IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 11 December 2024, accepted 1 January 2025, date of publication 8 January 2025, date of current version 17 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3527236

RESEARCH ARTICLE

Degrade or Super-Resolve to Recognize? Bridging the Domain Gap for Cross-Resolution Face Recognition

KLEMEN GRM[1], (Member, IEEE), BERK KEMAL ÖZATA[2,3], ALPEREN KANTARCI[2], VITOMIR ŠTRUC[1], (Senior Member, IEEE), AND HAZIM KEMAL EKENEL[2]

[1]Faculty of Electrical Engineering, University of Ljubljana, 1000 Ljubljana, Slovenia
[2]Department of Computer Engineering, Istanbul Technical University, 34485 Istanbul, Türkiye
[3]ASELAN Inc., 06200 Ankara, Türkiye

Corresponding author: Klemen Grm (klemen.grm@fe.uni-lj.si)

This work was supported in part by the Bilateral Slovenian Research Agency (ARRS), in part by the Scientific and Technological Research Council of Türkiye (TUBITAK) funded project "Low Resolution Face Recognition (FaceLQ)" TUBITAK under Project 120N011, and in part by the ARRS Research Programme "Metrology and Biometric Systems" under Grant P2-0250(B).

ABSTRACT In this work, we address the problem of cross-resolution face recognition, where a low-resolution probe face is compared against high-resolution gallery faces. To address this challenging problem, we investigate two approaches for bridging the quality gap between low-quality probe faces and high-quality gallery faces. The first approach focuses on degrading the quality of high-resolution gallery images to bring them closer to the quality of the probe images. The second approach involves enhancing the resolution of the probe images using face hallucination. Our experiments on the SCFace and DroneSURF datasets reveal that the success of face hallucination is highly dependent on the quality of the original images, since poor image quality can severely limit the effectiveness of the hallucination technique. Therefore, the selection of the appropriate face recognition method should consider the quality of the images. Additionally, our experiments also suggest that combining gallery degradation and face hallucination in a hybrid recognition scheme provides the best overall results for cross-resolution face recognition with relatively high-quality probe images, while the degradation process on its own is the more suitable option for low-quality probe images. Our results show that the combination of standard computer vision approaches such as degradation, super-resolution, feature fusion, and score fusion can be used to substantially improve performance on the task of low resolution face recognition using off-the-shelf face recognition models without re-training on the target domain.

INDEX TERMS Biometrics, image processing, machine learning.

I. INTRODUCTION

Face recognition (FR) systems have been used in a wide range of real-world applications, such as surveillance and biometric authentication. The rapid development of deep learning algorithms and the availability of large datasets have made remarkable advancements in the field of face recognition research [1]. However, despite these advancements, the performance of face recognition systems is still affected when they are deployed on low-resolution face images, which is a common occurrence in real-world scenarios [2]. One of the biggest issues in this setting is the difficulty of the cross-resolution comparison tasks, caused by the mismatch in the resolution of the gallery/enrollment and probe/test images. For instance, in a surveillance system, the low-quality face images captured by the cameras need to be compared with high-resolution images in the gallery. This mismatch in resolution can cause a significant degradation in recognition performance compared to same-resolution recognition problems. In this paper, we address this challenging

The associate editor coordinating the review of this manuscript and approving it for publication was Larbi Boubchir.

cross-resolution face recognition/comparison task that, for convenience, is also illustrated in Fig. 1.

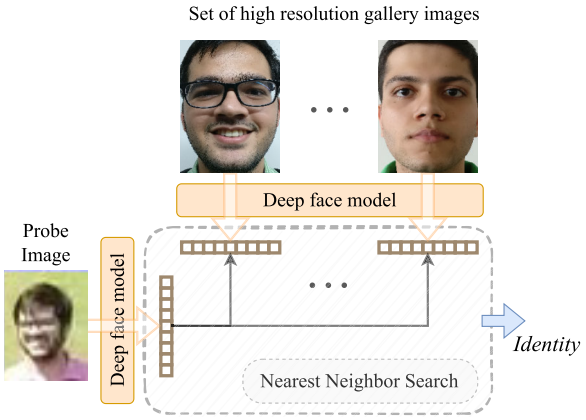


FIGURE 1. Cross-resolution face recognition. In this task, a low-resolution probe face image is compared with a set of high-resolution gallery face images.

Existing techniques for cross-resolution face recognition can in general be grouped into three categories: (*i*) resolution-invariant methods, (*ii*) face hallucination-based methods, and (*ii*) degradation-based methods. While a considerable amount of work has been done on resolution-invariant recognition techniques that aim to make face recognition models robust to varying face quality/resolution distributions [2], our work focuses on investigating face hallucination and degradation-based methods that strive to bring the distributions of the low-quality and high-quality images closer, by either degrading the typically high-quality enrollment/gallery images or enhancing the low-quality probe/test image, thus, making recognition easier and more accurate. These latter two groups of techniques also exhibit several highly desirable characteristics, such as:

- **Universality:** They are model agnostic and can, therefore, be applied with any FR model capable of extracting face representations from the given input images for a comparison procedure. In other words, these methods operate at the image-preprocessing level and are, therefore, universally applicable with arbitrary face recognition models without the need for model fine-tuning.
- **Simplicity:** Unlike resolution-invariant recognition models, hallucination and degradation-based methods typically require a significantly lower amount of training data to work effectively. Additionally, they can be well-described by explicit mathematical models that have fewer degrees of freedom than contemporary heavily-parameterized FR models, leading to (data) efficient learning procedures.
- **Interpretability:** Because the techniques are typically applied at the preprocessing level and produce observable (degradation/hallucination) results that are later fed to the FR model, they allow for an easier interpretation of the recognition decisions compared to standard cross-resolution comparison procedures, where model decisions due to the different characteristics of the input images are more difficult to understand.
- **Complementarity:** Hallucination and degradation-based techniques can be applied in conjunction with resolution-invariant models and have the capacity to further improve results by reducing the resolution induced mismatch between the probe and gallery images. Thus, these groups of techniques are complementary to methods aiming to design resolution-invariant FR approaches.

Due to the outlined characteristics, multiple studies explored degradation and face-hallucination techniques with the goal of improving cross-resolution FR performance [2], [3]. However, ensuring consistent improvements in recognition performance with either approach (degradation or hallucination) remains a challenging (open) research problem, with effective solutions for real-world data still largely missing from the literature. In this study, we aim to address this gap and explore three distinct strategies towards cross-resolution face recognition designed around novel degradation and face hallucination techniques, as illustrated in Fig. 2, i.e.:

- **Degrade-to-Compare (DtC):** With this strategy, we investigate the impact of degrading gallery images, so they viably mimic the characteristics of the low-resolution probes, as shown in Fig. 2(a). To be able to implement the DtC strategy, we propose a scale-wise degradation method, in which different types of degradations are applied at multiple scales. The proposed method allows us to model a wider variety of degradation types, making the generated low-resolution face images more realistic and representative of the real-world challenges faced by cross-resolution face recognition systems.
- **Hallucinate-to-Compare (HtC):** With this strategy, we study the effect of generating a high-resolution image from the low-resolution probes that aligns better with the resolution and quality characteristics of the high-resolution gallery images, as presented in Fig. 2(b). To analyze the feasibility of hallucination techniques for cross-resolution face recognition, we propose a novel multi-scale and multi-hypothesis face-superresolution approach. The approach involves upscaling the low-resolution probe images to multiple scales, i.e., $2\times$, $4\times$, and $8\times$. Additionally, at each scale, multiple hypotheses are reconstructed from different versions of the original low-resolution image to capture potential variations in the degradations encountered during the acquisition of the low-resolution probes.
- **Degrade-and-Hallucinate-to-Compare (DHtC):** With the last strategy, we explore the feasibility of hybrid schemes that combine both gallery degradations and probe hallucinations to bridge the gap between the

distributions of low-resolution and high-resolution images. The main idea behind this scheme, shown in Fig. 2(c), is to simultaneously improve the resolution of the input probes and degrade the quality of the high-resolution gallery images in a sort-of meet-in-the-middle solution. Specifically, in this paper, we propose an approach that combines the multi-scale degradation process from the DtC strategy with the multi-scale, multi-hypothesis hallucination technique from the HtC strategy into a hybrid procedure using various fusion approaches. These fusion approaches aggregate the information from the multi-scale comparisons into a single similarly score that can ultimatelly be used for identity inference.

The research, presented in this paper, builds on our preliminary work from [4], but extends it in multiply aspects, i.e., (*i*), it systematically analyzes the impact of face image quality and resolution on the different strategies towards cross-resolution face recognition; (*ii*), it presents a significantly more comprehensive experimental evaluation that includes two diverse datasets (SCFace, DroneSURF) for cross-resolution face comparison; Third, it studies the interplay between the proposed degradation and hallucination approaches and examines their impact under different face-image quality conditions, and (*iii*), it reports a new state-of-the-art on the SCFace and DroneSURF datasets.

The rest of the paper is structured as follows. In Section II, we review closely related work and position our research within the existing literature. In Section III, we provide details on the three studied strategies (DtC, HtC and HDtC) and describe in-depth the novel multi-scale degradation techniques, the multi-hypothesis and multi-scale face hallucination method as well as the joint hybrid scheme with the corresponding fusion approaches. We evaluate and study the behavior of all three strategies to cross-resolution face recognition on the SCFace and DroneSURF datasets in Section IV, and, finally, conclude the paper with a summary of the main findings and some directions for future work in Section V.

II. RELATED WORK

In this section, we review related prior work with the goal of providing context for our research. Specifically, we first discuss existing super-resolution and face hallucination models, then elaborate on modern face recognition techniques and, finally, explore cross-resolution recognition problems.

A. SUPER-RESOLUTION AND FACE HALLUCINATION

Recently, there has been a surge of interest in utilizing modern deep learning techniques to tackle the problem of super-resolution. Typically, supervised learning methods involve creating a dataset of low-resolution and high-resolution image pairs, where the high-resolution images serve as targets, i.e. ground truth. The training inputs are then derived by subjecting each image to a predetermined degradation process. Models such as Convolutional Neural Networks (CNNs), are then trained to upscale the artificially degraded input images by minimizing a pixel reconstruction error, such as the mean square error (MSE) or the mean absolute error (MAE) [5], [6], [7].

Most of the recent advances in super-resolution have focused on using more complex loss functions that go beyond simple pixel-wise differences. For example, some methods use perceptual loss functions [8] that take into account higher-level semantics to guide the learning process. Others use adversarial learning objectives [9], [10], [11], where a discriminator is trained to distinguish between generated and real images, to further improve the realism of the generated images. These advancements have led to significant progress in the field of super-resolution.

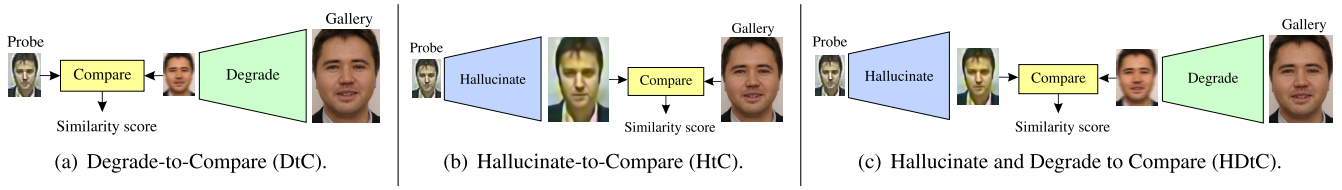
Super-resolution techniques that are used for upscaling human face images are often referred to as *face hallucination* techniques. Unlike general super-resolution methods, which are restricted by the information contained in the input image, face hallucination techniques are able to achieve better reconstructions at higher magnification factors, up to 8 times the resolution of the input image. This is because they are specifically trained on a limited domain of objects, i.e., human faces, which acts as an additional regularizer for the hallucination process. In contrast, most of the existing general super-resolution methods are usually limited to magnification factors of up to 4× [12], [13], [14], [15], [16]

B. FACE RECOGNITION

Recent advancements in large-scale face recognition have involved the collection of large face datasets. Typical examples include VGGFace2 [17], DeepID [18], MS-Celeb-1M [19], WebFace260M [20], Glint360k [21], and others. Modern datasets typically contain thousands of subjects and millions of images, in order to capture a large amount of inter-class and intra-class variance. Model architectures have not been the main focus of recent face-recognition research [3], with most state-of-the-art approaches using the above datasets to train a ResNet-based backbone [22]. These models are commonly trained using classification and metric-learning loss functions, which enable them to learn to extract discriminative features from face images for face identification or verification purposes. In recent years, researchers have focused on developing novel loss functions that can combine classification and metric learning objectives, such as CosFace [23] and ArcFace [24]. Furthermore, researchers have also been working on developing loss functions that explicitly account for the quality of the input image, such as AdaFace [25]. These advances have demonstrated significant potential to improve face recognition performance, especially in challenging conditions, where image quality is poor.

C. CROSS-RESOLUTION FACE RECOGNITION

Cross-resolution face recognition refers to a specific FR problem, where the resolution of the images to be compared during the comparison process differs significantly. Existing



(a) Degrade-to-Compare (DtC). (b) Hallucinate-to-Compare (HtC). (c) Hallucinate and Degrade to Compare (HDtC).

FIGURE 2. **In this paper, we investigate three distinct strategies (i.e., DtC, HtC and HDtC) to cross-resolution face recognition and propose new multi-scale degradation and multi-hypothesis hallucination techniques for their implementation. Additionally, we study the impact of low-resolution probe quality on the behavior of the three considered strategies.**

approaches to this problem can, in general, be categorized into three main groups: (*i*) resolution-invariant methods [26], [27], [28], [29], [30], (*ii*) face-hallucination based methods [12], [13], [31], [32], and (*iii*) degradation-based methods [33], [34].

Resolution-invariant methods aim to minimize the difference between the feature representation of low-resolution and high-resolution face images. One such method is the Deep Coupled ResNet (DCR) model, proposed by Lu et al. [26], which consists of one trunk network and two branch networks. The trunk network is first trained with face images of different resolutions, then the two branch networks are trained to learn coupled-mappings between low-resolution and high-resolution face images. Other knowledge distillation based models [27], [28], [29], [30] distill the information from a *Teacher* network, which is pre-trained with high-resolution face images, to the *Student* network, which is trained on images of different resolutions.

Face hallucination based methods reconstruct high-resolution face images from low-resolution ones and target face recognition in the high-resolution domain. In [12], an identity preserving face hallucination method is proposed. It utilizes a super-identity loss that penalizes the identity difference between high-resolution and super-resolved face images. A similar idea is also presented in [13], where identity priors in the form of pretrained face recognition models are used to steer the face-hallucination process. The Feature Adaptation Network (FAN), presented in [31], disentangles the features into identity and non-identity components and performs face normalization, while improving the resolution, facilitating cross-resolution recognition tasks.

In contrast to the face hallucination based methods, degradation based methods transform high-resolution faces into low-resolution ones. In [33], it is shown that a simple resolution comparison technique that downsamples high-resolution gallery images to the resolution of the low-resolution probe images improves the cross-resolution face recognition performance. Another approach, i.e., the Resolution Adaption Network from [34], employs a Generative Adversarial Network (GAN) that realistically transforms high-resolution images into the low-resolution domain and uses a feature adaption network to extract low-resolution information from the high-resolution embedding.

While hybrid schemes that combine both face degradations and face hallucinations have also been explored in the literature before, e.g., [4], work on this topic is still limited, with studies trying to understand the benefits and behavior of such schemes and their relation to the quality of the input low-resolution images being extremely scarce. We fill this gap with the techniques and analyses presented in this paper.

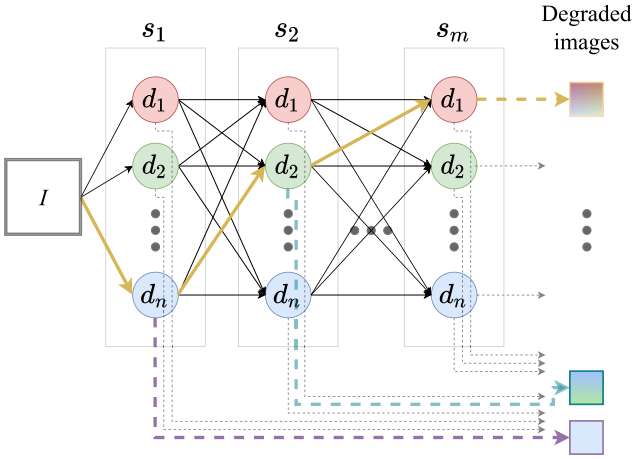


FIGURE 3. **Scale-wise degradation process overview. In the graph above, there are *n* different degradation options to be applied at each step. All the possible paths in the graph generate a degraded gallery image and all of them are used in the recognition pipeline. Note that a downsampling operation is applied between each degradation. The highlighted yellow lines represent a combination of three degradations. The blue and purple dashed lines show specific combinations of one and two degradations, respectively.**

D. FEATURE SELECTION AND FUSION

Feature fusion and score fusion are well-established approaches in pattern recognition. As such, there is a large body of existing work on feature selection and feature fusion approaches in machine learning in general [35], [36] and in the field of biometrics specifically. Existing works in the domain of palmprint recognition have shown more complex approaches can work as well. In [37], the authors show discriminative power analysis as a feature selection tool can improve palmprint recognition when using the DCT coefficients as features. Furthermore, in [38], low correlation between features is used as a criterion. In comparison to these approaches, our proposed feature fusion method uses simple feature concatenation and averaging, while using more capable underlying feature extraction models.

III. METHODOLOGY

In this section, we present our three solutions for the Degrade-to-Compare (DtC), Hallucinate-to-Compare (HtC) and Degrade-and-Hallucinate-to-Compare (DHtC) strategies towards cross-resolution face recognition. We note that all strategies start by cropping the gallery and probe images using the bounding boxes provided by a face detector, so the input to the various models are always cropped faces. Additionally, all strategies use the same pretrained FR model ψ, which, given an input face image I, produces a D-dimensional embedding (feature representation/face template) t, i.e., $t = \psi(I) \in \mathbb{R}^D$.

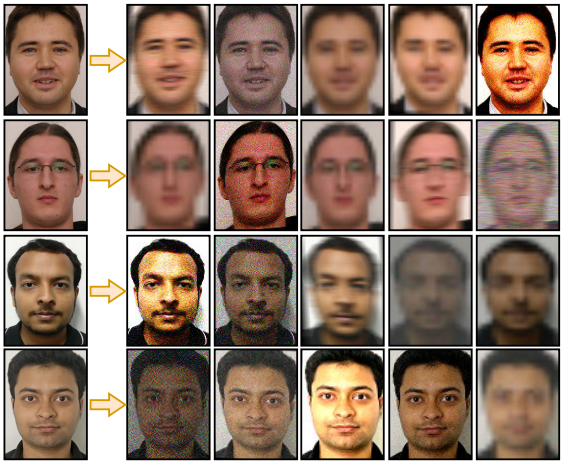


FIGURE 4. Sample degraded gallery images using the proposed multi-scale degradation method. Five degraded examples (in columns) are presented for four distinct gallery images (in rows).

A. DEGRADE-TO-COMPARE WITH MULTI-SCALE DEGRADATIONS

1) MULTI-SCALE DEGRADATIONS

Previous work [33], [39] has shown that matching the resolution of the images alone is insufficient to significantly improve the comparison capabilities. We hypothesize that state-of-the-art face recognition models are sensitive to various *image quality factors*, which differ greatly when comparing a high-quality gallery image to a low quality probe image. In order to match the quality between the gallery and probe images more closely, we propose a stochastic degradation-based approach for the DtC strategy.

Specifically, we propose a Multi-Scale degradation method, as illustrated in Fig. 3. The proposed method involves generating multiple degraded versions of each face image in the gallery set by using a set of n degradation functions $\mathcal{D} = \{d_1, d_2, \ldots, d_n\}$. These functions correspond to various image degradations, e.g., blur, noise addition, and are applied across multiple scales. The applied degradation functions are listed in Table 1 and model common image distortions that appear due to challenging imaging conditions during the (probe) data capture process. Let G be the input gallery image and $\mathcal{C}$ be a combination of k elements of the degradation functions. Here, k takes all values from the set $\{1, 2, \ldots, m\}$ and represents the number of considered image scales s. The total number of downsampling steps, m, is computed individually for each gallery image, and selected in order to match the resolution of either the low-resolution probe image, or the super-resolved images. Thus, the combination $\mathcal{C}$ is defined as: $\mathcal{C} = \{d_{s_1}, d_{s_2}, \ldots, d_{s_k}\}$, where the i-th element d_{s_i} corresponds to a selected degradation function from $\mathcal{D}$ applied at the s_i-th scale. The degradations in the combination $\mathcal{C}$ are applied sequentially to the gallery image G. Furthermore, downsampling is performed between each degradation operation. This process is carried out as shown below:

$$\begin{aligned} G^1 &= \downarrow_s (d_{s_1}(G)), \\ G^2 &= \downarrow_s (d_{s_2}(G^1)), \\ &\vdots \\ G^k &= d_{s_k}(G^{k-1}). \end{aligned} \tag{1}$$

In the above equation, the operator represented by $\downarrow_s$ reduces the resolution by half. The proposed method creates multiple versions of each face image in the gallery by generating all possible degradation combinations. In this approach, if a probe image with low resolution (and/or quality) is received, the comparison procedure will compare it to the degraded gallery images, and some of the degraded images will have a similar quality to the probe image, making the cross-resolution recognition task simpler. A few illustrative examples of degraded gallery images for 4 distinct input samples are presented in Fig. 4. The proposed non-deterministic degradation process is applied multiple times on each of the gallery images. Specifically, a random combination of the degradation functions listed in table 1 is applied, each with randomly sampled parameters.

Note that the actual set of degraded images produced from a single gallery image can be arbitrarily large based on the combinatorial space of possible degradations from Table 1. In practice, we find that generating a set of 1024 degraded images from each gallery image is sufficient to capture the range of possible degradations, and the improvements to performance past that point reach diminishing returns.

2) RESOLUTION MATCHING

The multi-scale degradation process, described above, produces a set of degraded gallery images $\mathcal{G} = \{G_i\}_{i=1}^{M}$. The images in $\mathcal{G}$ are by definition of different resolutions and (in general) differ from the resolution of the probe images P. We therefore introduce an additional (optional) mapping that rescales all degraded galleries to the resolution of the given probe P before resizing them for the targeted FR model. While some of the images in $\mathcal{G}$ require downsampling, others require upsampling. We use bilinear interpolation for both scenarios and refer to the described procedure as *resolution matching* in the reminder of the paper.

TABLE 1. List of degradation functions.

Degradation	Description
Additive Gaussian noise	Additive Gaussian noise with mean 0 and variance 0.02
Speckle noise	Multiplicative Gaussian noise is added to the image with mean 0 and variance 0.02
Color jitter	Randomly change hue and saturation
Brightness jitter	Randomly change brightness and contrast
Motion blur	Horizontal motion blur is applied with a window size 20
Gaussian blur	Gaussian blur with sigma 1.1 and window size 5
Disk blur	Disk blur with radius 5
Perspective transform	Random perspective transform
Shear mapping	Random shear transformation
Upscale following downscale	First downscale the image, then upscale back to the original resolution
Patch shuffle	Random patch shuffle [40]

3) SIMILARITY-SCORE CALCULATION

The goal of the comparison procedure is to produce a scalar similarity score for each given comparison between a probe image P and a given gallery G. Because the proposed multi-scale degradation method produces multiple degradation hypotheses $\mathcal{G} = \{G_i\}_{i=1}^{M}$, as shown in Fig. 5, we determine the final similarity score r by taking the highest computed score, i.e.:

$$r = \max_i(\{\varphi(\psi(G_i), \psi(P))\}_{i=1}^{M}), \tag{2}$$

where ψ is a pretrained FR model and φ is the cosine similarity.

B. HALLUCINATE-TO-COMPARE WITH MULTI-HYPOTHESIS FACE SUPER-RESOLUTION

1) MULTI-HYPOTHESIS FACE-SUPERRESOLUTION

In order to add high-resolution details to real-life low-resolution face images, we train a variant of the EDSR [7] super-resolution convolutional neural network (CNN) exclusively on face images. By limiting the training set to face images, as opposed to general computer vision datasets, such as ImageNet [41] or DIV2K [42] typically used for super-resolution training, the network is able to learn to upsample human faces in more detail, which enables a higher magnification factor ($8\times$) than is typically used for general super-resolution methods (up to $4\times$). Our super-resolution network is trained on a variant of the VGGFace2 [17] dataset with 3M images. Specifically, the dataset requirements for training super-resolution models differ somewhat from the requirements for training face recognition methods, the ostensible purpose of the dataset. In super-resolution training, the dataset images represent the target model outputs, and as such we want all dataset images to be as high-resolution and sharp as possible. To that end, we pick the 1M images from the VGGFace 2 dataset with the highest resolutions as our training set. We verify that all 8631 subjects from the dataset are present in this subset of training data, to maintain image diversity. The chosen 1M images then represent the target model outputs, and the training inputs are derived by applying a degradation (downsampling) pipeline to the full-resolution images.

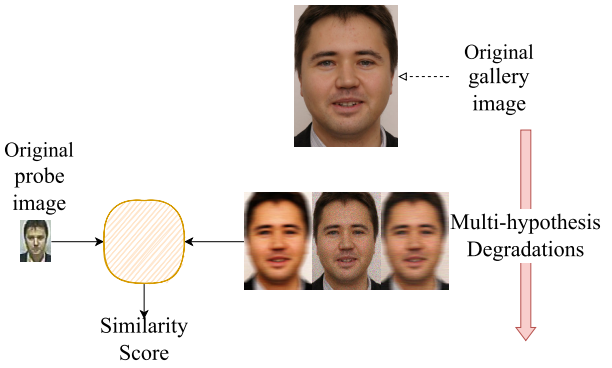


FIGURE 5. Similarity-Score Calculation with the Degrade-to-Compare strategy. The proposed multi-scale degradation method generates several degradation hypotheses from the high-resolution gallery image. These hypotheses are then used during the comparison procedure to calculate a scalar similarity score for a given probe-gallery pair.

Given a super-resolution model ξ_{SR} capable of upsampling an input low-resolution probe image P_{LR}, i.e., such that $P_{HR} = \xi_{SR}(P_{LR})$ where $P_{LR} \in \mathbb{R}^{h\times w\times 3}$, $P_{HR} \in \mathbb{R}^{sh\times sw\times 3}$, and s is the upsampling factor, we design a *multi-hypothesis* upsampling approach. We notice that many low-resolution images are corrupted beyond their limited spatial resolution, e.g., by noise and sampling artifacts. In order to alleviate these artifacts, we blur the input image to different extents. Specifically, we generate 16 versions of the input images by blurring them using a Gaussian kernel with $\sigma = 0$ through $\sigma = 1$. Each of the images is then upsampled separately using our super-resolution model. We present examples of the multi-hypothesis super-resolution for an upscaling factor of $8\times$ in Fig. 6. We note that for most real-world low-resolution images, applying the super-resolution mode without pre-processing the image at all, i.e., the $\sigma = 0$ case, results in a suboptimal reconstruction, since the super-resolution model amplifies its noise and artifacts to an extent. On the other hand, if the input image is blurred too much, this results in a blurry reconstruction.

We note that the face template extraction models require a fixed input image resolution of $112px$. Thus, in all cases where we do not use face super-resolution, such as gallery degradation and resolution matching, the probe and gallery images are re-sampled using interpolation.

2) MULTI-SCALE PROCESSING

The performance of face super-resolution models typically depends on the initial resolution of (and, in turn, the information content contained in) the input probe images and the desired magnification factor. For example, for a low-resolution input image of 24×24 pixels and a magnification factor of $2\times$, the super-resolution model needs to predict 75% of the pixel values in the upscaled image, while this percentage increases to 98.4% for an upscaling

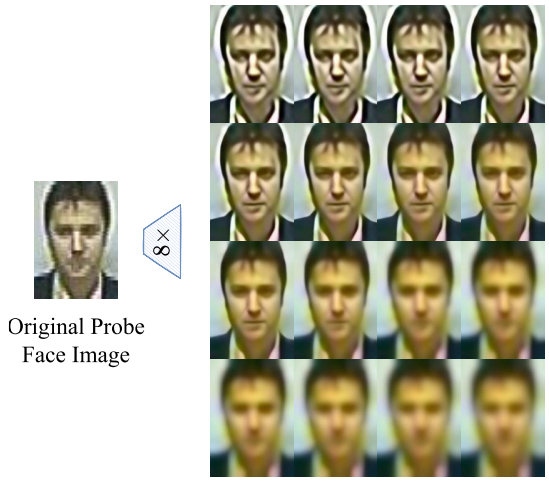


FIGURE 6. Examples of the super-resolved hypotheses for a sample low-resolution probe images using a trained face super-resolution model for an upscaling factor of 8×. By manipulating the high-frequency characteristics of the input probe, our approach generated different variants of the upscaled images that can be used for similarity score calculation. The upper-left sample corresponds to applying the super-resolution model in the single-hypothesis regime without modifying the low-resolution input image.

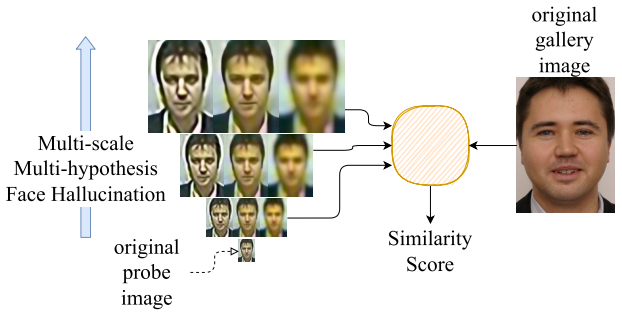


FIGURE 7. Similarity-Score Calculation with the Hallucinate-to-Compare strategy. The proposed multi-scale multi-hypothesis super-resolution method generates several versions of upsampled probe images from the provided low-resolution probe. These hypotheses are then used during the comparison procedure to calculate a scalar similarity score for the given probe-gallery pair.

factor of 8×. Thus, significantly more information needs to be hallucinated for larger magnification factors. In our approach, we therefore adopt a multi-scale scheme and generate multiple super-resolved hypotheses for each of the three upscaling factors considered, i.e., 2×, 4× and 8×. This procedure allows us to produce higher-resolution probe images with varying amounts of hallucinated information for the comparison procedure.

3) SIMILARITY-SCORE CALCULATION

The multi-scale multi-hypothesis face supper-resolution approach, presented above, produces a set of N upscaled hypotheses $\mathcal{P} = \{P_i\}_{i=1}^{N}$ from the provided input probe P. Here, we note that we do not attempt to determine, which of the reconstruction hypotheses represents the most desirable output that most closely resembles the characteristics of the gallery images. Instead, we use all of the recovered hypotheses as input to the similarity-score calculation step. To that end, we extract an embedding from each of the reconstructed hypotheses using a pretrained face recognition model ψ, and again consider the maximum similarity between a probe P and gallery G as the final comparison score r:

$$r = \max_{i}(\{\varphi(\psi(G), \psi(P_i))\}_{i=1}^{N}), , \tag{3}$$

where φ is again the cosine similarity.

C. HALLUCINATION-AND-DEGRADE-TO-COMPARE WITH MULTI-SCALE DEGRADATIONS AND MULTI-HYPOTHESIS SUPER-RESOLUTION

1) HYBRID HALLUCINATION-DEGRADATION SCHEME

We combine the multi-scale degradation method and the multi-hypothesis hallucination procedure into a hybrid scheme with the goal of compensating for their individual shortcomings and to further bridge the cross-resolution domain gap. As illustrated in Fig. 8, it is likely that some of the images from the large set of degraded gallery images and super-resolved probe images will be closer in quality than the original pair, since various quality/resolution hypotheses are created with our approach for both, the initial gallery as well as the initial probe image.

The face hallucination method produces N superresolved faces $\mathcal{P} = \{P_i\}_{i=1}^{N}$ from the given probe P, while the multi-scale degradation method produces M degraded versions of each gallery face G, i.e., $\mathcal{G} = \{G_i\}_{i=1}^{M}$. To compute a single scalar similarity score for a comparison between G and P from the sets $\mathcal{G}$ and $\mathcal{P}$, we utilize various fusion techniques over the corresponding embeddings. These fusion techniques can, in general, be implemented either on the feature level or the similarity score level. Feature level fusion involves representing the hypotheses of a single face image using a single face feature vector. In contrast, similarity score based fusion aims to reduce the similarity scores between the hypotheses of a probe image and those of a gallery image into a single similarity score. Details on the two fusion types are given below.

2) FEATURE FUSION

We consider two types of feature-level fusion for the implementation of our hybrid scheme, namely, feature addition (T_{add}) and feature concatenation (T_{concat}). The feature vectors obtained from a face image's hypotheses are denoted as $\{t_i \in \mathbb{R}^D\}$, where D represents the dimensionality of the feature vector, and i denotes the hypothesis index over either the probes in $\mathcal{P}$ or galleries in $\mathcal{G}$. Feature addition is carried out by calculating the sum of the face features:

$$t_{acc} = \sum_{i \in \mathcal{P} \vee \mathcal{G}} t_i, \tag{4}$$

where $t_{acc} \in \mathbb{R}^D$ is the accumulated face feature. On the other hand, feature concatenation involves concatenating the face

features into a single vector, as shown below:

$$t_{acc} = \mathop{\#}_{i} (t_i)\,, \tag{5}$$

where $\#$ is the concatenation operator and merges the feature vectors. Note that, in order for the concatenated probe features and gallery features to be comparable, their concatenated sizes must be identical. To ensure this, hypotheses are first summed up using feature addition at each scale, assuming both probe and gallery hypotheses have matching scales as shown in Fig. 2(c), and then the resulting face features are concatenated across those scales.

3) SCORE FUSIO

To fuse the similarity scores between probe and gallery image hypotheses, we obtain the feature vectors from these hypotheses, as described in previous sections. Then, we calculate each possible similarity score between the probe hypotheses' feature vectors and the gallery hypotheses' feature vectors. To fuse these similarity scores, two options are available: (*i*) adding them up (S_{add}) or (*ii*) defining the similarity between the given probe and gallery image as the maximal similarity between any pair of hypotheses generated (S_{max}). By fusing the similarity scores using these strategies, we obtain a single similarity score that represents the overall similarity for a given gallery-probe pair.

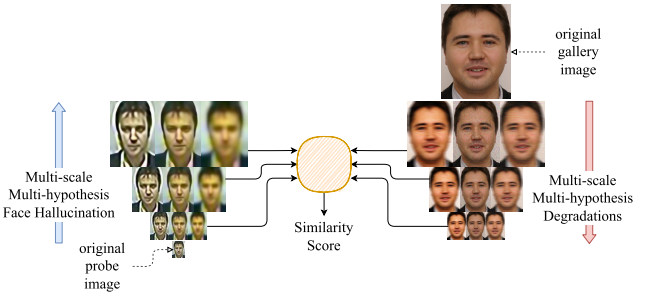


FIGURE 8. Similarity-Score Calculation with the Degrade-and-Hallucinate-to-Compare strategy. The multi-scale degradation method and the multi-hypothesis super-resolution methods generate multiple versions of probe and gallery images, respectively, that are used to generate a single scalar score in the comparison procedure for a given gallery-probe pair.

The use of the maximal similarity score is motivated by the fact that face recognition models are trained such that false positive matches are much less likely than false negatives. Thus, degrading the gallery image is extremely unlikely to increase its similarity with any given probe image, unless that probe image contains the same person, and degrading the gallery image only brings the quality of the two images closer together. On the other hand, the use of the sum of similarity scores is motivated by the interpretation of face feature vectors containing information (signal) related to the identity of the person of the image, and noise related to irrelevant factors such as image quality, as well as pose, background, etc. The idea is that adding up similarity scores from a large set of images where the noise factors differ while the identity remains the same will cause the noise factors to average out, dampening the noise while amplifying the signal.

D. IMPLEMENTATION DETAILS

The different degradation and hallucination procedures presented in this section all rely on comparison in the embedding space of a pretrained FR model ψ and require no fine-tuning. To analyze the impact of different FR models on the presented procedures, we use publicly available state-of-the-art (SOTA) models for our experiments. Specifically, we select the ArcFace [24] models with ResNet-50 and ResNet-101 backbones adopted from the InsightFace repository,[1] trained on the MS1M [19] and Glint360k [21] datasets. All models take 112×112 resolution images as input and extract $D = 512$ dimensional feature vectors. If the resolution of the given face image is different from 112×112, then we resize it to the target resolution using bilinear interpolation.

To extract features from test images, we first crop the images using face detection coordinates provided by the dataset authors. Then, the images are subjected to the preprocessing procedure provided by the authors of the face recognition models, and passed through the models to obtain the face feature vectors.

We note that, without fine-tuning on the target domain, the performance of our proposed method is highly reliant on the face recognition models used. The models used here were selected for their performance and large training set size, which enables a degree of robustness to image quality factors such as resolution, blur, lighting, and head pose.

IV. EXPERIMENTS

A. DATASETS

We select two diverse and challenging datasets with cross-resolution comparison problems for the experiments, i.e., the SCFace [43] and the DroneSURF [44] datasets. Details on the two datasets are provided below.

- **The SCFace Dataset:** There are 130 subjects in the SCFace dataset, each having one high-quality frontal image corresponding to the *gallery* face images, and multiple low-resolution images corresponding to the *probe* face images. Probe face images are captured using five different surveillance cameras and from three different distances, d1: 4.2m, d2: 2.6m and d3: 1.0m. Sample images from the dataset are shown in Fig. 9(a),(b). In the experiments on this dataset, we report Rank-1 identification rate (IR) results for all 130 subjects. In order to compare the obtained results with the ones from the previous works, we also report the mean of Rank-1 IR for 10 Repeated Random Sub-Sampling Validation (RRSSV) experiments on 80 subjects, which is the common benchmark on this dataset. Please note that, in most of the previous works, the remaining 50 subjects are used for training purposes, however, our proposed approach does not require any training

[1] https://github.com/deepinsight/insightface

or fine-tuning on the same dataset. In the SCFace experiments, faces are detected using MTCNN [45], then cropped by enlarging the bounding boxes with a scale factor of 1.3 following the findings in [33].

- **The DroneSURF Dataset:** The dataset contains 200 videos of 58 subjects captured with drones, of which 24 subjects are used for testing purposes. Videos are captured in two types of surveillance settings: active and passive. In the active scenario, subjects are actively monitored, therefore, the camera-to-subject distance is relatively constant. In the passive scenario, a drone monitors an area or event while its position and orientation remain fixed. Therefore, the distance between the drone and subjects changes. Both scenarios are captured under two different day-times, during the day and before sunset, and at two locations, in the park and on the terrace of a building. The dataset also contains a gallery set of frontal face images captured using smartphones in constrained environments. In Fig. 9(c), sample images captured under the active and passive settings are given in the first and second rows, respectively. As can be seen, in the passive scenario, the face resolution is very low. In the DroneSURF experiments, gallery faces are detected using MTCNN [45]. Annotations for probe face bounding boxes are provided in DroneSURF [44], however, most of them contain a significant portion of background information. Therefore, in order to obtain tightly cropped faces, we detected probe face images using TinyFace [46]. We chose TinyFace [46] over MTCNN [45] due to its ability to detect low-resolution face images. Still, TinyFace [46] could not detect all the probe faces. In these cases, we directly use the provided bounding box annotations. The detected faces are then cropped by enlarging the bounding boxes with a scale factor of 1.3. No fine-tuning of the deep face recognition models is performed on this dataset.

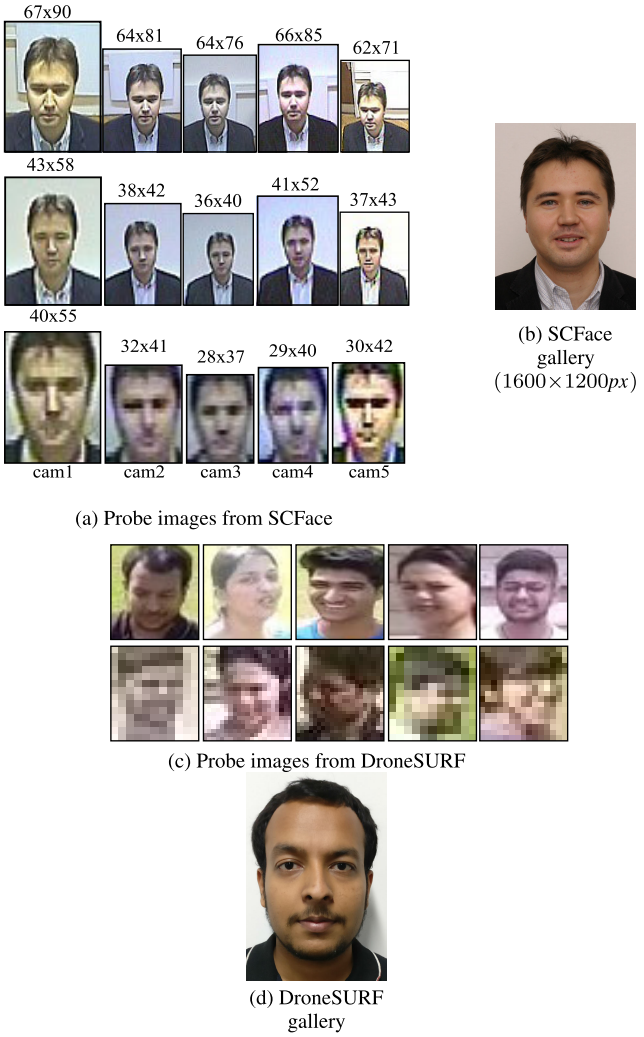


FIGURE 9. Example probe and gallery images from the SCFace and DroneSURF datasets. In (a), probe face images from the SCFace dataset are given. The probe images in (a) belong to the same subject and are captured at the same distance of 4.2m (bottom row), 2.6m (middle row), and 1m (top row), with five different cameras. Faces are localized with MTCNN and cropped with a scale factor of 1.3. Although they are captured at the same distance, we can see that face resolution differs across cameras. Column (b) shows the gallery image of the subject given in (a). In (c), we see example probe images from the DroneSURF dataset, where the first row consists of the images captured under active surveillance and the second row consists of the images captured under the passive surveillance scenario. In (d), a sample gallery image from the DroneSURF dataset is shown.

B. DATA ANALYSIS

We start the experimetal section with an initial analysis of the impact of the degradation and hallucination procedures on the quality characteristics of the SCFace and DroneSURF data. To this end, we examine the Stochastic Embedding Robustness Face Image Quality (SER-FIQ) [47] score distribution of the face images before and after applying the degradation and hallucination processes. In the top row of Fig. 10, we show SER-FIQ score distributions calculated for the SCFace dataset. As can be seen, the image quality distribution obtained from the degraded gallery face images in Fig. 10(c) becomes more similar to that of the original low-resolution probe images (Fig. 10(a)) due to the applied degradation process. Conversely, the SER-FIQ score distribution of hallucinated probe face images (Fig. 10(d)) shows that the proportion of high-quality face images is increased, suggesting that the hallucination process may be beneficial for the comparison procedure on this dataset.

The same analysis is carried out also for the active surveillance scenario of the DroneSURF dataset. As expected, the image quality of the face images decreases due to gallery degradations, as seen in Fig. 10(c). However, no improvement is observed in terms of image quality for the hallucinated probe face images, which is likely due to the quality of the original DroneSURF probe images, which is too low to recover sufficient face details using the proposed hallucination approach.

TABLE 2. Baseline Rank-1 IR (%) results on the SCFace and DroneSURF datasets.

Model			
SCFace	d1 (%)	d2 (%)	d3 (%)
MS1M-RetinaFace-R50	39.84	86.92	93.69
MS1M-RetinaFace-R101	50.61	91.84	95.84
Glint360k-R50	64.00	98.15	**100.00**
Glint360k-R101	**74.61**	**99.38**	**100.00**
DroneSURF		Active (%)	Passive (%)
MS1M-RetinaFace-R50		21.29	11.75
MS1M-RetinaFace-R101		26.14	13.98
Glint360k-R50		37.15	19.38
Glint360k-R101		**43.60**	**21.61**

TABLE 3. Rank-1 IR results (%) on SCFace dataset with gallery degradations and (with and without) resolution matching.

Model	d1 (%)	d2 (%)	d3 (%)
Gallery Degradations – w/o Resolution Matching			
MS1M-RetinaFace-R50	54.30	96.00	98.15
MS1M-RetinaFace-R101	66.61	96.61	98.92
Glint360k-R50	76.76	99.07	100.00
Glint360k-R101	87.38	99.69	100.00
Gallery Degradations – w/ Resolution Matching			
MS1M-RetinaFace-R50	59.53	95.84	97.69
MS1M-RetinaFace-R101	71.53	97.38	98.46
Glint360k-R50	82.61	98.61	99.84
Glint360k-R101	89.69	99.69	100.00

TABLE 4. Rank-1 IR (%) results on DroneSURF dataset with gallery degradations.

Model	Active (%)	Passive (%)
Gallery Degradations		
MS1M-RetinaFace-R50	31.58	16.22
MS1M-RetinaFace-R101	36.41	17.77
Glint360k-R50	43.97	22.42
Glint360k-R101	**47.21**	**21.71**
Gallery Degradations w/ Resolution Matching		
MS1M-RetinaFace-R50	35.18	18.34
MS1M-RetinaFace-R101	39.39	20.04
Glint360k-R50	48.23	21.75
Glint360k-R101	**51.55**	**26.84**

C. BASELINE RECOGNITION RESULTS

In the first series of recognition experiment, we evaluate the performance of four off-the-shelf deep face recognition models, described in Section III-D, on the SCFace and DroneSURF datasets without applying any degradation or hallucination technique. The results of this experiment are presented in Table 2. Notably, on the SCFace dataset, the accuracies achieved by the face recognition models surpass 90% for the closer distances, indicating the effectiveness of deep face models when the face resolutions are relatively high. Hence, we focus our analysis on the face images captured from the d1 distance, which represents the farthest range leading to the lowest resolution of the face images in the SCFace dataset. The recognition accuracies on the DroneSURF dataset exhibit a significant variation. For the passive scenario, all the models achieved approximately only half of the performance that were obtained in the active scenario. It is important to note here that, compared to the face resolution in the active scenario, in the passive scenario, the resolution is much lower, which causes this significant accuracy difference.

D. GALLERY DEGRADATION RESULTS

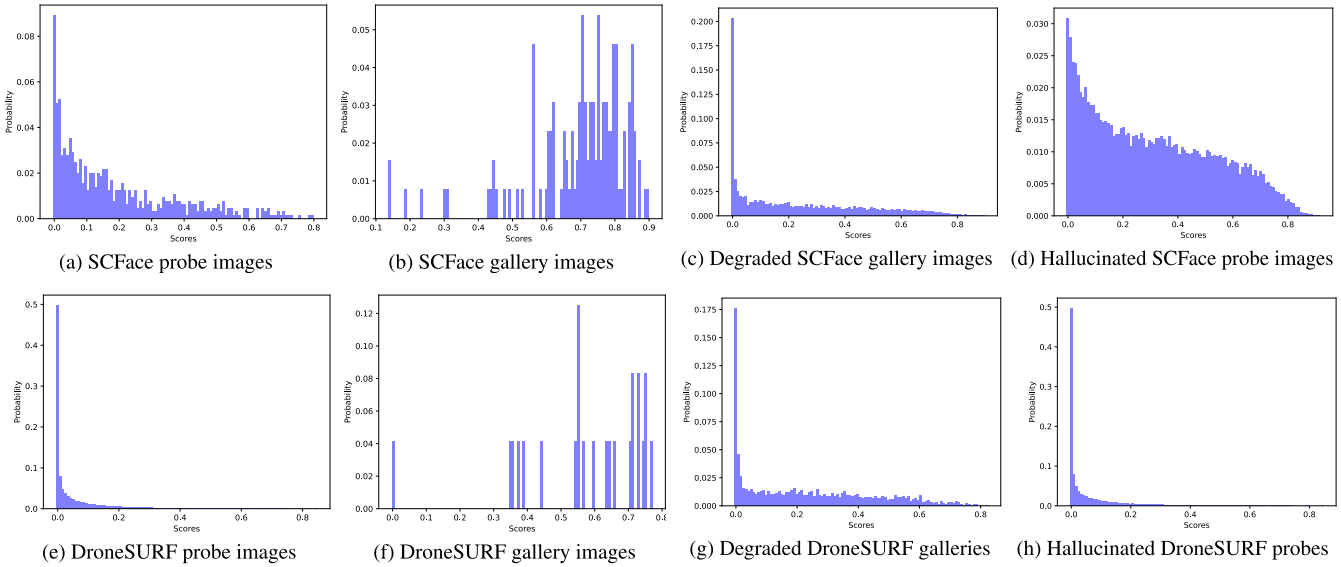
Next, using the method introduced in Section III-A for the DtC strategy, we apply degradations on the gallery face images of both SCFace and DroneSURF datasets. Thus, in this experiment, we study how degrading high-resolution gallery face images, with the goal of comparison the characteristics of the low-resolution probes, impacts the cross-resolution face recognition performance. Each gallery image is degraded by applying the degradations from Table 1, using the random process in Eq. (2), so as to produce a large set of degraded images, which are then compared to the probes.

The results obtained on the SCFace dataset, given in Table 3, point to significant improvements in the Rank-1 Identification Rate (%) compared to the baseline experiment results. Specifically, the model trained on the MS1M-RetinaFace dataset with a ResNet-50 backbone yields a 36% relative increase in accuracy. Moreover, the model trained on the Glint360k dataset with a ResNet-101 backbone achieves a recognition accuracy of 87.38%, approaching the results obtained at closer distances. Further reducing the domain gap by matching the resolution of gallery faces and probe faces results in even higher accuracies, which can be seen from the lower half of Table 3. The most successful model achieves a 89.69% identification rate at the d1 distance, but more importantly, al models consistently improve in performance at d1, when degradation and resolution matching are used, compared to using degradations only.

We employ the same gallery degradation strategy also on the DroneSURF dataset. The results in Table 4, show similar behavior as observed with the SCFace experiments, that is, gallery degradations lead to significant improvements over the baseline recognition accuracies, which is observed in both, the active and passive scenarios. Resolution matching applied with the gallery degradations further improves the face recognition performance in the cross-resolution setting.

E. FACE HALLUCINATION RESULTS

In the next series of experiments, we explore the impact of the HtC face hallucination strategy, introduced in Section III-B, on the cross-resolution face recognition performance. We investigate three different strategies: (*i*) single-hypothesis, (*ii*) multi-hypothesis, and (*iii*) multi-scale multi-hypothesis. In the *single-hypothesis* approach, we enhance each low-resolution face image on an individual



(a) SCFace probe images (b) SCFace gallery images (c) Degraded SCFace gallery images (d) Hallucinated SCFace probe images

(e) DroneSURF probe images (f) DroneSURF gallery images (g) Degraded DroneSURF galleries (h) Hallucinated DroneSURF probes

FIGURE 10. SER-FIQ score distributions on the SCFace and DroneSURF datasets before and after applying the proposed degradation and hallucination schemes.

TABLE 5. Rank-1 IR (%) results on SCFace dataset with face hallucination at d1=4.2m distance.

Model	d1 (%)		
Single-Hypothesis SR	2×	4×	8×
MS1M-RetinaFace-R50	41.23	32.92	27.69
MS1M-RetinaFace-R101	50.15	41.07	36.92
Glint360k-R50	67.07	55.07	47.53
Glint360k-R101	**73.07**	60.92	52.92
Multi-Hypothesis SR	2×	4×	8×
MS1M-RetinaFace-R50	50.92	53.07	52.61
MS1M-RetinaFace-R101	56.00	58.46	60.15
Glint360k-R50	67.84	74.15	72.00
Glint360k-R101	74.15	**76.61**	76.30
Multi-Hypothesis Multi-scale SR	S_{max}	S_{add}	T_{add}
MS1M-RetinaFace-R50	63.53	65.69	64.92
MS1M-RetinaFace-R101	72.30	72.30	75.69
Glint360k-R50	83.38	85.07	85.07
Glint360k-R101	87.38	**88.15**	87.69

basis, without taking into account any image artifacts. For the *multi-hypothesis* approach, we create several enhanced versions of each probe face image. Each hypothesis in the multi-hypothesis approach exhibits artifacts to varying degrees. These approaches are compared with each other using different upscaling factors, i.e., 2×, 4×, and 8×. Additionally, we explore a *multi-hypothesis and multi-scale* approach, where the probe images are enhanced using the multi-hypothesis method at different scales. The resulting multi-scale multi-hypothesis versions of the probe images are fused together using the fusion methods described in Section III-C.

Table 5 presents the face hallucination results for the d1=4.2m distance of the SCFace dataset. The upper part of the table displays the results for single-hypothesis face hallucination, while the middle section presents the results for multi-hypothesis face hallucination. The reported results suggest that single-hypothesis face hallucination does not lead to improvements in recognition accuracies. However, multi-hypothesis face hallucination results in better performance. This suggests that naively enhancing a low-quality face image without addressing image degradations does not result in significant improvements for the cross-resolution face recognition task on this dataset. Instead, generating multiple hypotheses by removing high-frequency artifacts before applying super-resolution can aid the recognition process. In the last section of Table 5, we combine the multi-hypothesis probe images of different scales, rather than measuring their performance at separate scales. This further improves the cross-resolution comparison performance and leads to significant performance gains for all tested models.

Due to the inadequate quality of probe face images in the passive surveillance scenario of the DroneSURF dataset, we conduct face hallucination experiments only for the active scenario. In Table 6, we present the results for the multi-scale multi-hypothesis face hallucination approach that performed overall best of SCFace. As can be seen, these results show a slight reduction in the face identification rates compared to the baseline experiment results from Table 2, suggesting that face hallucination is not able to sufficiently improve the face image quality in active surveillance scenarios due to unsuitable (low) quality characteristics of the DronSURF probes. These results are also consistent with the SER-FIQ generated quality score distributions from Fig. 10 that already suggested limited quality impact of the hallucination process.

TABLE 6. Rank-1 IR (%) results on DroneSURF dataset with face hallucination.

Model	Active (%)
MS1M-RetinaFace-R50	18.37
MS1M-RetinaFace-R101	20.58
Glint360k-R50	33.12
Glint360k-R101	**36.10**

To provide a qualitative comparison and offer additional insight into the behavior of the face hallucination approach, we include hallucinated images from both the SCFace and DroneSURF datasets in Fig. 11. It is worth noting that the super-resolved DroneSURF face images have significantly more artifacts than the SCFace examples. This is the case for most of the samples in the DroneSURF dataset due to poorer image quality and lower face resolution, which appear to have a significant impact on the success of face hallucination strategies applied as preprocessing steps to cross-resolution face comparison.

(a) Face hallucination results of a probe image from SCFace dataset

(b) Face hallucination results of a probe image from DroneSURF dataset

FIGURE 11. Multi-hypothesis multi-scale face hallucination results. For the purpose of visualization, all the images are resized to a fixed size of 112 × 112, which is the model input size. As shown, the images generated for the DroneSURF dataset exhibit more severe artifacts than those of the SCFace dataset.

F. COMBINING GALLERY DEGRADATIONS AND PROBE HALLUCINATIONS

In this section, we investigate the performance of a combined approach that integrates both, the gallery degradations and the probe face hallucinations, into a hybrid DHtC procedure. Our goal with the combined approach is to bridge the gap between the distributions of low-resolution and high-resolution face images using a joint degradation-hallucination scheme, as illustrated in Fig. 2(c).

In Table 7, we combine the gallery degradation method with the probe face hallucinations for the d1=4.2m distance of the SCFace dataset. By merging these two schemes with the fusion strategies from Section III-C, we are able to achieve a face identification rate exceeding 90% for 130 subjects. While all of the fusion strategies lead to comparable cross-resolution recognition performance, the highest recognition accuracy of 91.07% is obtained with the feature concatenation strategy (T_{add}) using the Glint360k-R101 face recognition model.

TABLE 7. Rank-1 IR results (%) on SCFace dataset with combined gallery degradation and face hallucination.

Model	S_{max}	S_{add}	T_{add}	T_{concat}
MS1M-RetinaFace-R50	78.46	72.00	70.61	71.69
MS1M-RetinaFace-R101	78.92	75.84	79.38	79.23
Glint360k-R50	86.76	88.15	88.61	89.23
Glint360k-R101	90.15	90.76	90.15	**91.07**

TABLE 8. Rank-1 IR results (%) on DroneSURF dataset with combined gallery degradation and face hallucination.

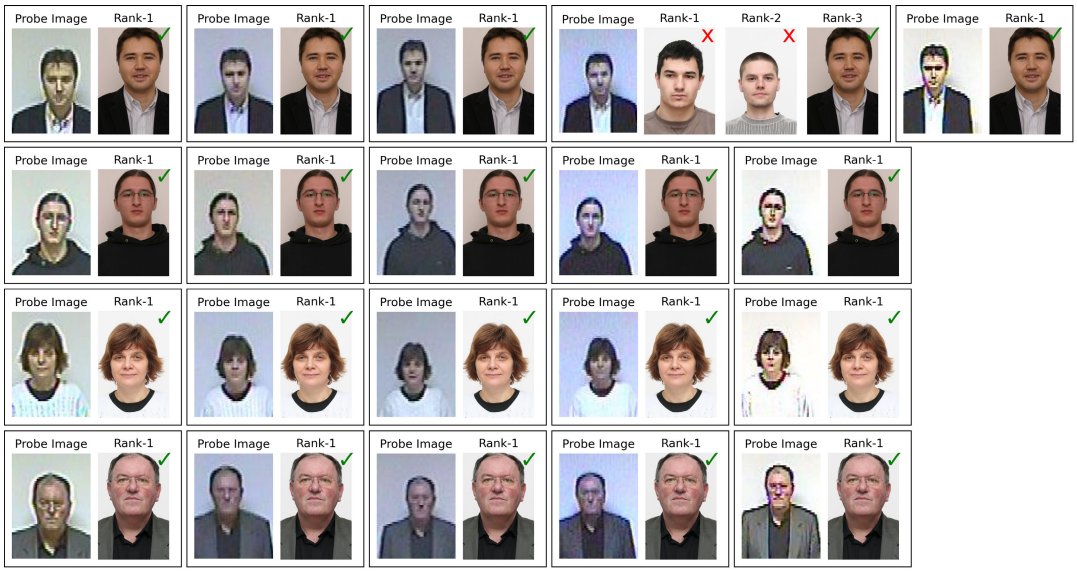
Model	Active (%)
MS1M-RetinaFace-R50	28.87
MS1M-RetinaFace-R101	30.95
Glint360k-R50	39.65
Glint360k-R101	**42.27**

In the case of the DroneSURF dataset, the performance of the combined method is impacted by the poor performance of face hallucination scheme. However, the combination still outperforms the baseline results and the hallucination-only method. The corresponding results can be found in Table 8, where the maximum similarity score strategy (S_{max}), which was found to work best, was employed when generating the reported results.

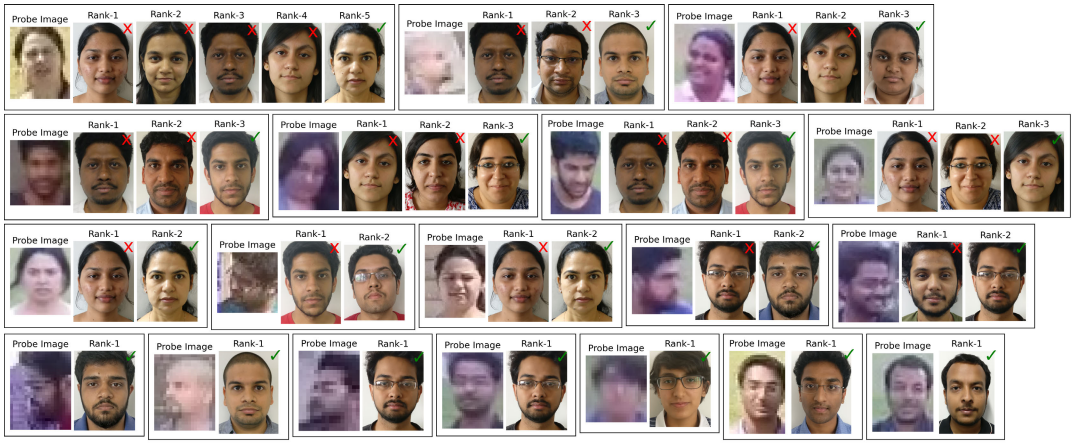
G. RESOLUTION IMPACT

In this section, we examine the impact of probe-image resolution on the face recognition performance. Table 9 presents the results of the DtC strategy, which employs the Glint360k-R101 model with gallery degradations and resolution matching (but no face hallucination), for each camera individually, together with the corresponding average face sizes per camera. The results in the table suggest that the performance of the face recognition model is directly proportional to the resolution of the face images. Fig. 9 presents a subject's probe face images taken by five distinct cameras, which shows the disparities in resolution and degradations caused by each camera. It is straight forward to see the differences in the probe image quality and information content in the images, captured by the five cameras, that are reflected in the reported recognition rates.

The resolution of the face images also varies between the different surveillance settings of the DroneSURF dataset,



(a) Success and failure cases for SCFace



(b) Success and failure cases for DroneSURF

FIGURE 12. **Top-10 predictions for probe face images on SCFace and DroneSURF datasets, including success and failure cases.**

namely active and passive surveillance, as exemplified in Fig. 9. Table 4 reports the face recognition performance for both scenarios, and we can observe that the performance in the passive scenario, where the faces have lower resolution, is lower than the performance of the active scenario.

H. SUCCESS AND FAILURE CASES

In compliance with the terms of the SCFace release agreement, we display the success and failure cases for the Glint360k-R101 model with gallery degradations and resolution matching in the DtC scheme. We use only the subjects whose images are cleared for publication in Fig. 12(a). Failures typically arise with lower-quality images that contain a high degree of uncertainty about the subject's identity. However, the proposed method consistently aligns its predictions with discernible attributes such as gender, haircut, skin, or hair color. Even if the top prediction isn't always accurate, the correct identification is generally found within the top four, suggesting that the proposed method can successfully extract and apply key facial features in its identification process.

Fig. 12(b) demonstrates the model's ability to tackle the face identification task in the challenging environment of the DroneSURF dataset. The probe images shown in Fig. 12(b) include many low-quality and non-frontal face images, captured in an unconstrained setting. Similarly to the previous findings, it aligns predictions with key observable attributes such as gender, haircut, hair color, and skin tone. In comparison fo SCFace, our proposed method performs worse overall on DroneSURF in terms of the per-frame rank-1 identification rate measure. The uncontrolled drone

TABLE 9. Influence of camera resolution on cross-resolution face matching. Here, $w \times h$ denotes the average width and height of the face bounding boxes per camera and IR denotes the Rank-1 identification results. These results belong to the Glint360k-R101 model utilized with gallery degradations and resolution matching.

Cam	d1=4.2m		d2=2.6m		d3=1.0m	
	w × h	IR	w × h	IR	w × h	IR
1	**26×36**	**96.92**	**46×62**	**100.0**	**81×112**	**100.0**
2	22×30	93.84	38×51	100.0	70×95	100.0
3	20×26	86.92	34×46	100.0	64×88	100.0
4	20×27	86.15	35×48	100.0	66×91	100.0
5	20×28	84.61	35×48	98.46	60×82	100.0
avg	**22×29**	**89.69**	**37×51**	**99.69**	**68×94**	**100.0**

footage with very low-resolution faces present in the dataset demonstrates the limits of its capability. These attributes remain present across the top-10 predictions. We note that this is the most challenging experiment setting for DroneSURF, compared to the per-video setting where close-up frames can be used to identify the subjects for the entire video sequence.

I. COMPARISON WITH THE STATE-OF-THE-ART

In this section, we present a comparison with the state-of-the-art on the SCFace and DroneSURF datasets.

1) SCface RESULTS

In Table 10, we compare our results on the SCFace dataset with those of previous state-of-the-art works. To make our results comparable with those of the prior methods, we report the mean of 10 RRSSV results for 80 randomly selected subjects. Our model, Glint360k-R101, incorporates both gallery degradations and face hallucination. Existing methods that perform fine-tuning on 50 randomly selected subjects are marked in Table 10 with a check mark in the FT column. Among the models that do not perform fine-tuning, our model, Glint360k-R101, applied within the joint degradation-hallucination scheme achieves the highest performance with 95.4% accuracy at the d1=4.2m distance. The second-best result, 88.3%, is achieved by [27]. Remarkably, our approach also performs better than the competing techniques that utilize a portion of the dataset for training purposes, despite not requiring any training or fine-tuning on the target dataset.

2) DroneSURF RESULTS

In Table 11, DroneSURF Rank-1 IR (%) results are reported for the active and passive scenarios under the *frame-wise* protocol. The third column indicates whether face recognition is carried out on tightly cropped probe faces that are obtained using face detectors or on the bounding boxes provided with the dataset. The table also specifies whether the models are fine-tuned on the target dataset or not. Unlike the SCFace experiments, our model only utilizes the proposed degradation method along with resolution matching and does not involve face hallucination. In the active scenario,

TABLE 10. Comparison of Rank-1 IR results on the SCFace dataset with previous works. Models fine-tuned on the SCFace dataset are denoted with a checkmark on FT column. The average of 10 RRSSV for 80 subjects out of 130 subjects is reported for our models.

Model	FT	d1 (%)	d2 (%)	d3 (%)
Martinez et al. [48]	✗	68.3	97.0	99.8
Fang et al. [34]	✗	70.5	96.0	98.0
Aghdam et al. [33]	✗	78.5	98.4	**99.8**
Lai et al. [49]	✗	79.7	95.7	98.2
Khalid et al. [50]	✗	85.7	**99.1**	99.1
Khalid et al. [27]	✗	**88.3**	98.3	98.6
Sun et al. [51]	✓	65.5	87.2	98.7
DCR [26]	✓	73.3	93.5	98.0
TCN [52]	✓	74.6	94.9	98.6
FAN [31]	✓	77.5	95.0	98.3
Fang et al. [34]	✓	81.3	97.8	98.8
Huang et al. [28]	✓	86.8	98.3	98.3
Li et al. [53]	✓	90.4	98.0	98.0
Lai et al. [49]	✓	93.0	98.5	98.5
Martinez et al. [54]	✓	**95.3**	**100.0**	**100.0**
Ours	✗	**95.4**	**99.8**	**100.0**

TABLE 11. Comparison of Rank-1 IR results on the DroneSURF dataset with previous works. Models that perform face detection before performing face recognition denoted in the column *Tight-Crop*.

Model	FT	Tight-Crop	Active	Passive
Kalra et al. [44]	✗	✗	14.36	5.08
Ferro et al. [56]	✗	✗	24.25	2.61
DeriveNet [55]	✓	✗	36.33	**27.81**
Amato et al. [30]	✗	✓	39.16	13.04
Ours	✗	✓	**51.55**	**26.84**

TABLE 12. Comparison of the computational complexity of our method with previous works.

Model	SR fwd. passes	FR fwd. passes	Training
Kalra et al. [44]	0	1	✗
Ferro et al. [56]	0	1	✗
DeriveNet [55]	1	1	✓
Amato et al. [30]	0	1	✓
Ours	48	96	✗

we achieve the highest accuracy of 51.55%, which is the best result among all the approaches listed. The second-best result belongs to the method proposed in [30], which also employs a face detector to crop probe face images and does not fine-tune on the target dataset. In the passive surveillance scenario, we obtain the highest accuracy among the approaches that do not use the target dataset for training, with an accuracy of 26.84%. In [55], they use 34 subjects of the dataset for training purposes and achieve an accuracy of 27.81%. This shows the limitations of our training-free method, since fine-tuning directly on the target domain can still improve the performance somewhat on the lower end of resolution and image quality.

We compare the computational complexity of the methods evaluated on DroneSURF in table 12. Complexity evaluation for deep learning-based methods is typically split into training and inference time constraints. Here, we note only which methods require training to begin with. Furthermore, the inference computational complexity is measured in terms of the forward passes through face recognition and super-resolution networks required to compare a gallery and probe image. The time complexity is sub-linear with regards to the number of forward passes, due to the effects of GPU batch processing, as processing singleton batches is highly inefficient. The space complexity is affected only by the need to keep the expanded dataset in memory. No extra storage is required. We note that our proposed method has far higher inference complexity, while not requiring training. This indicates utility in cases where insufficient target domain data exists for extensive training or fine-tuning. Furthermore, we note that while increased computational complexity is obviously undesirable, methods that make efficient us of increased computational budgets tend to be better at eliciting improved performance in the long run [57], [58], which mirrors our findings.

V. CONCLUSION

In this work, we have addressed the challenge of cross-resolution face recognition and investigated two strategies for improving recognition accuracy, i.e., gallery degradation and face hallucination. We have proposed a multi-scale degradation method for the high-resolution galleries and a multi-scale and multi-hypothesis face hallucination method to improve the quality of the probe images. We have also explored the combination of these two methods using score-level and feature-level fusion techniques. Our experiments on the SCFace and DroneSURF datasets have shown that both methods can improve cross-resolution face recognition accuracy. However, face hallucination was not useful on DroneSURF due to poor image quality. Our findings emphasize the importance of considering image quality when selecting face recognition methods. The combination of gallery degradation and face hallucination is likely to provide the best results for cross-resolution face recognition with relatively high-quality probe images, while degradation alone may be more appropriate for low-quality probe images. Our proposed strategies are agnostic with respect to the deep face recognition model being used and do not require any fine-tuning on the target dataset.

However, in the worst-case scenario of face resolution and image quality, our proposed method is still marginally outperformed by previous approaches that include fine-tuning on the target domain. Thus, if extensive footage from the target domain exists, using it for training results in better performance so far. However, our proposed method presents a promising approach for scenarios where this is not the case, and could be used as a basis for long-range recognition applications in those domains.

As part of our future work, we plan to explore naively super-resolution methods that are robust to unknown degradations, particularly for the case of low-quality face images. This would address the challenge of poor quality face images and potentially improve the performance of the face hallucination method in such cases. Additionally, we intend to extend the ideas presented in this work to multi-frame super-resolution models that are capable of inferring high-frequency face information from a sequence of low-resolution frames, instead of hallucinating it from a single input face. Such strategies are expected to further address the challenges of cross-resolution face recognition.

REFERENCES

[1] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, Mar. 2021.
[2] P. Li, L. Prieto, D. Mery, and P. Flynn, "Face recognition in low quality images: A survey," 2018, *arXiv:1805.11519*.
[3] G. Guo and N. Zhang, "A survey on deep learning based face recognition," *Comput. Vis. Image Understand.*, vol. 189, Dec. 2019, Art. no. 102805.
[4] K. Grm, B. K. Özata, V. Štruc, and H. K. Ekenel, "Meet-in-the-middle: Multi-scale upsampling and matching for cross-resolution face recognition," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. Workshops (WACVW)*, Jan. 2023, pp. 1–10.
[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, Feb. 2016.
[6] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 1646–1654.
[7] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1132–1140.
[8] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. ECCV*, Jan. 2016, pp. 694–711.
[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 105–114.
[10] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. ECCV Workshops*, Jan. 2019, pp. 63–79.
[11] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1905–1914.
[12] K. Zhang, Z. Zhang, C.-W. Cheng, W. H. Hsu, Y. Qiao, W. Liu, and T. Zhang, "Super-identity convolutional neural network for face hallucination," in *Proc. ECCV*, Jan. 2018, pp. 196–211.
[13] K. Grm, W. J. Scheirer, and V. Štruc, "Face hallucination using cascaded super-resolution and identity priors," *IEEE Trans. Image Process.*, vol. 29, pp. 2150–2165, 2020.
[14] C. Chen, D. Gong, H. Wang, Z. Li, and K. K. Wong, "Learning spatial attention for face super-resolution," *IEEE Trans. Image Process.*, vol. 30, pp. 1219–1231, 2021.
[15] T. Lu, Y. Wang, Y. Zhang, Y. Wang, L. Wei, Z. Wang, and J. Jiang, "Face hallucination via split-attention in split-attention network," in *Proc. 29th ACM Int. Conf. Multimedia*, Oct. 2021, pp. 1–9.
[16] S. Sharma, A. Dhall, and V. Kumar, "Frequency aware face hallucination generative adversarial network with semantic structural constraint," *Comput. Vis. Image Understand.*, vol. 223, Oct. 2022, Art. no. 103553.
[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in *Proc. 13th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2018, pp. 67–74.

[18] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Dec. 2014, pp. 1988–1996.

[19] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "MS-Celeb-1M: A dataset and benchmark for large-scale face recognition," in *Proc. ECCV*, Jan. 2016, pp. 87–102.

[20] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, and J. Zhou, "WebFace260M: A benchmark unveiling the power of million-scale deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 10487–10497.

[21] X. An, X. Zhu, Y. Gao, Y. Xiao, Y. Zhao, Z. Feng, L. Wu, B. Qin, M. Zhang, D. Zhang, and Y. Fu, "Partial FC: Training 10 million identities on a single machine," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 1445–1449.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[23] H. Wang, Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, Z. Li, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5265–5274.

[24] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4685–4694.

[25] M. Kim, A. K. Jain, and X. Liu, "AdaFace: Quality adaptive margin for face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 18729–18738.

[26] Z. Lu, X. Jiang, and A. Kot, "Deep coupled ResNet for low-resolution face recognition," *IEEE Signal Process. Lett.*, vol. 25, no. 4, pp. 526–530, Apr. 2018.

[27] S. S. Khalid, M. Awais, Z.-H. Feng, C.-H. Chan, A. Farooq, A. Akbari, and J. Kittler, "Resolution invariant face recognition using a distillation approach," *IEEE Trans. Biometrics, Behav., Identity Sci.*, vol. 2, no. 4, pp. 410–420, Oct. 2020.

[28] Y. Huang, P. Shen, Y. Tai, S. Li, X. Liu, J. Li, F. Huang, and R. Ji, "Improving face recognition from hard samples via distribution distillation loss," in *Proc. ECCV*, Jan. 2020, pp. 138–154.

[29] F. V. Massoli, G. Amato, and F. Falchi, "Cross-resolution learning for face recognition," *Image Vis. Comput.*, vol. 99, Jul. 2020, Art. no. 103927.

[30] G. Amato, F. Falchi, C. Gennaro, F. V. Massoli, and C. Vairo, "Multi-resolution face recognition with drones," in *Proc. 3rd Int. Conf. Sensors, Signal Image Process.*, Oct. 2020, pp. 13–18.

[31] X. Yin, Y. Tai, Y. Huang, and X. Liu, "FAN: Feature adaptation network for surveillance face recognition and normalization," in *Proc. ACCV*, Jan. 2021, pp. 301–319.

[32] K. Grm, M. Pernu, L. Cluzel, W. J. Scheirer, S. Dobriek, and V. Štruc, "Face hallucination revisited: An exploratory study on dataset bias," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2405–2413.

[33] O. A. Aghdam, B. Bozorgtabar, H. K. Ekenel, and J.-P. Thiran, "Exploring factors for improving low resolution face recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 2363–2370.

[34] H. Fang, W. Deng, Y. Zhong, and J. Hu, "Generate to adapt: Resolution adaption network for surveillance face recognition," in *Proc. ECCV*, Jan. 2020, pp. 741–758.

[35] U. Mangai, S. Samanta, S. Das, and P. Chowdhury, "A survey of decision fusion and feature fusion strategies for pattern classification," *IETE Tech. Rev.*, vol. 27, no. 4, p. 293, 2010.

[36] J. Deng, S. Bei, S. Shaojing, and Z. Zhen, "Feature fusion methods in deep-learning generic object detection: A survey," in *Proc. IEEE 9th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC)*, vol. 9, Dec. 2020, pp. 431–437.

[37] L. Leng, M. Li, C. Kim, and X. Bi, "Dual-source discrimination power analysis for multi-instance contactless palmprint recognition," *Multimedia Tools Appl.*, vol. 76, no. 1, pp. 333–354, Jan. 2017.

[38] L. Leng and J. Zhang, "PalmHash code vs. PalmPhasor code," *Neurocomputing*, vol. 108, pp. 1–12, May 2013.

[39] K. Grm, V. Štruc, A. Artiges, M. Caron, and H. K. Ekenel, "Strengths and weaknesses of deep learning models for face recognition against image degradations," *IET Biometrics*, vol. 7, no. 1, pp. 81–89, Jan. 2018.

[40] G. Kang, X. Dong, L. Zheng, and Y. Yang, "PatchShuffle regularization," 2017, *arXiv:1707.07103*.

[41] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[42] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 1122–1131.

[43] M. Grgic, K. Delac, and S. Grgic, "SCface–surveillance cameras face database," *Multimedia Tools Appl.*, vol. 51, no. 3, pp. 863–879, Feb. 2011.

[44] I. Kalra, M. Singh, S. Nagpal, R. Singh, M. Vatsa, and P. B. Sujit, "DroneSURF: Benchmark dataset for drone-based face recognition," in *Proc. 14th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2019, pp. 1–7.

[45] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016.

[46] P. Hu and D. Ramanan, "Finding tiny faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1522–1530.

[47] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, and A. Kuijper, "SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 5650–5659.

[48] Y. Martínez-Díaz, M. Nicolás-Díaz, H. Méndez-Vázquez, L. S. Luevano, L. Chang, M. Gonzalez-Mendoza, and L. E. Sucar, "Benchmarking lightweight face architectures on specific face recognition scenarios," *Artif. Intell. Rev.*, vol. 54, no. 8, pp. 6201–6244, Dec. 2021.

[49] S.-C. Lai and K.-M. Lam, "Deep Siamese network for low-resolution face recognition," in *Proc. Asia–Pacific Signal Inf. Process. Assoc. Annu. Summit Conf. (APSIPA ASC)*, Dec. 2021, pp. 1444–1449.

[50] S. Safwan Khalid, M. Awais, C.-H. Chan, Z. Feng, A. Farooq, A. Akbari, and J. Kittler, "NPT-loss: A metric loss with implicit mining for face recognition," 2021, *arXiv:2103.03503*.

[51] J. Sun, Y. Shen, W. Yang, and Q. Liao, "Classifier shared deep network with multi-hierarchy loss for low resolution face recognition," *Signal Process., Image Commun.*, vol. 82, Mar. 2020, Art. no. 115766.

[52] J. Zha and H. Chao, "TCN: Transferable coupled network for cross-resolution face recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3302–3306.

[53] P. Li, S. Tu, and L. Xu, "Deep rival penalized competitive learning for low-resolution face recognition," *Neural Netw.*, vol. 148, pp. 183–193, Apr. 2022.

[54] Y. Martínez-Díaz, H. Méndez-Vázquez, L. S. Luevano, L. Chang, and M. Gonzalez-Mendoza, "Lightweight low-resolution face recognition for surveillance applications," in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 5421–5428.

[55] M. Singh, S. Nagpal, R. Singh, and M. Vatsa, "DeriveNet for (Very) low resolution image classification," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 10, pp. 6569–6577, Oct. 2022.

[56] E. Ferro, C. Gennaro, A. Nordio, F. Paonessa, C. Vairo, G. Virone, A. Argentieri, A. Berton, and A. Bragagnini, "5G-enabled security scenarios for unmanned aircraft: Experimentation in urban environment," *Drones*, vol. 4, no. 2, p. 22, Jun. 2020.

[57] C. Burns, H. Ye, D. Klein, and J. Steinhardt, "Discovering latent knowledge in language models without supervision," 2022, *arXiv:2212.03827*.

[58] R. Sutton, "The bitter lesson," *Incomplete Ideas (Blog)*, vol. 13, no. 1, p. 38, 2019.

KLEMEN GRM (Member, IEEE) received the M.Sc. and Ph.D. degrees from the University of Ljubljana, in 2015 and 2020, respectively. He is currently a Research Assistant with the Machine Intelligence Laboratory, Faculty of Electrical Engineering, University of Ljubljana. His current work focuses on explainable AI in the field of face recognition. His research interests include image processing, biometrics, and machine learning.

BERK KEMAL ÖZATA received the master's degree in computer science from Istanbul Technical University (ITU), Türkiye, in 2023. During the master's program, he worked on low-resolution face recognition, super-resolution, and video object tracking. Currently, he works as a Computer Vision Engineer at ASELAN Inc., focusing on image processing and computer vision solutions in embedded systems.

ALPEREN KANTARCI received the bachelor's and master's degrees in computer science from Istanbul Technical University (ITU), Türkiye. He is currently pursuing the Ph.D. degree with Goethe University Frankfurt. During the bachelor's and master's studies, he worked on face biometrics, signature verification, and anomaly detection in industrial production. His current work focuses on explainable and transparent AI systems for education and understanding cognitive processes.

VITOMIR ŠTRUC (Senior Member, IEEE) is currently a Full Professor with the University of Ljubljana, Slovenia. His research interests include biometrics, computer vision, image processing, pattern recognition, and machine learning. He co-authored more than 150 research papers for leading international peer-reviewed journals and conferences in these and related areas. He has served in different capacities on the organizing committees of IEEE Face and Gesture, ICB, WACV, and IJCB. He is a member of IAPR, EURASIP, and Slovenia Ambassador for the European Association for Biometrics, and the Former President and a Current Executive Committee Member of the Slovenian Branch of IAPR. He is also the VP Technical Activities for the IEEE Biometrics Council. He is a Senior Area Editor of IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, a Subject Editor of *Signal Processing* (Elsevier), and an Associate Editor of *Pattern Recognition*, *EURASIP Journal on Image and Video Processing*, and *IET Biometrics*. He served as the Area Chair for WACV, ICPR Eusipco, and FG, and the Program Chair for ISPA, IWBF, and IJCB. He currently acts as the General Co-Chair for IJCB 2023 and the Program Co-Chair for FG 2024.

HAZIM KEMAL EKENEL received the Ph.D. degree in computer engineering from the University of Karlsruhe (TH). He is currently a Full Professor with the Department of Computer Engineering, Istanbul Technical University. He has over 20 years of experience in computer vision and machine learning with a focus on face analysis and human perception. He has received the IEEE Türkiye Chapter Research Award, in 2019; the Parlar Foundation Research Award; and the Outstanding Young Scientist Award from the Science Academy, Türkiye, in 2018. In 2008, he received the EBF European Biometric Research Award. He is one of the co-organizers of the IEEE International Conference on Automatic Face and Gesture Recognition (FG 2024); the International Workshop on Face and Gesture Analysis for COVID-19@FG 2021; the MULTI-modal Imaging of FOREnsic Science Evidence Tools Conference, in 2021; the Face Analysis for Advanced Driver Assistance Systems@FG 2019; the Turkish German Multimodal Interaction Summit, in 2014; the Affective Computing for Mobile HCI Workshop, in 2012; and the Benchmarking Facial Image Analysis Technologies Workshops, in 2011 and 2012.