A Lightweight End-to-End Pipeline for Low-Resolution Facial Recognition

Daniel Szurek Brandon Nguyen

***Abstract*—Low-resolution facial recognition (LRFR) in real-world surveillance and security systems is difficult because security cameras often capture subjects at very small spatial resolutions (e.g., 16×16 or 32×32), which modern face recognition models are not designed to handle. We propose an end-to-end pipeline that first super-resolves low-resolution faces using an identity-aware Deep Super-Resolution (DSR) network, and then recognizes them using EdgeFace, a lightweight backbone optimized for edge deployment. Unlike two-stage approaches that train SR and recognition separately, we use a cycle-style training strategy: first fine-tune EdgeFace on DSR outputs, then retrain DSR with the fine-tuned EdgeFace as the identity supervisor. This co-adaptation helps close the domain gap between super-resolved faces and high-resolution gallery embeddings. On CMU Multi-PIE with 32×32 probes, the pipeline improves rank-1 accuracy over a bicubic+recognition baseline while remaining feasible for embedded devices.**

I. Introduction

Facial recognition is widely used in surveillance, access control, and public safety, but most deployed camera systems do not capture faces at 112×112 (or larger) resolutions expected by current facial recogination models. Cameras placed far from subjects, low-cost sensors, and poor lighting produce very-low-resolution (VLR) faces where high-frequency, identity-discriminative details are lost. Simply upsampling these VLR images with a basic interpolation method like bicubic rarely recovers the information needed for accurate matching, and it often introduces different artifacts.

At the same time, many real deployments target edge platforms, so the recognition model must remain small and fast. This creates a problem: we need to *add* information via super-resolution (SR), but we cannot afford a huge model.

In this project we explore the following idea: if we train SR with the recognition model in the loop, the SR network can learn to reconstruct exactly the features the recognizer needs, not just visually pleasing pixels. Essentially, we pair a Deep Super-Resolution Color (DSR) network with EdgeFace XXS—a lightweight, 1.24M-parameter recognition backbone—and train them in a cycle so that (i) EdgeFace adapts to DSR outputs and (ii) DSR learns what EdgeFace is sensitive to.

Our goals are:

- to build a VLR → HR front-end based on DSR that is *identity-aware*;
- to keep the recognition stage lightweight enough for embedded deployment (EdgeFace);
- to show, via small-scale experiments on CMU Multi-PIE, that cycle-style co-training improves rank-1 accuracy over simple bicubic+recognition baselines.

II. Related Work

A. Super-Resolution for Faces

Classical deep SR models such as SRCNN, VDSR, and EDSR focus on pixel-level or perceptual metrics and are trained on generic images. These approaches improve PSNR/SSIM but do not guarantee preservation of identity-dependent structures. Face-specific SR methods (e.g., FSRNet, SPARNet) inject facial priors such as landmarks or parsing maps to reconstruct plausible facial parts, but they still optimize mostly for appearance, not recognition.

B. Low-Resolution Face Recognition

Earlier LRFR work tried to learn a common space for HR and LR faces or to do coupled SR+FR. More recent papers add identity or feature-matching losses to force the SR output to be recognizable. A recurring problem is a *domain gap*: the recognizer is usually trained on clean, high-res faces, while at test time it gets SR-generated faces that lie slightly off the training manifold.

C. Lightweight/Edge Face Recognition

MobileFaceNet, ShuffleFaceNet, and, more recently, EdgeFace target mobile or edge hardware using depthwise convolutions and compact bottlenecks. These models are a good match for our setting because the SR stage already consumes some compute, so we cannot afford a big recognition backbone.

III. Proposed Approach

Our draft pipeline has four parts.

A. 1) DSR Front-End

We use a residual SR network that:

- takes a 32×32 RGB face crop;
- maps it to a 112-channel feature space;
- passes it through ∼16 residual blocks; and
- upsamples via pixel shuffle to 112×112

The loss is a weighted sum of: pixel L1, multi-scale perceptual (VGG), identity (cosine) loss using EdgeFace embeddings, and an optional feature-matching loss on intermediate EdgeFace layers. The identity term has the highest weight so the SR network does not “hallucinate” pixels that break recognition.

B. 2) EdgeFace XXS Recognition Backbone

We adopt EdgeFace XXS as the recognizer. It takes the 112×112 SR output, produces a 512-D embedding, and is trained/fine-tuned with ArcFace so that embeddings of the same person cluster together while different people are angularly separated.

C. 3) Cycle / Two-Stage Training

If we train DSR against a fixe recognizer, DSR must match that recognizer's feature space, which was learned on HR images. To reduce this mismatch we do:

1) **Cycle 0:** Train DSR using the original EdgeFace XXS to give identity supervision.
2) **Cycle 1:** Fine-tune EdgeFace XXS on DSR outputs using ArcFace, so EdgeFace learns the SR distribution.
3) **Cycle 2:** Retrain DSR, now using the fine-tuned EdgeFace XXS as supervisor.

This loop makes SR and recognition co-adapt, which is the main point we are claiming in the project.

D. 4) Evaluation Plan

For the draft, our plan is:

- **Dataset:** CMU Multi-PIE, using HR frontal images as gallery, and downsampled 32×32 probes as query.
- **Metrics:** rank-1 accuracy, TAR/FAR at a fixed threshold, and (optionally) PSNR/SSIM just to show SR quality.
- **Baselines:** (i) Bicubic $\rightarrow$ EdgeFace, (ii) DSR (trained once) $\rightarrow$ EdgeFace, (iii) our cycle-trained DSR+EdgeFace.
- **Success Criterion:** cycle-trained model beats single-stage DSR by a clear margin.

References

[1] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, 2016.

[2] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPRW*, 2017.

[3] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *CVPR*, 2019.

[4] G. F. Boutros *et al.*, "EdgeFace: Efficient face recognition model for edge devices," 2023.

[5] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image and Vision Computing*, 2010.