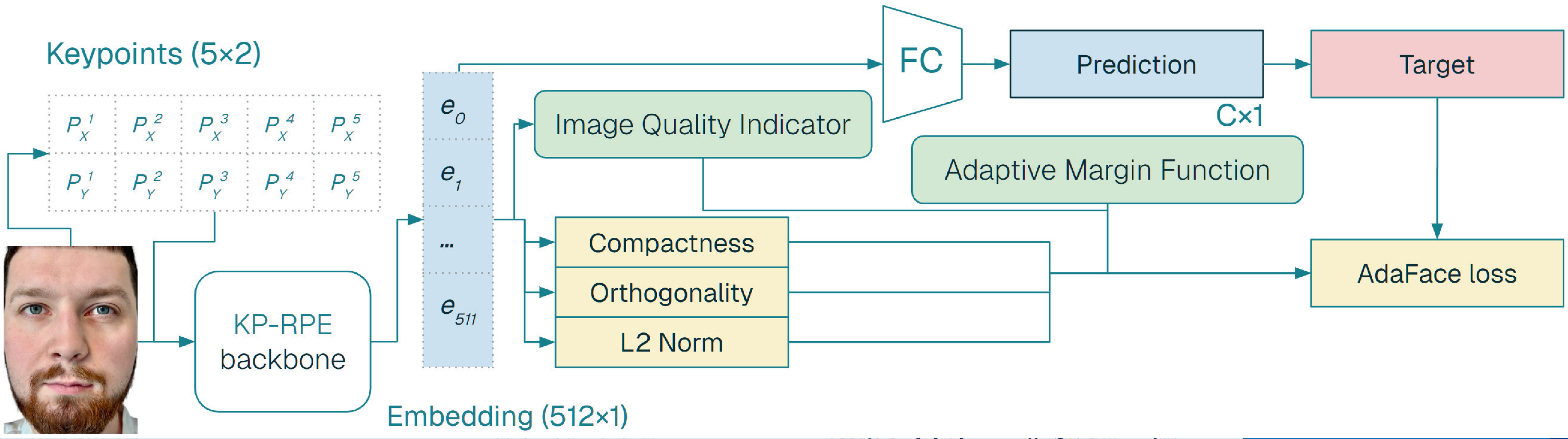
FACECLUSTER: INTERACTIVE PHOTO ORGANIZATION WITH ENHANCED FACE RECOGNITION

Alexander Filonenko, Ilya Makarov, Andrey V. Savchenko

fall into ML 2025
4th conference on machine learning & AI

IREX AIRI ISP RAS ITMO SBER AI Lab



Abstract

FaceCluster is an interactive photo management system that leverages our enhanced KP-RPE face recognition model with Embedding Statistical Regularization to organize personal photo collections automatically. Unlike existing cloud-based systems that raise privacy concerns, FaceCluster operates entirely locally while demonstrating high performance across multiple challenging benchmarks, including IJB-C (97.25% TAR@0.01%), TinyFace (74.14% Rank-1), and AgeDB (97.78% accuracy) when trained on the WebFace4M dataset. The demo showcases real-time face detection, clustering, and organization capabilities through an intuitive web interface, enabling users to effortlessly manage large photo collections with a single-command Docker deployment.

Methodology

The approach enhances the KP-RPE transformer backbone with Embedding Statistical Regularization by applying three direct constraints during training: compactness loss to tightly cluster embeddings of the same identity, orthogonality loss to improve feature disentanglement, and L2 regularization to control embedding magnitude. Combined with AdaFace loss for quality-adaptive margins, these regularization components (weights: 1.0 for compactness, 0.5 for orthogonality, 0.01 for L2) optimize the 512-dimensional embedding space to create more discriminative representations without adding inference overhead.

$$\mathcal{L}_{\text{compact}} = \frac{1}{|S_c|}\sum_{i\in S_c}\|\mathbf{e}_i - \boldsymbol{\mu}_c\|_2^2$$

$$\mathcal{L}_{\text{ortho}} = \frac{1}{n(n-1)}\sum_{i=1}^{n}\sum_{j=1, j\neq i}^{n}(\mathbf{e}_i^T\mathbf{e}_j)^2$$

$$\mathcal{L}_{L2} = \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{e}_i\|_2$$

Demo Application

FaceCluster implements a microservices architecture with a FastAPI backend, a React frontend, and Redis caching. The enhanced KP-RPE model generates embeddings resulting in high clustering performance without additional inference costs.

The system utilizes GPU-accelerated ONNX models with automatic CPU fallback. FaceCluster demonstrates effortless deployment through Docker Compose, orchestrating a FastAPI backend, a React frontend, a Redis cache, and a SQLite database.

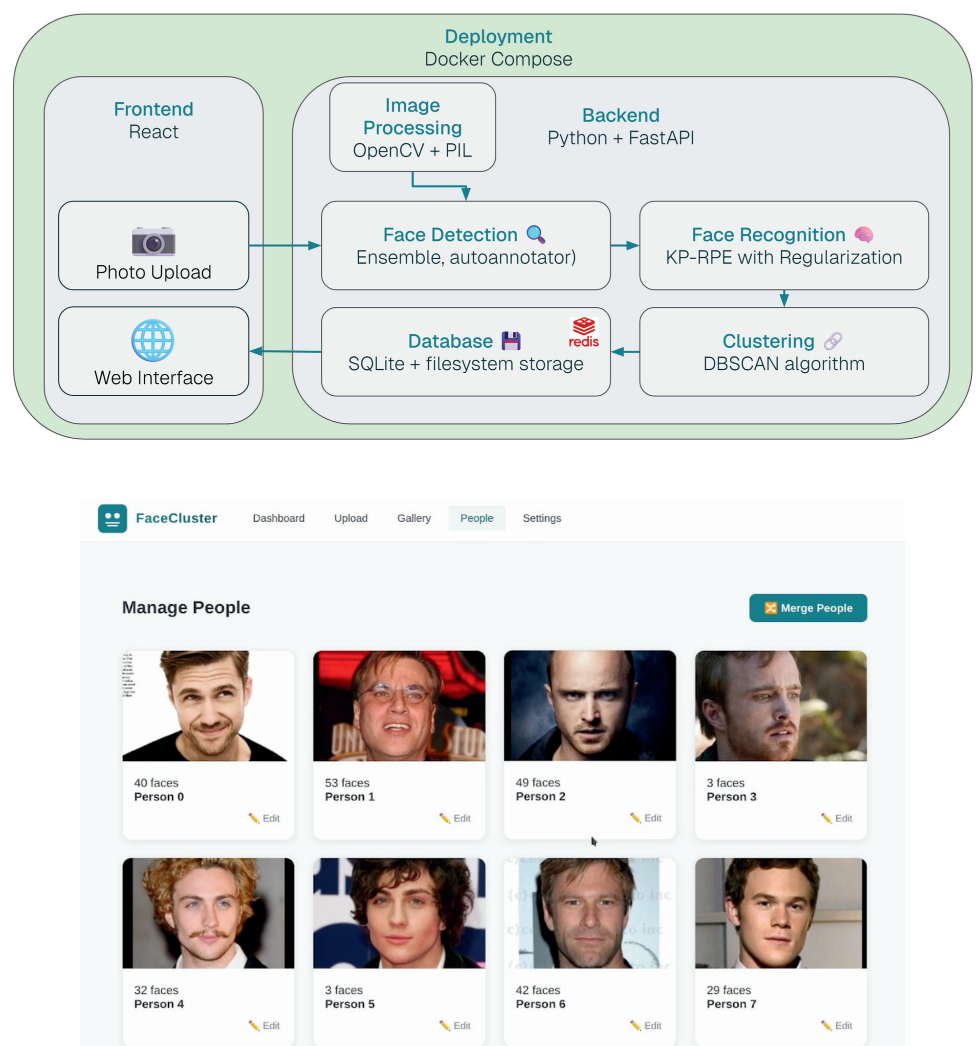
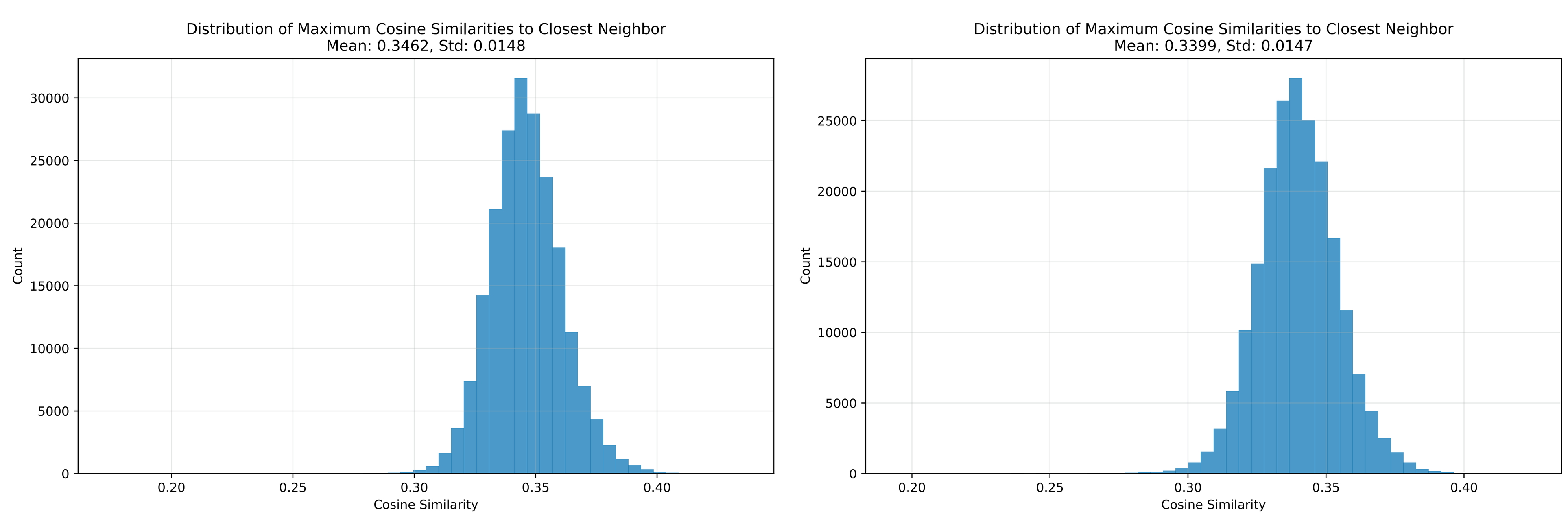


Figure 1: Deployment scheme and sample UI of our demo application

Results

The enhanced KP-RPE model with Embedding Statistical Regularization trained on WebFace4M demonstrates superior performance across multiple challenging benchmarks including low-resolution faces, age variations, and pose diversity. The system reliably groups images of the same person across decades from young adulthood to old age, demonstrating robust performance for real-world photo organization without adding inference overhead.



(a) (b)

Figure 2: Cosine similarities of face embeddings to the closest neighbor of (a) baseline and (b) our model.

Table 1. Comparison to KP-RPE ViT-Small baseline trained on WebFace4M

Method	TinyFace	AgeDB	CFP-FP	IJB-C	IJB-B	CPLFW	CALFW
	Rank-1	*Acc.*	*Acc.*	*TAR@0.01%*	*TAR@0.01%*	*Acc.*	*Acc.*
KP-RPE reported in [1]	69.88	-	96.60	94.20	-	-	-
KP-RPE + Regularization	74.14	97.78	98.94	97.25	95.75	95.33	96.08

This paper was accepted at ACM Multimedia 2025 demo track

alexander@filonenko.net