → embed (ArcFace/FaceNet) with model-specific normalization → L2-normalize → classify via (1) centroid-cosine, (2) kNN, (3) linear/SVM. *Unknown* via threshold τ tuned on val ROC to maximize F1.

Evaluation Protocol

- **Closed-set ID:** Top-1, precision/recall/F1 (macro/micro), confusion matrix.
- **Verification (optional):** ROC, EER, TPR@FPR.
- **Open-set:** Unknown precision/recall/F1 (val-tuned τ); report threshold.
- **Robustness (5 buckets):**
  - Lighting: brightness/contrast $\in \{\pm 20\%, \pm 40\%\}$.
  - Quality: Gaussian noise ($\sigma \in \{5, 15, 25\}$), blur (k=$\{3, 7, 11\}$), JPEG (q=$\{90, 50, 20\}$).
  - Occlusion: eye bar, mouth mask, full mask rectangles at 15/30/45% face height.
  - Explainability: top eigenfaces, LBP response maps; Grad-CAM on last conv of embedder or on a shallow classification head.
  - Bias: per-group accuracy/EER; report max–min gap & confidence intervals.
- **Crowd test:** On group photos:
  1. Detect all faces;
  2. For each face: embed $\rightarrow$ classify or Unknown(τ);
  3. Metrics: detection P/R/F1 (IoU$\geq$0.5), identification accuracy on matched faces, Unknown P/R; qualitative panels.

Optimization & Quality Controls

- **Alignment first:** 5-point align improves deep & classical accuracy.
- **Caching:** store aligned crops and embeddings (NPZ) to avoid recompute.
- **Efficient search:** FAISS cosine for large galleries; batch embedding; AMP mixed-precision on GPU.
- **Prototype stability:** average 5–20 train embeddings per ID; optionally shrinkage toward global mean.
- **Thresholding:** tune τ on val by maximizing F1 for Unknown; verify on test.
- **Sanity plots:** intra-class vs inter-class cosine histograms; per-ID support vs accuracy.
- **Determinism:** fix seeds; log versions; freeze randomness in dataloaders/augs.

Repo Layout (v2)

```
v2/
  data/ (symlinks or README on how to fetch)
  src/
    detect_align.py    # RetinaFace/MTCNN + 5-pt align
    classical.py       # PCA (Eigenfaces), LBP features, SVM/kNN
    deep.py            # embedding loaders, centroid/kNN/SVM heads
    eval_closedset.py  # ID metrics; (opt) verification pairs
    aug.py             # lighting/quality/occlusion transforms
    explain.py         # eigenfaces, LBP maps, Grad-CAM
    bias.py            # subgrouping + metrics
    crowd.py           # crowd inference + metrics + panels
    demo.py            # CLI/notebook demo
  configs/
    default.yaml       # dataset, model, augs, thresholds, paths
  scripts/
    build_crowd_dataset.py  # (optional) helper to assemble crowd set
  reports/             # figs, tables, paper.tex
```

Milestones

- **:** data subset ready; detection+alignment working; deep embed + centroid baseline (clean Top-1).
- **:** classical (Eigen/LBP) baseline; closed-set metrics and confusion matrices.
- **:** robustness buckets (lighting/quality/occlusion) + Δ% plots.
- **:** explainability figures (eigen/LBP/Grad-CAM).
- **:** bias analysis + subgroup charts; finalize Unknown threshold.
- **:** crowd dataset + evaluation; demo polish; report draft.

Success Criteria

- Clean closed-set Top-1: deep $\geq$ 70% on 20–40 IDs; classical $\geq$ 45%.
- Robustness: quantified Δ% with graceful degradation; occlusion sensitivity characterized.
- Bias: subgroup gaps reported with clear causes & mitigation ideas.
- Crowd: det. F1 $\geq$ 0.7; ID acc. on matched $\geq$ 0.6; strong Unknown precision.
- Reproducible demo and report with at least 8 clear figures.

Ethics & Licensing

Document data sources, licenses (Commons/Flickr CC), consent considerations, bias limitations, and safeguards against misuse (e.g., thresholding Unknown, no surveillance claims).