IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 24 September 2024, accepted 13 November 2024, date of publication 18 November 2024,
date of current version 26 November 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3500212

RESEARCH ARTICLE

Denoising Diffusion-Based Image Generation Model Using Principal Component Analysis

MYUNG KEUN SONG[1], ASIM NIAZ[1], MUHAMMAD UMRAIZ[1], EHTESHAM IQBAL[2], SHAFIULLAH SOOMRO[3], AND KWANG NAM CHOI[1]

[1]Department of Computer Science and Engineering, Chung-Ang University, Seoul 06974, Republic of Korea
[2]Advanced Research and Innovation Center (ARIC), Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates
[3]Department of Computer Science and Media Technology, Linnaeus University, 3016 Växjö, Sweden

Corresponding author: Kwang Nam Choi (knchoi@cau.ac.kr)

This work was supported by the Ministry of Science and Information and Communication Technology (ICT) and National IT Industry Promotion Agency (NIPA) through the High Performance Computing (HPC) Support Project.

ABSTRACT In recent years, advancements in GPU technology and increased data collection have significantly enhanced the performance of artificial intelligence and image generation models. However, in specific areas such as medical imaging or facial images, constraints in data collection and class imbalance issues have posed challenges to improving image quality. This study proposes the integration of Principal Component Analysis (PCA) into image generation models to address these challenges. Specifically, to overcome the limitations of conventional image generation models like GANs and VAEs, we utilize the Denoise Diffusion Probabilistic Model (DDPM) as the backbone, integrating it with PCA techniques. Using the CIFAR10 and FFHQ datasets, we evaluated the image quality of the proposed PCA-DDPM, the traditional DDPM, and DCGAN. As a result, the PCA-DDPM demonstrated superior image quality and efficiency. Notably, it maintained high performance even when trained with a limited amount of data. The findings of this research contribute significantly to the advancement of image generation technology and are expected to be applied in various domains.

INDEX TERMS Artificial intelligence, deep learning, denoising diffusion, image generation, principal component analysis.

I. INTRODUCTION

Artificial intelligence has made significant contributions in various fields of computer vision, including but not limited to video anomaly detection [1], [35], explainable AI [37], [38], and image segmentation [39], [40]. These advancements are largely due to the availability of graphical processing units (GPUs) at the consumer level. Advancements in GPU technology and the real-time collection of large datasets have recently driven significant progress in artificial intelligence (AI) models. Deep learning-based AI models are now utilized across various domains. In particular, advancements in natural language processing have moved beyond simple machine translation, with models like ChatGPT capable of generating creative text and engaging in complex conversations [36]. In particular, in the field of computer vision, achievements equal to or surpassing human performance have been realized in image recognition and segmentation [31], [32], [33], [34]. The emergence of generative models such as Generative Adversarial Networks (GANs) [2] has demonstrated the potential for high-quality image generation, with continued research into new models by Xiang et al. [3], Lee and Lee [4], and others.

However, like other AI models trained through deep learning, these image generation models require a substantial amount of high-quality data to accurately learn the complex distribution of real data (original images). Challenges such as securing facial datasets due to privacy laws and copyright issues, as well as data imbalance in medical imaging, pose significant obstacles to stable model training. The lack of high-quality training images exacerbates issues in GAN-based models, such as a lack of diversity due to mode

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy.

collapse, and limits the ability of models based on Variational Autoencoders (VAEs) to generate high-quality images [5].

Efforts are underway to address the issues of data scarcity and imbalance resulting from the time-consuming and costly processes of data collection and refinement. Data augmentation techniques, such as rotation, inversion, scaling, and cropping, artificially increase the quantity of data [6]. However, these techniques are insufficient as fundamental solutions due to the risks of model overfitting, increased computational costs, and the potential introduction of data bias.

Deep learning-based image generation networks leverage vast image datasets to learn the characteristics and distributions necessary for generating new images. For instance, Generative Adversarial Networks (GANs) [2] are neural networks that generate target data from randomly given noise vectors. GANs consist of a generator, which learns the desired data distribution and creates synthetic images, and a discriminator, which evaluates the authenticity of the generated data. The typical learning structure of adversarial generative networks is depicted in the upper part of Figure 1.

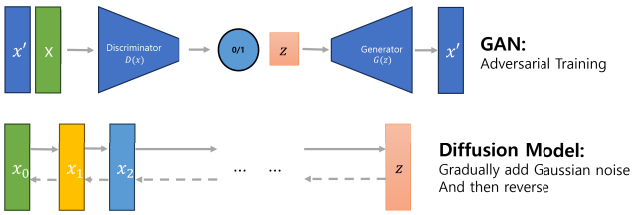

FIGURE 1. A visual comparison of the structural frameworks of Generative Adversarial Networks (GAN) and Diffusion models.

Diffusion Models [7] generate images by progressively adding and removing noise. Starting from the original image distribution, these models add noise in multiple stages and then progressively remove it to reconstruct the original images. Their general structure is depicted in the lower part of Figure 1.

This research focuses on the Denoising Diffusion Probabilistic Model (DDPM) [8], integrating Principal Component Analysis (PCA) [9] for universal feature extraction and application. This approach aims to efficiently generate high-quality images with limited data. For conditional generation based on specific labels, the CIFAR10 dataset [10], comprising 6,000 images across 10 classes, was used. For unconditional generation, the FFHQ (Flickr-Faces-HQ) dataset [11], consisting of 70,000 high-resolution face images, was employed.

10% of each training dataset was randomly selected for PCA, and the extracted features were applied to the DDPM's Forward Diffusion Process through multiplication. Experiments compared PCA-applied DDPM, standard DDPM, and GAN [2] models under identical conditions. Finally, the superiority of the proposed model was validated through qualitative and quantitative evaluations based on images sampled from each model trained on a 1/10 randomly sampled FFHQ dataset.

The main contributions of this study are as follows:

- Integration of PCA with DDPM to enhance image generation quality with limited data.
- Comparative analysis of PCA-DDPM with standard DDPM and GAN models.
- Verification of the proposed model's superiority through qualitative and quantitative assessments.

The remainder of this paper is structured as follows. Section II reviews the related background work, including an overview of existing image generation models and the theoretical foundations of Denoising Diffusion Probabilistic Models (DDPM) and Principal Component Analysis (PCA). Section III presents the proposed PCA-DDPM method, detailing how PCA is integrated into the DDPM framework to enhance image generation. Section IV outlines the experimental setup, including the datasets used, the hardware environment, and the comparative models. Section V discusses the results and analysis of the experiments, demonstrating the superiority of the PCA-DDPM model across various scenarios. Finally, Section VI offers concluding remarks and suggests directions for future work, highlighting the potential expansion of PCA-DDPM model applications and the exploration of alternative feature extraction methods.

II. RELATED WORK

A. PRINCIPAL COMPONENT ANALYSIS

In machine learning, Principal Component Analysis (PCA) [9] is one of the key techniques for dimensionality reduction. It provides a way to transform multidimensional data into a lower dimension while preserving as much important information as possible. As the number of features in data increases, so does the dimensionality. PCA projects this high-dimensional data onto new orthogonal axes while preserving variance, effectively reducing data complexity and removing noise.

PCA is often applied to mitigate issues of dimensionality and the risk of overfitting as the number of data features increases. In fields like image processing and computer vision, transforming high-dimensional image data into a lower dimension by removing less important features simplifies the data, thereby reducing the computational load required for model training and enhancing performance. In cases like high-dimensional facial image data, PCA plays a significant role in facial recognition technology. Studies like GANSpace [12] have analyzed the latent space of GANs using PCA to understand how each principal component controls different image features.

Recent studies have explored various methods to enhance the performance of image generation models by applying PCA. Cha et al. [13] and Han et al. [14] conducted research applying PCA to the generator of GANs. In this research, we proposes a method to combine PCA with DDPM [8], preserving key features of images while effectively

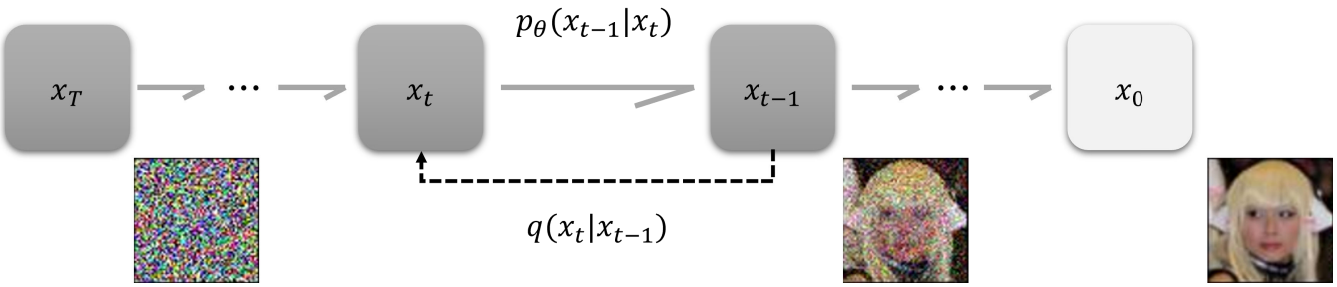

FIGURE 2. This diagram presents the two key phases of the Diffusion Model: the forward diffusion that introduces noise to the data, and the backward process that progressively restores the original data, elucidating its underlying structure.

generating them. It guides the diffusion and reverse diffusion processes in DDPM, improving the quality and efficiency of image generation.

B. DIFFUSION MODEL

The Diffusion Model [7] was first introduced in a study that presented a new approach to effectively extract and understand hidden structures and patterns in the data space. The central idea of the Diffusion Model is shown in [Figure 2]. The forward diffusion process, which adds noise to the data, and the reverse diffusion process, which gradually removes it, enable learning the original data distribution to uncover its intrinsic structure and pattern. This process is analogous to natural diffusion and utilizes concepts from statistical thermodynamics.

However, as the number of steps in the noise addition and removal processes increases, learning the small changes between each step (x_t and x_{t-1}) can compromise model stability and inevitably increase training time. Subsequent studies based on the Diffusion Model have proposed various methods to address or minimize these issues. According to the survey by Croitoru et al. [15], models such as the Denoising Diffusion Probabilistic Model (DDPM) [8], DDIM by Song et al. [16], IDDPM by Nichol and Dhariwal [17], and Stable Diffusion [18] are based on the fundamental concepts of diffusion models, while also incorporating efforts to improve these issues.

C. DENOISING DIFFUSION PROBABILISTIC MODEL

The original diffusion model attempted to simulate the diffusion process of data to learn complex data distributions. Its core idea was to add noise to data and then gradually remove it to learn the distribution of the original data. However, this model, which learns by finding the difference in the original data distribution at each stage of noise addition, could compromise training stability and result in inaccurate estimates, especially in complex data structures.

The Denoising Diffusion Probabilistic Model (DDPM) [8] emerged to solve these problems. The key idea of DDPM is to start with the original data and add a specific Gaussian noise distribution at each stage. The model then tries to remove the noise in the denoising process, learning how far it has deviated from the original data distribution and how the noise was added. Thus, by learning the distribution of noise at each stage, DDPM precisely understands the relationship between the noise distribution and the original data distribution at each stage.

Consequently, DDPM addresses the issues of training stability and inaccurate estimations in complex data structures of the Diffusion Model [7], while still enabling the generation of high-quality samples.

FIGURE 3. Overview of CIFAR10 and FFHQ datasets. On the left, CIFAR10 includes 60,000 32 × 32 pixel color images in 10 classes, each with 6,000 images, featuring animals and vehicles. On the right, FFHQ presents 70,000 high-quality, high-resolution human face images, showcasing a diverse range of ages, ethnicities, and backgrounds.

D. DATASET

Conditional sampling is a method of generating data based on specific conditions. The model is trained to generate data that matches the given conditions. For example, models like CGAN [19] or StyleGAN [20] can be requested to generate images of specific categories, such as specifying hair color, skin color, or gender. In contrast, unconditional sampling generates data without specific conditions. The model creates samples based on the overall distribution of the training data, yielding various results without specifying conditions. In this paper, two datasets were used for training conditional and unconditional sampling with DDPM [8].

The CIFAR10 [10] dataset consists of 60,000 color images of 32 × 32 pixels, categorized into 10 classes. These classes include airplanes, cars, birds, cats, deer, dogs, frogs, horses, ships, and trucks, with 6,000 images per class. Out of these,

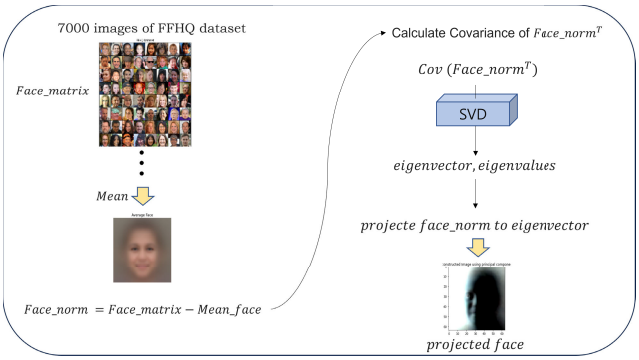

FIGURE 4. PCA Process and Feature Image Extraction. Image data is transformed into one-dimensional arrays for PCA, extracting three principal components. The most significant component is used for feature image creation, randomly applied to 10% of the training data to capture universal features and prevent overfitting.

50,000 images are used for training and 10,000 for testing. CIFAR10 is widely used as a benchmark dataset for computer vision research and performance evaluation of deep learning models. Particularly, it is frequently employed to assess the performance of deep learning models like ResNet [21] and VGGNet [11]. This study used CIFAR10 [10] for conditional sampling training with DDPM [8].

The FFHQ (Flickr-Faces-HQ) [11] dataset comprises 70,000 high-resolution human face images. These images are 1024×1024 pixels in resolution and include a variety of ages, ethnicities, and expressions. The dataset provides versions of various resolutions, including the original images. FFHQ is primarily used in research for generating high-resolution facial images, such as in StyleGAN [20] and StyleGAN2 [22], and is provided by NVIDIA. In this study, the FFHQ dataset was used for unconditional sampling training with DDPM.

III. PROPOSED METHOD

Research on image generation has consistently progressed, particularly in adversarial generative networks that combine Convolutional Neural Networks (CNN) [23] and U-Net [24] structures to create higher-quality images. This paper proposes a method that diverges from traditional GAN models [2] and is instead based on DDPM [8] generative models. It integrates Principal Component Analysis into the learning process to more clearly define the direction of generation objectives. This approach aims to capture complex structures and patterns in images more accurately, resulting in the creation of higher-quality and more diverse images.

A. FEATURE EXTRACTION

PCA [9] is a statistical method used to emphasize variability and capture strong patterns in datasets. It extracts the principal components that represent the most significant features of the dataset. These extracted features are integrated into the DDPM learning process through multiplication, ensuring that the generated images retain the core characteristics of the training data.

The datasets used for training the proposed model are FFHQ (Flickr-Faces-HQ) [11] and CIFAR10 [10], comprising 70,000 high-resolution facial images and 60,000 images across 10 classes, respectively. To extract principal components, 10% of the images from these datasets are randomly selected. Principal Component Analysis (PCA) is performed on each channel of these selected images to extract major features.

The process of principal component analysis and feature image extraction is shown in Figure 4. The image data for each channel is transformed into one-dimensional arrays and then PCA is conducted. From the PCA, three principal components are extracted for each channel. The first principal component, representing the most dominant pattern or structure and the direction of maximum variance in the data, generates feature images. Consequently, the final feature image extracted is of size $3 \times 64 \times 64$ [channel, height, width].

The extracted principal component feature images will be applied to the training dataset. However, to prevent overfitting, these features are randomly applied to only 10% of the training data. The feature images are integrated into the training data through multiplication. This approach captures universal features of the dataset, providing directions in the image generation process.

B. PROPOSED MODEL

The Denoise Diffusion Probabilistic Model (DDPM) [8] is one of the modern probabilistic approaches to image generation. DDPM characteristically generates images conditionally over various time steps, allowing for detailed control of the generation process. Furthermore, DDPM realizes high-resolution image generation through sophisticated learning of noise distributions at specific time steps. In this research, principal component features extracted through PCA [9] are integrated into the structure of DDPM, as shown in Figure 5. The forward diffusion process involves applying deterministic Gaussian noise to the original image, transforming it into a noisy image. The reverse diffusion process, given the noisy image and original image, predicts the noise from the previous diffusion process to remove it, thereby learning the distribution of the image data.

1) FORWARD DIFFUSION PROCESS

The traditional forward diffusion process is an initial stage of DDPM learning, where Gaussian noise [25] is progressively added to the original image, as shown in Equation 1. Here, the original image is denoted as x_0. The noise addition process, denoted as q, transforms x_0 into x_1, represented as $q(x_1 \mid x_0)$. Over time t, this can be represented as $q(x_t \mid x_{t-1})$. To prevent the variance from diverging during the noise addition, the noise is scaled by the diffusion rate β before being added. This ensures that when the forward diffusion ends, the noise follows a normal distribution $N(x_T; 0, I)$.

$$q(x_t|x_{t-1}) := N(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t I) \tag{1}$$

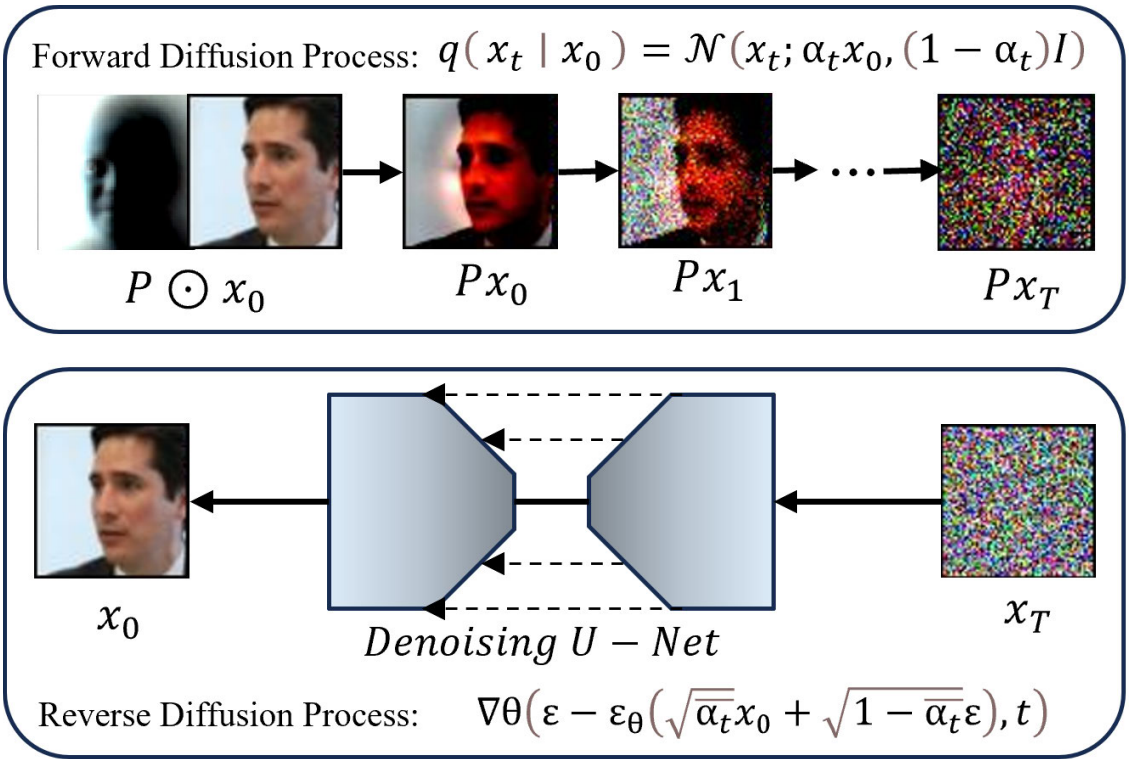

FIGURE 5. The Structure of the Proposed DDPM Model with PCA Integration. This illustration showcases the integration of principal component features extracted through PCA into the DDPM structure. The model utilizes a forward diffusion process, applying deterministic Gaussian noise to the original image, and a reverse diffusion process that predicts and removes noise, thereby learning the image data distribution. This integration allows for enhanced control and higher resolution in image generation, showcasing the unique capabilities of the combined PCA-DDPM model.

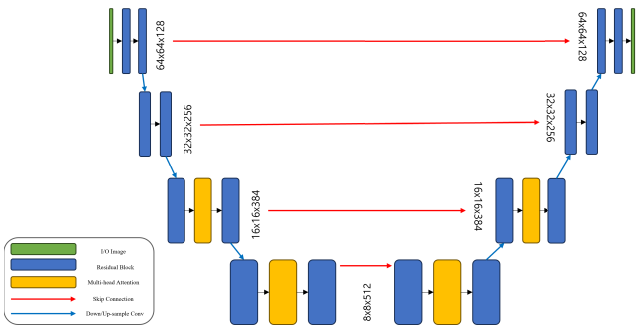

FIGURE 6. U-Net Architecture in the Proposed Model. This figure presents the U-Net structure used in our model, similar to the original DDPM [8]. It includes attention layers, enhancing focus on key image features, crucial for high-resolution image generation within the DDPM framework.

Based on this, the distribution of x_t for given image data x_0 can be represented by the following Equation 2:

$$q(x_t \mid x_0) = N(x_t; \sqrt{\bar{\alpha}_t}x_0, (1-\bar{\alpha}_t)I) \tag{2}$$

Here, $\alpha_t := 1 - \beta_t$, and $\bar{\alpha}_t$ represents the cumulative noise coefficient up to time t, formalized as $\bar{\alpha}_t := \prod_{s=1}^{t} \alpha_s$. In the proposed model's forward diffusion process, the principal component features P extracted from PCA are additionally multiplied with the original data, resulting in the noise-added image x'_t at time t, as shown in Equation 3.

$$x'_t = N(x'_t; \sqrt{\bar{\alpha}}x_0 \odot P, (1-\bar{\alpha}_t)I) \tag{3}$$

In this, x'_t represents the image with added noise and PCA features, and $\odot$ denotes element-wise multiplication. This new forward diffusion process allows for the retention of principal features of the original image while introducing noise and PCA features, playing a crucial role in maintaining the core features of the original image while increasing its diversity.

2) REVERSE DIFFUSION PROCESS

The reverse diffusion process (Reverse Diffusion Process) restores the noise-transformed image data x'_t back to the original image x_0. This process utilizes a Gaussian transition within a Markov chain [26], similar to the reverse diffusion process of the original DDPM [8]. The p_θ used for restoration estimates x'_{t-1} from x'_t, and this process can be expressed as follows (Equation 4):

$$\begin{aligned} p_\theta(x_{0:T}) &:= p(x_T) \prod_{t=1}^{T} p_\theta(x_{t-1} \mid x'_t) \\ p_\theta(x_{t-1} \mid x'_t) &:= N(x_{t-1}; \mu_\theta(x'_t, t), \Sigma_\theta(x'_t, t)) \end{aligned} \tag{4}$$

C. OBJECTIVE FUNCTION

DDPM differs from traditional diffusion models in that it focuses on predicting the value of applied noise at time t, rather than predicting the distribution difference between x_t and x_{t-1}. The objective function will be derived by getting Kullback-Leibler divergence by the sum of L_T and L_{t-1} and subtract L_0 which is as follows (Equation 5, 6, 7):

$$L_T = D_{KL}\left(q(x_T \mid x_0) \parallel p(x_T)\right) \tag{5}$$

$$L_{t-1} = \sum_{t>1} D_{KL}\left(q(x_{t-1} \mid x_t', x_0) \parallel p_\theta(x_{t-1} \mid x_t')\right) \tag{6}$$

$$L_0 = \log p_\theta(x_0 \mid x_1)_{L_0} \tag{7}$$

In Equation 5, the first part L_T represents the forward diffusion process, indicating the distribution difference between the noise x_T generated by p and q given the original image x_0. As x_T always follows a Gaussian distribution in DDPM, this process is tractable and usually approximated as a constant near zero, thus often ignored in the learning process.

The second part L_{t-1} signifies the distribution difference between the reverse and forward diffusion processes, represented by the Kullback-Leibler divergence between the actual forward process distribution $q(x_{t-1} \mid x_t', x_0)$ and the estimated forward process distribution $p_\theta(x_{t-1} \mid x_t')$. To compute L_{t-1}, the distribution of $q(x_{t-1} \mid x_t', x_0)$ needs to be determined, and Σ_θ and μ_θ for $p_\theta(x_{t-1} \mid x_t')$ are defined as follows (Equations 8 and 9):

$$\begin{aligned}\mu_\theta(x_t', t) &= \mu_t\left(x_t', \frac{1}{\alpha_t}(x_t' - (1-\alpha_t)\varepsilon_\theta(x_t'))\right) \\ &= \frac{1}{\alpha_t}x_t' - \frac{\beta_t}{1-\alpha_t}\varepsilon_\theta(x_t', t)\end{aligned} \tag{8}$$

$$\mathbb{E}_{x_0,\varepsilon}\left[\frac{\beta_t^2}{2\sigma_t^2\alpha_t(1-\alpha_t)} \parallel \varepsilon - \varepsilon_\theta(\alpha_t x_0 + (1-\alpha_t)\varepsilon, t)\right]^2 \tag{9}$$

Using the determined values of q, p_θ, and Σ_θ, the Kullback-Leibler divergence L_{t-1} can be calculated as shown in Equation 10:

$$L_{t-1} = \mathbb{E}_q\left[\frac{1}{2\sigma_t^2} \parallel \mu_t(x_t', x_0) - \mu_\theta(x_t', t)\right]^2 + C \tag{10}$$

This loss function can be expressed in a simplified form, focusing on the noise ε, as shown in Equation 11:

$$L_{\text{simple}}(\theta) := \mathbb{E}_{t,x_0,\varepsilon} \parallel \varepsilon - \varepsilon_\theta(\alpha_t x_0 + (1-\alpha_t)\varepsilon, t) \parallel^2 \tag{11}$$

Finally, the term L_0 is a likelihood function estimating the original data x_0 from the latent noise data x_1, with training directed towards maximizing this term. Consequently, the proposed model, like DDPM, aims to minimize the difference between the actual noise ε and the predicted noise ε_θ at each stage of the forward and reverse diffusion processes as per Equation 11. Table 1 gives a comprehensive overview of the equations of the proposed method.

The proposed approach is summarized by the Algorithm 1.

TABLE 1. Description of equations.

Equation	Description
(1)	Forward diffusion process equation.
(2)	Distribution of x_t for given image data x_0 in the forward diffusion process.
(3)	New forward diffusion process incorporating principal component features.
(4)	Reverse diffusion process equation for restoring the noise-transformed image data.
(5)	Objective function part representing the forward diffusion process.
(6)	Objective function part representing the distribution difference between the reverse and forward diffusion processes.
(7)	Likelihood function estimating the original data from the latent noise data.
(8)	Mean calculation for $p_\theta(x_{t-1}\|x_t')$.
(9)	Expectation calculation for $p_\theta(x_{t-1}\|x_t')$.
(10)	Calculation of L_{t-1} using determined values of q, p_θ, and Σ_θ.
(11)	Simplified loss function focusing on the noise.

Algorithm 1 Denoise Diffusion Probabilistic Model With Principal Component Features

1 **Initialization:** Set initial image x_0, diffusion rate β, and principal component features P.
2 **Forward Diffusion Process:**
   a. Initialize $x_t = x_0$, $\bar{\alpha} = 1$, $t = 1$.
   b. Iterate from $t = 1$ to T:
      i. Compute $x_t' = N(x_t'; \sqrt{\bar{\alpha}}x_0 \odot P, (1-\bar{\alpha})I)$.
      ii. Update $\bar{\alpha} = \bar{\alpha} \times (1-\beta_t)$.
      iii. Increment t.
3 **Reverse Diffusion Process:**
   a. Initialize $t = T$.
   b. Iterate from $t = T$ to 1:
      i. Estimate $x_{t-1}' = N(x_{t-1}'; \mu_\theta(x_t', t), \Sigma_\theta(x_t', t))$.
      ii. Decrement t.
4 **Objective Function (Loss Calculation):**
   a. Calculate L_T using Equation 5.
   b. Calculate L_{t-1} using Equation 10.
   c. Calculate L_0 using Equation 11.
   d. Compute the total loss $\mathcal{L} = L_T + L_{t-1} - L_0$.
5 **Summary:** The Denoise Diffusion Probabilistic Model with Principal Component Features integrates principal component analysis (PCA) into the traditional DDPM framework, facilitating high-resolution image generation while preserving essential image structures.

IV. EXPERIMENTS AND RESULTS

The experiments in this study were conducted using the Flickr-Faces-HQ (FFHQ) dataset [11], which consists of 70,000 high-resolution facial images, for the training of the unconditional model. For the training of the conditional model, the CIFAR10 dataset [10], comprising 60,000 images across 10 classes (airplane, automobile, bird, cat, deer, dog, frog, horse, ship, truck), was utilized. All experiments were

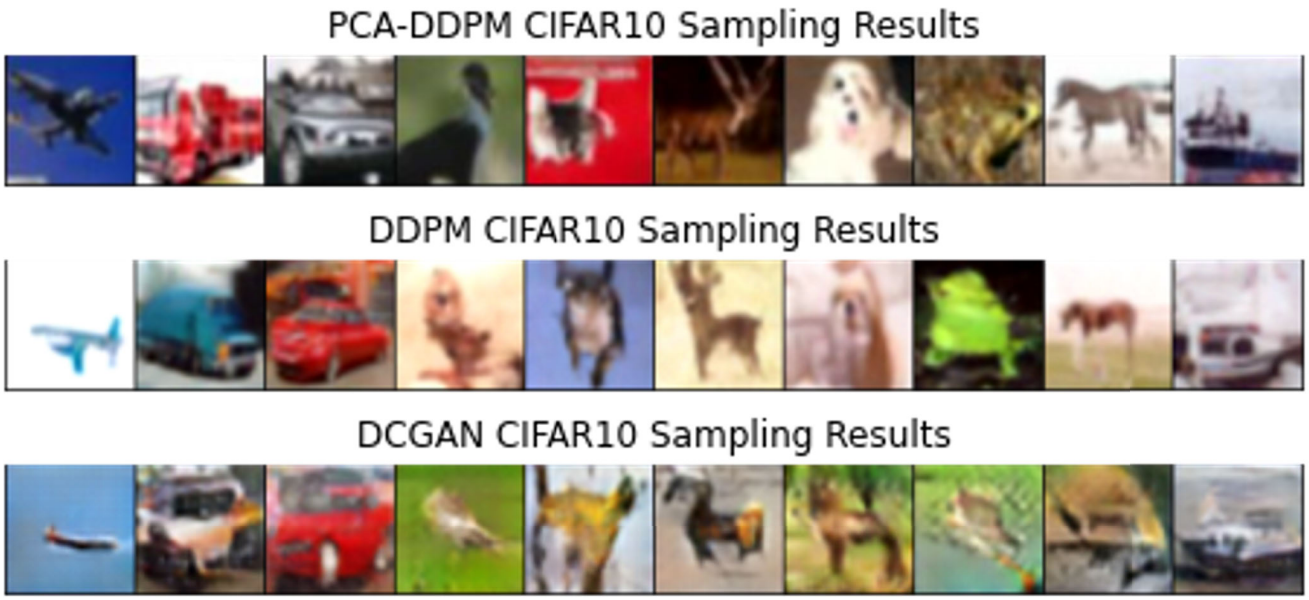

FIGURE 7. Generated Images on CIFAR10 Dataset. This figure displays images of an airplane, truck, car, bird, cat, deer, dog, frog, horse, and boat, created using the proposed model, DDPM, and DCGAN. The proposed model demonstrates superior natural blending and realism, particularly in the images of the deer, frog, and horse.

performed in an environment equipped with a GPU: Tesla V100 64G RAM and an Intel(R) Xeon(R) Gold 6132 CPU @ 2.60GHz. To ensure consistency and reproducibility in each experiment, a random seed value of 416 was set.

To compare performance with existing image generation models, Deep Convolutional Generative Adversarial Network (DCGAN) [27] was also trained under the same conditions. DCGAN, a variant of GAN [2] based on convolutional neural networks (CNN), has the advantage of stable training and high image quality compared to traditional GANs. However, it still faces challenges, such as training instability and reduced diversity in generated images, commonly referred to as the 'mode collapse' issue.

A. MODEL TRAINING AND EVALUATION

Image generation models aim to learn the distribution of target data and generate images that resemble real ones. There are various metrics to assess the quality and diversity of generated images. One of the quantitative evaluation metrics, the Inception Score [28], uses a pre-trained Inception-v3 model to predict the class probability distribution of images generated by the model. A high Inception Score indicates both good quality and diversity of the generated images. However, issues arise when the model generates images not present in the training dataset or when it replicates the training data itself, resulting in either a low or misleadingly high Inception Score, despite high-quality images.

Considering these challenges, this study utilizes the Frechet Inception Distance (FID) [29] to comprehensively evaluate the model's performance. Unlike the Inception Score, which evaluates only the generated images, FID measures the Frechet distance between the feature distributions of real and generated images, calculated from their respective mean and covariance matrices. This direct comparison between real and generated data provides a more reliable and accurate assessment, considering both the quality and diversity of the images.

The training methodology of Denoising Diffusion Probabilistic Models (DDPM) [8] involves gradually adding noise to the data, leading to a disordered image state, and then learning to restore the original image through a denoising process. This process is modeled using a Markov Chain [26] and diffusion processes, with the training objective being to effectively learn a denoiser that removes noise at each step.

In this paper, principal component analysis is initially performed on the images from the dataset to extract key features, which are then used in the noise addition process. During training, the U-Net model [24] takes the noise-added image and time information as input and learns to output the expected noise. [Figure 6] depicts the structure of the U-Net used in the training of the proposed model. The goal here is to restore the progressively noise-added images to their original state, using the Mean Squared Error (MSE) loss function to minimize the difference between the actual noise and the model's prediction.

For comparative analysis, the proposed model and the adversarial model, DCGAN, are trained on the CIFAR10 [10] and FFHQ [11] datasets. Both models undergo repeated training for a total of 100 epochs. The optimizer used for training is AdamW [30], identical to the backbone model DDPM [8]. Every 10 epochs, DCGAN [27] records the GAN

FIGURE 8. Comparison of Unconditionally Generated Images on FFHQ Dataset. From left to right: images by the proposed model, original DDPM, and DCGAN. The proposed model and DDPM produce sharp, high-resolution faces, with the former showing greater naturalness and diversity. DCGAN, however, struggles with more distortions and less realistic outputs.

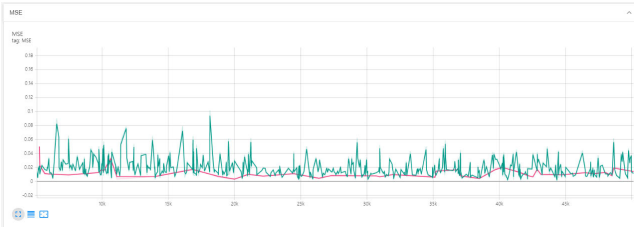

FIGURE 9. Early Training MSE Comparison of Proposed Model and Original DDPM on FFHQ Dataset. The x-axis marks training iterations, and the y-axis shows MSE values. The green and pink lines represent the original DDPM and the PCA-applied DDPM models, respectively. The proposed model exhibits a more gradual and stable MSE reduction compared to the original DDPM's steep decrease.

loss and generates target images through the generator. PCA-DDPM records the MSE loss and samples the target images. Throughout the training process, the model with the lowest loss value, compared to previously stored weights, is saved.

Furthermore, to compare the performance of the proposed model (PCA-DDPM) with the original DDPM [8], the proposed model is trained on a reduced FFHQ dataset, consisting of 7,000 images. This training corresponds to 1/10th of the total training data volume, demonstrating the efficiency of the proposed model.

Upon completion of training, qualitative evaluations of the images generated by each model are conducted by generating 70,000 images for the CIFAR10 dataset and 60,000 for the FFHQ dataset. From these, 64 images are randomly selected to compare the diversity and quality of generation. For quantitative evaluation, the FID score [22] is calculated to compare the performance of each model.

B. EXPERIMENTAL RESULTS

This study trained three models: the proposed PCA-DDPM, DDPM [8], and DCGAN [27], using the CIFAR10 [10] and FFHQ [11] datasets for 100 epochs each. Subsequently, both quantitative and qualitative evaluations of the images generated by these models were conducted to compare their performances and demonstrate the superiority of the proposed model.

1) QUALITATIVE EVALUATION

Figure 7 presents the conditionally generated images from the top-down order of the proposed model, the original DDPM, and DCGAN, using the final models trained on the CIFAR10 dataset. Starting from the left, images of an airplane, truck, car, bird, cat, deer, dog, frog, horse, and boat were generated. The proposed model produced images that blend more naturally with the background, particularly creating more realistic images of the deer (6th), frog (8th), and horse (9th) compared to the other models.

Figure 8 shows the unconditionally generated images from the final models trained on the FFHQ dataset. From left to right, the images are produced by the proposed model, the original DDPM, and DCGAN, respectively.

The proposed model and the original DDPM [8] generally produced high-resolution and sharp images. However, when comparing the images generated by the original DDPM with those of the proposed model, the latter was found to generate more natural-looking faces, covering various ages, genders, expressions, and features such as faces with glasses. On the other hand, the images generated by DCGAN [27] predominantly included distortions and noise, resulting in images that were distorted or not resembling faces.

2) QUANTITATIVE PERFORMANCE EVALUATION

For a quantitative evaluation of the models' performance, the quality of the generated images was directly compared visually, along with a qualitative performance evaluation using the FID Score [29]. Table 2 presents the FID Scores calculated based on the conditionally generated images from the CIFAR10 dataset, as shown in Figure 7, for the proposed

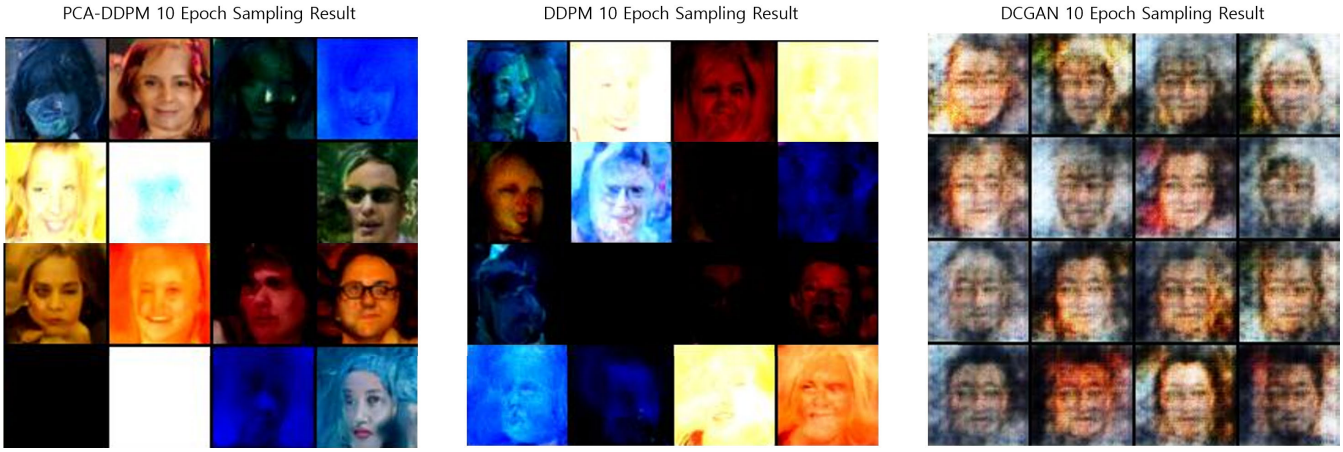

FIGURE 10. Comparison of Early Training Results at 10th Epoch on FFHQ Dataset. The image showcases samples from the proposed model, original DDPM, and DCGAN. At this stage, the proposed model demonstrates superior ability in generating more complete and diverse facial images, while DCGAN exhibits noise and less diversity, and the original DDPM shows a lower yield of complete faces.

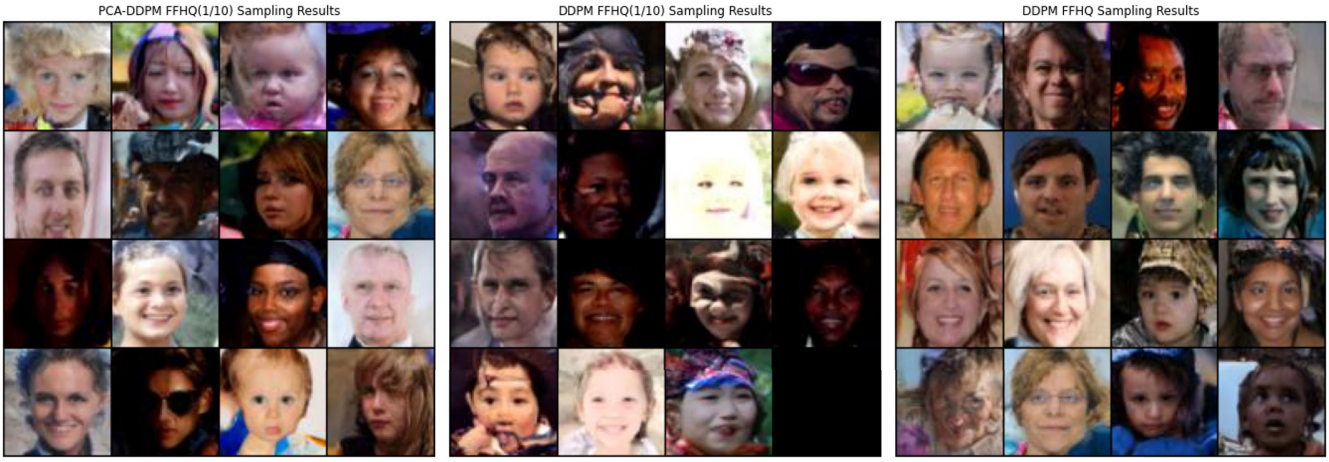

FIGURE 11. Comparative Image Generation with Limited Data. The figure illustrates the performance of the proposed PCA-DDPM and conventional DDPM when trained with a reduced dataset of 7,000 images, compared to the standard DDPM trained with the full 70,000-image dataset. Despite the data limitation, the proposed model successfully generates high-quality images with diverse characteristics, whereas the conventional DDPM shows limitations in generating standard facial images. This highlights the proposed model's efficacy in maintaining high-quality image generation even with limited data.

model, DDPM, and DCGAN. Additionally, the final error function (MSE) at the end of model training was also compared. The DCGAN model uses a log-likelihood function as its error function instead of MSE, hence it is not included in this comparison. The proposed model recorded a lower final MSE than the original DDPM. The order of the models from smallest to largest FID Score is the proposed model, DDPM, and then DCGAN, proving that the proposed model generates images most similar and diverse compared to the original data.

Similarly, Table 3 presents the FID Scores and the final Mean Squared Error (MSE) values for the unconditionally generated images from the FFHQ dataset, as shown in Figure 8, for the proposed model, DDPM, and DCGAN. As with the conditional image generation results, the proposed model, DDPM, and then DCGAN recorded the smallest FID Scores in that order, demonstrating that the proposed model was most effective in generating images similar and diverse compared to the original dataset.

TABLE 2. Comparison of conditional sampling results for each model.

Model	FID	MSE
DCGAN	17.02	-
DDPM	12.49	0.01739
DDPM with PCA (Ours)	11.86	0.01636

3) INITIAL TRAINING STAGE COMPARISON

Figure 9 illustrates the Mean Squared Error (MSE) values observed during the initial training stages of the proposed model and the original DDPM on the FFHQ dataset. The x-axis represents the number of training iterations, while the

TABLE 3. Comparison of unconditional sampling results for each model.

Model	FID	MSE
DCGAN	20.05	-
DDPM	16.42	0.1921
DDPM with PCA (Ours)	15.87	0.01004

TABLE 4. Comparison of training results with a limited dataset for the conventional DDPM and our model.

Model	Dataset	FID	MSE
DDPM	FFHQ	16.42	0.01921
DDPM	$\frac{1}{10}$FFHQ	22.36	0.02103
DDPM with PCA (Ours)	$\frac{1}{10}$FFHQ	19.43	0.01684

y-axis shows the recorded MSE values at each training stage. The green line represents the original DDPM model, and the pink line represents the proposed PCA-applied DDPM model. Throughout most of the stages, the original DDPM model shows a steep decrease in MSE, whereas the proposed model demonstrates a more gradual and stable reduction in MSE during the early stages of training.

Furthermore, the sample images generated by each model during the initial stages of training based on the FFHQ dataset were compared. Figure 10 showcases sample images generated by the proposed model, the original DDPM, and the DCGAN model after the completion of the 10th epoch of training on the FFHQ dataset. The proposed model generated a higher number of complete facial images compared to the original DDPM at this stage. The DCGAN model, while able to generate recognizable facial structures, produced images with significant noise and less diversity. These results indicate that the proposed model outperforms the other models in terms of stability and quality of image generation during the early stages of training.

4) ROBUSTNESS EVALUATION UNDER DATA LIMITATION

The proposed model demonstrated superior performance in both conditional and unconditional image generation after training under identical conditions compared to the models evaluated. Notably, from the initial stages, it exhibited high stability in learning and the ability to generate high-quality images. To assess whether the proposed model could still produce quality image data through stable training in cases of dataset deficiencies or imbalances, the FFHQ dataset was drastically reduced to one-tenth of its original size, and images were generated under the same training conditions.

Figure 11 compares images generated using the proposed model and the conventional DDPM trained only with 7,000 images, against the normal DDPM trained with the full 70,000-image dataset. The proposed model was able to generate high-quality images with a variety of races, genders, and expressions, whereas the conventional DDPM with the limited dataset struggled to produce standard facial images. Furthermore, when comparing the image generation results of the proposed model with the conventional DDPM trained with the full dataset, the proposed model was able to produce

TABLE 5. Comparison of unconditional sampling results for each model.

Model	FID	MSE
DCGAN	20.05	-
DDPM	16.42	0.1921
DDPM with PCA (Ours)	15.87	0.01004

images of comparable quality. This demonstrates the ability of the proposed PCA-DDPM to maintain high performance even with limited data.

Table 4 presents the FID scores and mean squared error (MSE) values measured based on the image generation results from Figure 11, using the limited FFHQ dataset for the proposed model and the conventional DDPM, and the normal FFHQ dataset for the conventional DDPM. The proposed model recorded a lower FID score compared to the DDPM trained with the limited dataset and a slightly higher score than the DDPM trained with the normal FFHQ dataset.

V. CONCLUSION AND FUTURE RESEARCH

This study proposed the incorporation of universal characteristics extracted through Principal Component Analysis (PCA) into the training of the Denoise Diffusion Probabilistic Model (DDPM). We confirmed that this approach can enhance the performance and efficiency of image generation networks.

The proposed PCA-DDPM model demonstrates superior performance compared to conventional DDPM and DCGAN models. It is particularly outstanding in generating high-resolution images and produces high-quality images even in the initial stages of training. Additionally, it maintains high performance even with limited training data, indicating that PCA-DDPM is efficient and effective even under data constraints.

The results of this study are considered to be a significant contribution to the advancement and efficiency improvement of image generation networks. However, the potential applications and areas of use for the proposed model can be further expanded. Future research will involve more thorough verification of PCA-DDPM's performance across various datasets and under different conditions. We also plan to explore the possibility of integrating PCA with other feature extraction methods.

This research is expected to accelerate the progress of image generation technologies and is anticipated to be applied in various practical applications, such as medical imaging and anomaly detection.

REFERENCES

[1] S. U. Amin, M. Ullah, M. Sajjad, F. A. Cheikh, M. Hijji, A. Hijji, and K. Muhammad, "EADN: An efficient deep learning model for anomaly detection in videos," *Mathematics*, vol. 10, no. 9, p. 1555, May 2022.

[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, 2014, pp. 1–9.

[3] P. Xiang, L. Wang, F. Wu, J. Cheng, and M. Zhou, "Single-image deraining with feature-supervised generative adversarial network," *IEEE Signal Process. Lett.*, vol. 26, no. 5, pp. 650–654, May 2019.

[4] I. Lee and W. Lee, "UniQGAN: Unified generative adversarial networks for augmented modulation classification," *IEEE Commun. Lett.*, vol. 26, no. 2, pp. 355–358, Feb. 2022.
[5] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.
[6] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
[7] J. Sohl-Dickstein, E. Weiss, N. Maheswaranathan, and S. Ganguli, "Deep unsupervised learning using nonequilibrium thermodynamics," in *Proc. Int. Conf. Mach. Learn.*, 2015, pp. 2256–2265.
[8] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 6840–6851.
[9] H. Abdi and L. J. Williams, "Principal component analysis," *Wiley Interdiscipl. Rev., Comput. Statist.*, vol. 2, no. 4, pp. 433–459, 2010.
[10] A. Krizhevsky and G. Hinton, "Learning multiple layers of features from tiny images," Univ. Toronto, Toronto, ON, Canada, Tech. Rep., 2009.
[11] 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. 4401–4410.
[12] E. Härkönen, A. Hertzmann, J. Lehtinen, and S. Paris, "Ganspace: Discovering interpretable GAN controls," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 9841–9850.
[13] G. S. Cha, U. Asim, M. K. Song, A. Niaz, and K. N. Choi, "Image generation network model based on principal component analysis," in *Proc. Asia Conf. Adv. Robot., Autom., Control Eng. (ARACE)*, Aug. 2022, pp. 76–80.
[14] S. H. Han, A. Niaz, and K. N. Choi, "A U-net based self-supervised image generation model applying PCA using small datasets," in *Proc. 2nd Asia Conf. Algorithms, Comput. Mach. Learn.*, Mar. 2023, pp. 450–454.
[15] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah, "Diffusion models in vision: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 9, pp. 10850–10869, Sep. 2023.
[16] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," 2020, *arXiv:2010.02502*.
[17] A. Q. Nichol and P. Dhariwal, "Improved denoising diffusion probabilistic models," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2021, pp. 8162–8171.
[18] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 10684–10695.
[19] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
[20] 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.
[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
[22] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of StyleGAN," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8110–8119.
[23] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
[24] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Munich, Germany. Cham, Switzerland: Springer, 2015, pp. 234–241.
[25] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans. Image Process.*, vol. 12, no. 11, pp. 1338–1351, Nov. 2003.
[26] J. R. Norris, *Markov Chains: 2*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
[28] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.
[29] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, "GANs trained by a two time-scale update rule converge to a local Nash equilibrium," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–12.
[30] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017, *arXiv:1711.05101*.
[31] M. Asim, F. Shamshad, and A. Ahmed, "Blind image deconvolution using deep generative priors," *IEEE Trans. Comput. Imag.*, vol. 6, pp. 1493–1506, 2020.
[32] Q. Jiang, H. Shi, L. Sun, S. Gao, K. Yang, and K. Wang, "Annular computational imaging: Capture clear panoramic images through simple lens," *IEEE Trans. Comput. Imag.*, vol. 8, pp. 1250–1264, 2022.
[33] O. Leong, A. F. Gao, H. Sun, and K. L. Bouman, "Discovering structure from corruption for unsupervised image reconstruction," *IEEE Trans. Comput. Imag.*, vol. 9, pp. 992–1005, 2023.
[34] A. M. Teodoro, J. M. Bioucas-Dias, and M. A. T. Figueiredo, "Image restoration and reconstruction using targeted plug-and-play priors," *IEEE Trans. Comput. Imag.*, vol. 5, no. 4, pp. 675–686, Dec. 2019.
[35] A. Niaz, S. U. Amin, S. Soomro, H. Zia, and K. N. Choi, "Spatially aware fusion in 3D convolutional autoencoders for video anomaly detection," *IEEE Access*, vol. 12, pp. 104770–104784, 2024.
[36] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 1877–1901.
[37] F. Xu, H. Uszkoreit, Y. Du, W. Fan, D. Zhao, and J. Zhu, "Explainable AI: A brief survey on history, research areas, approaches and challenges," in *Proc. CCF Int. Conf. Natural Lang. Process. Chin. Comput.*, Dunhuang, China. Cham, Switzerland: Springer, 2019, pp. 563–574.
[38] A. Niaz, S. Soomro, H. Zia, and K. Nam Choi, "Increment-CAM: Incrementally-weighted class activation maps for better visual explanations," *IEEE Access*, vol. 12, pp. 88829–88840, 2024.
[39] N. Ravi, V. Gabeur, Y.-T. Hu, R. Hu, C. Ryali, T. Ma, H. Khedr, R. Rädle, C. Rolland, L. Gustafson, E. Mintun, J. Pan, K. V. Alwala, N. Carion, C.-Y. Wu, R. Girshick, P. Dollár, and C. Feichtenhofer, "SAM 2: Segment anything in images and videos," 2024, *arXiv:2408.00714*.
[40] A. Niaz, E. Iqbal, A. A. Memon, A. Munir, J. Kim, and K. N. Choi, "Edge-based local and global energy active contour model driven by signed pressure force for image segmentation," *IEEE Trans. Instrum. Meas.*, vol. 72, pp. 1–14, 2023.

MYUNG KEUN SONG received the B.S. degree in computer science from Sahmyook University, South Korea, in 2021. He is currently pursuing the M.S. degree with the Department of Computer Science and Engineering, Chung-Ang University, Seoul, South Korea. He is a Graduate Research Student. Since 2021, he has been a Research Assistant with the AI Vision Laboratory, Chung-Ang University, under the supervision of Prof. Dr. Choi. His current research interests include object detection, image generation, and deep learning.

ASIM NIAZ received the B.S. degree in electrical (computer) engineering from the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2016, and the M.S. degree in computer science and engineering from Chung-Ang University, Seoul, South Korea, in 2020. He was a Research Assistant with the Visual Image Media Laboratory. Later, he visited INRIA Sophia Antipolis, France, followed by his research internship at Ericsson Canada, Montreal, Canada. He is currently a Research Assistant with the AI Vision Laboratory, Chung-Ang University. His current research interests include action recognition, video understanding, anomaly detection in images and videos, remote sensing, medical image analysis, and image segmentation.

MUHAMMAD UMRAIZ received the bachelor's degree in electrical engineering from COMSATS University Islamabad, Islamabad, Pakistan, in 2017, and the M.S. degree in electronics and information engineering from Jeonbuk National University, Jeonju, South Korea, in 2020. He is currently a Research Assistant with the AI Vision Laboratory, Chung-Ang University, Seoul, South Korea. His research interests include medical image processing, precision agriculture, and smart farming.

EHTESHAM IQBAL received the B.S. degree in electrical (computer) engineering from the COMSATS Institute of Information Technology, Pakistan, in 2017, and the M.S. degree from the Department of Computer Science and Engineering, Chung-Ang University, South Korea.

He was a Research Assistant with the Visual Image Media Laboratory, Chung-Ang University. He is currently a Research Associate with the Advanced Research and Innovation Center, Khalifa University of Science and Technology, United Arab Emirates. He has experience in industry and academia. His current research interests include industrial anomaly detection, medical image analysis, semantic segmentation, and generative modeling.

SHAFIULLAH SOOMRO received the Bachelor of Engineering (B.E.) degree from QUEST, Nawabshah, Sindh, Pakistan, in 2008, the Master of Engineering (M.E.) degree from MUET, Jamshoro, Sindh, in 2014, and the Ph.D. degree in computer science from Chung-Ang University, Seoul, South Korea, in 2018. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Media Technology, Linnaeus University, Sweden.

His research interests include motion tracking, object segmentation, and 3D image recognition.

KWANG NAM CHOI received the B.S. and M.S. degrees from the Department of Computer Science, Chung-Ang University, Seoul, South Korea, in 1988 and 1990, respectively, and the Ph.D. degree in computer science from the University of York, U.K., in 2002.

He is currently a Professor with the School of Computer Science and Engineering, Chung-Ang University. His current research interests include motion tracking, object categorization, and 3D image recognition.