IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 27 January 2024, accepted 16 February 2024, date of publication 20 February 2024, date of current version 26 February 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3367763

RESEARCH ARTICLE

Enhancing Upscaled Image Resolution Using Hybrid Generative Adversarial Network-Enabled Frameworks

R. GEETHA[1], (Member, IEEE), G. BELSHIA JEBAMALAR[1], S. ARUMAI SHINEY[1], NHU-NGOC DAO[2], (Senior Member, IEEE), HYEONJOON MOON[2], AND SUNGRAE CHO[3]

[1]Department of Computer Science and Engineering, S. A. Engineering College, Chennai 600077, India
[2]Department of Computer Science and Engineering, Sejong University, Seoul 05006, South Korea
[3]School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea

Corresponding authors: Nhu-Ngoc Dao (nndao@sejong.ac.kr) and Sungrae Cho (srcho@cau.ac.kr)

This work was supported in part by the Ministry of Science and ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information Communications Technology Planning Evaluation (IITP) under Grant IITP-2024-RS-2023-00258639; and in part by the National Research Foundation of Korea (NRF) funded by Korean Government (MSIT) under Grant RS-2023-00209125.

ABSTRACT Upscaling images typically depends on facial features, such as facial geometry details or references, to rebuild reasonable details. The low-quality input images cannot offer accurate geometric details, and the high-quality details are obscure, limiting performance in practical conditions. This work addresses the problem using an enhanced version of the generative adversarial network (GAN) model that uses rich and varied facial features incorporated into the pretrained StyleGAN2 for face restoration. The generated facial features are integrated into the facial restoration process using spatial feature transform layers to achieve facial details and color quality to improve reliability. In particular, the image background is upsampled using a modified enhanced super-resolution GAN trained in parallel to remove noise while recreating a high-resolution image from low-quality input. The upscale image resolution GAN is used to enlarge the image for a clear view without sacrificing the original image quality. Finally, by combining the upscaled background and restored faces, the hybrid GAN-enabled framework can obtain high-resolution upscaled images. An experimental comparison has been conducted using the Flickr-Faces-HQ Dataset collected from Kaggle. The findings indicate that the suggested framework outperforms existing methods in terms of reconstruction, adversarial, and facial component loss metrics as well as similarity indexes.

INDEX TERMS Image enhancement, generative adversarial network, upscaled image resolution.

I. INTRODUCTION

A. MOTIVATION

The goal of machine learning [1], a branch of artificial intelligence (AI) and computer science, is to simulate human learning using data and algorithms, gradually improving the model's accuracy. Deep learning [2] is a subtype of machine learning. Deep learning approaches can tackle any pattern recognition problem without the assistance of a person. Deep learning, the fastest-growing discipline of machine learning, is being used by more firms to develop new business models. In this regard, a generative adversarial network (GAN) [3], a kind of deep neural network, produces new data that looks exactly like the data that is trained. A generator, which is used to train, as a result, it generates new examples. On the other hand, a discriminator will take those examples and classify those examples as true and false based on the domain and generated. Typically, the generator and discriminator are the two sub-models of the supervised learning algorithm. The GANs is a realistic and exciting tool used to create objects,

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

TABLE 1. Literature review comparison.

Reference	Algorithm/ Methodology	Advantages	Disadvantages
[4]	SRGAN	Recover the finer texture details	computational complexity
[5]	Swin Transformer blocks (RSTB)	Several swin with better accuracy	Reduces memory
[6]	Swin Transformer	On different tasks Swinger beats the state-of-the-art methods by 0.14 to 0.45dBlevel	Performance is lower
[7]	(MRDGAN) Multi scale residual GAN	Effective optimization and improved clinical diagnostic value	Sometimes it results in unstable condition
[8]	Labeled wild face (LWF) and 3D Face reconstruction	Not only in Light it can even works in complete dark mode	It is worst in case of different scenarios
[9]	Postural Transformer Network and Face Generative Adversarial Networks	Reconstruction speed is fast	Reconstruction speed is fast
[10]	Semi-Cycled Generative Adversarial Networks (SCGAN)	Accurate in nature	It works on single output mapping
[11]	Wasserstein generative adversarial network (WGAN)	Stable	Does not compute higher dimension.
[12]	Deep convolutional neural networks	Computation is minimum	Cost is high
[13]	single-image super-resolution (SISR)	Consistent clarity	Flexibility is poor
[14]	Feedforward convolutional neural network	When compared to the optimization method, this is 3x faster	It takes time to produce a high-resolution image since it involves three steps.
[15]	Convolutional Neural Network	Visual improvements against the state of art methods	Accuracy is bit lower
[16]	DCNN	Restoration quality is good and achieves fast speed for practical online usage	Accuracy is less though it is fast
[17]	Traditional sparse-coding-based SR	Speed is improved by 40 times with much superior restoration quality	The output of this project is unrealistic.
[18]	Deconvolution layer	Diverse photorealistic output as scenes with multimodel synthesis are provided along with guided image synthesis	Still, application for multimodal and guided image synthesis is pending
[19]	MWCNN	Experiment results in the effectiveness of the MWCNN for JPEG images mainly for artifact removal, single image super resolution and image denoising	Very less training data set for image segmentation
[20]	WGAN-GP	Wasserstein GAN trains the various architecture and provides a metric with its trained progress	the depth of the network is increased
[21]	SISR model with DCNN	Fast and accurate where the local and global features are reconstructed with less computation	Handling large and non uniform blur kernels performance and requirements for relieving the face alignment is low.
[22]	SAR-GAN, OCTA Technology	Has wider utility range and robust completely	Difficulties is faced on the choroidal vascular network and Deep capillaries
[23]	CGAN	Algorithm is precise and showed improved in peak signal to noise ratio	Sometimes it contradicts the unsupervised model.

photorealistic images, and scenes. The tool provides realistic examples in all those problem domains. It can even do notable translations from image to image, for example, converts the images of summer to winter and vice versa, as well as day to night.

Facial restoration is the process of reconstructing high-quality facial images from low-quality inputs, which may suffer from such problems as low resolution, blurriness, artifacts, and noise, causing unsatisfaction in alignment with standard quality threshold. This occurs due to some kind of defect or because of poor durability. This task becomes particularly challenging in real-time scenarios due to the complex nature of facial deterioration, variations in expressions, and changes in facial positions [24], [25]. Recent research has significantly relied on critical geometric facial aspects, such as facial features, analysis maps, and facial temperature maps, to enhance facial precision and details during restoration [26], [27], [28], [29]. Although these methods obtain remarkable performance in processing well-defined facial images, most are negatively affected by the low-quality input encountered in real-world applications. Their practical utility is further limited by the absence of high-resolution reference data and dictionaries and the lack of diversity and comprehensive information.

B. LITERATURE REVIEW

The GAN, a kind of deep neural network, was introduced to cope with the mentioned problem and can produce realistic new face images based on the pretrained data. For instance, Dou et al. proposed a novel approach to get images with high quality by improving the perceptual loss functions built on high features retrieved from pretrained networks [4]. The proposed approach combines the advantages of strategies and feedforward networks for image transformation tasks using perceptual loss functions. A supervised feedforward convolutional neural network (CNN) was trained using a per-pixel loss characteristic to distinguish between output and ground-truth reality images. Liang et al. proposed the

SwinIR model, an image restoration method acquired from the swin transformer [5]. This model consists of three parts: high-quality image reconstruction,shallow feature extraction, and deep feature extraction. The deep feature extraction module is constituted by several residual swin transformer blocks, which are linked together by residual connections. In [6], low-resolution photos are used as the input for a deep convolutional neural network (DCNN) mapping algorithm, which uses a single image super-resolution (SISR) technique to output high-resolution images. In [7], the original structure of the GAN was improved along with the multiscale residual block introduction. In this model, the information is interleaved between the streams, and the network with a loss function structure is changed to obtain the relative probabilities. According to [8], Hangaragi et al. proposed a three-step framework to identify and recognize faces. The framework first detects human faces, and then uses the Blaze face model to calculate the face landmarks. Subsequently, 3D face reconstruction has been performed by the face mesh method. The produced 3D face meshes are fed into a DNN for face recognition. The labeled wild face (LWF) dataset photos and real-time photographs are utilized to verify the proposed framework, which exposes 94.23% accuracy in facial recognition. In [9], Wei et al. combined three techniques including head pose estimation, postural transformation, and face recognition for facial restoration from low-resolution images. The proposed model shows advantages in improving visual experience. In [10], the shared restoration branch regularized by both the forward and backward cycle-consistent learning processes allows the proposed semi-cycled GAN (SCGAN) model to achieve accurate and robust face restoration performance while mitigating the negative effects of the domain gap between the real-world low-resolution face images and the synthetic ones. In [11], Ma et al. proposed the rectified Wasserstein generative adversarial network (ReWaGAN) for image restoration. The ReWaGAN model modifies the original WGAN by rectifying the gradients in the generator training. Accordingly, a link between the ReWaGAN and the perception-distortion tradeoff component is established to optimize the restoration quality. A summary of related work comparison has been illustrated in Table 1. Although existing work has significantly improved the qualify of image restoration, most of them are a one-sided focus, which upscales either the whole image or human faces separately.

C. PAPER CONTRIBUTIONS

According to the analysis, an effective model for improving image resolution on both human faces and backgrounds with greater precision is considered an open challenge. Thus, this paper aims to enhance upscaled image resolution by proposing a hybrid GAN-enabled framework. The framework performs the upscaled background enhancement and recovers the face GAN models simultaneously to restore a high-resolution upscaled image. In particular, the upscaled background of the image is obtained using a modified enhanced super-resolution GAN (ESRGAN) to remove noise while producing a high-resolution image from low-quality input. In the next stage, the upscale image resolution GAN (UIRGAN) model enhances facial traits while upscaling, where the upscaling background zooms the content from the original image without affecting its current resolution. For low-resolution images, to display in larger formats, the gaps in the image pixels can be filled. Depending on the current computing capacity of the physical hardware, UIRGAN-based facial restoration is configured with higher priority, whereas the background upscaling operation is optional.

II. HYBRID GAN-ENABLED FRAMEWORK

A. DATA ENGINEERING

Dataset collection: The dataset collection plays an important role in this proposed work. The data were collected from the Flickr-Faces-HQ Dataset including 7200 images on Kaggle [30]. The training was conducted with 80% of the dataset, and 20% was used for the testing dataset.

Dataset preparation and preprocessing: In this phase, raw data were used for the deep learning model. A data preprocessing task was conducted to clean and format the data satisfactorily. The steps were to label the location of the eyes, nose, and mouth landmarks and then preprocess the images to refine colors and sharpness.

B. FRAMEWORK DESIGN

GAN generally has two minimax game players those are a discriminator and a generator [31]. The generative network G generates fresh data $G(x)$ from a noisy sample $x \sim p(x)$ which has been taken from a normal distribution, with a distribution p_g that should resemble the data distribution, p_{data}. In the meanwhile, the discriminating network D is used. $y \sim p_{data}(y)$ as the actual data sample and the sample $G(x)$ is approximately $p_g(G(x))$ produced. Within the initial GAN, the formulation of this adversarial training procedure was

$$\min_G \max_D \mathbb{E}_{x\sim p_{data}}[\log D(x)] + \mathbb{E}_{y\sim p_y}[\log(1 - D(G(y)))].$$

First, data engineering is conducted to prepare and preprocess the necessary dataset for GAN model training in the hybrid framework. Then, the corresponding GAN models for facial restoration and background upscaling are trained and merged into a final upscaled image, as illustrated in Figure 1. The framework is designed to prioritize the facial upscaling and restoration stage. In contrast, the background upscaling and refinement stage is optional, depending on computing resource availability, to adapt to differences in computing performance.

In the initial stage, the user uploads a low-resolution image with blurring, noise, and unwanted artifacts through the front end. As the graphics processing unit is available in the backend/server, this proposed system upscales and enhances the image background; if not, the system upscales and enhances only the faces in the image. The system detects faces in the uploaded image and crops the faces in the image,

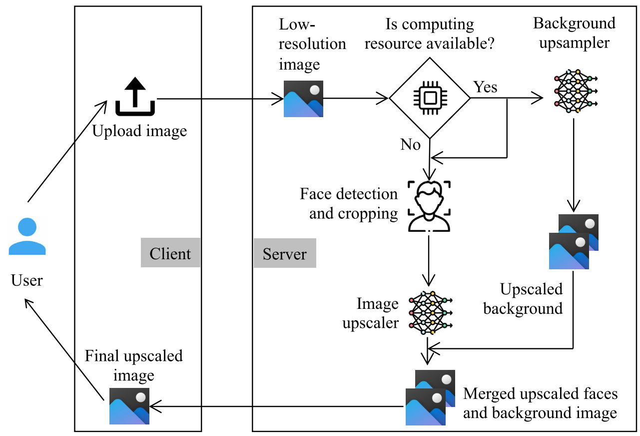


FIGURE 1. **Hybrid generative adversarial network (GAN)-enabled framework.**

Algorithm 1 UIR GAN Pseudocode

Input: Low-Resolution Facial Image

//Procedure//

Load image as input
Detect faces in the image given as input
Loop over each face detected in the input image
Extract facial landmarks in the current face
Crop the input image to contain only the current face
Remove image degradation using U-Net
Transform F-latent features into intermediate representation
Apply channel split feature transformation to F-spatial features
Feed W (Intermediate Image) to the pre-trained StyleGan2 model to generate high-quality faces
Merge the enhanced faces and the original background
Apply post-processing techniques
Replace the current face with the processed image

Output: Enhanced Upscaled Image

which are input to the image to be upscaled. The upscaled facial image provides upscaled faces, whereas the image background upscaling provides an upscaled background; thus, these images are merged accordingly and input to the front end. A pseudocode of the operation is described in Algorithm 1.

C. FACIAL RESTORATION

Model training: A combination of the U-Net CNN architecture and pretrained StyleGAN2 model is used to create a custom model with the mentioned dataset. The images are segmented for facial detection using the U-Net, and the faces in the image are cropped and resized to 512 × 512 pixels to maintain high-quality and photo-realistic. Bilinear upsampling has been regarded as one of the effective options for collecting features that are photo-realistic since it interpolates pixel values depending on nearby locations smoothly. Choosing a resolution of 512 × 512 offers a high pixel density, which makes it possible to capture detailed elements and helps produce a more accurate representation of the visual information. Moreover, bilinear upsampling makes it easier to preserve minute details like delicate textures and subtle subtleties, which results in a more realistic and nuanced representation of the image. In general, this pairing of 512×512 with bilinear upsampling enhances the fidelity of the image, especially for applications where maintaining both high and low-level details is crucial, such as in photo-realistic rendering and computer vision tasks [32]. The cropped faces are input into the StyleGAN2 model to restore the upscaled faces. Finally, the restored face is pasted to the upsampled background image (see Figure 2).

Recognizing facial landmarks: The faces are detected, and their landmarks are determined using the *facexlib* package with the combination of prior facial boxes and the eye distance threshold. The facexlib package aims to offer a convenient solution for face-related tasks by integrating state-of-the-art (SOTA) open-source methods. The package appears to focus on real-time facial landmark detection with high-quality predictions. In our model, the critical facial landmarks are the eyes and mouth. These features are then extracted as data points.

Enhancing and upscaling faces: This module contains four submodules: degradation removal, the pretrained StyleGAN2 with latent code mapping, the channel-split spatial feature transform, and loss functions.

- Degradation removal. The degradation removal module removes blur, low-resolution, noise, and JPEG artifacts and extracts the clean features of the image as F-latent and F-spatial features. The U-Net architecture is employed as a degradation removal module. The U-Net helps eliminate blur, increase receptive fields, and generate features of multiple resolutions. The F-latent features are transformed into an intermediate representation, W, and are input into the pretrained StyleGAN2. The F-spatial features are input into the corresponding channel-split feature transformation modules to scale and shift the output at each stage of the generative process.
- Pretrained StyleGAN2 with latent code mapping. Pretrained facial GANs offer numerous and rich facial elements for this task, capturing the distribution over faces using their convolutional weights. A standard method involves attempting to translate the image to the nearest latent coding Z before using a trained GAN to create the appropriate output. These approaches typically require time-consuming iterative optimization to maintain feature mapping. Instead of directly building the final image, we built transient convolutional characteristics on the closest face because it has more facial characteristics and can later be tuned and manipulated using input variables to produce superior facial features. With higher semantics, the coded vector of the input image is transferred to intermediate latent code W

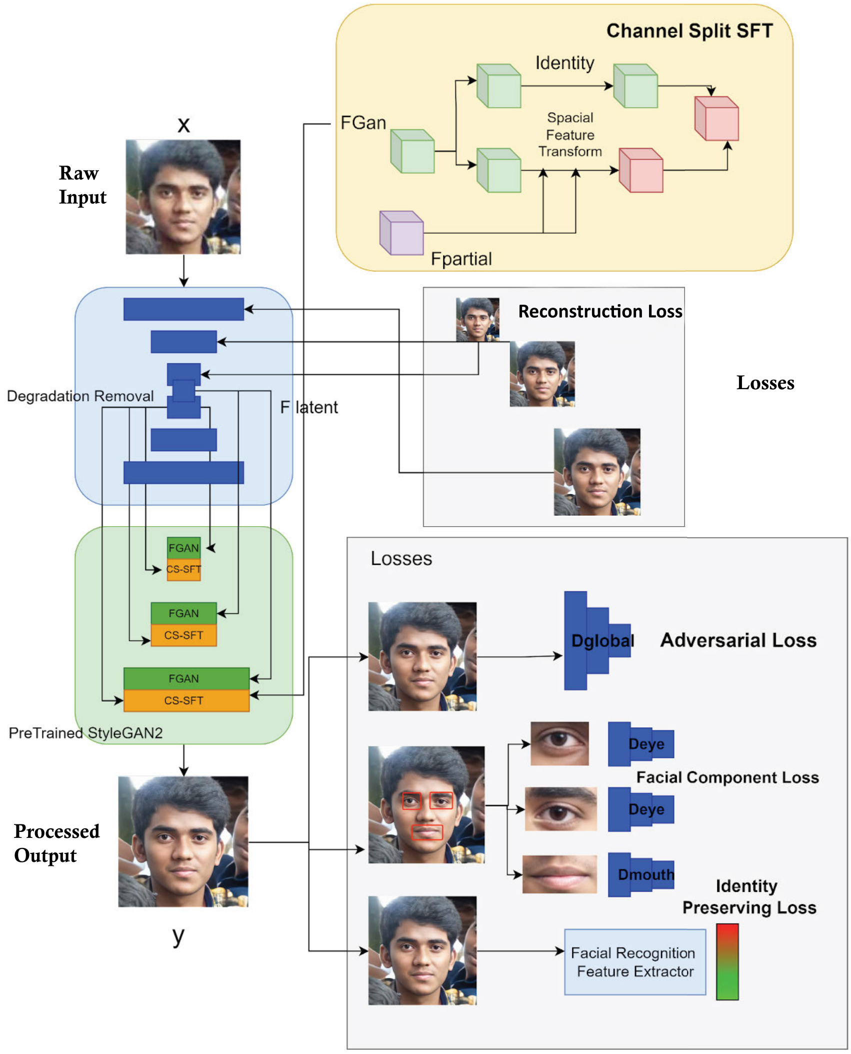


FIGURE 2. Facial image upscaling model architecture.

(i.e., the intermediate space is transformed to the latent code Z with several multilayer perceptron layers). The latent code W is subsequently delivered into each convolutional layer of the pretrained StyleGAN2 and generates GAN features for each resolution scale.

- Channel-split spatial feature transform. Preserving spatial data from the input is vital for face resampling and restoration because it generally requires neighboring characteristics to maintain facial features and adaptive restoration at distinct spatial regions of the face. Thus, we employed the spatial feature transform, creating affine transform parameters for spatially intelligent feature modulation, which is effective in incorporating numerous conditions in picture healing and photograph creation.
- Model objectives and loss functions: The learning objectives of the network training consist of the following: lossy reconstruction that limits near-ground-truth output, unfavorable loss for facial reconstruction, realistic textures, lossy facial components to further enhance facial features and details, and loss of identity preservation. Reconstruction loss (the most commonly used perception loss) and L1 loss are used as the reconstruction loss (L rec). Adversarial loss improves the resolution in favor of natural image diversity and realistic elements, which is similar to the logistic loss of StyleGAN2. Facial component loss increases the perceptually significant facial components and is characterized by the absence of a part, including local discriminators for the nose, left eye, right eye, and mouth. By alignment, the image is reduced to a region of interest. Small local discriminators are taught for each region to determine whether the recovered regions are real and near the natural facial components. Identity-preserving loss, similar to perception loss, is based on embedding facial features into a pretrained ArcFace face recognition model to capture the most vital and abundant features for identification. Because of the absence of

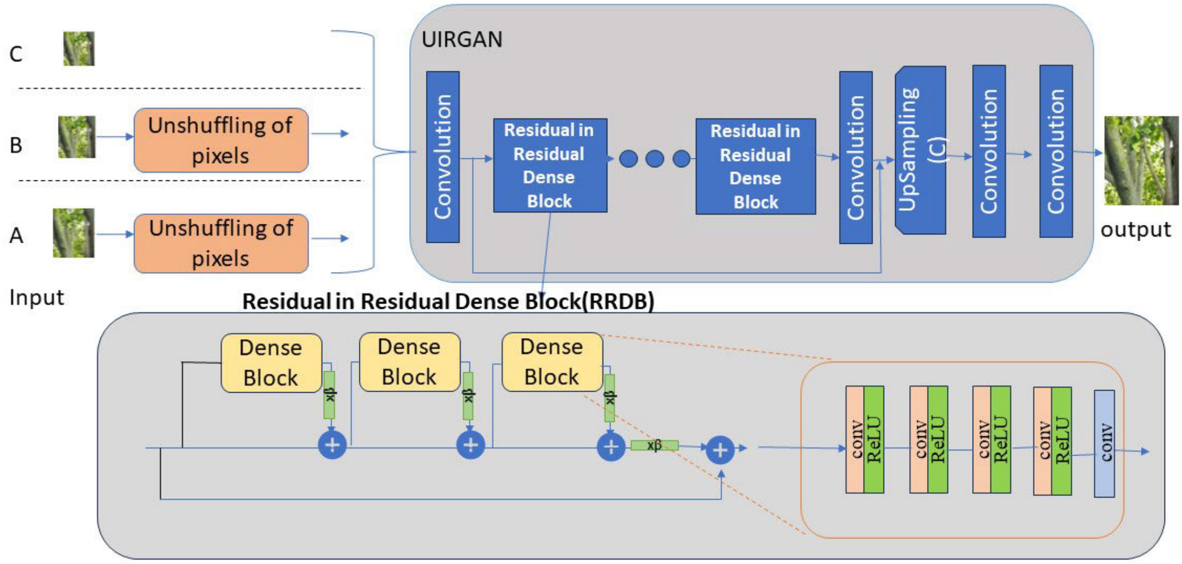


FIGURE 3. Background upscaling model architecture.

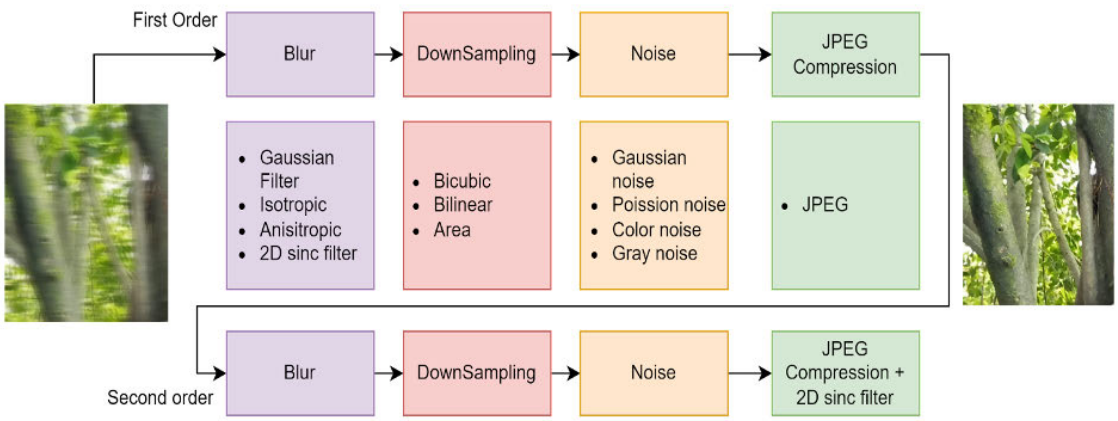


FIGURE 4. Background image consolidation.

identity conservation, the retrieved result must be similar to the genuine image in the deep feature space.

D. BACKGROUND UPSCALING

Networks and training: The generator is designed based on a modified ESRGAN with the addition of unshuffled pixels to reduce the local size and increase the channel size ahead of the main modified ESRGAN network. The U-Net discriminator is used in conjunction with the SN standard to obtain additional discriminative power with stable training. The upscaled cropped faces from the input image are pasted back into the upscaled background of the input image.

Enhancing and upscaling background: The modified ESRGAN architecture uses a deep network with multiple residual dense blocks. It uses the unshuffle-pixel operation to reduce the spatial size and reorganize the information according to the channel dimension for scale factors of 4, 2, and 1. The modified ESRGAN also improved the visual geometry group or VGG-style discriminator in the connection-skipping U-Net design. The U-Net provides precise feedback to the generator for each pixel and outputs realistic values for each pixel. The upscale image resolution procedure can boost the image resolution, making it sharper and more detailed, as illustrated in Figure 3. The smoother, less pixelated appearance of the upscaled image is more appealing. This procedure also views images on a large-format printer or a screen with high resolution. Grainy or fuzzy photographs can be effectively fixed to improve the viewing experience, which can make the image appear to have been captured using a high-resolution camera. Enhancing pictures that frequently lack clarity and depth ability increases the size of photographs without sacrificing the quality.

The first-order or classical degradation model consists of a Gaussian blur followed by a resizing or downsampling operation. Poisson noise, Gaussian noise, and color noise simulate camera-sensor artifacts. Additionally, JPEG

TABLE 2. Loss comparison.

Methods	Reconstruction loss	Adversarial loss	Facial component loss
DeBlurGAN v2	0.4001	52.69	4.917
HiFaceGAN	0.4770	66.09	4.916
PSFRGAN	0.4240	47.59	5.123
Proposed framework	0.3646	42.62	4.277

compression is used for lossy compression. A higher-order or second-order degradation model is applied to the image twice in succession. Near-sharp transition ringing artifacts appear to be a false edge among the ringing and overshoot artifacts. Near the boundaries, they appear as bands or ghosts. Overshoot artifacts are frequently coupled with sound artifacts, appearing as a further transition to the change in the edge. The primary cause of these aberrations is the limited signal, except for high frequencies.

The sync filter is used to remove high-frequency frequencies to integrate ringing and decoding artifact training pairs. The filter is used in the blurring process and the final step of merging, where its structure is randomly changed using JPEG compression, as depicted in Figure 4.

III. RESULTS AND DISCUSSION

To investigate the performance of the proposed framework, we compared the framework with the following models: DeBlurGAN v2 [33], HiFaceGAN [34], and PSFRGAN [35].

1) RECONSTRUCTION LOSS

Reconstruction loss is a vital metric that computes the similarity of a generated image to the original image, thereby quantifying how much the generated images match reality based on direct observations of the original images. After just 10 epochs, the reported loss value is 0.47 as shown in Figure 5(a). This starting value is remarkably low, partly because the model underwent pretraining, giving it a strong base for comprehending a variety of data. As the model is continually trained over time, the loss value's diminishing trend holds true. This is explained by the fact that the model was trained on a variety of datasets, honing its capacity to recognize minute details and minor differences in the photos. Notably, this loss function excels at detecting variations between images at the pixel level, making it especially skilled at detecting even slight alterations between the generated and original images. 50 complete epochs later, the reconstruction loss value has significantly decreased to 0.342. The model is becoming increasingly adept at producing images that closely resemble the reality shown in the original photographs, as evidenced by the loss value's significant decline. As seen by the diminishing reconstruction loss over time, it is clear that ongoing training and exposure to a variety of data have further honed the model's ability to create realistic images of high quality.

2) ADVERSARIAL LOSS

During the upscaling process, the adversarial loss plays a very important part in determining how realistic the recovered part of the textures are. Similarities exist between this loss function and the logistic loss function used in StyleGAN2, a well-known generative model for image synthesis. A continuous and exact downward trend can be seen in Figure 5(b), In the training session, it initially starts with an initial value of 67 and gradually declines to reach 42 at the end. It is significant that this loss function describes adversarial loss across all of the weights in the generator network. It emphasizes the value of facial details in the upscaling process by placing a greater emphasis on the evaluation of the average loss resulting from features taken from facial areas within the cropped images. This multifaceted strategy guarantees that the upscaling process maintains fidelity to the finer aspects of the content being generated.

3) FACIAL COMPONENT LOSS

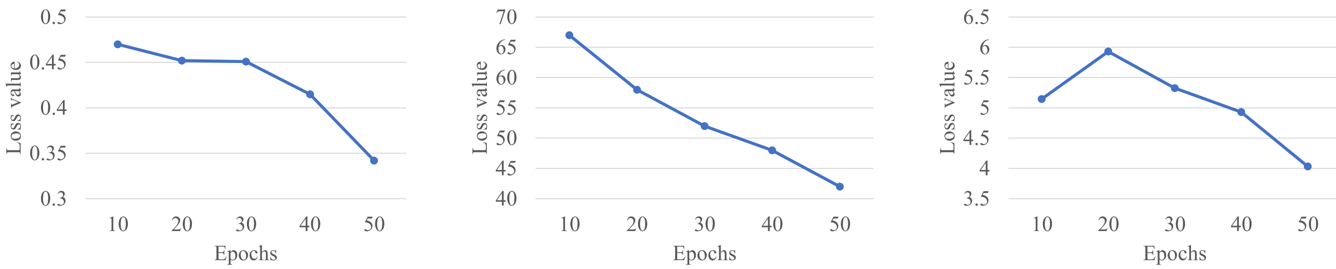
The left eye, right eye, and mouth were some of the examples of the local discriminators that are computed for the facial component loss, which is an essential factor in strengthening particular facial features. This loss primarily computes the difference between the local features of the face in the photos which has been not cropped. The trajectory of this loss, which begins at a starting value of 5.145 and slowly those are decreased to 4.032 at the end of the training period, can be shown in Figure 5(c). This loss function is used to provide a rough estimation of the losses experienced within the cropped faces based on areas of interest, highlighting the enhancement and preservation of certain specific facial features. It assures that the local face traits are preserved by observing and reducing these losses.

A. PERFORMANCE COMPARISON WITH EXISTING WORK

Table 2 compares the original image with the results of DeBlurGAN v2, HiFaceGAN, PSFRGAN, and the proposed framework concerning reconstruction, adversarial, and facial component loss. In particular, DeBlurGAN v2 tends to remove blurring but is unable to preserve the image features. In addition, HiFaceGAN produces good facial features, but the images contain blurring and artifacts. Further, PSFRGAN produces images without artifacts and blurring but lacks details in the region of interest of the face. The proposed framework has satisfactory facial features and improves the overall quality of the image due to a combination of the modified ESRGAN with the sync filter for upsampling nonfacial regions. The comparative quantitative performance analysis regarding the peak structural similarity index measure,signal-to-noise ratio and learned perceptual image patch similarity is illustrated in Figure 6. The result of the original image processed by the proposed framework was compared with the results of DeBlurGAN v2, HiFaceGAN, and PSFRGAN. The results of the proposed framework reveal significant improvement over the other GAN models.

1) PEAK SIGNAL TO NOISE RATIO

Peak Signal-to-Noise Ratio (PSNR) assesses quality of the image by comparing maximum signal intensity to interfering



(a) Reconstruction loss (b) Adversarial loss (c) Facial component loss

FIGURE 5. Loss evaluation.

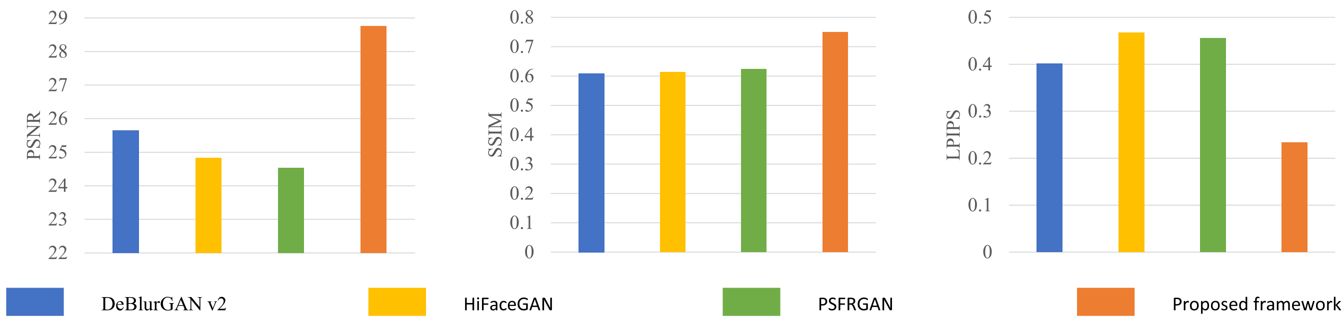


FIGURE 6. Quantitative performance analyses.

noise. The presence of the degradation reduction module within the framework is primarily responsible for the high PSNR value. This specialized component is essential in improving image quality by successfully minimizing various types of degradation such as blur, noise, and artifacts. This indicates a significant reduction in noticeable differences between the original and processed images. The reason for the high PSNR value is mainly because of the degradation removal module. The degradation removal module leads to a substantial improvement in image quality. The mentioned module's primary goal is to improve the quality of an input signal or image by minimizing or eliminating degradation. The mentioned module is built into the proposed framework's U-Net design. This module demonstrates a high level of proficiency in effectively reducing or eliminating undesired effects in images. Specifically, it excels in the tasks of minimizing blurriness, reducing noise, and eliminating artifacts, thereby significantly enhancing the overall quality of the image. As a result, the PSNR value exceeds that of other models. This results in a significant improvement in overall image quality. The U-Net's design, distinguished by its U-shaped architecture, excels at picture segmentation. Blurriness, noise, and unpleasant visual distortions are all addressed by the degradation removal. The images produced through this process exhibit a higher degree of accuracy in portraying the original image. The generated photos are more accurate representations compared to the original photo. This improvement is measured by the amplified PSNR value, which indicates a better approximation of the original value in the processed image. In comparative evaluations, the U-Net framework outperforms other models that lack such a dedicated degradation removal module.

2) STRUCTURAL SIMILARITY INDEX

The Structural SIMilarity (SSIM) index is used to evaluate image quality in a particular way. The U-Net model gives detailed feedback to the generator on a per-pixel basis, resulting in accurate values for each individual pixel. By upscaling the image's resolution, this painstaking procedure considerably improves its quality. The findings of this investigation clearly show that, when compared to other models, the similarity index is significantly higher. This represents a significant improvement in the similarity between the generated images and the original reference photographs. Because of the U-Net's capacity to provide detailed per-pixel feedback, it is possible to create visuals with a surprising level of realism and precision.

3) LEARNED PERCEPTUAL IMAGE PATCH SIMILARITY

The Learned Perceptual Image Patch Similarity (LPIPS) metric differs from traditional metrics such as PSNR or SSIM by focusing on high-level visual aspects, which better match how humans perceive picture quality. Unlike pixel-wise comparisons, LPIPS analyzes images based on their perceptual resemblance, which is more in line with human judgment. The improved Enhanced Super-Resolution

Generative Adversarial Network (ESRGAN) architecture, enhanced by the incorporation of several residual dense blocks, exhibits extraordinary skill in gathering detailed, fine-grained pixel-level information. This enhancement enables the model to recognize and recreate detailed features with amazing precision. The combined effect of using LPIPS and enhancing the ESRGAN architecture with these residual dense blocks has proven to be especially effective for tasks that require an accurate representation of pixel-level nuances, significantly contributing to the overall quality and fidelity of the generated image.

IV. CONCLUDING REMARKS

This paper proposed a hybrid GAN-enabled framework to enhance upscaled image resolution by subsequently processing the facial and background components. The experimental comparison demonstrates that the proposed framework outperforms prior state-of-the-art methods in combined facial restoration and background upscaling for real-world images. However, the framework suffers from certain limitations, such as dataset bias and hyperparameters. Future work can be extended to address sensitivity to hyperparameters using various data augmentation techniques and additional methods, such as automatic hyperparameter optimization techniques, and use transfer learning from pretrained models to overcome these problems.

REFERENCES

[1] R. Geetha, K. Ramyadevi, and M. Balasubramanian, "Prediction of domestic power peak demand and consumption using supervised machine learning with smart meter dataset," *Multimedia Tools Appl.*, vol. 80, no. 13, pp. 19675–19693, May 2021.

[2] R. Geetha, T. Thilagam, and T. Padmavathy, "Effective offline handwritten text recognition model based on a sequence-to-sequence approach with CNN-RNN networks," *Neural Comput. Appl.*, vol. 33, no. 17, pp. 10923–10934, Sep. 2021.

[3] A. Aggarwal, M. Mittal, and G. Battineni, "Generative adversarial network: An overview of theory and applications," *Int. J. Inf. Manage. Data Insights*, vol. 1, no. 1, Apr. 2021, Art. no. 100004.

[4] H. Dou, W. Zhang, and X. Liang, "Image super resolution reconstruction algorithm based on deep learning," in *Proc. IEEE 2nd Int. Conf. Inf. Commun. Signal Process. (ICICSP)*, Sep. 2019, pp. 306–310.

[5] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image restoration using Swin Transformer," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jun. 2021, pp. 1833–1844.

[6] O. Soufi, Z. Aarab, and F.-Z. Belouadha, "Benchmark of deep learning models for single image super-resolution (SISR)," in *Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET)*, Mar. 2022, pp. 1–8.

[7] J. Sun, Z.-Y. Li, P.-C. Li, H. Li, X.-W. Pang, and H. Wang, "Improving the diagnostic performance of computed tomography angiography for intracranial large arterial stenosis by a novel super-resolution algorithm based on multi-scale residual denoising generative adversarial network," *Clin. Imag.*, vol. 96, pp. 1–8, Apr. 2023.

[8] S. Hangaragi, T. Singh, and N. Neelima, "Face detection and recognition using face mesh and deep neural network," *Proc. Comput. Sci.*, vol. 218, pp. 741–749, Jan. 2023.

[9] T. Wei, Q. Li, Z. Chen, and J. Liu, "FRGAN: A blind face restoration with generative adversarial networks," *Math. Problems Eng.*, vol. 2021, pp. 1–14, Oct. 2021.

[10] H. Hou, J. Xu, Y. Hou, X. Hu, B. Wei, and D. Shen, "Semi-cycled generative adversarial networks for real-world face super-resolution," *IEEE Trans. Image Process.*, vol. 32, pp. 1184–1199, 2023.

[11] H. Ma, D. Liu, and F. Wu, "Rectified Wasserstein generative adversarial networks for perceptual image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 3, pp. 3648–3663, Mar. 2023.

[12] X. Hu, W. Ren, J. LaMaster, X. Cao, X. Li, Z. Li, B. Menze, and W. Liu, "Face super-resolution guided by 3D facial priors," in *Proc. 16th Eur. Conf. Comput. Vis. (ECCV)*, Glasgow, U.K. Cham, Switzerland: Springer, Aug. 2020, pp. 763–780.

[13] Y. Yan, W. Ren, X. Hu, K. Li, H. Shen, and X. Cao, "SRGAT: Single image super-resolution with graph attention network," *IEEE Trans. Image Process.*, vol. 30, pp. 4905–4918, 2021.

[14] Y. Zeng, S. Peng, and S. Ruan, "Research on image super resolution reconstruction model based on deep learning," in *Proc. IEEE Int. Conf. Adv. Electr. Eng. Comput. Appl. (AEECA)*, Aug. 2020, pp. 912–917.

[15] C. Song, Z. He, Y. Yu, and Z. Zhang, "Low resolution face recognition system based on ESRGAN," in *Proc. 3rd Int. Conf. Appl. Mach. Learn. (ICAML)*, Jul. 2021, pp. 76–79.

[16] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," 2018, *arXiv:1804.02815*.

[17] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 334–355.

[18] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2337–2346.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," 2015, *arXiv:1505.04597*.

[20] M. S. M. Sajjadi, B. Scholkopf, and M. Hirsch, "EnhanceNet: Single image super-resolution through automated texture synthesis," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 4491–4500.

[21] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," 2018, *arXiv:1803.03345*.

[22] X. Yuan, Y. Huang, L. An, J. Qin, G. Lan, H. Qiu, B. Yu, H. Jia, S. Ren, H. Tan, and J. Xu, "Image enhancement of wide-field retinal optical coherence tomography angiography by super-resolution angiogram reconstruction generative adversarial network," *Biomed. Signal Process. Control*, vol. 78, Sep. 2022, Art. no. 103957.

[23] Z. Song, D. Qiu, X. Zhao, D. Lin, and Y. Hui, "Channel attention generative adversarial network for super-resolution of glioma magnetic resonance image," *Comput. Methods Programs Biomed.*, vol. 229, Feb. 2023, Art. no. 107255.

[24] A. Kammoun, R. Slama, H. Tabia, T. Ouni, and M. Abid, "Generative adversarial networks for face generation: A survey," *ACM Comput. Surv.*, vol. 55, no. 5, pp. 1–37, May 2023.

[25] L. M. Dang, S. I. Hassan, S. Im, and H. Moon, "Face image manipulation detection based on a convolutional neural network," *Expert Syst. Appl.*, vol. 129, pp. 156–168, Sep. 2019.

[26] L. Feihong, C. Hang, L. Kang, D. Qiliang, Z. Jian, Z. Kaipeng, and H. Hong, "Toward high-quality face-mask occluded restoration," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 19, no. 1, pp. 1–23, Jan. 2023.

[27] Y. Hu, Y. Wang, and J. Zhang, "DEAR-GAN: Degradation-aware face restoration with GAN prior," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 9, pp. 4603–4615, Sep. 2023.

[28] Y. Luo, D. Pi, Y. Pan, L. Xie, W. Yu, and Y. Liu, "ClawGAN: Claw connection-based generative adversarial networks for facial image translation in thermal to RGB visible light," *Expert Syst. Appl.*, vol. 191, Apr. 2022, Art. no. 116269.

[29] H. Bouzid and L. Ballihi, "Facial expression video generation based-on spatio-temporal convolutional GAN: FEV-GAN," *Intell. Syst. with Appl.*, vol. 16, Nov. 2022, Art. no. 200139.

[30] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4396–4405.

[31] C. Wang, C. Xu, X. Yao, and D. Tao, "Evolutionary generative adversarial networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 6, pp. 921–934, Dec. 2019.

[32] Y. Guo, Q. Chen, J. Chen, Q. Wu, Q. Shi, and M. Tan, "Auto-embedding generative adversarial networks for high resolution image synthesis," *IEEE Trans. Multimedia*, vol. 21, no. 11, pp. 2726–2737, Nov. 2019.

[33] O. Kupyn, T. Martyniuk, J. Wu, and Z. Wang, "DeBlurGAN-v2: Deblurring (orders-of-magnitude) faster and better," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Jun. 2019, pp. 8878–8887.

[34] L. Yang, S. Wang, S. Ma, W. Gao, C. Liu, P. Wang, and P. Ren, "HiFace-GAN: Face renovation via collaborative suppression and replenishment," in *Proc. 28th ACM Int. Conf. Multimedia*, Oct. 2020, pp. 1551–1560.

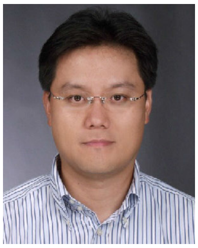
[35] C. Chen, X. Li, L. Yang, X. Lin, L. Zhang, and K. K. Wong, "Progressive semantic-aware style transformation for blind face restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 11891–11900.

R. GEETHA (Member, IEEE) received the B.E. degree in computer science and engineering from Madras University, in 1999, the M.E. degree in computer science and engineering from Anna University, in 2006, and the Ph.D. degree from the School of Computing and Information Technology, Vel Tech Rangarajan Dr. Sagunthala R&D Institute of Science and Technology, India, in 2017. She has over 23 years of teaching experience. She is currently with the S. A. Engineering College as a Professor and the Head of the Department. Her research interests include machine learning, wireless networks, and security schemes in wireless networks. She is the author/coauthor of several research papers in international conferences and journals. She is a Life Member of the Indian Society for Technical Education and a member of the Computer Society of India.

G. BELSHIA JEBAMALAR received the B.E. and M.E. degrees in computer science and engineering from Anna University, in 2019 and 2021, respectively, where she is currently pursuing the Ph.D. degree. She has over two and half years of teaching experience. She is currently with the S. A. Engineering College as an Assistant Professor. Her research interests include machine learning, deep learning, and cyber security. She is a Life Member of the Indian Society for Technical Education.

S. ARUMAI SHINEY received the B.E. and M.E. degrees in computer science and engineering from Anna University, in 2006 and 2008, respectively, where she is currently pursuing the Ph.D. degree. She has over 12 years of teaching experience. She is also with the S. A. Engineering College as an Assistant Professor. Her research interests include machine learning and artificial intelligence. She is a Life Member of the Indian Society for Technical Education.

NHU-NGOC DAO (Senior Member, IEEE) received the B.S. degree in electronics and telecommunications from the Posts and Telecommunications Institute of Technology, Hanoi, Vietnam, in 2009, and the M.S. and Ph.D. degrees in computer science from the School of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea, in 2016 and 2019, respectively. He is currently an Assistant Professor with the Department of Computer Science and Engineering, Sejong University, Seoul. Prior to joining Sejong University, he was a Visiting Researcher with the University of Newcastle, Callaghan, NSW, Australia, in 2019, and a Postdoctoral Researcher with the Institute of Computer Science, University of Bern, Switzerland, from 2019 to 2020. His research interests include network softwarization, mobile cloudification, communication security, intelligent systems, and the Intelligence of Things. He is a member of ACM. He is also an Editor of the *Scientific Reports* and *PLOS One* journals.



HYEONJOON MOON received the B.S. degree in electronics and computer engineering from Korea University, in 1990, and the M.S. and Ph.D. degrees in electrical and computer engineering from the State University of New York at Buffalo, in 1992 and 1999, respectively. From January 1996 to October 1999, he was a Senior Researcher with the Electro-Optics/Infrared Image Processing Branch, U.S. Army Research Laboratory (ARL), Adelphi, MD, USA. He developed a face recognition system evaluation methodology based on the Face Recognition Technology (FERET) Program. From November 1999 to February 2003, he was a Principal Research Scientist with Visage Technology, Littleton, MA, USA. His research interests include research and development is real-time facial recognition system for access control, surveillance, and big database applications. He has extensive background on still image and real-time video-based computer vision and pattern recognition. Since March 2004, he has been with the Department of Computer Science and Engineering, Sejong University, where he is currently a Professor and the Chairperson. His current research interests include image processing, biometrics, artificial intelligence, and machine learning.

SUNGRAE CHO received the B.S. and M.S. degrees in electronics engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in electrical and computer engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002. He is currently a Professor with the School of Computer Science and Engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an Assistant Professor with the Department of Computer Sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a Senior Member of the Technical Staff with the Samsung Advanced Institute of Technology (SAIT), Giheung-gu, South Korea, in 2003. From 1994 to 1996, he was a Research Staff Member with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a visiting professorship with the National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA. His current research interests include wireless networking, ubiquitous computing, and ICT convergence. He has been a Subject Editor of *IET Electronics Letters*, since 2018, and was an Area Editor of *Ad Hoc Networks* journal (Elsevier), from 2012 to 2017. He has served numerous international conferences as the Organizing Committee Chair, such as IEEE SECON, ICOIN, ICTC, ICUFN, TridentCom, and the IEEE MASS; and a Program Committee Member for IEEE ICC, GLOBECOM, VTC, MobiApps, SENSORNETS, and WINSYS.