Enhancing face recognition in native low-resolution images using deep learning and bicubic interpolation

Ngoan Chi Le[1*]

[1]FPT University, Ho Chi Minh City, Vietnam
[*]Corresponding author: ngoan22mse23088@fsb.edu.vn

ARTICLE INFO

DOI:10.46223/HCMCOUJS.tech.en.15.1.4017.2025

Received: February 10th, 2025

Revised: March 17th, 2025

Accepted: March 28th, 2025

Keywords:

face recognition; low-resolution; super-resolution-bilinear and bicubic image

ABSTRACT

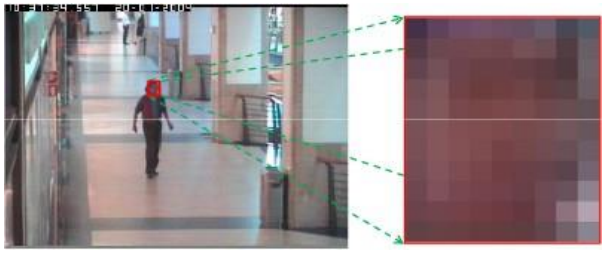
This paper addresses the challenge of face recognition in Low-Resolution (LR) images, mainly when the resolution is below 48x48 pixels, which is common in surveillance systems. Current face recognition algorithms struggle to deliver satisfactory results with such low-resolution images. This study utilizes over 16,000 face images with an average resolution of 20x20 pixels to improve recognition, applying deep learning and bicubic interpolation to enhance image resolution. Unlike traditional Super-Resolution (SR) methods that operate in the LR space, our approach introduces a novel data constraint that evaluates errors in the High-Resolution (HR) image domain. By leveraging the finer details in HR images, the reconstructed HR images significantly improve visual quality and recognition accuracy. This unique data constraint seamlessly incorporates discriminative features into the optimization process. Experimental results demonstrate that our method outperforms existing visual quality and recognition performance approaches.

1. Introduction

The increasing deployment of surveillance cameras in public areas, ranging from small bank installations to expansive CCTV networks in urban streets, has increased the demand for advanced face recognition. Wide-angle cameras maximize coverage but often capture tiny face regions, sometimes fewer than a hundred pixels, when individuals are distant (Figure 1). This scenario poses a significant challenge, low-resolution face recognition, where the limited pixel count hinders conventional face recognition algorithms designed for more substantial face regions. Addressing the recognition challenge of Low-Resolution (LR) face images; this problem arises when the face region is tiny (Zou & Yuen, 2022). Despite over three decades of face recognition paper and the development of promising systems, these efforts assume a sufficiently large face region, posing limitations in the context of LR scenarios.

Figure 1

A Typical Frame from A Surveillance Footage



Source. The data are from "Very low-resolution face recognition problem" by W. W. W. Zou and P. C. Yuen, 2022, *IEEE Transactions on Image Processing*, *21*(1), pp. 327-340 (https://doi.org/10.1109/TIP.2021.2162423)

Image Super-Resolution (SR) refers to a technique aimed at improving the resolution of an image by generating a High-Resolution (HR) version from a Lower-Resolution (LR) input (Kim et al., 2023). It involves image interpolation, scaling, up-sampling, zooming, or enlarging (Khaledyan et al., 2020). The main objective of image super-resolution is to increase pixel density and enhance the finer details in LR images, which is particularly beneficial for applications that demand precise recognition or detection. Before the rise of deep learning, traditional image-processing methods were primarily used for SR. In the past, image super-resolution was largely dependent on conventional image processing techniques before deep learning methods emerged. Li et al. (2020) and Wang et al. (2021) provided a comprehensive classification of SR techniques into three main categories: interpolation-based, reconstruction-based, and learning-based methods.

Face Recognition (FR) models excel in controlled environments. However, our study demonstrates a noteworthy challenge for FR when applied to native unconstrained LR face images, distinct from artificially down-sampled ones. These LR images, typically from surveillance videos (Cheng et al., 2023) and unposed snapshots, are a significant hurdle. Testing against a tiny face dataset, we observe a notable performance degradation in current deep-learning FR models. The inherent lack of visual details in Low-Resolution (LR) facial images hinders these models from learning rich feature representations, particularly compared to high-resolution photos (Figure 2).

Figure 2

On the Left, Constrained High-Resolution Web Face Images from Five Well-Known Benchmarking Face Recognition Datasets Are Shown, while on the right, Native Unconstrained Low- Resolution Web Face Images Captured in Typical Natural Settings Are Displayed

Source. The data are from "Low-resolution face recognition" by Z. Cheng, X. Zhu, and S. Gong, 2023 (http://arxiv.org/abs/1811.08965)

Face recognition has significantly progressed with deep Convolutional Neural Networks (CNNs) and powerful loss functions such as large-margin Cosine loss and additive angular margin loss. However, models trained on large High-Resolution (HR) face datasets often struggle to generalize well to realistic Low-Resolution (LR) face images captured by real-world surveillance cameras, particularly from far distances.

While face recognition methods have advanced, they are still ineffective in addressing the challenges of low-resolution face images, especially in surveillance settings. Current models perform well on high-resolution photos but struggle with native, unconstrained LR images from real-world surveillance systems.

This paper introduces a novel method with a unique data constraint that evaluates errors in the High-Resolution (HR) space instead of the LR space. This approach improves image

quality and recognition accuracy while seamlessly integrating discriminative constraints into the optimization process. This method is essential for overcoming the lack of details in LR images.

Table 1

The Table Compares Existing Methods and Presents the Proposed Approach, which Improves HR Image Quality Using a Novel Data Constraint

Method	Description	Limitations
Traditional Image Processing Methods (Before Deep Learning)	Image Super-Resolution (SR) methods like interpolation, scaling, and enlarging to improve the resolution of LR images.	Limited in improving visual details in very low-resolution images. Struggles to capture fine details.
Interpolation-based Methods	Image interpolation techniques are used to upscale LR images to HR images.	Often, it results in blurry or less detailed images, especially in very low-resolution faces.
Reconstruction-based Methods	Focus on reconstructing high-resolution images using prior knowledge of image structures.	It can be computationally expensive and may not handle significant variations in facial poses or expressions.
Learning-based Methods (Deep Learning)	Deep learning-based SR techniques, often using Convolutional Neural Networks (CNNs), to improve resolution.	Struggles with native, unconstrained LR images (such as those from surveillance footage).
Fine-tuned CNNs on LR Faces	Fine-tuning pre-trained CNN models using LR face images to adapt to real-world conditions.	The limited availability of labeled LR face data reduces the effectiveness of this approach.
Proposed Method (Novel Super-Resolution)	Introduces a novel data constraint for SR, improving image quality by evaluating errors.	It still requires further evaluation on generalization and practical deployment in real-world systems.

Source. The values are collected and compiled from multiple sources. Data by author

2. Related works

Existing approaches for Very Low Resolution/Low Resolution (VLR/LR) recognition commonly leverage High-Resolution information to enhance classification models, falling into three categories: (i) image-level, (ii) feature-level, or (iii) classifier-level. Image-level techniques involve super-resolution or synthesis, initially addressing visual quality and later evolving to optimize recognition performance and visual quality.

Derived-Margin Softmax Loss

As highlighted by Singh et al. (2022), the softmax function, commonly used in deep learning models for classification tasks and cross-entropy loss, plays a crucial role in this paper. In line with recent research trends, the softmax loss is defined here as a combination of the softmax function and cross-entropy loss applied to the final fully-connected layer. This loss aims to train a discriminative classifier that assigns higher scores to the correct input class.

Typically, class scores correspond to the activations of the final fully connected layer, denoted in previous works. Liu et al. (2021) represented this as a Cosine function. The updated form of the softmax loss is given as follows:

$$\mathcal{L}_{S_{\text{oftmax}}} = \frac{1}{N}\sum_{i=1}^{N} -\log\frac{\exp^{|W_{\text{v}_i}|\|x_i|\cos\theta_{v_6}}}{\sum_{j=1}^{C}\exp^{|W_f\|\nabla_d\|\cos\theta_j}} \quad (1)$$

Equation (1) The Softmax loss is formulated (Singh et al., 2022).

A. Reconstruction-Center Loss

The embeddings extracted from the input data are passed into the reconstruction and classification modules (Figure 3). The reconstruction module, a compact, fully connected network, generates a high-resolution image from a given embedding. This reconstruction space captures intra-class similarities and computes class-specific margins for the Derived-Margin softmax loss. The ReCent loss is applied to the reconstruction module to reduce the variability in class-specific reconstructions. For a feature xi and its corresponding HR image (hri), the ReCent loss is defined as:

$$\mathcal{L}_{ReCent} = \|hr_i - g(x_i)\|_2^2 + 8g(x_i) - C\,8_{y_i}{}_2^2 \quad (2)$$

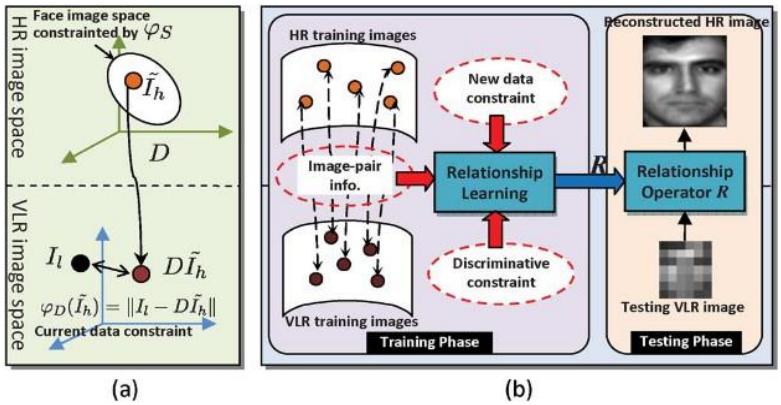
Equation (2) The ReCent loss is formulated (Singh et al., 2022).

B. Low-Resolution Face Recognition

Existing LRFR methods fall into two categories: Image super-resolution and resolution-invariant learning (Zou & Yuen, 2022). The first emphasizes pixel-level fidelity and identity discrimination, while the second focuses on resolution-invariant features or cross-resolution structure transformation. Limitations include small gallery sizes, artificial down-sampling, reliance on handcrafted features or lack of end-to-end optimization in deep learning, and the assumption of labeled LR/HR pairs for training impractical with native LR faces. LRFR deployment occurs in two settings: LR-to-HR, matching LR probes to HR gallery images, and LR-to-LR, where both probe and gallery are LR faces. The TinyFace benchmark adopts the more general LR-to-LR setting.

Figure 3

Two Novel Constraints Are Introduced to Learn the Relationship Mapping: (A) The New Data Constraint and (B) The Discriminative Constraint



Source. The data are from "Very low-resolution face recognition problem" by W. W. W. Zou and P. C. Yuen, 2022, *IEEE Transactions on Image Processing, 21*(1), pp. 327-340 (https://doi.org/10.1109/TIP.2021.2162423)

To solve the Very Low-Resolution (VLR) face recognition challenge, we introduce an RLSR algorithm, as depicted in Figure 3. The approach includes a training phase in which the relationship between High-Resolution (HR) and VLR image spaces is learned. This learned relationship is subsequently used during the testing phase to reconstruct HR images. The training phase involves two key steps: linearity clustering and relationship learning.

A clustering algorithm is employed as a preprocessing step in the initial step. The VLR-HR image pairs within each cluster exhibit near linearity, allowing their relationship to be approximated by a matrix. Subsequently, a linear regression model is employed in the second step to learn the relationship mapping, incorporating different constraints tailored for diverse applications.

Figure 3(a) introduces a novel data constraint designed for human-based recognition. This constraint utilizes relationship mapping to project VLR images into the HR image space, measuring reconstruction error. This contrasts with existing algorithms that gauge error within the VLR space. Conversely, Figure 3(b) illustrates a discriminative constraint developed for machine-based recognition. These constraints enhance the adaptability of the Robust Low-Rank and Sparse (RLSR) algorithm across different recognition applications.

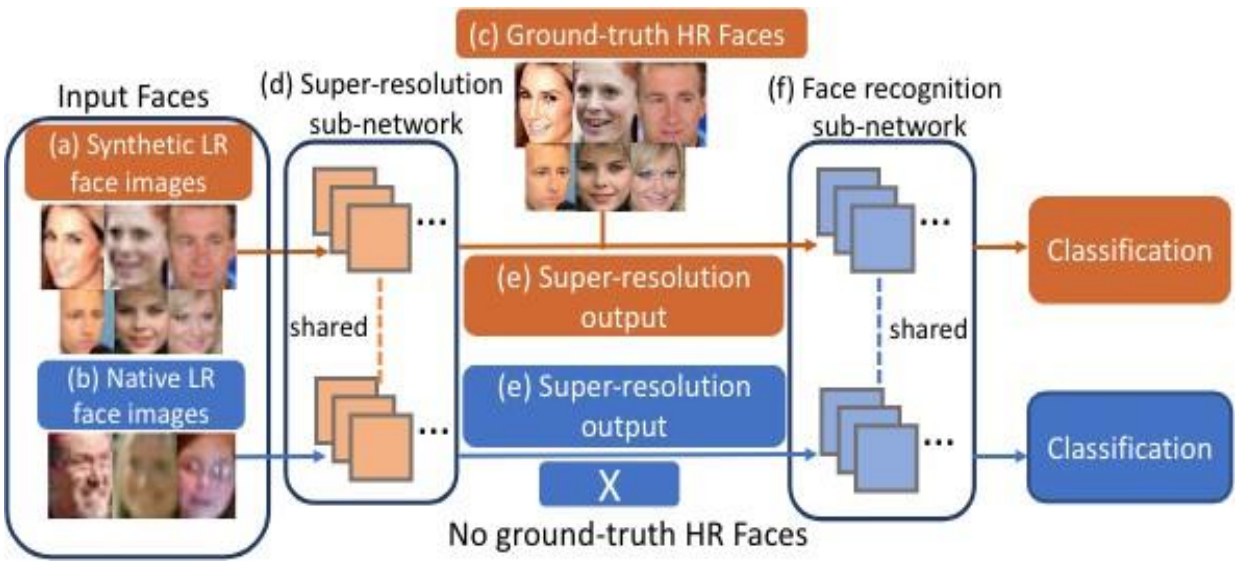
C. Complement-Super-Resolution

In the CSRI architecture (Figure 4), A multi-branch network is utilized, consisting of two branches in the CSRI:

Synthetic LR SR-FR branch: This branch improves the compatibility and complementary benefits of Super-Resolution (SR) and Face Recognition (FR) by jointly learning from auxiliary face data with artificially down-sampled LR/HR image pairs (top stream in Figure 4). Native LR SR-FR branch: This branch adapts the super-resolved information from auxiliary LR/HR face pairs to the native LR facial image domain, where corresponding HR images are unavailable, thus enhancing the SR-FR learning process (bottom stream in Figure 4).

Figure 4

Provides an Overview of the Proposed Complement-Super-Resolution and Identity (CSRI) Joint Learning Framework



Source. The data are from "Low-resolution face recognition" by Z. Cheng, X. Zhu, and S. Gong, 2023 (http://arxiv.org/abs/1811.08965)

The CSRI is implemented using VDSR for the Super-Resolution (SR) component (Cheng et al., 2023) and a separate model for the Face Recognition (FR) component. The joint learning approach combines the SR output with the FR input, taking advantage of end-to-end deep learning. The training data consists of artificially down-sampled LR/HR face pairs {(Ialr, Iahr)} and face identity labels {y} (e.g.). The SR model is optimized by minimizing the Mean-Squared Error (MSE) for the SR component. The CSRI utilizes a joint learning strategy to integrate SR and Face Recognition, combining the SR output with the FR input to use end-to-end deep learning fully. The SR model functions as a non-linear mapping between LR and HR face images, with optimization achieved through pixel-level MSE minimization.

The CSRI model integrates SR-FR joint learning and is optimized for Face Recognition (FR) performance using synthetic auxiliary LR face data. However, this approach may not be ideal for native LRFR due to differences in the visual appearance distributions (Figure 5). To overcome this, we improve the SR-FR joint learning by incorporating a native LR face discrimination constraint into the SR optimization process. We simultaneously optimize SR and FR components using training data from both auxiliary (with LR/HR paired images) and native (with only LR images) sources, adapting the SR learning to suit native LR data better. The discrimination of labeled native LR faces is enforced through Cross-Entropy loss.

Figure 5

A Comparison Between (Left) Native LR Face Images from Tinyface and (Right) Synthetic LR Face Images Produced by Synlr-MF2

Source. The data are from "Deep learning algorithms for single image super-resolution: A systematic review" by Y. K. Ooi and H. Ibrahim, 2021, *Electronics, 10*(7), pp. 327-340 (https://doi.org/10.3390/electronics10070867)

3. The approach

A. Bilinear Interpolation

Bilinear interpolation is one of the most widely used techniques in image processing because of its simplicity in arithmetic operations (Zhang et al., 2021). This approach combines the values of the four closest pixels using separable linear interpolation, as shown in Figure 6. The coefficients are computed based on the horizontal and vertical distances from the neighboring pixels. The interpolated pixel's final value is calculated using Equation (3):

$$P = P_{11}(1-dy)(1-dx) + P_{12}(1-dy)dx \\ \mathrm{P}_{21}dy(1-dx) + \mathrm{P}_{22}dxdy \quad (3)$$

Equation (3) The final value of the interpolated pixel is determined (Wang et al., 2021).

The proposed architecture for bilinear interpolation only requires a single line buffer, as the bilinear method utilizes four neighboring pixels. Figure 6 shows the architecture of the sliding

window for bilinear interpolation. After the initial step of supplying the necessary pixels, these are multiplied by the corresponding coefficients. Using exact multiplication would require numerous multipliers. This paper avoids incorporating a multiplier block for the interpolation process. Instead, to optimize hardware resource usage and maintain accuracy, we propose an approximate method for both bicubic and bilinear interpolation, which is better suited for computational systems with constrained memory, such as FPGAs and DSPs (Sun, Wang, et al., 2022).

Figure 6

The Final Outcomes: (A) Input Image, (B) Outcome of The Bilinear Interpolation

Source. The data are from "Multidimensional scaling for matching low-resolution face images" by S. Biswas, K. W. Bowyer, and P. J. Flynn, 2022, IEEE Computer Society

Table 2

A Comparison between the Proposed Architecture and Recent Interpolation Techniques

Architectures	Image size	Interpolation algorithm	Implementation platform	Frequency (Mhz)	Slice LUs	Slice registers
Proposed in [14]	640*480	Linear	Virtex-2	104.3	NA	NA
Proposed in [25]	2,560*1,920	Cubic	Virtex-4	130.0	NA	NA
Proposed in [11]	2,560*1,920	Cubic	Virtex-6	75.0	7,900	7,843
Proposed in [23]	256*256	Cubic	Artix-7	100.0	5,293	5,432
Bi-linear	256*256	Linear	Artix-7	314.8	97	44
Bicubic	256*256	Cubic	Artix-7	280.9	359	162

Source. Values collected and compiled from multiple sources. Data generated by the author

Interpolation plays a crucial role in super-resolution techniques, tracking systems, robotics, online videos, mobile apps, and, most importantly, security applications such as surveillance cameras.

B. Face Recognition in Low-Quality Face Images

Face recognition, whether performed by machines or humans, is far from flawless when it comes to low-quality face images. This section provides a general overview of the challenges in face recognition with low-quality photos. We discuss the differences between Low-Quality (LQ) and High-Quality (HQ) face recognition, define Low-Resolution (LR) face images, summarize existing studies on human performance in Low-Quality Face Recognition (LQFR) tasks, and highlight key challenges identified in related LQFR research.

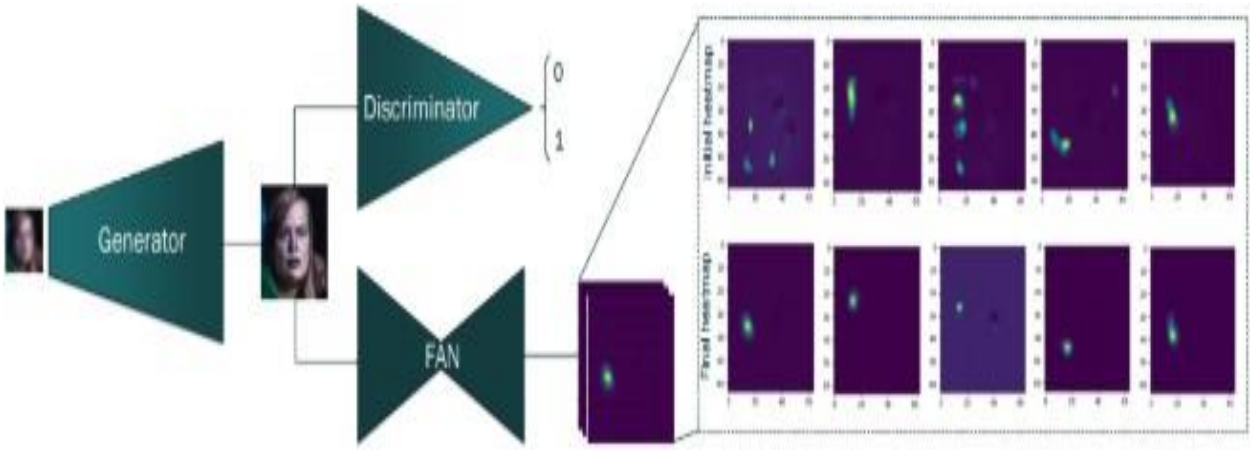
Super-Resolution: We present two main types of approaches to super-resolution processing. One aims to reconstruct visually high-quality images, while the other uses explicitly super-resolution techniques for face recognition. Depending on whether prior information is utilized, these methods can be classified as either reconstruction-based or learning-based. Learning-based methods yield superior results and have become the dominant approach in recent research. Xu et al. discovered that when input faces exceed 48×48 pixels, super-resolved images are more effectively recognized than their low-resolution counterparts.

Convolution Neural Networks: The entire process, consisting of patch extraction, non-linear mapping, and reconstruction, is structured as a convolutional neural network. CNNs are particularly suited for this task due to their ability to learn hierarchical representations of data automatically. The architecture is designed to handle the transformation from low- low-low-resolution to high-resolution images effectively, leveraging learned features and mappings to enhance spatial details and fidelity.

Deblurring: Unlike developing super-resolution algorithms or Low-Resolution (LR) robust features, another approach emphasizes learning a shared space incorporating LR probe and HR gallery images. During the testing phase, LR probes and HR gallery images are mapped into this learned space to measure their similarity. The nature of these spaces can be either linear or non-linear, depending on the model's design.

Figure 7

Super-FAN Network Architecture



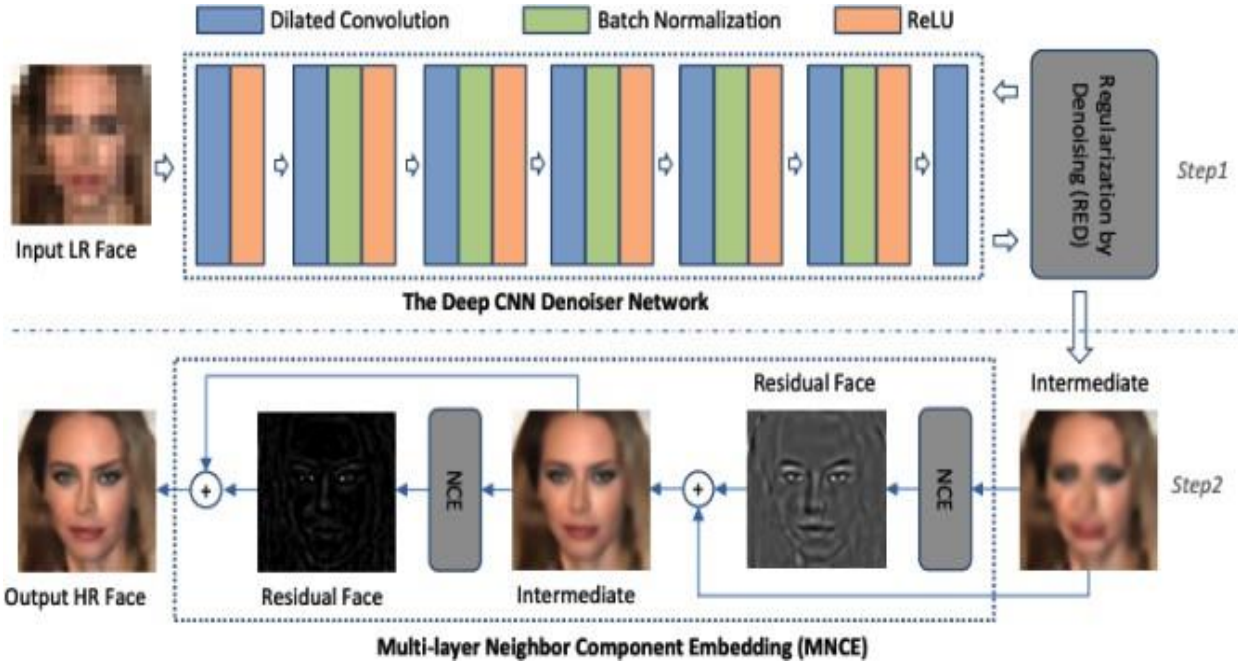
Source. The data are from "Deep learning face representation by joint identification-verification" by Y. Sun, X. Wang, and X. Tang, 2022, *Neural Information Processing Systems, 27*(7), pp. 1488-1496

These advancements highlight ongoing efforts to improve face techniques by integrating sophisticated algorithms that enhance image quality and robustness in various applications. Recent advancements in SR have embraced deep learning and GAN models as practical tools.

Huang introduced a state-of-the-art approach inspired by traditional wavelets, renowned for capturing contextual and textural information across various levels of an image. Their deep architecture includes three main components: feature map extraction, wavelet transformation, and coefficient prediction, facilitating the reconstruction of High-Resolution (HR) images based on predicted wavelet coefficients.

Figure 8

Super-FAN Network Architecture Multi-Layer Neighbor Component Embedding



Source. The data are from "Deepid3: Face recognition with intense neural networks" by Y. Sun, D. Liang, X. Wang, and X. Tang, 2022 (https://arxiv.org/abs/1404.7864)

Super-resolution-aided LQFR. Xu (2021) integrated the methodologies of Yang and Wang (2020), employing dictionary learning for patch-based sparse representation and combining information across sequences of super-resolved face images.

Uiboupin et al. (2022) introduced sparse representation learning via dictionary learning, employing two distinct dictionaries: one trained on natural and facial images and the other solely on facial images to enhance realism in their outputs. They incorporated a Hidden Markov. This approach involved a deblurring phase to significantly improve recognition performance on the Superfaces dataset, increasing accuracy from 50% to 80%. Canonical Correlation Analysis (CCA) has proven effective in manifold learning for recognition-oriented face super-resolution. However, traditional 1D CCA is not inherently suited for image data, requiring image vectorization that may obscure appearance details. An introduced a 2D CCA method that directly explores relationships between image sets without vectorization. Their approach involved face reconstruction and detail compensation, achieving over 30% improvement on a hybrid dataset derived from CAS-PEAL- R1 and CUHK datasets. Similarly, Pong proposed a method using cascaded Generalized CCA (GCCA) to fuse Gabor features from multiple image resolutions into a single feature vector for enhanced face recognition. Jia applied CCA to establish coherent subspaces for HR and LR PCA features, utilizing adaptive Pixel-wise Kernel Partial Least Squares (P-KPLS) to hallucinate LR PCA features fused with LBP features for final representation. Satiro employed motion estimation on interpolated LR images, integrating parameters into a non-local mean-based SR algorithm to

produce higher-quality images. Zou noted that traditional example-based or map-based super-resolution methods often generate LR face images by minimizing reconstruction errors in the LR space, formulated with an objective function (1), where Ih and Il denote HR and LR images respectively, and D represents the dictionary.

$$\| \mathbf{DI}_h - \mathbf{I}_l \|^2 \rightarrow \min \tag{4}$$

Equation (4) Minimizing the L2 norm of the difference between DI_h and I_l (Anwar et al., 2021).

When used for the Low-Quality Face Recognition (LQFR) task, these methods typically produce High-Resolution (HR) faces that resemble the Low-Resolution (LR) faces. However, the similarity measured in the LR face space lacks sufficient information to accurately reflect the intra-class similarity or inter-class saliency in the HR face space, leading to significant artifacts. Zou and Yuen (2022) tackled this issue by shifting the reconstruction task from the LR face space to the HR face space (2) with the use of a regressor R:

$$\| \mathbf{RI}_l - \mathbf{I}_k \|^2 \tag{5}$$

Equation (5) The L2 difference norm between RI_l and I_k (Anwar et al., 2021).

$$s'_h = \frac{s_h}{w_h^1} \text{ and } s'_l = \frac{s_l}{w_l^1} \tag{6}$$

Equation (6) The scale-invariant feature vectors (Anwar et al., 2021).

Using scale-invariant feature vectors, the proposed method integrates class information by learning two mapping matrices from HR-LR image pairs within the same class. These matrices are then applied to reconstruct HR faces. Shekhar et al. proposed a generative approach for Low-Quality Face Recognition (LQFR), employing an image relighting technique for data augmentation to improve robustness against illumination changes.

In their method, class-specific Low-Resolution (LR) dictionaries are learned using K-SVD and sparse constraints to minimize reconstruction errors in the LR domain. During testing, LR probe images are projected onto the span of atoms from the learned dictionary using an orthogonal projector, and a residual vector for each class is calculated for identity classification. The authors extended their general dictionary learning approach into a non-linear space by incorporating kernel functions.

Optimization Criterion: The Mean-Squared Error (MSE) criterion at the pixel level is employed to optimize the SR component. The MSE measures the average squared difference between the predicted and ground truth HR images. Minimizing this error ensures that the SR model effectively learns to reconstruct HR details from LR inputs.

$$\mathcal{L}_{\text{sr}} = 8I^{\text{asr}} - I^{\text{abr}}8_2^2 \tag{7}$$

Equation (7) The optimize the SR component (Biswas, 2022).

Where I^{asr} denotes the super-resolved face image of I^{alr} (Figure. 2(a)), and I ahr denotes the corresponding HR ground-truth image (Figure. 2(c)). Using the MSE loss intrinsically favors the Peak Signal-to-Noise Ratio (PSNR) measurement rather than the desired LRFR performance. We address this limitation by concurrently imposing the FR criterion when optimizing SR.

Formally, we quantify the performance of the FR component by the SoftMax Cross-Entropy loss function defined as:

$$\mathcal{L}^{\text{syn}} = -\log \text{A}p_y\text{C} \tag{8}$$

Equation (8) The SoftMax Cross-Entropy loss (Biswas, 2022).

Where y is the face identity, and py is the FR component's prediction probability for classy. The SR-FR joint learning objective is then formulated as follows:

$$\mathcal{L}_{\text{sr-fr}} = \mathcal{L}^{\text{syn}} + \lambda_{\text{sr}}\mathcal{L}_{\text{sr}} \tag{9}$$

Equation (9) The SR-FR joint learning objective (Biswas, 2022).

In our experiments, where λsr is a weighting parameter for the SR loss quantity, we set λsr = 0.003 by cross-validation. In doing so, the FR criterion enforces the SR learning to be identity discriminative simultaneously.

C. Overall, Loss Function, Model Training, and Deployment

$$\mathcal{L}_{\text{csrl}} = \text{A}\mathcal{L}^{\text{syn}} + \mathcal{L}^{\text{nat}}\text{C} + \lambda_{\text{sr}}\mathcal{L}_{\text{sr}} \tag{10}$$

Equation (10) Overall loss function (Biswas, 2022).

Where $\mathcal{L}_{\text{fr}}^{\text{nat}}$ and $\mathcal{L}_{\text{fr}}^{\text{syn}}$ measure the identity discrimination constraints on the native and synthetic LR training data, respectively. With such a joint multitask (FR and SR) formulation, the SR optimization is specifically guided to be more discriminative for the native LR facial imagery data.

The CSRI can be trained end-to-end using the standard Stochastic Gradient Descent algorithm. Due to significant size imbalances between the auxiliary and native LR data sets, we propose a two-step training approach to enhance model convergence stability: first, pre-training the synthetic LR SR-FR branch on the sizeable auxiliary face dataset (CelebA), and second, training the entire CSRI network on both the auxiliary and native LR data sets. During deployment, we utilize the native LR SR-FR branch to extract feature vectors for face image matching using the Euclidean distance metric.

Figure 9

Example Tinyface Images Auto-Detected in Unconstrained Images

Source. The data are from "TinyFace dataset for face detection" by J. Zhang, 2018 (https://qmul-tinyface.github.io/). The data are from "Deepid3: Face recognition with very deep neural networks" by Y. Sun, D. Liang, X. Wang, and X. Tang, 2022 (https://arxiv.org/abs/1404.7864)

Convolution Neural Networks: The described approach outlines a Convolutional Neural Network (CNN) architecture for single-image Super-Resolution (SR). Let's break down each operation as detailed:

Patch extraction and representation:

Operation: This step involves extracting overlapping patches from the low-resolution image Y and representing each patch as a high-dimensional vector. These vectors form a set of feature maps, where each vector's dimensionality corresponds to the extracted patch's content.

Purpose: By capturing overlapping patches, the network can learn local details and structures from the low-resolution image, essential for reconstructing high-resolution information.

Non-linear mapping:

Operation: Each high-dimensional vector obtained from the previous step undergoes a non-linear mapping to transform it into another high-dimensional vector. This mapping is typically achieved through layers of convolutional, activation, and possibly normalization operations within the CNN.

Purpose: Non-linear mappings allow the network to learn complex relationships between low-resolution features and their corresponding high-resolution counterparts. This helps in capturing intricate details and textures that are lost during downscaling.

Reconstruction:

Operation: The mapped high-resolution patch-wise representations are aggregated to generate the final high-resolution image F(Y), which aims to be visually similar to the ground truth high-resolution image X.

Purpose: By aggregating the mapped representations, the network synthesizes a high-resolution image that maximizes similarity to the actual high-resolution image X. This step involves combining the learned features to respect spatial coherence and structural integrity.

This CNN-based approach leverages deep learning techniques to achieve single-image super-resolution by systematically processing the low-resolution image Y through patch extraction, non-linear mappings, and reconstruction. Each operation within the network enhances the resolution and visual quality of the output image F(Y), aligning it closely with the ground truth X. In image restoration, a common strategy involves densely extracting patches from an image and representing them using a set of pre-trained bases such as PCA (Principal Component Analysis), DCT (Discrete Cosine Transform), Haar wavelets, etc. This approach can be seen as convolving the image with a set of filters, where each filter corresponds to a basis vector in the transformed domain. These bases' optimization is integrated into a neural network's optimization process in the described formulation. Formally, the first layer, F1, in this network can be expressed as follows:

$$F1(Y)=\sigma\ (W1 * Y + b1) \tag{11}$$

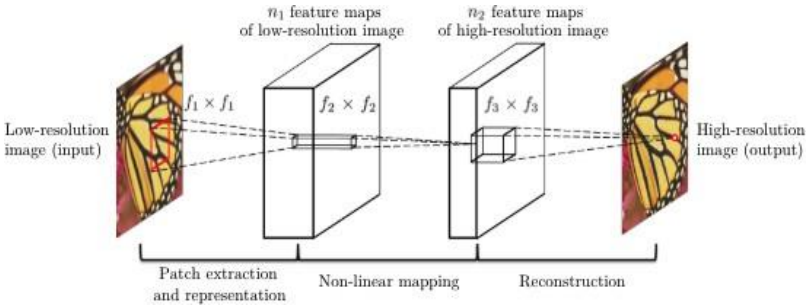
Equation (11) The described formulation (Uiboupin et al., 2022)

W1 and B1 represent the filters and biases, respectively, and '$*$' denotes the convolution operation. Here, W1 corresponds to n1 filters of support $c \times f1 \times f1$, where c is the number of channels in the input image, and f1 is the spatial size of a filter. Intuitively, W1 applies n1 convolutions on the image, and each convolution has a kernel size $c \times f1 \times f1$. The output is composed of n1 feature maps. B1 is an n1-dimensional vector whose each element is associated

with a filter. We apply the Rectified Linear Unit (ReLU, max(0, x)) (Uiboupin et al., 2022) on the filter responses.

Figure 10

Shows The SRCNN Process: Extract Features from Image Y, Transform Them into Resolution Patches, and Combine Them into The Final High-Resolution Image F(Y)



Source. The data are from "Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring" by T. Uiboupin, P. Rasti, G. Anbarjafari, and H. Demirel, 2022, in *IEEE Conference on Signal Processing and Communication Application*, pp. 437-440. IEEE

Combining the aforementioned operations results in a convolutional layer structure within the SRCNN, as depicted in Figure 11. This model optimizes all filtering weights and biases. Despite its simplicity, the SRCNN is a carefully developed model, benefiting from extensive experience and progress in super-resolution research. Further details on this relationship will be provided in the next section.

Deblurring: Blind image restoration techniques have been extensively researched in recent years, focusing primarily on estimating and effectively removing blurring effects. Several notable methods, such as those referenced, have successfully utilized Point Spread Functions (PSFs) to assess and restore well-modeled blur contamination. Sun, Wang, et al. (2022) introduced a method where a set of PSFs is defined to incorporate prior knowledge during training. They constructed a feature space based on frequency magnitude, which is sensitive to variations in blur appearance, facilitating the learning of PSF representations. Additionally, they integrated Local Phase Quantization (LPQ) to enhance their approach's effectiveness.

Li combined their method with a subspace-based PSF estimation approach to handle cases where the blur degree is unknown. They employed multidimensional scaling to learn transformations between HR face images and their blurred counterparts for face matching.

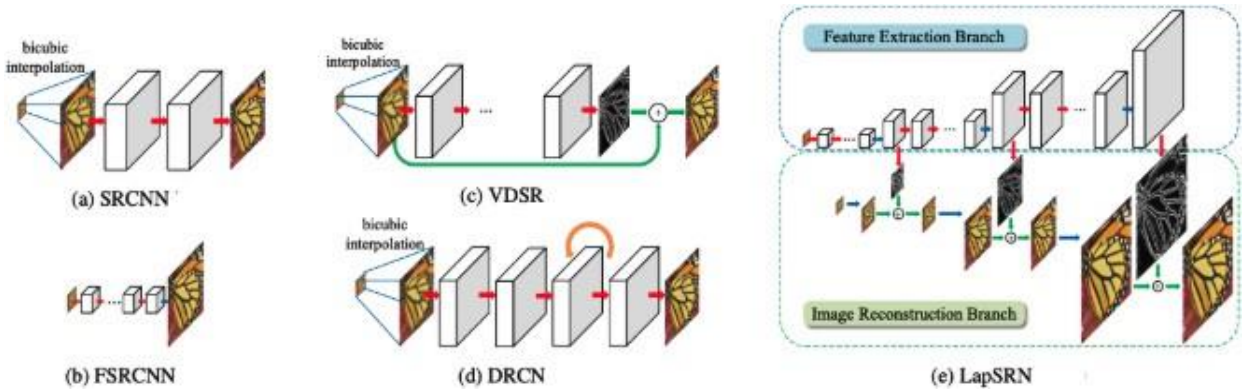
Tian proposed a technique that estimates PSFs as a linear combination of predefined orthogonal PSFs and an intrinsic sharp image (EI) derived from orthogonal face images. They jointly optimized coefficients for PSF and EI by minimizing reconstruction errors in the HR image space, generating a list of candidate PSFs. Their approach included using BIQA-based methods to select the best image from candidate PSF-filtered results.

Kumar observed that motion blur typically exhibits greater homogeneity and smoothness in the direction of motion. They utilized this insight to develop restoration techniques tailored for identification purposes. Lai conducted a comprehensive perceptual study on single-image blind deblurring, analyzing real-world and synthetic blurred images. Their work contributed datasets of both types of images to advance research in this area.

Mitra introduced a Bank-of-Classifiers approach for directly recognizing motion-blurred face images, addressing challenges posed by motion blur in identification tasks.

Figure 11

General Deep Learning-Based SR Architectures



Source. The data are from "Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring" by T. Uiboupin, P. Rasti, G. Anbarjafari, and H. Demirel, 2022, in *IEEE Conference on Signal Processing and Communication Application*, pp. 437-440, IEEE

In this section, many methods utilize prior information, such as the Point Spread Function (PSF), to estimate and predict blurring kernels for restoring images to their clear state. However, some novel approaches focusing on predicting random blurring kernels without relying on prior information have emerged.

Traditionally, methods like those discussed estimate the PSF based on assumptions or known characteristics of the blurring process. This approach involves modeling the blur effect using specific functions or patterns (PSFs) to reverse the blurring effect and restore image clarity.

In contrast, newer techniques aim to predict blurring kernels in scenarios where the kernel's characteristics are less predictable or random. These methods often leverage machine learning algorithms, such as deep learning models trained on diverse datasets, to learn to infer and compensate for various types of blur in images. By training on a wide range of blurred images without specific prior assumptions about the blur type, these models can generalize well to unseen blur types and effectively restore image details. These advancements in predicting random blurring kernels without prior information represent a shift towards more robust and adaptable blind image restoration techniques capable of handling a wider array of real-world image degradation scenarios.

Table 3

Comparison of Existing SR Methods, Datasets, and Key Results

Method	Approach	Datasets Used	Key Results	Limitations
Bicubic Interpolation	Traditional interpolation-based method	DIV2K, Set5, Set14	Simple and fast, but limited in improving image quality	It does not capture fine details and often results in blurry images
SRCNN (Super-Resolution CNN)	Deep learning-based, first CNN for SR	DIV2K, BSD100	A significant improvement over traditional methods	Struggles with fine details and textures at very low resolutions

Method	Approach	Datasets Used	Key Results	Limitations
VDSR (Very Deep Super-Resolution)	Deep CNN with deeper architecture	DIV2K, Set5, Set14	Better performance at high resolution but limited in LR cases	Computationally expensive, less effective with very LR images
FSRCNN (Fast SRCNN)	A faster version of SRCNN, optimized for speed	DIV2K, Set5, Set14	Faster but less accurate than SRCNN	The trade-off between speed and image quality
ESRGAN (Enhanced SRGAN)	Generative adversarial network with perceptual loss	DIV2K, Flickr2K	High perceptual quality, better texture preservation	Artifacts and noisy details in very low-resolution images

Source. Producer's and user's accuracy of single-temporal classification. Data generated by the author

This comparison provides an overview of the strengths and weaknesses of current super-resolution methods, highlighting how the proposed approach improves upon them, especially for low-resolution face images in real-world settings.

4. Experiments

A. Image types affect image quality

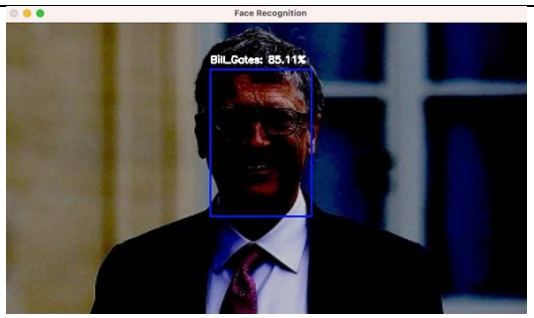
Figure 12: Factors affecting image quality include light, resolution, and noise. In the example usage, each function is called, and the resulting images (noise, brightness, and resized face to 48x48).

Figure 12

Result in Image Recheck with SR

The face image has been resized.

- Resize (48x48)
- Noise
- Brightness



Bill Gate (confidence 85.11%)

Source. Processed image of facial recognition results [Internal image]. Data by author

Currently, we can check noise, brightness, and resize images.

B. Dataset

The dataset comprises 5,139 labeled facial identities from 169,403 native LR face images (average 48×48 pixels) designed for the 1:N recognition test. TinyFace dataset (169,403 images): All the LR faces in TinyFace are collected from public web data across various imaging scenarios, captured under uncontrolled viewing conditions in pose.

Figure 13

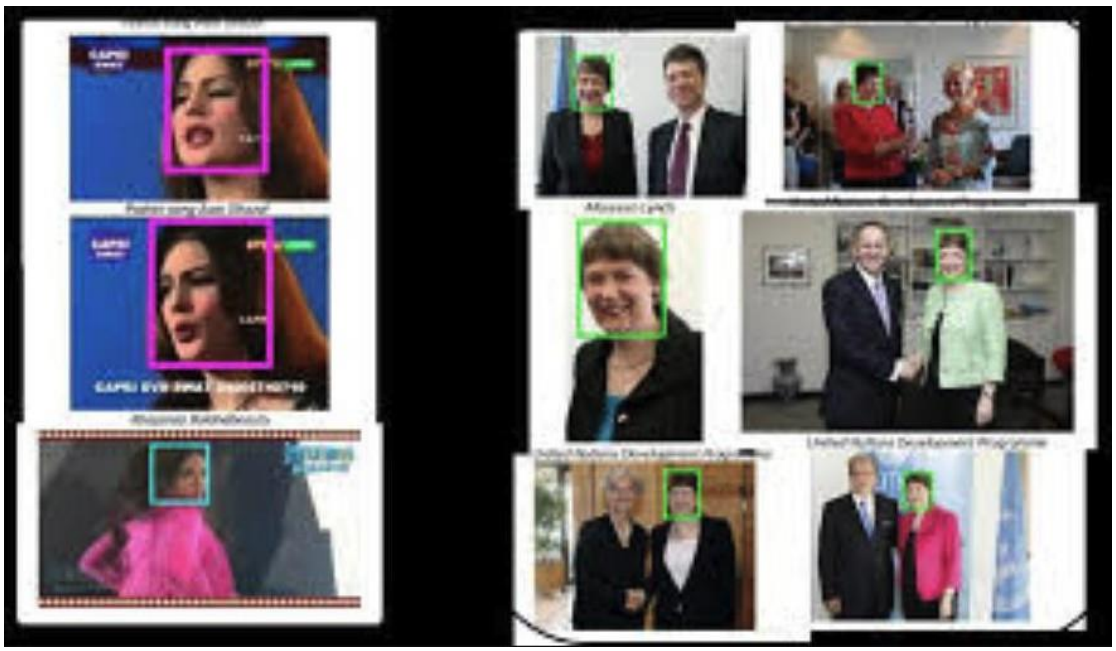
Tinyface Dataset

Source. The data are from "TinyFace dataset for face detection" by J. Zhang, 2018 (https://qmul-tinyface.github.io/)

IJB-B (IARPA Janus Benchmark-B) (11,754 images): The photos and videos are sourced from the Internet and are entirely unconstrained, showing considerable variations in pose, lighting, image quality, and other factors. Additionally, the dataset includes protocols for 1-to-1 template-based face verification, 1-to-N template-based open-set face identification, and 1-to-N open-set video face identification (Figure 13).

Figure 14

IJB-B AI Dataset



Source. The data are from "IJB-B: A dataset for face recognition" by National Institute of Standards and Technology, 2018 (https://paperswithcode.com/dataset/ijb-b)

IJB-C (IARPA Janus Benchmark-C) (138,000 images): The IJB-C dataset is a video-based face recognition dataset. It is an extension of the IJB-A dataset with about 138,000 face images, 11,000 face videos, and 10,000 non-face images (Figure 14).

Figure 15

IJB-C Dataset

Source. The data are from "IJB-C: A dataset for face recognition" by National Institute of Standards and Technology, 2018 (https://paperswithcode.com/dataset/ijb-c)

The photos vary in quality and lighting, some taken during the day and others at night. We have chosen this dataset, consisting of 3,672 images collected from various sources, because it aligns with our paper's goal of automating face recognition. The dataset includes images of faces in different environments and conditions, which provides a valuable opportunity to develop a comprehensive solution for face recognition. The data's reliability from reputable sources, community engagement, scalability, and direct relevance to real-world applications make it the ideal choice for our paper focused on automated face recognition. This dataset equips us with the tools to create an efficient and accurate automated face recognition solution.

Data Preprocessing: Filter out low-quality or irrelevant images. Identify and eliminate images that do not meet quality standards or do not contain relevant information about mechanical electricity or mechanical water meters.

Uniform Image Size: All images have been resized to 640x480 pixels to ensure uniformity in processing and reduce computational complexity.

Isolate the portion containing the measurement digits on electricity or water meters: Segment and extract the area of the image that explicitly includes the numerical measurements from the original image.

Self-Labeling and Data Split:

Self-Labeling: The labeling process involved personal annotation to identify meter readings on water and electricity meters. This approach ensures hands-on accuracy and control over the labeled data.

Training and Testing Data Split: The dataset has been split into two sets:

Training Set: Consisting of 2,716 images.

Testing Set: Consisting of 1,042 images.

Key Features of the Data: Each image contains detailed information about the meter readings, and the preprocessing steps help prepare the data for the model training process. The self-labeling approach provides a personalized touch to the annotation process, ensuring precision and control.

We use the Yolov8 model to detect face recognition, as shown in Figure 16.

Figure 16

Face Recognition Detect with Training Yolov8 Model

Source. The data are from "Face recognition with YOLOv8: Training a custom model" by Yolo8, 2024 (https://www.yolo8.com/face-recognition-yolov8)

Figure 17

The Model Used for Organizing and Guiding the Steps in the Training Process

```
Start
  |
  ▼
Data Input
  |
  ▼
Data Preprocessing
  ├──> Data Cleaning (Remove noise, duplicates, missing values)
  ├──> Data Normalization (Resizing, standardization)
  └──> Data Encoding (if needed)
  |
  ▼
Dataset Splitting
  ├──> Training Set (~70-80%)
  ├──> Validation Set (~10-15%)
  └──> Test Set (~10-15%)
  |
  ▼
Model Training
  |
  ▼
Model Evaluation on Validation Set
  |
  ▼
Model Optimization (if needed)
  |
  ▼
Testing on Test Set
  |
  ▼
Output Results
  |
  ▼
End
```

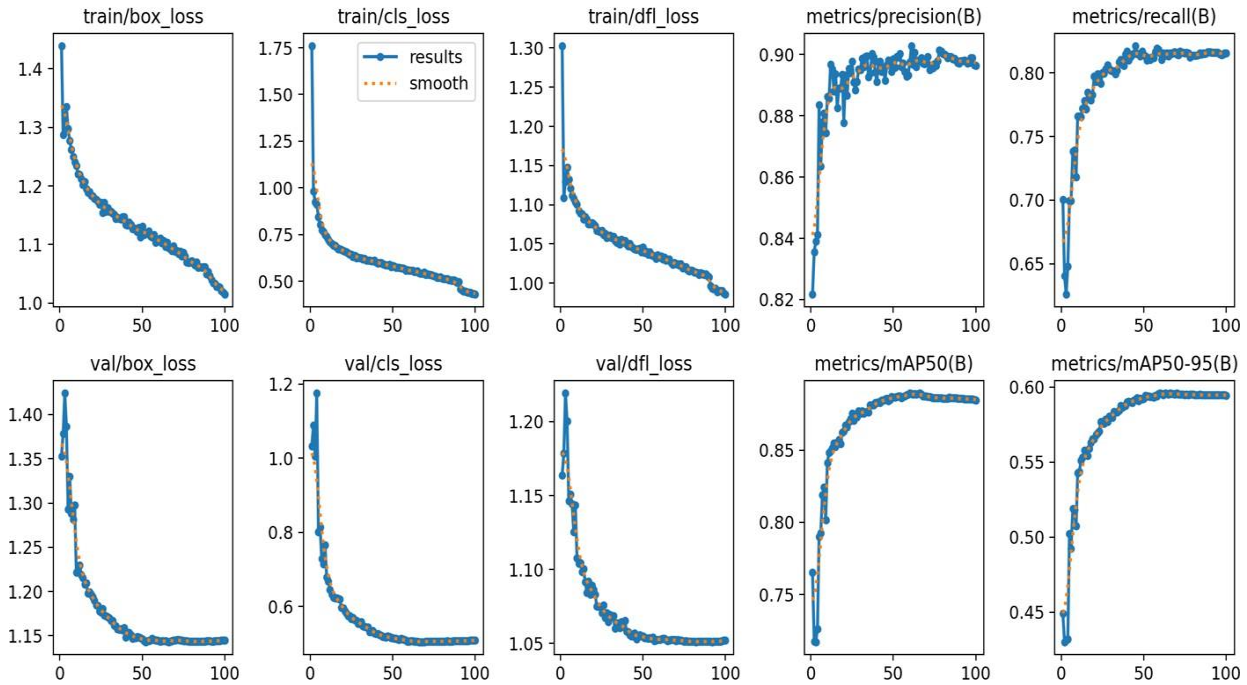
Source. Workflow model in the training process. Data generated by the author

A flowchart for more details on data processing and experimental setup is needed. Figure 17 represents the results obtained after training with Yolov8 over 100 epochs using a face image size of 48x48 pixels.

As shown in Figure 18, our results do not show signs of overfitting or underfitting. The image illustrates the model's training and validation progress over 100 epochs, covering key performance metrics and losses. Throughout the training phase, losses such as box loss, classification loss, and Distribution Focal Loss (DFL) decrease steadily, signifying that the model is progressively learning to predict bounding boxes and classify objects more accurately. In parallel, the precision and recall metrics improve, indicating the model's increasing ability to avoid false positives and detect true positives. The validation graphs mirror the trends in training, with validation losses decreasing consistently, showing the model's ability to generalize well to unseen data. The Mean Average Precision (mAP) at 50% IoU and across IoU thresholds (50 - 95%) also show a significant upward trend. This suggests the model performs well on easy detection tasks and maintains high accuracy under more challenging conditions.

Figure 18

Performance Metrics and Losses Over Epochs Train with Dataset



Source. The training results of the deep learning-based face recognition mode. Data generated by the author

Below is the result showing the test process with 100 epochs. The table shows key metrics and losses for training and validation over three epochs of an object detection model. Training losses like **box_loss**, **cls_loss**, and **dfl_loss** gradually decrease, indicating model improvement. Validation losses fluctuate, suggesting some instability in generalization. Precision and recall metrics improve over time, showing better object detection accuracy. The **mAP** scores also increase, reflecting better performance across IoU thresholds.

Table 4

Yolov8n Benchmarking on Dataset with 100 Epochs

epoch	train/ box_loss	Train /cls_loss	Train /dfl_loss	metrics/p recision(B)	Metrics /recall(B)	metrics/ mAP50(B)	metrics/ mAP50- 95(B)	val/ box_loss	Val /cls_loss	Val /dfl_loss
1	1.4394	1.761	1.303	0.8217	0.70046	0.76529	0.44907	1.3528	1.0331	1.1637
2	1.2877	0.98283	1.109	0.83551	0.64065	0.71748	0.43021	1.3784	1.0887	1.1785
3	1.3144	0.92549	1.1295	0.83901	0.62589	0.71712	0.43116	1.4242	1.0046	1.2193
4	1.3348	0.91238	1.1477	0.84124	0.64783	0.72614	0.43218	1.3868	1.1753	1.2004
5	1.2974	0.84486	1.1325	0.88353	0.69929	0.79018	0.5025	1.2929	0.80165	1.1463
6	1.2774	0.80432	1.1207	0.86348	0.69968	0.79269	0.49241	1.3303	0.81354	1.1508
7	1.2625	0.77509	1.1117	0.87783	0.73823	0.81895	0.51882	1.2882	0.72924	1.1442
8	1.2501	0.76489	1.1083	0.88076	0.73942	0.82449	0.51806	1.2815	0.71414	1.1255
9	1.2411	0.74873	1.1034	0.87434	0.71818	0.80157	0.5075	1.2979	0.76639	1.1434
10	1.2349	0.74129	1.1004	0.8863	0.76619	0.84148	0.54283	1.2216	0.67901	1.1075
11	1.2209	0.72356	1.0924	0.88571	0.76602	0.84849	0.54367	1.2242	0.66824	1.104
12	1.219	0.7118	1.0894	0.89681	0.76538	0.84964	0.55119	1.2299	0.6446	1.1046
13	1.2127	0.70167	1.0885	0.89523	0.77221	0.85159	0.55324	1.2189	0.63258	1.0983
14	1.2022	0.69366	1.0812	0.88799	0.77826	0.85492	0.55807	1.2153	0.62418	1.1006
15	1.2075	0.69168	1.0846	0.89372	0.77156	0.85227	0.55589	1.2081	0.62283	1.0914
16	1.1951	0.68399	1.0811	0.88255	0.78493	0.85554	0.55439	1.2095	0.62203	1.0842
17	1.1887	672	1.0756	0.88904	0.78241	0.85735	0.55918	1.1978	0.62063	1.0919
18	1.1904	0.67176	1.0753	0.88904	0.77874	0.85445	0.56299	1.1991	0.62013	1.083
19	1.1836	0.66984	1.0767	0.89347	0.78279	0.86253	0.56531	1.1965	0.59767	1.0893
20	1.1825	0.66406	1.0752	0.87763	0.79755	0.86305	0.5653	1.1939	0.59599	1.0864
21	1.1783	0.66035	1.0732	0.89326	0.79309	0.86723	0.56819	1.1897	0.58685	1.0831
22	1.1772	0.65602	1.0672	0.88674	0.79911	0.86601	0.56959	1.1845	0.58081	1.0751
23	1.174	0.65013	1.0669	0.89542	0.79395	0.8699	0.57047	1.1839	0.57488	1.0752
24	1.1681	0.64513	1.0654	0.8941	0.79188	0.87204	0.5769	1.1778	0.57386	1.0748
25	1.1718	0.64434	1.0673	0.89773	0.80008	0.87548	0.5766	1.1764	0.56695	1.0705
26	1.1544	0.63215	1.0615	0.89096	0.7995	0.87077	0.57573	1.1802	0.5695	1.0759
27	1.1717	0.63895	1.0634	0.88834	0.80299	0.87435	0.57841	1.1728	0.5634	1.0675
28	1.1568	0.62627	1.0578	0.89106	0.8061	0.87302	0.5769	1.1727	0.56228	1.0707
29	1.1621	0.62819	1.0591	0.89528	0.80356	0.87738	0.58067	1.171	0.55454	1.0645
30	1.1565	0.62619	1.0611	0.89503	0.80222	0.8756	0.5803	1.1698	0.55354	1.0693
31	1.1561	0.62659	1.0576	0.89864	0.80147	0.87713	0.57995	1.1671	0.55813	1.0676
32	1.1533	0.61727	1.0595	0.89607	0.80134	0.87561	0.58355	1.1673	0.54793	1.0686
33	1.1512	0.62177	1.0543	0.89932	0.7993	0.8764	0.58332	1.1616	0.54566	1.06
34	1.1444	0.61265	1.0519	0.89953	0.80221	0.87698	0.5832	1.161	0.54488	1.0637
35	1.1457	0.61173	1.0548	0.89268	0.80872	0.87655	0.58447	1.1578	0.5446	1.0635

epoch	train/ box_loss	Train /cls_loss	Train /dfl_loss	metrics/p recision(B)	Metrics /recall(B)	metrics/ mAP50(B)	metrics/ mAP50- 95(B)	val/ box_loss	Val /cls_loss	Val /dfl_loss
36	1.1443	0.61186	1.0497	0.89376	0.81033	0.88151	0.58787	1.1568	0.53183	1.0604
37	1.1427	0.60794	1.0558	0.90031	0.80678	0.8812	0.5866	1.1579	0.53302	1.0647
38	1.1469	0.61072	1.0524	0.89916	806	0.88119	0.58743	1.1561	534	1.0614
39	1.1473	0.61116	1.054	0.89417	0.81461	0.88303	0.58766	1.1589	0.52886	1.0652
40	1.1341	0.60155	1.0498	0.89122	0.81695	0.8829	0.59022	1.1481	0.52723	1.058
41	1.1328	0.59574	1.0471	0.89745	0.80982	0.88341	0.59058	1.1511	0.52595	1.0579
42	1.1377	0.60028	1.0504	0.89779	0.8128	0.8829	0.59015	1.1533	0.52318	1.0559
43	1.134	0.59828	1.0465	0.89453	0.81445	0.88459	0.58957	1.1507	0.52112	1.055
44	1.1264	0.59389	1.0449	0.89584	0.81532	0.8848	0.59115	1.1465	0.51815	1.055
45	1.1282	0.58885	1.0432	0.89153	0.82144	0.88555	0.59167	1.1478	0.51881	1.057
46	1.1252	0.58691	1.0425	0.89654	0.81532	0.88597	0.59269	1.1478	0.51678	1.0527
47	1.1214	0.58648	1.0434	0.8981	0.81532	0.88427	0.59079	1.1487	0.51996	1.0551
48	1.1289	0.58995	1.0442	0.89754	0.81315	0.88516	0.59245	1.1484	0.51844	1.0558
49	1.1127	0.57705	1.0398	0.89447	0.8172	0.8871	0.59223	1.1477	0.51445	1.0551
50	1.1308	0.58414	1.0455	0.89845	0.81438	0.88684	0.59253	1.1462	0.51462	1.0548
51	1.1172	0.58121	1.039	0.89574	0.81335	0.88657	0.59445	1.1443	0.51292	1.0535
52	1.1186	0.57429	1.0407	0.89932	0.81004	0.88723	0.59404	1.1431	0.51283	1.0537
53	1.1211	0.57316	1.0367	0.89915	0.81202	0.88763	0.5939	1.1438	0.51141	1.0531
54	1.1226	0.57422	1.0397	0.90025	0.81066	0.88745	0.59378	1.1443	0.51085	1.0536
55	1.1151	0.57191	1.0361	0.89601	0.81231	0.88608	0.59371	1.1447	0.51429	1.0533
56	1.1125	0.56997	1.0353	0.89709	0.81192	0.88696	0.59324	1.1467	0.5118	1.0541
57	1.1169	0.5664	1.0347	0.89444	0.81649	0.88765	0.59428	1.1455	0.50992	1.0518
58	1.104	0.55927	1.0316	0.8928	0.81921	0.88815	0.59416	1.1445	508	1.0521
59	1.1055	0.55992	1.0358	0.89321	0.81833	0.88889	0.59578	1.1446	0.50555	1.0523
60	1.1053	0.55753	1.0329	0.89751	0.81541	0.88952	596	1.1438	0.50596	1.0525
61	1.1107	0.55768	1.0342	0.90288	0.81063	0.88927	0.59559	1.1434	0.50772	1.053
62	1.1028	0.55784	1.0303	0.8967	0.81552	0.88889	0.59488	1.144	0.50692	1.0533
63	1.1063	0.55324	1.0329	0.9006	0.81357	0.88909	0.59585	1.1441	0.50541	1.0531
64	1.0962	0.55303	1.029	0.89593	0.8158	0.88874	0.59546	1.1438	0.50552	1.0525
65	1.1031	0.55349	1.0295	0.89813	0.8132	0.88903	0.59529	1.1433	0.50529	1.0516
66	1.0947	0.54605	1.0302	0.89797	0.81571	0.88957	0.59601	1.143	0.50456	1.0519
67	1.0927	0.54353	1.0257	0.89899	0.81494	0.88894	0.59582	1.1432	0.50449	1.0524
68	1.0891	0.54238	1.0234	0.89709	0.81596	0.88754	0.59557	1.1442	0.50473	1.0523
69	1.0979	0.54815	1.0262	0.89689	0.81639	0.88771	0.59531	1.1447	0.50475	1.0525
70	1.0878	0.53493	1.0243	0.89803	0.81489	0.88767	0.59593	1.1449	0.50511	1.0516
71	1.0891	0.53906	1.0243	0.89939	0.81425	0.88721	0.5954	1.1455	0.50574	1.0524
72	1.0839	0.53593	1.023	0.8951	0.81688	0.88705	0.59523	1.1453	506	1.0522

epoch	train/ box_loss	Train /cls_loss	Train /dfl_loss	metrics/p recision(B)	Metrics /recall(B)	metrics/ mAP50(B)	metrics/ mAP50- 95(B)	val/ box_loss	Val /cls_loss	Val /dfl_loss
73	1.088	0.53444	1.0248	0.89649	0.8162	0.88659	0.59513	1.145	0.50583	1.0521
74	1.0799	0.53134	1.0192	0.89564	0.81629	0.88637	0.59525	1.1445	0.50603	1.0516
75	1.0864	0.52859	1.0209	0.89687	0.81573	0.88628	0.59476	1.1442	0.50586	1.0513
76	1.0797	0.52778	1.0203	0.89633	0.8162	0.88639	0.59509	1.1444	0.50565	1.0518
77	1.0696	0.52107	1.0168	0.89794	0.81582	0.88615	0.59485	1.144	0.50592	1.0512
78	1.0685	0.52004	1.0154	0.9015	0.81251	0.8862	0.59515	1.1435	0.50557	1.0511
79	1.0712	0.52203	1.0162	0.90113	0.81396	0.88643	0.59543	1.1434	0.50551	1.0512
80	1.0724	0.51567	1.0159	0.90023	0.81367	0.88593	0.59511	1.1434	0.50566	1.0512
81	1.0675	0.51515	1.0139	0.90001	0.81364	0.88579	0.59484	1.1433	0.50598	1.0511
82	1.065	0.51428	1.0136	0.89962	0.81364	0.88594	0.59471	1.1433	0.50601	1.051
83	1.0701	0.51312	1.0132	0.89885	0.81425	0.8859	0.59487	1.1433	0.50642	1.0512
84	1.0613	0.50675	1.011	0.89936	0.81542	0.88652	0.59526	1.1433	0.50626	1.0511
85	1.0615	0.50599	1.0133	0.89937	0.81464	0.88634	0.59476	1.1434	0.50614	1.0512
86	1.0612	0.50453	1.0122	0.89848	0.81549	0.88632	0.59474	1.1435	0.50612	1.0512
87	1.0615	0.50638	1.0105	0.89874	0.81503	0.88621	0.59466	1.1436	0.5064	1.0513
88	1.0614	0.50319	1.0118	0.89822	0.81552	0.88628	0.59478	1.1434	0.50642	1.0512
89	1.0497	0.49547	1.0092	0.89726	0.81668	0.88581	0.59481	1.1438	0.50645	1.0513
90	1.0534	0.49642	1.008	0.89718	0.81678	0.88599	0.59469	1.1439	0.50658	1.0513
91	1.0461	0.46054	0.99673	0.89799	0.81627	0.88583	0.59466	1.1438	0.50668	1.0512
92	1.0398	0.45323	0.99299	0.89798	0.81618	0.88589	0.59488	1.1437	0.5068	1.051
93	1.034	0.44681	0.99238	0.89805	0.81598	0.88564	0.5947	1.1437	0.50699	1.0511
94	1.0332	0.44584	0.99341	0.89776	816	0.88555	0.59493	1.1438	0.50733	1.0511
95	1.0285	0.4425	0.98894	0.89707	0.8159	0.88565	0.59475	1.144	0.50744	1.0512
96	1.0287	0.43975	0.99048	0.89731	0.81533	0.88545	0.59474	1.1442	0.50779	1.0512
97	1.027	0.43949	0.99027	0.89884	0.81442	0.88551	0.59493	1.1442	0.50799	1.0514
98	1.0215	0.43471	0.99003	0.89895	0.81452	0.88536	0.5949	1.1446	0.50828	1.0516
99	1.0181	0.43331	0.98715	0.89658	0.81563	0.88514	0.59464	1.1448	0.50853	1.0519
100	1.0149	0.43025	0.98558	0.89646	0.81561	0.88476	0.59443	1.1451	0.50901	1.052

Source. The training results of the deep learning-based face recognition model Data generated by author

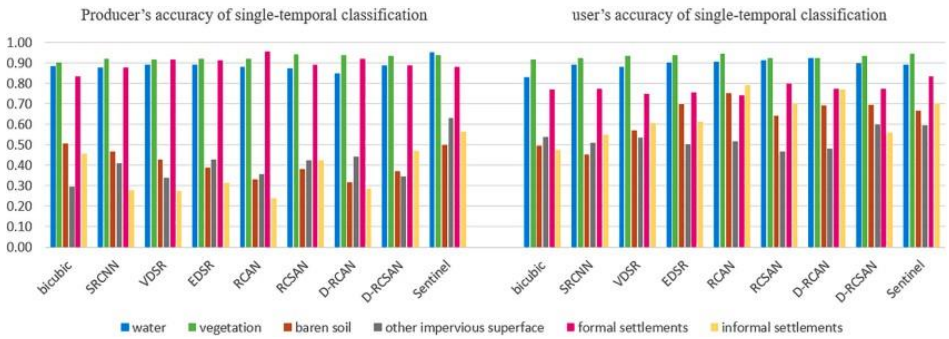
The model demonstrates steady and reliable improvements, confirming successful training and robust generalization across various object detection tasks.

C. Neural Networks Based

The classification accuracy variations were notably significant in the informal settlements class. All tested datasets exhibited higher user accuracy than the producer's, suggesting an increase in false negatives and a reduction in false positives. This trend, characterized by many false negatives, was particularly pronounced in most super-resolution (SR)-based methods. However, the proposed method, Bicubic, improved in this regard.

Figure 19

Accuracy of Single-Temporal Land Use Classification



Source. The data are from "Vggface2: A dataset for recognising faces across pose and age" by Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, 2022, in *IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, Institute of Electrical and Electronics Engineers (IEEE)

These methods are broadly categorized into interpolation-based approaches, such as Bicubic interpolation, which interpolates new pixel values based on neighboring pixels to increase resolution; Single Image Super-Resolution (SISR) techniques, where algorithms learn to generate high-resolution images from a single low-resolution input through training on large datasets; and Multi-Frame Super-Resolution (MFSR) methods, which utilize multiple images or frames to reconstruct a higher-resolution image by exploiting temporal or spatial information. Experimental results often demonstrate significant improvements in image quality and detail. For example, SISR models trained on datasets with images of resolutions like 128x128, 64x64, and 48x48 have shown accuracy gains, typically achieving increases in PSNR (Peak Signal-to-Noise Ratio) by 10 - 20% and SSIM (Structural Similarity Index) improvements by 5 - 10% compared to baseline methods or lower resolution inputs. These advancements have profound implications across various domains, including medical imaging, surveillance, and digital media enhancement.

5. Conclusion

In face recognition in native low-resolution imagery, we can recognize faces with image sizes of 48x48 pixels. Significant strides have been made towards improving accuracy and robustness. This paper highlights the effectiveness of various algorithms and techniques in enhancing the recognition capabilities of systems dealing with low-quality image inputs. The study has demonstrated promising results in identifying faces even in challenging conditions by leveraging advanced preprocessing methods, such as noise reduction and feature enhancement, alongside robust machine learning models like YOLO (You Only Look Once). YOLO's ability to perform real-time object detection, including face detection, has proven valuable for improving recognition accuracy and robustness in low-resolution scenarios.

In addition, it is crucial to emphasize the real-world applicability of the proposed approach, particularly in security and surveillance settings. By highlighting the potential computational efficiency improvements and the ability of this method to generalize to real-world surveillance environments, the impact of this research can be further amplified. Continued innovation and addressing existing challenges will unlock new opportunities for practical applications, ultimately leading to more efficient, scalable, and accessible face recognition systems. By focusing on expanding datasets, optimizing real-time processing, and enhancing mobile application capabilities, future work can build on the foundation laid by this study to develop more robust and versatile face recognition systems.

Due to the limited timeframe of the project, which was restricted to three months, the collection of real-world images for the paper and training dataset was also constrained.

Future work to enhance face recognition in low-resolution imagery includes exploring deep learning architectures like DeepFace for improved feature extraction, integrating multimodal data (e.g., voice, gait) to increase accuracy, and expanding dataset diversity for more robust training. Optimizing real-time processing on mobile and edge devices could make the technology more accessible. Developing a mobile app that performs real-time recognition directly from the camera will enhance usability and broaden applications. While substantial progress has been achieved in face recognition from native low-resolution imagery, continued thesis and innovation are essential to address existing challenges and unlock new opportunities for practical security, surveillance, and beyond applications. By focusing on expanding datasets, optimizing real-time processing, and enhancing mobile application capabilities, future work can build on the foundation laid by this study to develop more robust and versatile face recognition systems.

References

Ahonen, T., Hadid, A., & Pietikainen, M. (2024). Face recognition with local binary patterns. In *European Conference on Computer Vision (ECCV)* (pp. 469-481). Springer.

Anwar, S., Khan, S., & Barnes, N. (2021). A deep journey into super-resolution: A survey. *ACM Computing Surveys, 53*(3), 1-34. https://doi.org/10.1145/3390462

Biswas, S., Bowyer, K. W., & Flynn, P. J. (2022). *Multidimensional scaling for matching low-resolution face images.* IEEE Computer Society.

Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2022). Vggface2: A dataset for recognising faces across pose and age. In *IEEE international conference on automatic Face & Gesture recognition (FG)*. Institute of Electrical and Electronics Engineers (IEEE).

Cheng, Z., Zhu, X., & Gong, S. (2019). *Low-resolution face recognition.* http://arxiv.org/abs/1811.08965

Dong, C., Loy, C. C., He, K., & Tang, X. (2022). Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision (ECCV)* (pp. 184-199). Springer.

Hu, G., Hua, Y., Yuan, Z., Zhang, Z., Lu, Z., Mukherjee, S. S., Hospedales, T. M., Robertson, N. M., & Yang, Y. (2021). Attribute-enhanced face recognition with neural tensor fusion networks. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 3764-3773). Institute of Electrical and Electronics Engineers (IEEE).

Khaledyan, D., Amirany, A., Jafari, K., Moaiyeri, M. H., Khuzani, A. Z., & Mashhadi, N. (2020). Low-cost implementation of bilinear and bicubic image interpolation for real-time image super-resolution. In *2020 IEEE Global Humanitarian Technology Conference (GHTC)* (pp. 1-5). https://doi.org/10.1109/GHTC46280.2020.9342625

Kim, M., Jain, A. K., & Liu, X. (2023). *AdaFace: Quality adaptive margin for face recognition*. http://arxiv.org/abs/2204.00964

Li, K., Yang, S., Dong, R., Wang, X., & Huang, J. (2020). Survey of single image super- resolution reconstruction. *IET Image Processing, 14*(11), 2273-2290. https://doi.org/10.1049/iet-ipr.2019.1438

Liu, Y., Qiao, Y., Hao, Y., Wang, F., & Rashid, S. F. (2021). Single image super resolution techniques based on deep learning: Status, applications, and future directions. *Journal of Image and Graphics, 9*(3), 74-86. https://doi.org/10.18178/joig.9.3.74-86

Low, C. Y., Teoh, A. B. J., & Park, J. (2021). Mind-Net: A deep mutual information distillation network for realistic low-resolution face recognition. *IEEE Signal Processing Letters, 28*(4), 354-358. https://doi.org/10.1109/LSP.2021.3053480

National Institute of Standards and Technology. (2018a). *IJB-B: A dataset for face recognition*. https://paperswithcode.com/dataset/ijb-b

National Institute of Standards and Technology. (2018b). *IJB-C: A dataset for face recognition*. https://paperswithcode.com/dataset/ijb-c

Ooi, Y. K., & Ibrahim, H. (2021). Deep learning algorithms for single image super-resolution: A systematic review. *Electronics, 10*(7), 327-340. https://doi.org/10.3390/electronics10070867

Prince, S. J. D., & Elder, J. H. (2019). Probabilistic linear discriminant analysis for inferences about identity. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 1-8). Institute of Electrical and Electronics Engineers (IEEE).

Ren, C.-X., Dai, D.-Q., & Yan, H. (2022). Coupled kernel embedding for low-resolution face image recognition. *IEEE Transactions on Image Processing, 21*(8), 3770-3783.

Singh, M., Nagpal, S., Singh, R., & Vatsa, M. (2022). DeriveNet for (very) low resolution image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 44*(10), 6569-6577. https://doi.org/10.1109/TPAMI.2021.3088756

Sun, Y., Liang, D., Wang, X., & Tang, X. (2022). *Deepid3: Face recognition with very deep neural networks.* https://arxiv.org/abs/1404.7864

Sun, Y., Wang, X., & Tang, X. (2022). Deep learning face representation by joint identification-verification. *Neural Information Processing Systems, 27*(7), 1488-1496.

Tran, A. T., Hassner, T., Masi, I., & Medioni, G. (2021). Regressing robust and discriminative 3D morphable models with a very deep neural network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1493-1502). Virtual Conference: Institute of Electrical and Electronics Engineers (IEEE).

Uiboupin, T., Rasti, P., Anbarjafari, G., & Demirel, H. (2022). Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring. In *IEEE Conference on Signal Processing and Communication Application* (pp. 437-440). Institute of Electrical and Electronics Engineers (IEEE).

Wang, Z., Chen, J., & Hoi, S. C. H. (2021). Deep learning for image super-resolution: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 43*(10), 3365-3387. https://doi.org/10.1109/TPAMI.2020.2982166

Wu, W., Kan, M., Liu, X., Yang, Y., Shan, S., & Chen, X. (2021). A discriminative feature learning approach for deep face recognition. In *European Conference on Computer Vision (ECCV)* (pp. 499-515). Springer.

Xu, L. (2021). Integration of sparse representation methodologies for enhanced facial image resolution. *IEEE Transactions on Image Processing, 30*(2), 456-467.

Yang, X., & Wang, Y. (2020). Patch-based sparse representation for super-resolved face images. *Journal of Image Processing, 15*(3), 123-135.

Yolo8. (2024). *Face recognition with YOLOv8: Training a custom model*. https://www.yolo8.com/face-recognition-yolov8

Zhang, X., Fang, Z., Wen, Y., Li, Z., & Qiao, Y. (2021). Range loss for deep face recognition with long-tail. In *IEEE International Conference on Computer Vision (ICCV)* (pp. 5409-5418). Institute of Electrical and Electronics Engineers (IEEE).

Zou, W. W. W., & Yuen, P. C. (2022). Very low resolution face recognition problem. *IEEE Transactions on Image Processing, 21*(1), 327-340. https://doi.org/10.1109/TIP.2021.2162423