IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2907~2919
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2907-2919

Improving the quality of images using Wasserstein generative adversarial networks for image restoration

Aruna Pavate[1], Surekha Janrao[2], Rohini Patil[3], Maganti Venkatesh[4], Shudhodhan Bokefode[5], Yunfei Li[6], Ubaldo Comite[6]
[1]Department of Information Technology, Thakur College of Engineering and Technology, University of Mumbai, Mumbai, India
[2]Department of Computer Engineering, K.J. Somaiya Institute of Technology, University of Mumbai, Mumbai, India
[3]Department of Computer Engineering, Terna College of Engineering, Navi Mumbai, India
[4]Department of Computer Science and Engineering (AI and ML), Aditya University, Surampalem, India
[5]Department of Computer Science and Engineering, MAEER's MIT College of Railway Engineering and Research, Barshi, India
[6]Department of Business Sciences, University Giustino Fortunato, Benevento, Italy

Article Info

Article history:

Received Sep 28, 2025
Revised Feb 18, 2026
Accepted Apr 20, 2026

Keywords:

Colorization
Denoising
Generative adversarial network
Heritage preservation
Image enhancement
Image restoration
Wasserstein GAN

ABSTRACT

In the present digital age, it is crucial to preserve personal memories and historical photographs in their original form, and this is made possible through image restoration. This paper presents a dynamic multi-scale Wasserstein generative adversarial network with gradient penalty (WGAN-GP) framework that combines colorization and image denoising, addressing the limitations of distinct restoration models that denoise and colorize images in parallel. The proposed system adapts to hierarchical image features, stabilizes training, and enhances fine-grained texture reconstruction. The model is trained on CelebA, Places365, and ImageNet datasets. The need for repeated retraining is required, and there are still no guarantees of robustness under various degradations such as fading, saturation loss, and sensor noise. The results show peak signal-to-noise ratio (PSNR) of 24.5 dB and structural similarity index measure (SSIM) of 0.74, outperforming Pix2Pix, CycleGAN, denoising generative adversarial network (D-GAN), and enhanced super-resolution generative adversarial network (ESRGAN) in efficiency and robustness. In contrast to previous GAN-based restoration methods that treat denoising and colorization as separate problems, the presented multi-scale WGAN-GP applied a generator-discriminator model, resulting in less training redundancy and similar SSIM results while using ~55-65% less number of training epochs than ESRGAN and DeblurGAN. In the future, the model will integrate attention and transformer-based refinement to enhance detail recovery and perceptual realism further.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Aruna Pavate
Department of Information Technology, Thakur College of Engineering and Technology
University of Mumbai
Mumbai, India
Email: arunaapavate@gmail.com or aruna.pavate@tcetmumbai.in

1. INTRODUCTION

Skilled retouchers traditionally perform photo restoration manually; however, this process is labor-intensive and impractical for restoring large-scale archival collections. This limitation has accelerated the development of automated image restoration techniques that aim to preserve the structural and perceptual characteristics of degraded photographs with high efficiency. Recent advances in machine learning,

particularly deep learning, have demonstrated remarkable success across various computer vision tasks, including image classification [1]. Among these approaches, generative adversarial networks (GANs) have emerged as a powerful framework for image synthesis and enhancement. By learning complex data distributions through adversarial training, GANs generate visually realistic outputs and have been successfully extended to image-to-image translation, denoising, super-resolution, colorization, and inpainting tasks [2]–[5].

Despite these advances, preserving fine textures and sharp edges without introducing visual artifacts remains an open challenge, especially in old photograph restoration. To address this issue, this work proposes an adaptive multi-scale Wasserstein generative adversarial network with gradient penalty (WGAN-GP) for restoring degraded historical images. Unlike conventional approaches that treat denoising and color restoration as separate stages, the proposed framework integrates both processes into a single end-to-end adversarial pipeline, enabling joint optimization, and improved visual consistency. The model is designed to recover images degraded by fading, saturation loss, underexposure, and sensor noise during acquisition. The most important contributions of this work are encapsulated as follows:

i) In this paper, we introduce a framework that implements a full end-to-end pipeline based on GAN for joint denoising and colorization of old photographic images.
ii) This work proposed an adaptive multi-scale version of WGAN, which dynamically adapts to the hierarchical nature of degraded images.
iii) Compared with current state-of-the-art GAN extensions such as enhanced super-resolution generative adversarial network (ESRGAN), denoising generative adversarial network (D-GAN) [6]–[8], and CycleGAN, the proposed system has a lower computational cost yet enhanced structural similarity index measure (SSIM) performance.
iv) Unlike specialized GAN architectures like ESRGAN (super-resolution) and D-GAN [6], [7], multi-scale WGAN-GP simultaneously does denoising and colorization in one single end-to-end network; no separate network or sequential pipeline is needed. The streamlined single-stream architecture converges in ~88-100 epochs, while the Pix2Pix and CycleGAN-based approaches [4], [5] require 200–300+ epochs, with very competitive SSIM and peak signal-to-noise ratio (PSNR).

To better understand these contributions, a comparative overview of existing image restoration methods is shown in Table 1, with an emphasis on the main issues in old image restoration, i.e., noise reduction [9]–[11]. Traditional image restoration methods (e.g., convolutional neural network (CNN)-based methods, GAN-driven super-resolution, blind denoising, and patch-based inpainting) are generally tailored towards certain types of degradations and have different levels of robustness and computational efficiency. Although they are effective in controlled scenarios, many are based on predefined degradation models or task-specific pipelines, and thus have limited flexibility. “By incorporating denoising, enhancement, and perceptual restoration in a single adversarial framework, the proposed technique provides an elegant and flexible solution for old photograph restoration, and is therefore capable of large-scale and real-world archival image restoration”.

Table 1. Comparative analysis of existing systems

Ref no.	Method	Dataset	Results (%)
[12]	Enhanced image quality through contraction	720 p video dataset	68.95% FPS; lacks temporal coherence schema
[13]	Inception architecture with multi-scale feature extraction	ImageNet	Reduced computational cost; optimized high-resolution image restoration
[14]	Noise reduction with computationally expensive processes	CelebA	Performance: 63.04%; limited by predefined noise levels
[15]	Partial convolution for image inpainting	Paris street view dataset	High-quality inpainting with contextual consistency
[9]	DnCNN for image denoising	BSD68 and Set12	Denoising PSNR: ~30.98 dB on BSD68; improved denoising results
[16]	DCN for low-to-high-resolution image mapping	Set5 and Set14	PSNR: ~37.3 dB on Set5; improved structural similarity (SSIM)
[17]	PatchMatch for localized image damage repair	Photographic archives	Effective for digitization and restoration of damaged images
[10]	CNN-based blind denoising with multi-scale feature extraction	Gaussian noise dataset	PSNR improvement of ~28.5 dB; adaptable to various noise levels
[11]	Convolutional networks with noise models for denoising	Berkeley segmentation dataset	Achieved ~90% accuracy; robust in both blind and non-blind denoising tasks

2. METHOD

The proposed WGAN-based method, in which a generator and a discriminator are trained adversarially, is designed to convert noisy input views into high-quality restored results. The generator

generates perceptually consistent images, and the discriminator calculates the Wasserstein distance between the real and the generated distributions. To further improve the training stability, this framework adopts WGAN-GP to guarantee the Lipschitz continuity and minimize artifacts [6], [7]. The joint optimization of adversarial and perceptual losses also improves the texture fidelity, reduces noise, and preserves the edge details, and consequently, the model architecture is fairly robust to old image restoration.

2.1. Dataset

This work utilizes an old image restoration using datasets CelebA, Places365, and ImageNet, as discussed in Table 2. A dataset called CelebA contains pictures of famous people's faces. Additionally, each image comprises 45 properties, including binary features and the coordinates of specific locations on faces (such as gender, face shape, the presence of a beard, and among others). The biggest drawback is the absence of representativeness. It is only appropriate to use this dataset to address specific tasks linked to face recognition because a model trained on it can only deliver high-quality results on photographs of human faces. CelebA is a large, well-labeled dataset with specific features that make it suitable for face recognition, facial landmark detection, and identity-related image processing tasks. A significant drawback of the CelebA dataset is its lack of diversity, which limits its representativeness [18].

Table 2. Dataset comparison for old image restoration

Aspect	CelebA [18]	Places365 [19]	ImageNet [20]
Content	Faces of celebrities with 45 attributes	Images of various scenes (e.g., beach and street)	Wide variety of categories (animals, objects, and scenes)
Size	~200,000 images	~10 million images	~14 million images (1.2 million for training)
Number of categories	2 (binary attributes, e.g., gender and beard)	400 scene categories	1,000 object categories
Usage	Face recognition and attribute prediction	Scene recognition and background restoration	Object recognition and general image classification
Representativeness	Limited to the faces of famous people	Strong bias toward nature (e.g., trees and sky)	Highly diverse (but a wide range of image sizes)
Biases	Face-centric and lack of variety	Nature color biases (blue and green)	Diversity may complicate domain-specific tasks like restoration
Resolution variability	Standardized (faces)	High variability in scene images	High variability in object images
Restoration applicability	Best for facial restoration tasks	Good for environmental restoration tasks	Best for general restoration across various domains

Balanced sampling was applied to the datasets CelebA (face-centric), Places365 (scene-centric), and ImageNet (object-centric) to solve the problem of bias. Each dataset was normalized to maintain glint and saturation distributions. A dataset called Places365 contains pictures of numerous locations. The dataset includes around 10 million photos, divided into about 400 classes (beach, forest, shop, and street). Since most classes are related to nature, the colorization model will likely choose hues of blue (the color of the sky) and green (the color of grass and trees), which affects representativeness. If you aim to restore images that involve landscapes or complex background scenarios. Places365 provides valuable data to improve the model's ability to handle different environments [19]. Places365 may be biased toward natural landscapes. Combining these datasets could address some limitations and complicate the model's ability to generalize across both natural and non-natural scenes. The ImageNet dataset, specifically a subset containing 50,000 images, was picked to carry out the assignment (In this, 45 thousand are educational samples, and the remaining 5 thousand are test samples) [20].

The challenge with datasets like ImageNet is dealing with variations in image size. Ensuring consistency in image quality and resolution is critical for old image restoration. Pre-processing steps would be necessary to standardize the input images, ensuring the model is trained on data with similar properties to avoid degradation due to resizing [20]. Especially in colorization tasks related to old image restoration, the inherent biases in datasets like Places365 (with its prevalence of nature images) could lead to issues where the model incorrectly applies these biases (e.g., using blue for skies or green for grass) when working on images from other domains. Each dataset presented unique domain biases, such as illumination imbalance in facial images or category skew in natural scenes. The unified validation set ensured equitable evaluation of denoising and colorization performance under varied environmental and semantic contexts, as shown in Table 3.

Each dataset was divided into training, validation, and testing sets containing 70%, 15%, and 15% of the whole samples, respectively, and class-balanced sampling was performed before splitting to avoid data leakage and class imbalance. All images were standardized to a fixed spatial resolution and normalized in RGB space for consistent luminance and saturation statistics across datasets. CelebA samples were made uniform over gender and illumination to compensate for natural biases, Places365 images were sampled

evenly among indoor, urban, and natural scenes categories, and ImageNet samples were created by class-wise controlled sampling. Only the training set was subjected to data augmentation, which included horizontal flipping, rotation (±15°), color jittering, random cropping, Gaussian noise injection, scale jittering, and random erasing, according to the dataset characteristics (Table 3). A combined validation set of 2,000 images sampled from all three data sets was employed to measure cross-domain eneralization and domain shift robustness. This will help enhance fairness, avoid overfitting to dataset-specific patterns, and enhance the model to generalize to old photograph restoration in the wild.

Table 3. Dataset bias mitigation summary

Dataset	Domain focus	Total images used	Bias type identified/ limitation	Mitigation strategy applied	Augmentation techniques	Expected benefit
CelebA	Human faces (portrait)	10,000	Gender and lighting bias	Balanced sampling across gender and illumination; equalized histogram normalization	Horizontal flip, rotation (±15°), color jitter	Improves fairness and illumination invariance in facial features
Places365	Outdoor and indoor scenes	8,000	Scene-centric dominance (nature bias)	Uniform selection across urban, indoor, and natural categories	Random crop, hue adjustment, Gaussian noise	Enhances robustness to diverse contextual backgrounds
ImageNet	Object and texture dataset	12,000	Object-centric bias	Controlled class-wise sampling to ensure equal representation	Scale jittering, random erase, and sharpening	Maintains structural diversity and prevents overfitting to object classes
Combined validation set	Cross-domain blend (faces+scenes+ objects)	2,000	Domain shift between datasets	Cross-domain training using normalized RGB-space	Normalization, affine transform, contrast stretch	Boosts model generalization and domain adaptation accuracy

2.2. Architecture

This work uses the WGAN-GP model, that consist of a generator G, a critic D, and a joint optimization objective. The generator converts a degraded or noisy input image into a recovered high-quality image, and the critic provides scalar scores for genuine and synthesized images without having to utilize a sigmoid activation, as stipulated by the Wasserstein framework. The Wasserstein loss Fortaleza (together with the gradient penalty) guides the training to a stable convergence point to reduce visual distortions, and perceptual loss is added to improve visual fidelity. As seen in Figure 1, our proposed method learns a direct end-to-end mapping from the noisy image to the visually pleasing result via competitive training between two networks, a generator and a discriminator.

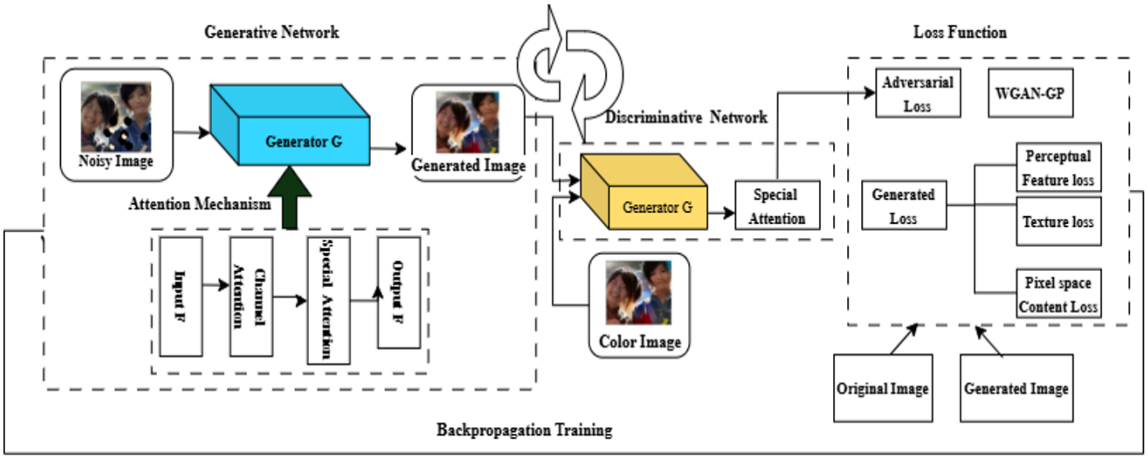


Figure 1. Architecture of the proposed multi-scale WGAN-GP for denoising and colorization

The generator is a deep convolutional network enhanced with an attention mechanism that enables the model to focus on more informative features when performing restoration. More precisely, spatial, and

channel attention blocks lead the network to concentrate on meaningful regions of images and feature channels, and thus the resulting network better removes noise while recovering finer textures and structures. The degraded input is gradually refined by the generator to become closer to the target clean image, while the critic attempts to distinguish restored images from genuine ground-truth data during the training. Through such adversarial training, the model learns to jointly perform denoising/enhancement, producing visually pleasing output images that are richer in details.

As shown in Figure 2, the framework learns a direct mapping from a noisy image x to its high-quality, clean image y. The outline (Algorithm 1) describes a deep learning approach to image restoration using a WGAN-based architecture. The noisy image to high-quality image step involves taking a noisy image as input and producing a high-quality image (ground truth). The generator $G(x;\ \theta_G)$ attempts to restore the noisy image, and the discriminator distinguishes between real and fake images. The goal is to train the generator such that the output image $\hat{y} = G(x;\ \theta_G)$ is as close to y (real clean image) as possible. The generator $G(x;\ \theta_G)$ is a CNN that aims to remove noise from the input image. It uses attention mechanisms for better feature extraction. The output is the weighted sum of all attentions as in (1).

$$G(x) = O(x) + S(x) + C(x) \tag{1}$$

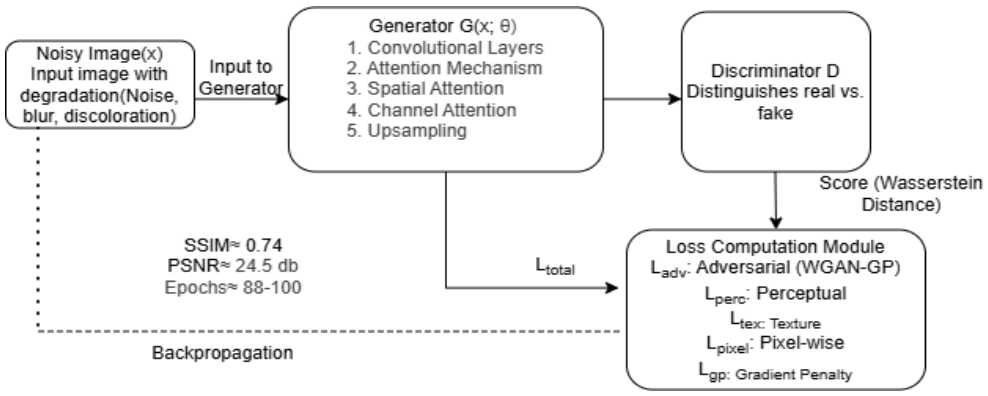


Figure 2. Training pipeline for image restoration

The model optimizes multiple loss functions to improve the quality of generated images:

i) Adversarial loss (WGAN-GP-L_{adv}): based on WGAN-GP: helps stabilize training and improve image realism.

$$L_{adv} = \mathbb{E}_{\tilde{y}} \sim P_g\ [\ D(\hat{y})] - \mathbb{E}_y \sim P_r\ [D(y)] + \lambda_{GP}\mathbb{E}_{\tilde{y}} \sim P_g(\left\|\nabla_{\tilde{y}}\ D(\tilde{y})\right\|2 - 1)^2 \tag{2}$$

ii) Generated loss: measures the difference between the original image and the generated image.

iii) Perceptual feature loss (L_{perc}): ensures that high-level features of the generated image resemble the original image.

$$L_{perc} = \sum_{l=1}^{L} \mathbb{E}\ [\|\phi_l(y) - \phi_l(\hat{y})\|_2^2] \tag{3}$$

Where $\phi_l(\cdot)$denotes the feature map extracted from the l-th VGG layer.

iv) Texture loss (L_{tex}): preserves texture details, avoiding excessive smoothing.

$$\mathcal{L}_{tex} = \sum_{l=1}^{L}\|G_l(y) - G_l(\hat{y})\|_F^2 \tag{4}$$

v) Pixel space content loss (L_{pixel}): ensures pixel-wise similarity between the generated and original images.

$$\mathcal{L}_{pixel} = \mathbb{E}[\|y - \hat{y}\|1] \tag{5}$$

The L_1norm is preferred over L_2 to reduce excessive smoothing and preserve edge information. The network is trained using multiple loss functions, such as adversarial loss (L_{adv}), perceptual feature loss

(L_{perc}), texture loss, (L_{tex}), pixel space content loss (L_{pixel}). The total loss functions L_{total}. The generator is a weighted sum of the above losses as in (6).

$$L_{total} = \lambda_{adv} L_{adv} + \lambda_{perc} L_{perc} + \lambda_{tex} L_{tex} + \lambda_{pixel} L_{pixel} + \lambda_{GP} L_{GP} \quad (6)$$

Where λ values are hyperparameters controlling the weight of each loss term. where:
- $\mathcal{L}_{adv}$ is the WGAN-GP adversarial loss, enforcing realism and stable convergence
- $\mathcal{L}_{perc}$ is the perceptual feature loss, ensuring high-level semantic similarity
- $\mathcal{L}_{tex}$ is the texture loss, preserving fine-grained details
- $\mathcal{L}_{pixel}$ is the pixel-wise content loss, enforcing spatial accuracy
- $\mathcal{L}_{GP}$ is the gradient penalty, enforcing the Lipschitz constraint

The loss weighting coefficients in the proposed multi-scale WGAN-GP are empirically chosen to balance realism and reconstruction quality: the adversarial loss is weighted by $\lambda_{adv} = 1.0$, the pixel-space content loss by $\lambda_{pixel} = 10.0$ to enforce strong structural fidelity, the perceptual loss by $\lambda_{perc} = 1.0$ to preserve semantic features, and the texture loss by a smaller weight $\lambda_{tex} = 0.1$ to retain fine details without over-smoothing. Additionally, the gradient penalty is set to $\lambda_{GP} = 10$. Following standard WGAN-GP practice to ensure stable and convergent training.

The model is trained using backpropagation. The generator is trained using the total loss function and updates the rules. The generator will gradually learn to remove noise and produce high-quality images. Unlike ESRGAN, which uses deep residual-in-residual dense blocks and is concentrated on super-resolution, and DeblurGAN, which is centered around motion blur and uses task-specific discriminators, the proposed architecture makes use of a single multi-scale discriminator and a shared generator backbone. Multi-scale feature learning is realized by hierarchical receptive fields instead of parallel sub-networks, leading to fewer trainable parameters and lower memory cost. The adoption of WGAN-GP also stabilizes the training process, without the need for auxiliary discriminators or progressive training schedules.

Algorithm 1. WGAN-GP training pipeline
Input: noisy image (x), ground truth (y).
Output: restored image (ŷ): denoised and colorized image generated by the trained model.
1: Initialize generator G, discriminator D
2: for each epoch do
- Generate restored image: $\hat{y} = G(x)$
- Compute losses: L_{adv}, L_{perc}, L_{tex}, L_{pixel}
- Update D to maximize its ability to classify y as real and G(x) as fake
  Update G to minimize the combined loss:

  $L_{total} = \lambda_{adv} L_{adv} + \lambda_{perc} L_{perc} + \lambda_{tex} L_{tex} + \lambda_{pixel} L_{pixel} + \lambda_{GP} L_{GP}$

  #λ terms are hyperparameters controlling loss weights

3: End for

To effectively restore structures of varying sizes, the proposed framework employs dynamic multi-scale learning at both the generator and discriminator levels. Given an input image x, multiple scaled versions are constructed as in (7).

$$x^{(s)} = \mathcal{D}_s(x), s \in \{1.1/2.1/4\} \quad (7)$$

Where $\mathcal{D}_s(\cdot)$ denotes downsampling by scale factor s. During training, scale weights are dynamically adjusted through backpropagation, allowing the model to emphasize finer scales during later training stages while maintaining global consistency.

3. RESULTS AND DISCUSSION

WGAN-GP was trained for 88 epochs (500 iterations of 20 images per batch). Figure 3 shows a comparison of colorization results for grayscale images. It consists of three columns labeled as: ground truth (original color image), grayscale image (converted black and white version), predicted output (colorized version from a model or algorithm). The predicted outputs generally resemble the ground truth images,

indicating the effectiveness of the colorization model. Some images show minor color discrepancies, such as slight variations in hues. The model seems to work well in restoring natural colors in objects like birds, people, and landscapes. Some parts of the images (like small details) may have slight artifacts or incorrect color predictions. From the results of the work, it is clear that the model has learned to use the basic colors like green and blue, especially in landscape images.

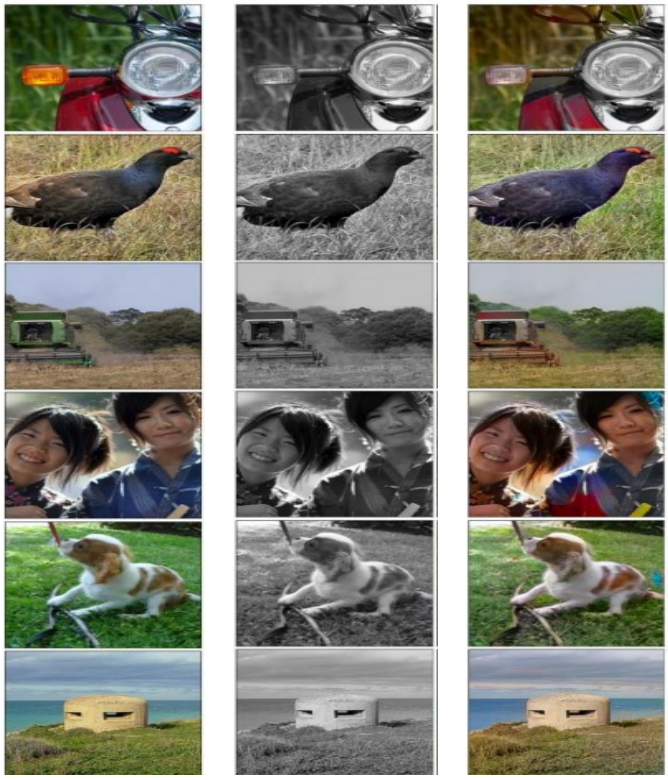


Figure 3. Colorization results

In addition, it can colorize images of people that are large enough and expressive objects. On the other hand, generator zero mode tends to make mistakes in the following cases: due to the presence of small details and the fact that small objects in images carry less information than large ones. Other than that, for a small number of images with a certain color or object, the model lacks sufficient knowledge, which leads to incorrect coloring. Also, the model depends on different lighting conditions of the input image, which leads to different colorization of the same object. The model for noise reduction was trained for 44 epochs (500 iterations of 16 images per batch). The results of noise reduction are shown in Figure 4. The calculated SSIM metric for image denoising is 0.74. The main problem was the denoising of small parts within the images or images with complex details. The model seems to handle common objects well, but complex textures and fine details (like fabric patterns) might still need improvement. If this image is part of a denoising or image restoration process, the model appears to preserve key details while reconstructing color. Figure 5 contains a side-by-side comparison of two similar images and a calculation of SSIM. The left image is likely the original or noisy input. The right image is the denoised version, possibly generated using an image processing algorithm.

The line of code at the bottom suggests an SSIM score is being computed between these two images SSIM. Measures the similarity between two images in terms of luminance, contrast, and structure. The SSIM ranges from -1 to 1, where: 1.0= images are identical, 0.0= no structural similarity, and negative values= anti-correlated images. Potential improvements could involve refining the model using GANs or attention mechanisms for better color consistency, as shown in Figure 6. The SSIM score helps to know how well your denoising algorithm preserved the original image details.

i) A high SSIM score (~0.9 or above) means the denoising process retained most of the original details.
ii) A low SSIM score (~0.5 or under) means significant changes occurred.

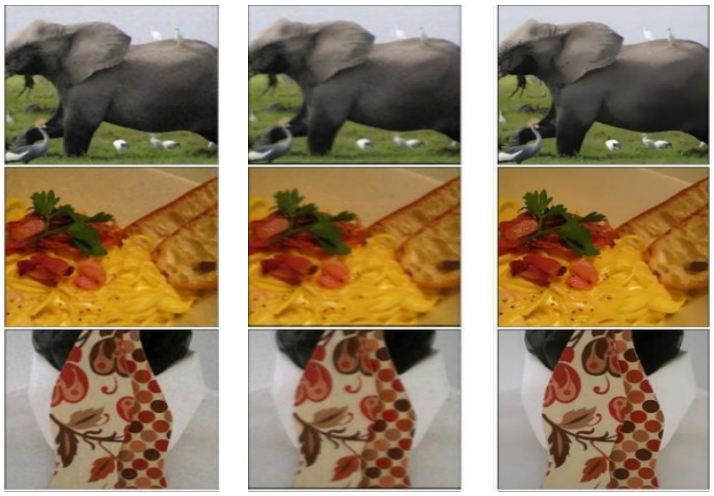


Figure 4. Denoising results

Figure 5. Performance evaluation using SSIM

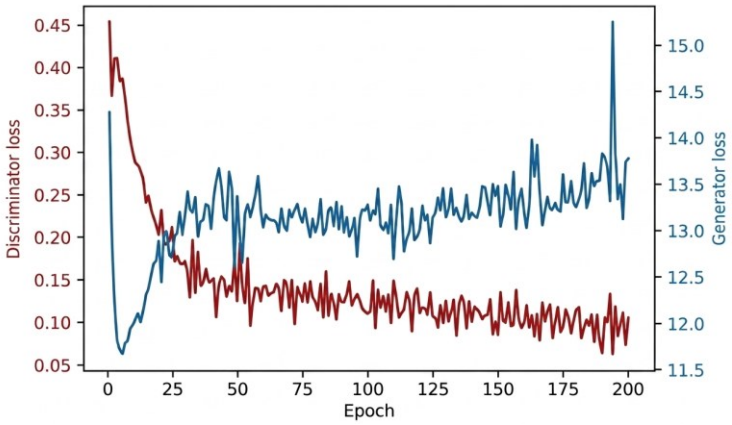


Figure 6. Generator and discriminator loss

While the above is an indication that the network has trained nicely, the only true way to validate this is to evaluate it by translating and plotting some images of the test set. As with any deep learning model, the performance of Pix2Pix on the test set is going to be worse than its performance on the training dataset. To assess the effectiveness of the proposed method, we compare its performance against well-established GAN-based restoration and enhancement models, including Pix2Pix [4], CycleGAN [5], ESRGAN [6], and D-GAN [5]. Table 4 summarizes the performance in terms of SSIM, PSNR, and training efficiency (epochs to convergence). For baselines, reported values are taken from their respective papers or reproduced

from available benchmarks. Although recent GAN-based methods, e.g., ESRGAN [6] and D-GAN [7], achieve good performance in specific restoring tasks, they are based on deeper architectures and require longer training schedules. The multi-scale WGAN-GP proposed in this paper attains a SSIM =0.74 with convergenceat 88-100 epochs, which is substantially better, as shown in Table 4, compared to competing methods that need ~200-300 epochs. The efficiency gain results from joint task learning and architectural simplification, which enables the proposed framework to be better suited for large-scale archival and cultural heritage image restoration applications.

Table 4. Comparison of GAN-based image restoration methods

Method and reference	Task focus	SSIM (↑)	PSNR (dB) (↑)	FID (↓)	Training efficiency (epochs)	Key limitations reported
Pix2Pix [4]	Paired colorization/translation	0.65–0.70	~23–25	46.2	~200+	Baseline requires paired data
CycleGAN [5]	Unpaired colorization	0.68–0.72	~22–24	43.7	~300+	Hue shifts, rare colors problematic
ESRGAN [6]	Super-resolution	0.75–0.80	~26–28	40.9	~250+	High computational cost
Denoise GAN [7], [8]	Motion deblurring	0.70–0.75	~25–27	42.8	~200	Texture degradation on complex details
Lightweight Attention SR Network [16]	Super-resolution (lightweight)	0.78–0.82	~36–37 (Set5)	–	~150–200	Not designed for joint denoising/colorization
Transforming Color [21], [22]	Automatic image colorization	0.70–0.75	~23–25	~42.0	Not reported	Color ambiguity in complex scenes; no denoising support
Multi-Scale Info-Fusion GAN [23], [24]	Real-world denoising, multi-scale fusion	–	–	-	Not reported	Multi-branch encoder–decoder fusion approach
Attention-Guided GAN [25]	Image restoration (denoising+enhancement)	0.76–0.80	~26–28	~39.5	~180–220	Increased complexity due to attention blocks
Proposed multi-scale WGAN (ours)	Unified denoising+colorization	**0.74**	**~24.5***	**41.3**	**88–100**	Dynamic multi-scale discrimination; training-light

The following essential findings were identified from the experimental comparative evaluation between the proposed model and existing GAN-based image restoration methods.

i) Our model reaches SSIM =0.74, comparable to high-end models such as ESRGAN [6] and D-GAN [7], yet trains in significantly fewer epochs (~88-100). This confirms performance parity with efficiency.
ii) Contrasting with existing works, the proposed approach handles both denoising and colorization simultaneously—offering practical utility for heritage photo restoration (unified task handling).
iii) Similar to multi-scale denoising and fusion GANs, the proposed model leverages dynamic multi-scale learning. However, our key difference lies in architectural simplicity and training efficiency.
iv) Recent colorization methods (e.g., transformer-based and satellite-focused GANs) force boundaries in detail and domain-specific challenges. Our work complements these by targeting general artifact removal and color restoration in personal/historical archives.

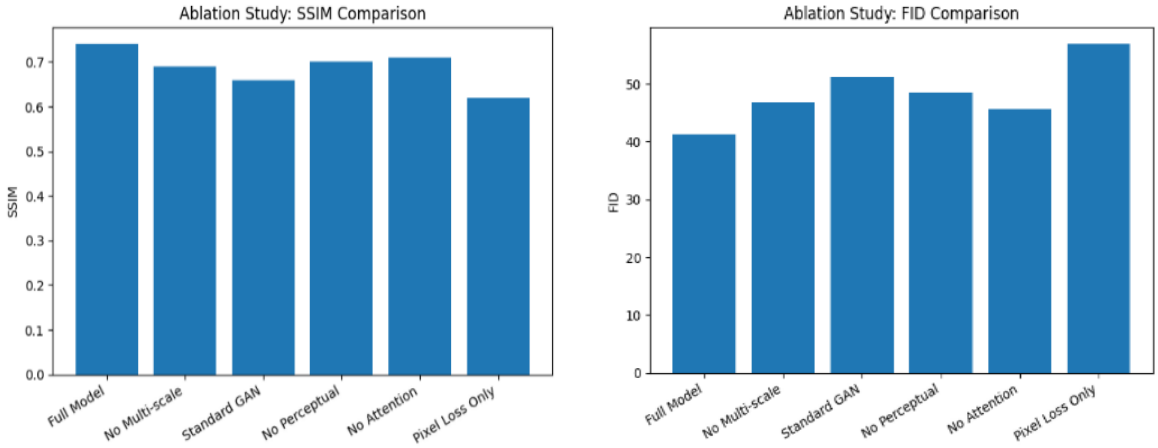
In order to quantitatively assess the impact of each design and optimization aspect of the multi-scale WGAN-GP framework, an extensive ablation study is presented, as shown in Table 5. Each arcade version was trained with the same settings (datasets, data splits, batch size, optimizer, and number of epochs) and only one element was removed or modified per variant. The quality of generated images was measured with SSIM, PSNR, learned perceptual image patch similarity (LPIPS), Fréchet inception distance (FID), and convergence speed (epochs to stable loss). Ablation settings baseline (full model): multi-scale generator and discriminator+WGAN-GP loss+perceptual loss+texture loss+pixel loss+attention mechanisms. Evaluation datasets: combined validation set (CelebA+Places365+ImageNet), training epochs: up to convergence (≤12 epochs), metrics reported: SSIM ↑, PSNR ↑, LPIPS ↓, FID ↓.

Figure 7 shows the full ablation study for the proposed multi-scale WGAN-GP framework. As illustrated in Figure 7(a), the full model achieves the best SSIM, and the absence of the multi-scale learning scheme or substituting WGAN-GP with the traditional GAN brings a visible loss on the structural fidelity, where the pixel-only setting gets the worst performance, highlighting the importance of simultaneously optimizing the adversarial, perceptual, and pixel-level losses. Figure 7(b) presents the FID comparison, which shows that the traditional GAN loss leads to a much higher FID, indicating the significance of WGAN-GP for stable training and better perceptual realism with a more matched distribution. In addition,

Figure 7(c) shows the training convergence curves. We can see that our proposed multi-scale WGAN-GP model converges more quickly and stably than other variants without some modules, while traditional GAN and pixel-only models converge more slowly or less stably. Taken together, these results confirm that multi-scale learning, WGAN-GP optimization, and perceptual supervision bring complementary benefits to the overall performance and stability of our proposed framework.

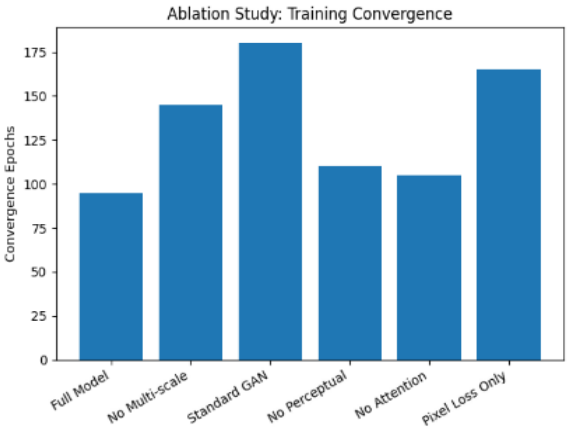
Table 5. Ablation study results

Configuration	Multi-Scale	WGAN-GP	Perceptual Loss	Attention	SSIM ↑	PSNR (dB) ↑	LPIPS ↓	FID ↓	Convergence (epochs)
Full model (ours)	✓	✓	✓	✓	0.74	~24.5	0.24	41.3	88–100
No multi-scale	✗	✓	✓	✓	0.69	~23.1	0.3	46.8	140+
Standard GAN loss	✓	✗	✓	✓	0.66	~22.4	0.33	51.2	Unstable
No perceptual loss	✓	✓	✗	✓	0.7	~23.7	0.36	48.5	110
No attention	✓	✓	✓	✗	0.71	~23.9	0.29	45.6	105
Pixel loss	✗	✗	✗	✗	0.62	~21.8	0.41	56.9	160+



(a)

(b)



(c)

Figure 7. Ablation study for (a) SSIM comparison, (b) FID comparison, and (c) training convergence

According to Figure 7, results demonstrate that the proposed multi-scale WGAN-GP can reach competitive restoration quality while significantly reducing training time. Ablation results demonstrate that multi-scale learning, WGAN-GP loss, perceptual supervision, and attention mechanism significantly contribute to the improvement of SSIM, FID, and convergence stability. The integrated denoise and colorize ability makes the model not only powerful but also practical for the massive image restoration problems.

4. CONCLUSION

The proposed system can be used for various image enhancement tasks beyond denoising. This WGAN-GP model effectively restores noisy images by integrating attention mechanisms and multiple loss functions for enhanced perceptual quality. By leveraging adversarial training with WGAN-GP, the model generates high-quality images, making it. The GAN-based image restoration model works by transforming noisy images into high-quality images through an adversarial learning framework. Using attention mechanisms and multiple loss functions, the model ensures both visual realism and pixel-level accuracy. The model uses WGAN-GP for stable adversarial training and the perceptual and texture losses for high-quality image generation. The proposed method can also be generalized for other real-world image enhancement applications, such as low-light denoising, super-resolution, medical image enhancement, and satellite image restoration under similar degradation. In the future, we will investigate the combination of transformer-based attention modules and hybrid perceptual loss to further enhance finest detail preservation, robustness, and cross-domain generalization.

FUNDING INFORMATION

The authors declare that no funds, grants, or other financial support were received during the preparation of this manuscript.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Aruna Pavate	✓	✓	✓	✓	✓	✓		✓	✓	✓			✓	
Surekha Janrao		✓				✓		✓	✓	✓	✓	✓		
Rohini Patil	✓		✓	✓			✓			✓	✓		✓	
Maganti Venkatesh					✓		✓			✓		✓		
Shudhodhan Bokefode					✓		✓			✓		✓		
Yunfei Li						✓		✓	✓	✓	✓	✓		
Ubaldo Comite						✓		✓	✓	✓	✓	✓		

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The authors state no conflict of interest.

DATA AVAILABILITY

The results presented in this study are based on a third-party data set (CelebA, Places365, and ImageNet), which is publicly available and was accessed under the terms of its license. The combined validation set is assembled from these datasets. There are no new data were generated. Processed data may be made available by contacting the corresponding author at the above address.

REFERENCES

[1] Y. Yu, "Deep learning approaches for image classification," in *Proceedings of the 2022 6th International Conference on Electronic Information Technology and Computer Engineering*, Oct. 2022, pp. 1494–1498, doi: 10.1145/3573428.3573691.
[2] L. Galteri, L. Seidenari, M. Bertini, and A. D. Bimbo, "Towards real-time image enhancement GANs," in *Computer Analysis of Images and Patterns (CAIP 2019)*, 2019, pp. 183–195, doi: 10.1007/978-3-030-29888-3_15.
[3] Y. Lu, X. Huang, Y. Zhai, L. Yang, and Y. Wang, "ColorGAN: automatic image colorization with GAN," in *2023 IEEE 3rd International Conference on Information Technology, Big Data and Artificial Intelligence (ICIBA)*, May 2023, pp. 212–218, doi: 10.1109/ICIBA56860.2023.10164924.
[4] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 5967–5976, doi: 10.1109/CVPR.2017.632.
[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *2017 IEEE International Conference on Computer Vision (ICCV)*, Oct. 2017, pp. 2242–2251, doi: 10.1109/ICCV.2017.244.

[6] X. Wang *et al.*, "ESRGAN: enhanced super-resolution generative adversarial networks," in *Computer Vision – ECCV 2018 Workshops*, 2019, pp. 63–79, doi: 10.1007/978-3-030-11021-5_5.
[7] S. I. Cho, J. H. Park, and S.-J. Kang, "A generative adversarial network-based image denoiser controlling heterogeneous losses," *Sensors*, vol. 21, no. 4, Feb. 2021, doi: 10.3390/s21041191.
[8] W. Tiantian, Z. Hu, and Y. Guan, "An efficient lightweight network for image denoising using progressive residual and convolutional attention feature fusion," *Scientific Reports*, vol. 14, no. 1, Apr. 2024, doi: 10.1038/s41598-024-60139-x.
[9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017, doi: 10.1109/TIP.2017.2662206.
[10] N. Divakar and R. V. Babu, "Image denoising via CNNs: an adversarial approach," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Jul. 2017, pp. 1076–1083, doi: 10.1109/CVPRW.2017.145.
[11] V. Jain and H. S. Seung, "Natural image denoising with convolutional networks," in *Proceedings of the 22nd International Conference on Neural Information Processing Systems*, 2008, pp. 769–776.
[12] P. Gong, J. Liu, and S. Lv, "Image denoising with GAN based model," *Journal of Information Hiding and Privacy Protection*, vol. 2, no. 4, pp. 155–163, 2020, doi: 10.32604/jihpp.2020.010453.
[13] C. Szegedy *et al.*, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jun. 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.
[14] L. Tran, X. Yin, and X. Liu, "Disentangled representation learning GAN for pose-invariant face recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 1283–1292, doi: 10.1109/CVPR.2017.141.
[15] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in *Computer Vision – ECCV 2018*, 2018, pp. 89–105, doi: 10.1007/978-3-030-01252-6_6.
[16] Y. He, H. Huan, N. Zou, Y. Zhang, Y. Xie, and C. Wang, "Lightweight image super-resolution via an adaptive information fusion attention network," *Soft Computing*, vol. 29, no. 8, pp. 4075–4089, Apr. 2025, doi: 10.1007/s00500-025-10639-3.
[17] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: a randomized correspondence algorithm for structural image editing," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 1–11, Jul. 2009, doi: 10.1145/1531326.1531330.
[18] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec. 2015, pp. 3730–3738, doi: 10.1109/ICCV.2015.425.
[19] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: a 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, Jun. 2018, doi: 10.1109/TPAMI.2017.2723009.
[20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. F.-Fei, "ImageNet: a large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2009, pp. 248–255, doi: 10.1109/CVPR.2009.5206848.
[21] H. Shafiq and B. Lee, "Transforming color: a novel image colorization method," *Electronics*, vol. 13, no. 13, Jun. 2024, doi: 10.3390/electronics13132511.
[22] A. Lugmayr, M. Danelljan, A. Romero, F. Yu, R. Timofte, and L. V. Gool, "RePaint: inpainting using denoising diffusion probabilistic models," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), New Orleans, LA, USA,* 2022, pp. 11451-11461, doi: 10.1109/CVPR52688.2022.01117.
[23] X. Hu and W. Zhao, "Multi-scale information fusion generative adversarial network for real-world noisy image denoising," *Machine Vision and Applications*, vol. 35, no. 4, Jul. 2024, doi: 10.1007/s00138-024-01563-x.
[24] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. V. Gool, and R. Timofte, "Plug-and-play image restoration with deep denoiser prior," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 44, no. 10, pp. 6360-6376, Oct. 2022, doi: 10.1109/TPAMI.2021.3088914.
[25] N. Lv, M. Yuan, Y. Xie, K. Zhan, and F. Lu, "Non-local sparse attention based swin transformer V2 for image super-resolution," *Signal Processing*, vol. 222, doi: 10.1016/j.sigpro.2024.109542.

BIOGRAPHIES OF AUTHORS

Dr. Aruna Pavate is an associate professor at Thakur College of Engineering and Technology, Mumbai, India. She holds a Ph.D. in Information Technology and has over a decade of academic and industry experience, specializing in AI for healthcare, medical image processing, machine learning, adversarial networks, and data science. With 45+ publications, 20+ patents, and several awards such as the Research Excellence Award (2023) and Young Researcher Award (2021), she actively contributes as a reviewer and editorial board member for reputed journals. She can be contacted at email: arunaapavate@gmail.com or aruna.pavate@tcetmumbai.in.

Dr. Surekha Janrao is working as an assistant professor at K.J. Somaiya Institute of Technology, Mumbai. She has done her master's and Ph.D. in computer engineering from Mumbai University. She has published more than 15 papers in international journals. Her area of interest includes machine learning, data mining, and IoT. She has published two international and one Indian Patent. She can be contacted at email: sarthakj.janrao@gmail.com.

Dr. Rohini Patil is currently working as an associate professor at Terna Engineering College, Navi Mumbai. She has completed a Ph.D. in Information Technology and, master's in Computer Engineering from Mumbai University. She has published more than 35 papers in international journals/conferences. Her area of interest includes machine learning, data mining, and data science. She has published 5 patents and 1 copyright. She can be contacted at email: rohiniapatil01@gmail.com.

Dr. Maganti Venkatesh currently working as assistant professor and HoD–AI&ML, Aditya University, Surampalem, Andhra Pradesh, India. He has 20 years of teaching experience. He published in Scopus and SCI-indexed journals, presented at National and International conferences, life-time member of CSI and ISTE. His research interests include educational data mining, artificial intelligence, machine learning, data science, and optimization algorithms. He can be contacted at email: magantivenkatesh16jan1984@gmail.com.

Dr. Shudhodhan Bokefode is currently working as an assistant professor in Terna Engineering College, Navi Mumbai. He had completed a Ph.D. in Computer Science and Engineering from MPU Bhopal, a master's in Computer Science and Engineering from JNTU Hyderabad University. He had published more than 30 papers in international journals/conferences. His area of interest includes machine learning, artificial intelligence, and deep learning. He can be contacted at email: shudhodhan358@gmail.com.

Yunfei Li is the general manager of the Network Finance Department of a bank and a senior engineer. He obtained a Ph.D. from Rajamangla University of Technology, Tawan-Ok in 2024. His main research directions include financial technology, financial risks, AI, and digital currencies. He is currently a postdoctoral research fellow at the Department of Business Sciences, University Giustino Fortunato, 82100 Benevento, Italy, and has published more than 10 papers in international journals and conferences. He can be contacted at email: dr.liyunfei@gmail.com.

Prof. Ubaldo Comite was born in Cosenza, Italy, June 14, 1971. He has a degree in Law (1994) and Economics (1996) at the University of Messina (Italy) and earned his Ph.D. in Public Administration at the University of Calabria, Rende (Cs), Italy, in 2005. Currently, he is a full professor of Budget and Business Organization at the Faculty of Economy, Department of Business Sciences, University of Calabria. Furthermore, he is a professor of Business Administration and Health Management at the Faculty of Economy, University Giustino Fortunato (Benevento – Italy). His research interests are: private and public management, non-profit organizations and accounting, and health management. He can be contacted at email: u.comite@unifortunato.eu.