Electronic Research Archive

https://www.aimspress.com/journal/era

ERA, 34(8): 5166–5183.
DOI: 10.3934/era.2026229
Received: 03 April 2026
Revised: 19 May 2026
Accepted: 05 June 2026
Published: 16 June 2026

Research article

ANeRF: Adversarial neural radiance fields for efficient 3D-aware synthesis

Yinghao Peng, Fangqing Gu* and Lei Chen

School of Mathematics and Statistics, Guangdong University of Technology, Guangdong, China

* **Correspondence:** Email: fqgu@gdut.edu.cn.

Abstract: Conventional 3D image reconstruction algorithms suffer from fundamental limitations in disentangling 3D scene geometry from rendering processes, which results in poor generalization to novel viewpoints and object poses. To address this, we propose adversarial neural radiance fields (NeRF) for efficient 3D-aware synthesis. By reformulating NeRF as the generator within an adversarial paradigm, our approach bypasses traditional encoding-decoding pipelines through a direct coordinate-to-pixel transformation, thus significantly reducing the computational overhead. This bidirectional enhancement enables a generative adversarial network (GAN) to leverage NeRF's efficient 3D-aware rendering, while NeRF benefits from adversarial regularization to overcome scene-specific overfitting. Extensive evaluations demonstrate that the proposed adversarial NeRF (ANeRF) performs well on benchmark datasets. ANeRF achieves a 3 dB improvement over traditional NeRF in terms of the peak signal-to-noise ratio (PSNR) upon convergence, while accelerating convergence to 6×10^{-3} loss in 34.4% fewer iterations. The framework maintains exceptional efficiency with 31.42 minutes per epoch and a memory footprint of 5.25 GB. It represents a 43% memory reduction compared to the baseline methods. This explicit 3D representation enables robust multiview consistency and precise scene manipulation, thus establishing a high-efficiency architectural foundation for potential Augmented Reality/Virtual Reality (AR/VR) content generation and digital twin simulations.

Keywords: neural radiance field; generative adversarial networks; adversarial training; 3D-aware image synthesis; coordinate-based generative modeling

1. Introduction

Image reconstruction has emerged as a major research frontier at the intersection of computer vision and graphics, demonstrating transformative impact across diverse application domains. This rapidly advanced field encompasses a broad spectrum of applications, from biometric systems, including face recognition [1] and face reconstruction [2], to sophisticated 3D-aware portrait synthesis with rich details [3] and high-efficiency talking avatar generation [4]. These advancements underscore the field's

shift towards the restoration and synthesis of complex real-world environments, including objects [5] and scenes [6]. In particular, medical imaging stands out as a critical application, where image reconstruction technologies enable clinicians to accurately identify and diagnose pathological regions [7], thus directly contributing to improved health care outcomes. However, conventional 3D reconstruction approaches face inherent limitations in decoupling physical 3D scene formation from rendering processes, while exhibiting poor generalization to novel viewpoints and object poses. This motivates the development of more sophisticated 3D-aware reconstruction methods that can bridge the gap between 2D observations and their underlying 3D representations.

Generative adversarial networks (GANs) [8] have revolutionized generative modeling through their game-theoretic formulation. The optimization framework of GANs [9, 10] employs a game-theoretic formulation based on a two-player zero-sum game that converges to a Nash equilibrium through adversarial training [11]. This elegant paradigm has spawned extensive architectural innovations in various domains. GANs have demonstrated exceptional performances across diverse deep learning applications. In the field of computer vision, super-resolution GANs (SRGAN) [12] and tabular GANs (TGAN) [13] have been used in advanced super-resolution, which can improve the quality of satellite remote sensing imagery. Dual resolution GANs (DRGAN) [14] have enabled pose-invariant image synthesis. Social GANs (SGAN) [15] have improved texture synthesis for 3D reconstruction. Multi-tasking triplet GANs (MTGAN) [16] have been used to improve object detection. The dense resolution network (DRNET) [17] has pioneered video generation. Beyond visual domains, the text-conditioned auxiliary classifier GAN (TAC-GAN) [18] has expanded to natural language processing.

A key advantage of GANs lies in their implicit generative modeling capability. This enables the synthesis of novel samples beyond the training distribution without requiring explicit probabilistic assumptions. This sharply contrasts with traditional generative models, such as autoencoders and variational autoencoders, which require explicit probability distributions and rely on explicit encoding-decoding pipelines. In these architectures, decoders are inherently constrained to only reconstruct those representations produced by their paired encoders, limiting their generalization capacity. Given the intrinsic complexity and multi-modal nature of real world data distributions, such explicit probabilistic assumptions often prove inadequate. However, despite the superior generative flexibility of GANs, they primarily operate at the pixel level [19, 20], incurring substantial computational costs and lacking explicit 3D scene understanding limitations that motivate our integration with neural radiance fields (NeRF) to achieve efficient 3D-aware generation.

NeRF represent a groundbreaking paradigm at the intersection of computer vision and graphics, thus fundamentally transforming how we model and synthesize 3D scenes [21]. NeRF have demonstrated exceptional capabilities across diverse applications [22]. Based on implicit representation, the scene is modeled as a continuous function that outputs color and density by entering 3D coordinates and the observation direction. Representative algorithms include the original NeRF [23] and multum in parvo NeRF (Mip-NeRF) [24], which enhance high-resolution continuity and are suitable to model complex scene details. Furthermore, based on explicit expressions, the scene is directly represented by some discrete geometric elements (voxels, point clouds), with representative algorithms such as 3D Gaussian splatting (3DGS) [25] and Plenoptic Voxels (Plenoxels) [26]. These methods have high computational efficiency and can be edited, but may compromise the fine-grained detail preservation. The last type is a fusion of implicit and explicit expressions, represented by algorithms such as point-based NeRF (PointNeRF) [27] and instant neural graphics primitives (InstantNGP) [28]. This type of method can

balance the drawbacks disclosed under two different scene representations, thereby balancing both the detail accuracy and the computational efficiency [29].

Despite their respective successes, both GANs and NeRF exhibit fundamental limitations that constrain their practical deployment. As powerful generators, GANs exclusively operate at the pixel level [19], which incurs substantial computational overhead. Recent analyses [30] further reveal that this computational burden is not merely due to discriminator training, but significantly from the generator's pixel-level optimization process. The traditional encoding-decoding paradigm further exacerbates these inefficiencies, thus motivating the need for alternative generative architectures. In contrast, NeRF suffer from severe scene specificity [31], a fundamental limitation that arises from its supervised learning paradigm. The tight coupling between network parameters and scene-specific training data means that any variation in scene characteristics, whether changes in object geometry or lighting conditions [32], completely invalidates the learned representation [33]. This requires exhaustive retraining for each new scene, thus resulting in prohibitive computational costs for multiscene applications [34]. These complementary limitations are, GAN's computational inefficiency and NeRF's lack of generalization. This integrated architecture synergistically addresses both challenges through adversarial training of neural radiance fields.

In recent years, there has been increasing efforts to combine GANs and NeRF to address real-world challenges by leveraging their complementary strengths. While recent works have explored diffusion-based priors for rich-detail portrait generation [3] or integrated specialized modules for dynamic talking head synthesis [4], adversarial NeRF (ANeRF) focus on the fundamental challenge of coordinate-based generative efficiency. Several notable approaches have emerged from this synergistic integration. For example, GRAF [35] is the introduction of the first framework that integrates NeRF-based implicit neural representation generators into an unsupervised adversarial paradigm. It enables 3D-aware image synthesis from unstructured 2D image collections by discovering semantic attributes through layer-wise subspace learning. π-GAN [36] lies in its use of periodic implicit neural representations to capture high-frequency scene details and ensure the view consistency. Because of its robust representation, it is often utilized as a feature extractor where its latent codes serve as conditioning signals for complex 3D-aware tasks.

While pioneering works such as GRAF [35] and π-GAN [36] have successfully integrated NeRF with GANs, they often face critical bottlenecks in real-world deployment. GRAF's patch-based sampling can lead to unstable training dynamics, whereas π-GAN's reliance on periodic activation functions (SIREN) incurs prohibitive memory and computational costs. ANeRF fundamentally differentiates itself from these approaches by focusing on the resource efficiency. By leveraging standard positional encoding rather than SIRENs and by stabilizing the generation process via a continuous Jensen-Shannon optimization, ANeRF significantly reduces the memory footprint to 5.25 GB while preserving robust multi-view consistency and high-fidelity rendering.

Therefore, we propose ANeRF, a novel framework for efficient 3D-aware synthesis that achieves end-to-end 3D perception through direct integration rather than knowledge distillation. Unlike existing approaches such as NeRF-GAN Distillation, which employs pre-trained models to transfer 3D knowledge to 2D convolutional generators through teacher-student knowledge distillation, our method fundamentally differs in its architectural philosophy. Knowledge distillation approaches freeze pre-trained teacher models and transfers learned representations to student networks, essentially representing a form of model compression. Critically, the reconstruction quality of such methods depends on the encoding capabilities of pre-trained model latent spaces rather than developing native 3D generation

abilities. In contrast, ANeRF directly employs NeRF as the generator within the adversarial framework from initialization, thus endowing GANs with inherent 3D representational capabilities. The core innovation of ANeRF lies in its bidirectional enhancement mechanism, which transforms the inherent limitations of both GANs and NeRF into complementary strengths through carefully designed architectural fusion. Rather than treating this as simple hybridization, ANeRF fundamentally reimagines the generative pipeline by positioning NeRF's volume rendering as the generator within an adversarial framework. This design eschews the computationally expensive pixel-level optimization of traditional GANs by leveraging NeRF's efficient coordinate-based scene representation.

By embedding volume rendering within an adversarial training regime, the discriminator guides the generator toward photorealistic outputs while simultaneously enhancing NeRF's generalization beyond single-scene overfitting. Adversarial training provides implicit regularization for NeRF, thereby preventing overfitting to high-frequency artifacts and enhancing robustness across diverse scenes, thus effectively serving as latent space data augmentation. Conversely, NeRF's inherent 3D inductive bias, particularly its volume density constraints and multiview consistency, injects physical plausibility into the generation process. This addresses common GAN artifacts such as geometric distortions and perspective inconsistencies, thus resulting in a globally coherent scene synthesis. Through this bidirectional enhancement, ANeRF achieves what neither framework could independently accomplish: efficient, generalizable, and physically grounded 3D-aware image generation that maintains both computational efficiency and high-fidelity rendering quality.

The main contributions of this paper are threefold.

- 3D-Aware Adversarial Generation: We pioneered the integration of NeRF as a generator in GANs, thereby allowing adversarial networks to operate in a 3D scene space rather than a 2D pixel space, thus fundamentally improving their understanding and synthesis of 3D structures.
- Generalization through Adversarial Regularization: We transform NeRF from a scene-specific reconstruction model into a generalizable generative framework through adversarial training, thus eliminating the need for per-scene retraining while maintaining a high-fidelity rendering quality.
- Computational Efficiency: Our approach achieves a 34.4% reduction in terms of the training time and a significantly lower memory consumption compared to baseline methods, while simultaneously improving the generation quality, demonstrating that architectural synergy can yield both performance and efficiency gains.

The remainder of this paper is organized as follows: Section 2 provides the mathematical foundations of GANs and NeRF; Section 3 presents the ANeRF architecture, thereby detailing the adversarial volume rendering pipeline and enhanced positional encoding; Section 4 demonstrates the experimental results, including comparative evaluations and ablation studies; and Section 5 concludes with key insights and future research directions.

2. Preliminaries

2.1. Generative adversarial networks

GANs consist of two neural networks engaged in a competitive game-theoretic framework: a generator G that synthesizes data from random noise and a discriminator D that distinguishes real samples from generated ones. Figure 1 illustrates the adversarial training.

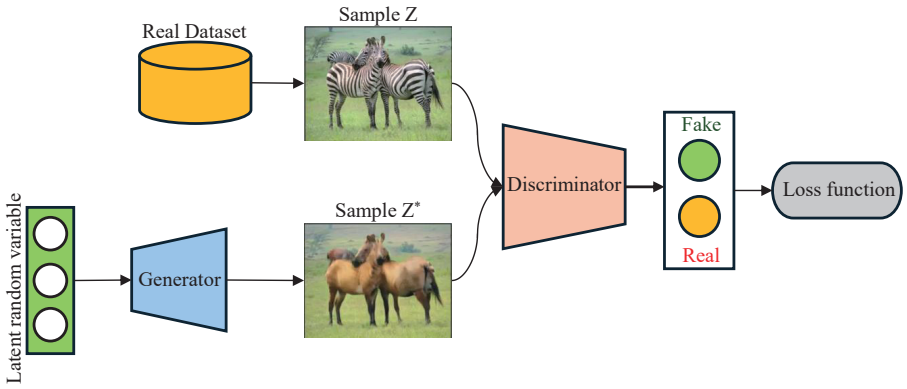


Figure 1. Illustration of the generative adversarial network architecture. The generator G transforms random noise z into synthetic data, while the discriminator D learns to distinguish real from generated samples through adversarial training.

Let $p_{data}(\mathbf{x})$ denote the true data distribution and $p_{\mathcal{Z}}(\mathbf{z})$ denote the prior noise distribution. The generator $G : \mathcal{Z} \to \mathcal{X}$ maps the noise vectors $\mathbf{z} \sim p_{\mathcal{Z}}(\mathbf{z})$ to the data space, thereby inducing a distribution $p_g(\mathbf{x})$. The discriminator $D : \mathcal{X} \to [0, 1]$ outputs the probability that a sample comes from p_{data} rather than p_g. The adversarial training objective is formulated as a minimax game as follows:

$$\min_G \max_D \mathcal{V}(D, G) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[\log D(\mathbf{x})] + \mathbb{E}_{\mathbf{z}\sim p_{\mathcal{Z}}(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))], \tag{2.1}$$

where the discriminator maximizes the probability of correctly classifying real and generated samples, while the generator minimizes $\log(1 - D(G(\mathbf{z})))$ to fool the discriminator.

At convergence, this game reaches a nash equilibrium where the generated data distribution p_g aligns with the real data distribution p_{data} and the output of the discriminator $D(\mathbf{x}) = \frac{1}{2}$ for all $\mathbf{x}$. However, in practice, GANs suffer from training instabilities and mode collapse, particularly when operating in a high-dimensional pixel space, which are limitations that motivate our integration with neural radiance fields for more stable 3D-aware generation.

2.2. Neural radiance fields

NeRF represent 3D scenes as continuous volumetric functions that map spatial locations and viewing directions to radiance and density values. Figure 2 illustrates the NeRF mechanism. Specifically, NeRF models a scene as a function $F : (\mathbf{x}, \mathbf{d}) \to (\mathbf{c}, \sigma)$, where $\mathbf{x} = (x, y, z) \in \mathbb{R}^3$ denotes the 3D spatial position, $\mathbf{d} = (\theta, \phi) \in \mathbb{R}^2$ represents the viewing direction in the unit sphere, $\mathbf{c} = (r, g, b) \in [0, 1]^3$ is the emitted RGB color, and $\sigma \in \mathbb{R}^+$ is the volume density.

This continuous representation is parameterized by a multilayer perceptron (MLP) network F_Θ as follows:

$$F_\Theta : (\gamma(\mathbf{x}), \gamma(\mathbf{d})) \to (\mathbf{c}, \sigma), \tag{2.2}$$

where Θ denotes the network parameters, and $\gamma(\cdot)$ is a positional encoding function that maps inputs to a higher dimensional space as follows:

$$\gamma(p) = \left[\sin(2^0\pi p), \cos(2^0\pi p), \ldots, \sin(2^{L-1}\pi p), \cos(2^{L-1}\pi p)\right]^T. \tag{2.3}$$

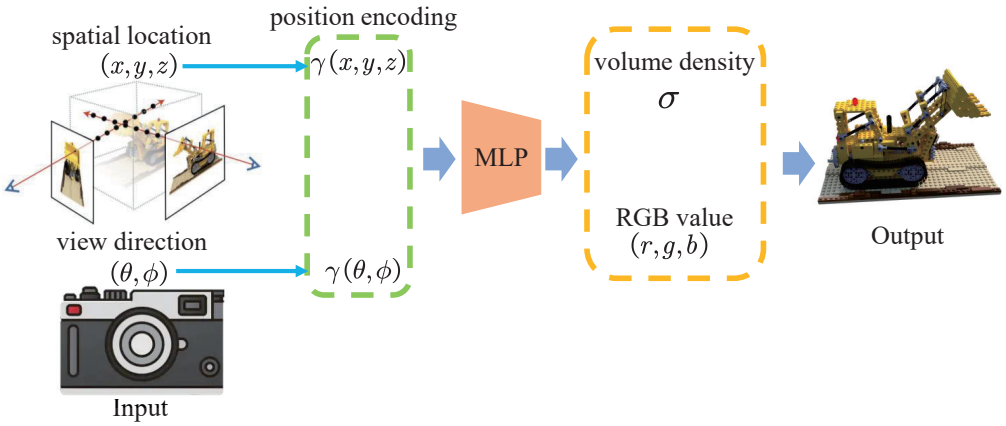


Figure 2. Neural radiance field architecture. The MLP network F_Θ maps positionally-encoded coordinates and viewing directions to color and density values.

This encoding enables the network to capture high-frequency details that would otherwise be smoothed by the spectral bias of deep networks toward low-frequency functions. The choice of L controls the resolution of the representable details.

To render a pixel color, NeRF uses volume rendering along camera rays. For a ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ with an origin of $\mathbf{o}$ and a direction of $\mathbf{d}$, the accumulated color is computed as follows:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot \mathbf{c}(\mathbf{r}(t), \mathbf{d})\, dt, \tag{2.4}$$

where $T(t) = \exp\left(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))\, ds\right)$ represents the accumulated transmittance along the ray, and $[t_n, t_f]$ defines the integration bounds from the near to the far planes. $\mathbf{c}(\mathbf{r}(t), \mathbf{d})$ denotes the directional color.

Despite achieving a remarkable reconstruction quality for individual scenes, NeRF suffers from severe scene specificity that requires complete retraining for each new environment. This fundamental limitation, combined with the computational efficiency of its coordinate-based representation, makes NeRF an ideal candidate for integration with adversarial training to achieve generalizable 3D-aware generation, which is the core motivation for our ANeRF framework.

3. ANeRF: The proposed method

In this section, we present ANeRF, a novel framework that synergistically combines NeRF with adversarial training to achieve an efficient and generalizable 3D-aware image synthesis. Unlike conventional NeRF approaches that require scene-specific optimization and precise camera poses, the proposed method learns a category level generative model capable of synthesizing diverse novel views from unstructured image collections without explicit geometric supervision. Unlike traditional NeRF, which requires a dense set of multi-view images to reconstruct a single specific scene, ANeRF learns the underlying 3D distribution of an entire object class. Once trained, it can generate an infinite number of novel identities and scenes by simply sampling from the latent space, thereby completely bypassing the need for per scene optimization or multi-view reconstruction.

3.1. Overview

The key insight of ANeRF lies in leveraging the complementary strengths of both paradigms: NeRF's efficient 3D scene representation provides structural inductive bias for generation, while adversarial training enables generalization beyond single-scene reconstruction. As illustrated in Figure 3, our architecture consists of three principal components.

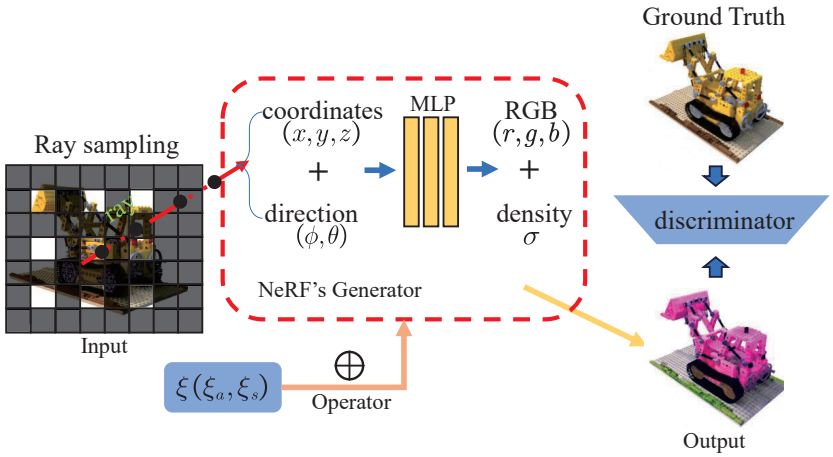


Figure 3. Architecture of ANeRF. The framework integrates a NeRF-based generator G that maps spatial coordinates and viewing directions to radiance fields, a discriminator D that distinguishes real from synthesized views, and a latent conditioning mechanism ξ that enables scene-specific control. Specifically, ξ_a controls the appearance and ξ_s controls the geometry of the object.

3.1.1. Formal definition of the generative mapping

Let $Z \subset \mathbb{R}^d$ denote the latent space, where a latent code $\xi \in Z$ is sampled from a prior distribution p_z. Unlike reconstruction-based frameworks that rely on an explicit inference network or encoder to compute latent representations from input images, ANeRF directly utilizes these sampled codes to span the generative manifold. We define the ANeRF generator as the following composite operator:

$$G = \mathcal{R} \circ \mathcal{F}_\Theta, \tag{3.1}$$

where $\mathcal{F}_\Theta : \mathcal{Z} \times \mathbb{R}^3 \times \mathbb{S}^2 \to \mathbb{R}^4$ represents the continuous 3D scene. For a given ξ, it defines a spatial function $\mathbf{f}_\xi(\mathbf{x}, \mathbf{d}) = (\mathbf{c}, \sigma)$, thereby mapping a 3D coordinate $\mathbf{x}$ and viewing the direction $\mathbf{d}$ to its emitted color $\mathbf{c}$ and volume density σ. Then, $\mathcal{R}$ is an operator that projects the 3D radiance field into the 2D image manifold $\mathcal{I}$. Given a camera pose $\mathbf{p} \sim p_{pose}$, the rendered image is obtained via $I = \mathcal{R}(\mathbf{f}_\xi, \mathbf{p})$. In our framework, the network parameters Θ are fixed after the adversarial training phase. Each distinct 3D scene or object is uniquely determined by the latent code ξ. Consequently, synthesizing a new scene does not involve any additional training or multi-view data; it is a straightforward pass through the learned generator.

3.2. Generator architecture and operator design

The generator G_Θ replaces traditional pixel-level convolution with coordinate-based radiance field generation to bypass computational bottlenecks.

3.2.1. Neural radiance field operator

The core mapping $\mathcal{F}_\Theta$ is parameterized by a multi-layer perceptron (MLP). To resolve high-frequency details, we apply a periodic positional encoding $\gamma(\cdot)$. The radiance field for a scene conditioned on ξ is represented as follows:

$$\mathbf{f}_\xi(\mathbf{x}, \mathbf{d}) = \mathrm{MLP}_\Theta(\gamma(\mathbf{x}), \gamma(\mathbf{d}); \xi), \tag{3.2}$$

where $\gamma(\cdot)$ projects inputs into a higher-dimensional spectral space. The latent code ξ is randomly partitioned into two disjoint segments: $\xi = (\xi_a, \xi_s)$. This partitioning occurs during the initialization of the generative process, where ξ_a modulates the radiance attributes (appearance) and ξ_s controls the volumetric density distribution (geometric structure) within the MLP-based radiance field operator.

3.2.2. Differentiable volume rendering operator

To synthesize the color $C(\mathbf{r})$ of a pixel corresponding to a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, the rendering operator $\mathcal{R}$ evaluates the following integral:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t) \cdot \sigma(\mathbf{r}(t)) \cdot \mathbf{c}(\mathbf{r}(t), \mathbf{d}) dt, \tag{3.3}$$

where the accumulated transmittance $T(t)$ is defined by the following exponential decay function:

$$T(t) = \exp\left(- \int_{t_n}^{t} \sigma(\mathbf{r}(s)) ds \right). \tag{3.4}$$

The differentiability of $\mathcal{R}$ with respect to the radiance field parameters Θ allows for end-to-end gradient propagation from the 2D discriminator back to the 3D scene representation.

3.3. Adversarial optimization and training objective

The training process is formulated as a minimax game between the generator G_Θ and a discriminator D_Φ. We define the adversarial value function $\mathcal{V}(D_\Phi, G_\Theta)$ as follows:

$$\min_\Theta \max_\Phi \mathcal{V}(D_\Phi, G_\Theta) = \mathbb{E}_{I \sim p_{data}}[\log D_\Phi(I)] + \mathbb{E}_{\xi \sim p_z, \mathbf{p} \sim p_{pose}}[\log(1 - D_\Phi(\mathcal{R}(\mathcal{F}_\Theta(\xi), \mathbf{p})))]. \tag{3.5}$$

To enhance training stability and prevent mode collapse, we adopt the Jensen-Shannon (JS) divergence as the optimization criterion for the generator as follows:

$$\mathcal{L}_G = JS(p_{data} \parallel p_g), \tag{3.6}$$

where p_g is the distribution of images induced by the operator $G_\Theta = \mathcal{R} \circ \mathcal{F}_\Theta$. This objective ensures that the discriminator provides informative gradients, thereby guiding the 3D generator to capture the underlying data distribution while maintaining physical plausibility through the volume rendering constraint.

Algorithm 1 ANeRF Training Procedure

Input: Dataset $\mathcal{D}$, epochs E, learning rates η_G, η_D
Output: Trained generator G_Θ

Initialize generator G_Θ (NeRF) and discriminator D_Φ
for $e = 1$ to E **do**
  for batch $\mathcal{B} \in \mathcal{D}$ **do**
    // Discriminator Update
    Sample real images: $\mathcal{I}_{real} \sim \mathcal{B}$
    Sample camera poses: $\{\mathbf{x}_i, \mathbf{d}_i\} \sim p_{camera}$
    Generate fake images: $\mathcal{I}_{fake} = \text{VolumeRender}(G_\Theta, \mathbf{x}_i, \mathbf{d}_i)$
    Compute discriminator outputs: $y_{real} = D_\Phi(\mathcal{I}_{real})$, $y_{fake} = D_\Phi(\mathcal{I}_{fake})$
    Update discriminator: $\mathcal{L}_D = -\log(y_{real}) - \log(1 - y_{fake})$
    $\Phi \leftarrow \Phi - \eta_D \nabla_\Phi \mathcal{L}_D$
    // Generator Update
    Sample new poses: $\{\mathbf{x}_j, \mathbf{d}_j\} \sim p_{camera}$
    Generate images: $\mathcal{I}_{gen} = \text{VolumeRender}(G_\Theta, \mathbf{x}_j, \mathbf{d}_j)$
    Update generator: $\Theta \leftarrow \Theta - \eta_G \nabla_\Theta \mathcal{L}_G$
  end for
end for

4. Experiments

4.1. Experimental setup

4.1.1. Datasets

We evaluated ANeRF on four benchmark datasets that comprehensively assess the reconstruction quality and the generalization capability.

- NeRF-Synthetic [23]*: Eight synthetic objects with complex geometry and materials, each containing 100 training views and 200 test views at 800×800 resolution.
- 3D Chairs [37]†: A collection of 1393 Computer-Aided Design (CAD) chair models rendered from multiple viewpoints, which tests the shape diversity and geometric consistency across different object categories.
- CelebA [38]‡: Large-scale face dataset with 202,599 images, which evaluates the performance on complex facial features and expressions.
- LLFF [39]§: Real-world forward-facing scenes captured with handheld cameras, which assesses the robustness to natural imaging conditions and camera pose variations.

4.1.2. Evaluation metrics

We employ three complementary metrics to comprehensively evaluate the generation quality:

*https://www.kaggle.com/datasets/nguyenhung1903/nerf-synthetic-dataset
†https://www.di.ens.fr/willow/research/seeing3Dchairs/data/rendered_chairs.tar
‡https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html
§http://cseweb.ucsd.edu/ viscomp/projects/LF/papers/SIG19/lffusion/testscene.zip

The Fréchet Inception Distance (FID) [40] measures the distributional similarity between real and generated images using feature statistics from a pre-trained Inception-v3 network. Lower values indicate better quality. Given the mean feature vectors μ_r, μ_g and the covariance matrices Σ_r, Σ_g for real and generated distributions, respectively, the FID is computed as follows:

$$\text{FID} = \|\mu_r - \mu_g\|^2 + \text{Tr}(\Sigma_r + \Sigma_g - 2(\Sigma_r\Sigma_g)^{1/2}). \tag{4.1}$$

The Maximum Mean Discrepancy (MMD) [41] quantifies the distance between distributions in a reproducing kernel Hilbert space. For samples $X = \{x_i\}_{i=1}^m$ and $Y = \{y_j\}_{j=1}^n$ with kernel $k(\cdot, \cdot)$,

$$\text{MMD}^2 = \frac{1}{m^2}\sum_{i,j}^{m} k(x_i, x_j) + \frac{1}{n^2}\sum_{i,j}^{n} k(y_i, y_j) - \frac{2}{mn}\sum_{i=1}^{m}\sum_{j=1}^{n} k(x_i, y_j). \tag{4.2}$$

The Peak Signal-to-Noise Ratio (PSNR) [42] evaluates the quality of reconstructed images compared to the ground truth. The PSNR is expressed in decibels (dB), with higher values indicating a better reconstruction quality. For original and reconstructed images I and K of size $m \times n$, the PSNR is computed as follows:

$$\text{PSNR} = 10 \cdot \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right), \tag{4.3}$$

where $\text{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i, j) - K(i, j)]^2$, and MAX_I represents the maximum possible pixel value. In our experiments, the PSNR is primarily employed within the ablation studies to quantify the specific improvements in the 3D scene reconstruction fidelity against the vanilla NeRF baseline. For cross-model generative evaluations, we prioritize distributional metrics such as FID and MMD.

4.2. Qualitative results

In this section, we provide a visual analysis of ANeRF's generative capabilities. We select the Lego and Fern scenes as the primary representatives [43] for qualitative evaluation, as they provide a rigorous test of the model's ability to handle both rigid geometric complexity and intricate, natural textures.

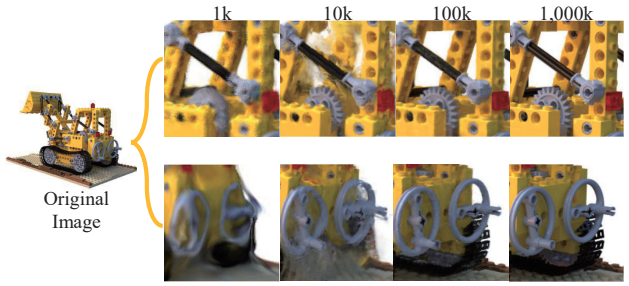
4.2.1. Progressive rendering quality

Figure 4 demonstrates the progressive refinement of the rendering quality during ANeRF training on the Lego and Fern scenes from the NeRF-Synthetic dataset. The visualization reveals three distinct phases that illustrate the effectiveness of our adversarial training approach: 1) initial coarse geometry formation (1 k iterations), where basic 3D structure emerges; 2) detail emergence and texture refinement (10–100 k iterations), characterized by improved surface details and material properties; and 3) final high-frequency detail capture (1000 k iterations), achieving photorealistic rendering with fine-grained textures. This progression validates our adversarial training's effectiveness in guiding the generator toward photorealistic synthesis while maintaining the geometric consistency and 3D structural integrity throughout the learning process.

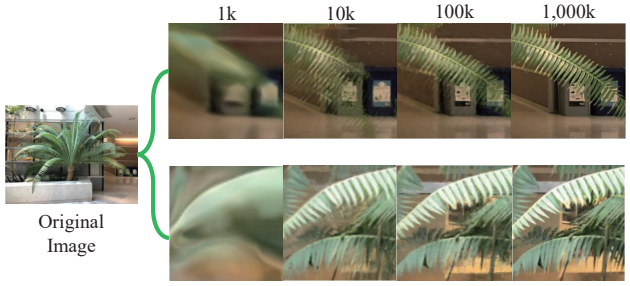
4.2.2. Disentangled appearance control

Figure 5 showcases ANeRF's capability to manipulate disentangled appearances through the learned latent space conditioning mechanism. By modulating the appearance latent code ξ_a while preserving the

geometric latent code ξ_s, our framework enables photorealistic color transformations without affecting the structural integrity, which is a key advantage of our coordinate-based generation approach. The Lego model demonstrates vivid color variations (pink, red, green) while maintaining precise brick geometry and structural coherence, and the Fern exhibits natural color gradients (yellow, orange, light green) with preserved leaf structures and fine details. This disentanglement validates our architecture's ability to separate the appearance from the geometry in the latent representation.

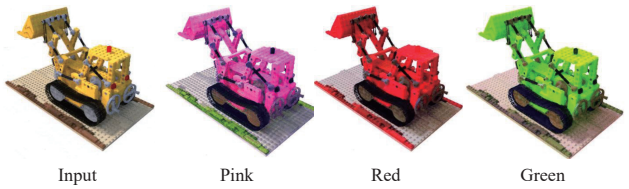


(a) Lego: Progressive refinement from coarse geometry to fine details

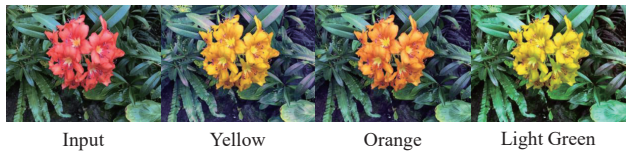


(b) Fern: Evolution from blurry structure to sharp leaf textures

Figure 4. Progressive rendering quality improvement during ANeRF training.



(a) Lego: Color variations with preserved geometric structure



(b) Fern: Natural color transformations maintaining leaf details

Figure 5. Disentangled appearance editing in ANeRF.

4.3. Comparative evaluation

4.3.1. Baseline methods

We compare ANeRF against three state-of-the-art generative rendering frameworks:

2DGAN [44]: A stabilized GAN variant that addresses convergence issues in non-continuous distributions through regularization-based gradient penalties. This method achieves stable training on low-dimensional manifolds, but lacks explicit 3D understanding and geometric consistency.

Platonic GAN(PGAN) [45]: A progressive GAN framework that discovers latent 3D structures from 2D collections via differentiable rendering. It employs visual shell rendering and ray tracing for 3D reconstruction but requires extensive computational resources and suffers from limited scalability.

HoloGAN (HGAN) [46]: A holistic GAN that learns explicit 3D representations with rigid-body transformations for viewpoint control. In terms of achieving a view consistency, it suffers from a high memory overhead and limited generalization capability across diverse scenes.

4.3.2. Quantitative results

Table 1 presents comprehensive quantitative comparisons in four benchmark datasets. While this table focuses on the distributional quality of synthesized samples using FID and MMD, the underlying reconstruction precision of the ANeRF architecture is further validated through a PSNR analysis in the subsequent ablation study (Section 4.4). ANeRF achieves state-of-the-art performances in terms of all metrics, with particularly striking improvements in 3D Chairs (FID: 57.4% reduction) and Birds (FID: 50% reduction) compared to the best baseline methods. These results demonstrate the effectiveness of our coordinate-based generation approach to achieve a superior image quality while maintaining computational efficiency.

Table 1. Quantitative comparison on benchmark datasets. Best results in **bold**, second-best underlined.

Method	3D Chairs		Birds (LLFF)		Cars (LLFF)		CelebA	
	FID↓	MMD↓	FID↓	MMD↓	FID↓	MMD↓	FID↓	MMD↓
2DGAN	47	1.64	24	1.29	66	1.88	15	1.67
PGAN	199	2.81	179	2.16	169	3.62	321	3.98
HGAN	59	2.77	78	1.07	134	2.43	25	2.19
ANeRF	**20**	**0.36**	**12**	**0.28**	**15**	**0.33**	**13**	**0.74**

These substantial improvements in both FID and MMD metrics validate our central hypothesis that integrating neural radiance fields with adversarial training creates a synergistic effect, thereby enabling superior modeling of complex visual distributions while maintaining 3D structural consistency. The coordinate-based generation paradigm effectively bridges the gap between efficient 3D representation and a high-quality image synthesis, thus addressing the fundamental limitations of both traditional GANs and vanilla NeRF approaches.

4.3.3. Computational efficiency

Table 2 demonstrates the significant computational advantages of ANeRF over existing methods. Our method achieves a 2.0–2.6× speedup in training time and a 43% reduction in memory consumption compared to baseline methods. These efficiency gains are attributed to our coordinate-based generation paradigm, which bypasses the expensive [47] encoding-decoding pipeline and pixel-level optimization inherent in traditional GANs. Specifically, our method achieves an average training time of 31.42 minutes per epoch with a memory footprint of only 5.250 GB. The substantial computational efficiency improvements validate one of our key contributions: the synergistic integration of neural radiance fields and adversarial training not only enhances the generation quality but also reduces the computational overhead. This efficiency gain, combined with the 34.4% reduction in the training convergence time demonstrated in our ablation studies, positions ANeRF as a practical solution for scalable 3D-aware content generation in resource-constrained environments. The consistent performance gains across both synthetic objects (3D Chairs) and high-diversity real-world data (CelebA) further confirm that ANeRF is a robust, category level generative framework rather than a scene-specific editing tool.

While the current implementation focuses on training-phase optimization, ANeRF's architecture inherently favors high-efficiency inference. By eliminating the computationally expensive encoding-

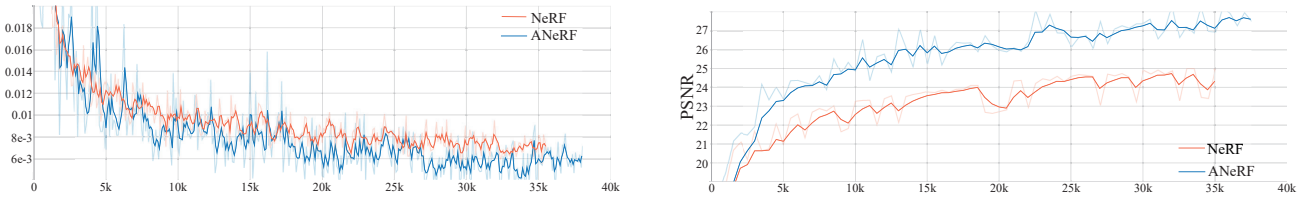
decoding pipelines common in traditional GANs and reducing the memory footprint by 43% (as shown in Table 2), the framework significantly lowers the hardware requirements for 3D-aware synthesis. This reduction to 5.25 GB of memory and the use of direct coordinate-to-pixel transformation provide a promising path toward real-time deployment on resource-constrained mobile Augmented Reality/Virtual Reality (AR/VR) devices.

Table 2. Computational efficiency comparison on NVIDIA RTX 3090.

Method	Time per Epoch↓		Memory Usage↓	
	Maximum	Average	Maximum	Average
2DGAN	86.73 min	69.94 min	8593 MB	7521 MB
PGAN	84.51 min	64.57 min	7677 MB	6691 MB
HGAN	75.01 min	62.82 min	9234 MB	7937 MB
ANeRF	**33.73 min**	**31.42 min**	**5504 MB**	**5250 MB**

4.4. Ablation study

To validate the contribution of adversarial training to our framework, we conduct ablation experiments that compare vanilla NeRF against our full ANeRF model on the NeRF-Synthetic dataset. This study isolates the impact of adversarial regularization on both the training dynamics and the quality of the generation.



(a) Training loss convergence **(b)** PSNR evolution

Figure 6. Ablation study comparing NeRF baseline with ANeRF. The curves illustrate the training dynamics for a representative scene (Lego) from the NeRF-Synthetic dataset. (a) The MSE convergence shows 34.4% faster optimization with adversarial training. (b) The PSNR demonstrates consistent 3+ dB improvement in reconstruction quality.

Figure 6a reveals that ANeRF achieves a significantly faster convergence despite higher initial oscillations from adversarial dynamics. To ensure a rigorous and consistent comparison between the generative ANeRF framework and the supervised NeRF baseline, we utilize the MSE as a unified monitoring metric for both models during the training process. Although their internal optimization objectives differ, thereby evaluating them against a common reconstruction error proxy allows for an objective assessment of their learning efficiency.

The model reaches 6×10^{-3} losses in 15 k iterations. It has 7 k iterations faster than vanilla NeRF and maintains a 27.4% lower average loss throughout the training process. This accelerated optimization comes from the discriminator providing informative gradients that guide the generator toward realistic distributions.

Figure 6b demonstrates consistent quality improvements, with ANeRF achieving 34 dB PSNR compared to NeRF's 31 dB plateau.

Figure 6b demonstrates consistent quality improvements throughout the training process. While the absolute PSNR values vary across scenes, with the Lego scene in Figure 6b reaching approximately 28 dB, the full ANeRF model achieves a peak average PSNR of 34.1 dB on the CelebA dataset (as detailed in Table 3). The 3+ dB gain (equivalent to 2× reduction in reconstruction error) validates that adversarial regularization enhances the model's capacity to capture fine details while preventing overfitting to training views. This improvement directly results from our synergistic architecture, where adversarial training provides implicit regularization for NeRF, thus enabling better generalization and higher fidelity synthesis compared to scene-specific optimization alone.

To further validate the contribution of individual components to the framework's cross-scene generalization capability, we conducted an extensive ablation study on the CelebA dataset, which contains high-diversity facial data. We evaluated four variants: 1) w/o Positional Encoding, 2) w/o Latent Conditioning, 3) w/o Adversarial Training, and 4) the full ANeRF model. The quantitative results are summarized in Table 3.

Table 3. Component-level ablation study on the CelebA dataset.

Variant	Dataset	FID ↓	PSNR ↑
w/o Positional Encoding	CelebA	28.5	29.2
w/o Latent Conditioning	CelebA	84.2	22.1
w/o Adversarial	CelebA	56.1	27.4
ANeRF	CelebA	13.0	34.1

The performance degradation that occurred when the latent conditioning mechanism was removed. The FID surged to 84.2 while the PSNR dropped to 22.1 dB. In the absence of the latent code ξ, the network is forced to map thousands of distinct identities and lighting conditions into a single set of weights. This confirms that latent conditioning is the primary driver for ANeRF's ability to distinguish and synthesize diverse scenes without individual retraining. Compared to the w/o Adversarial variant, the full ANeRF model achieved a 76.8% reduction in the FID. The adversarial discriminator serves as a powerful implicit regularizer, thereby providing informative gradients that guide the generator to bypass local optima and capture the true underlying data distribution. Removing the periodic positional encoding function $\gamma(\cdot)$ resulted in a 5 dB drop in the PSNR. Without positional encoding, the generator fails to capture high-frequency facial textures such as pores and fine hair, thus verifying the spectral bias of deep networks toward low-frequency functions.

5. Conclusions

In this paper, we presented ANeRF, a novel neural rendering framework that synergistically combines adversarial learning with NeRF to overcome the fundamental limitations of both paradigms. This method significantly enhances the generalization ability beyond the synthesis of constrained view scenes and effectively reduces the loss of pixel-level information in the rendering process by adding a periodic position encoding function. The proposed algorithm achieves a 34.4% reduction in the training convergence time and a 43% reduction in the memory consumption through compact latent representations, thus establishing a new efficiency benchmark for realistic neural rendering. Our

extensive experiments across four benchmark datasets demonstrated state-of-the-art performances in both quantitative metrics and qualitative assessments, thus validating the effectiveness of our synergistic approach.

While ANeRF establishes a new paradigm for efficient neural rendering, several promising avenues remain: 1) extending to dynamic scenes with temporal consistency; 2) incorporating a semantic understanding for controllable generation; and 3) scaling to city-scale environments for digital twin applications. The framework's efficiency and quality position adversarial radiance fields as a promising candidate for a high-efficiency application in AR/VR, autonomous navigation, and immersive content creation, thus positioning adversarial radiance fields as a cornerstone technology for next-generation visual computing.

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgment

This work was supported by the National Natural Science Foundation of China under Grant 62476063, and the Natural Science Foundation of Guangdong Province 2025A1515011293.

Conflict of interest

All authors declare no conflicts of interest in this paper.

References

1. A. Chatziagapi, G. G. Chrysos, D. Samaras, MI-NeRF: Learning a single NeRF for multiple identities, in *European Conference on Computer Vision*, (2025), 451–469.
2. S. Basak, P. Corcoran, R. McDonnell, M. Schukat, 3D Face-model reconstruction from a single image: A feature aggregation approach using hierarchical transformer with weak supervision, *Neural Networks*, **156** (2022), 108–122. https://doi.org/10.1016/j.neunet.2022.09.019
3. H. Wei, W. Han, X. Dong, J. Shen, Towards high-fidelity 3D portrait generation with rich details by cross-view prior-aware diffusion, in *Proceedings of the AAAI Conference on Artificial Intelligence*, **40** (2026), 10521–10529. https://doi.org/10.1609/aaai.v40i13.38024
4. M. Wang, S. Zhao, X. Dong, J. Shen, High-fidelity and high-efficiency talking portrait synthesis with detail-aware neural radiance fields, *IEEE Trans. Visual Comput. Graphics*, **31** (2024), 6022–6035. https://doi.org/10.1109/TVCG.2024.3488960
5. T. Tsesmelis, L. Palmieri, M. Khoroshiltseva, A. Islam, G. Elkin, O. I. Shahar, et al., Re-assembling the past: the RePAIR dataset and benchmark for real world 2D and 3D puzzle solving, in *Advances in Neural Information Processing Systems*, **37** (2024), 30076–30105.
6. C. Huang, Y. Hou, W. Ye, D. Huang, X. Huang, B. Lin, et al., NeRF-Det++: Incorporating semantic cues and perspective-aware depth supervision for indoor multi-view 3D detection, *IEEE Trans. Image Process.*, **34** (2025), 2575–2587. https://doi.org/10.1109/TIP.2025.3560240

7. P. Ticha, A. Sukop, Patient-reported outcomes in bilateral prophylactic mastectomy with breast reconstruction: A narrative review, *Breast*, **73** (2024), 103602. https://doi.org/10.1016/j.breast.2023.103602
8. I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, et al., Generative adversarial nets, in *Advances in Neural Information Processing Systems*, **27** (2014), 2672–2680.
9. X. Qin, H. Shi, X. Dong, S. Zhang, A new method based on generative adversarial networks for multivariate time series prediction, *Expert Syst.*, **41** (2024), e13700.
10. U. Demirbaga, Advancing anomaly detection in cloud environments with cutting-edge generative AI for expert systems, *Expert Syst.*, **42** (2025), e13722.
11. F. Han, Y. Liang, Q. Ling, H. Han, An improved conditional wasserstein GAN with gradient penalty for gene expression profiling data augmentation based on data segmentation and depth feature constraint, *IEEE Trans. Comput. Biol. Bioinf.*, **22** (2025), 1401–1414. https://doi.org/10.1109/TCBBIO.2025.3560097
12. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. P. Aitken, A. Tejani, et al., Photo-realistic single image super-resolution using a generative adversarial network, in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (2017), 105–114.
13. H. Xu, Q. Li, J. Chen, Highlight removal from a single grayscale image using attentive GAN, *Appl. Artif. Intell.*, **36** (2022), 1988441. https://doi.org/10.1080/08839514.2022.2127598
14. L. Tran, X. Yin, X. Liu, Disentangled representation learning GAN for pose-invariant face recognition, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2017), 1415–1424.
15. Z. Cai, Z. Xiong, H. Xu, P. Wang, W. Li, Y. Pan, Generative adversarial networks: A survey toward private and secure applications, *ACM Comput. Surv.*, **54** (2021), 1–38. https://doi.org/10.1145/3459992
16. Y. Bai, Y. Zhang, M. Ding, B. Ghanem, SOD-MTGAN: Small object detection via multi-task generative adversarial network, in *Proceedings of the European Conference on Computer Vision (ECCV)*, (2018), 206–221.
17. E. Denton, V. Birodkar, Unsupervised learning of disentangled representations from video, in *Proceedings of the 31st International Conference on Neural Information Processing Systems*, (2017), 4417–4426.
18. S. Gu, D. Chen, J. Bao, F. Wen, B. Zhang, D. Chen, et al., Vector quantized diffusion model for text-to-image synthesis, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 10696–10706.
19. J. Lin, Z. Zhou, Q. Qiu, R. Liu, J. Shen, Y. Zhang, Microatoll-GAN: A generative adversarial network for microatoll detection in assisting monitoring Marine Fishery, in *2025 28th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, (2025), 189–194.
20. K. Li, K. Deb, Q. Zhang, S. Kwong, An evolutionary many-objective optimization algorithm based on dominance and decomposition, *IEEE Trans. Evol. Comput.*, **19** (2015), 694–716.
21. X. Feng, Y. He, Y. Wang, C. Wang, Z. Kuang, J. Ding, et al., ZS-SRT: An efficient zero-shot super-resolution training method for neural radiance fields, *Neurocomputing*, **590** (2024), 127714.

22. X. Li, K. Li, G. Zhang, Z. Zhu, P. Wang, Z. Wang, et al., SREGS: Sparse-view gaussian radiance fields with geometric regularization and region exploration, *Neural Networks*, **191** (2025), 107820. https://doi.org/10.1016/j.neunet.2025.107820

23. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, R. Ng, NeRF: Representing scenes as neural radiance fields for view synthesis, *Commun. ACM*, **65** (2021), 99–106.

24. J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, R. Martin-Brualla, P. P. Srinivasan, Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields, in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, (2021), 5835–5844.

25. B. Kerbl, G. Kopanas, T. Leimkuehler, G. Drettakis, 3D gaussian splatting for real-time radiance field rendering, *ACM Trans. Graphics*, **42** (2023), 1–14. https://doi.org/10.1145/3592455

26. S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, B. Recht, A. Kanazawa, Plenoxels: Radiance fields without neural networks, in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, (2022), 5491–5500.

27. Q. Xu, Z. Xu, J. Philip, S. Bi, Z. Shu, K. Sunkavalli, et al., Point-NeRF: Point-based neural radiance fields, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2022), 5438–5448.

28. T. Müller, A. Evans, C. Schied, A. Keller, Instant neural graphics primitives with a multiresolution hash encoding, *ACM Trans. Graphics*, **41** (2022), 102:1–102:15. https://doi.org/10.1145/3528223.3530184

29. J. Liu, H. Cheng, S. Wang, F. Zhao, M. Li, NeRF dynamic scene reconstruction based on motion, semantic information and inpainting, *Neurocomputing*, **630** (2025), 129653. https://doi.org/10.1016/j.neucom.2025.129653

30. R. Mishan, X. Fu, C. Hingu, P. Fajri, Analyzing frequency spectrum and total harmonic distortion for high switching frequency operation of GAN-based filter-less multilevel cascaded H-bridge inverter, *e-Prime Adv. Electr. Eng. Electron. Energy*, **11** (2025), 100906. https://doi.org/10.1016/j.prime.2025.100906

31. M. Hassan, F. Forest, O. Fink, M. Mielle, ThermoNeRF: A multi-modal neural radiance field for joint RGB-thermal novel view synthesis of building facades, *Adv. Eng. Inf.*, **65** (2025), 103345. https://doi.org/10.1016/j.aei.2025.103345

32. I. Petrovska, B. Jutzi, Vision through obstacles—3D geometric reconstruction and evaluation of neural radiance fields (NeRFs), *Remote Sens.*, **16** (2024), 1188. https://doi.org/10.3390/rs16071188

33. Y. Zhou, Z. Zeng, A. Chen, X. Zhou, H. Ni, S. Zhang, et al., Evaluating modern approaches in 3D scene reconstruction: NeRF vs gaussian-based methods, in *2024 6th International Conference on Data-driven Optimization of Complex Systems (DOCS)*, (2024), 926–931.

34. X. Wen, K. Sun, T. Chen, Z. Wang, J. She, Q. Zhao, et al., A NeRF-based technique combined depth-guided filtering and view enhanced module for large-scale scene reconstruction, *Knowledge-Based Syst.*, **316** (2025), 113411. https://doi.org/10.1016/j.knosys.2025.113411

35. K. Schwarz, Y. Liao, M. Niemeyer, A. Geiger, GRAF: Generative radiance fields for 3D-aware image synthesis, in *Advances in Neural Information Processing Systems*, **33** (2020), 20154–20166.

36. E. R. Chan, M. Monteiro, P. Kellnhofer, J. Wu, G. Wetzstein, pi-GAN: Periodic implicit generative adversarial networks for 3D-aware image synthesis, in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2021), 5799–5809.
37. M. Aubry, D. Maturana, A. A. Efros, B. C. Russell, J. Sivic, Seeing 3D chairs: Exemplar part-based 2D-3D alignment using a large dataset of CAD models, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 3762–3769.
38. Z. Liu, P. Luo, X. Wang, X. Tang, Deep learning face attributes in the wild, in *Proceedings of the IEEE International Conference on Computer Vision*, (2015), 3730–3738.
39. X. He, Y. Peng, Fine-grained visual-textual representation learning, *IEEE Trans. Circuits Syst. Video Technol.*, **30** (2019), 520–531. https://doi.org/10.1109/TCSVT.2019.2892802
40. M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, S. Hochreiter, GANs trained by a two time-scale update rule converge to a local nash equilibrium, in *Advances in Neural Information Processing Systems*, (2017), 6626–6637.
41. Y. Sun, J. Fan, MMD Graph kernel: Effective metric learning for graphs via maximum mean discrepancy, in *International Conference on Representation Learning* (eds. B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, Y. Sun), (2024), 22274–22307.
42. M. G. Martini, Measuring objective image and video quality: On the relationship between SSIM and PSNR for DCT-based compressed images, *IEEE Trans. Instrum. Meas.*, **74** (2025), 1–13. https://doi.org/10.1109/TIM.2025.3529045
43. A. Gupta, J. Cao, C. Wang, J. Hu, S. Tulyakov, J. Ren, et al., LightSpeed: Light and fast neural light fields on mobile devices, in *Advances in Neural Information Processing Systems*, **36** (2023), 31021–31037.
44. L. Mescheder, A. Geiger, S. Nowozin, Which training methods for GANs do actually converge, in *International Conference on Machine Learning*, PMLR, (2018), 3481–3490.
45. P. Henzler, N. J. Mitra, T. Ritschel, Escaping plato's cave: 3D shape drom adversarial rendering, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 9984–9993.
46. T. Nguyen-Phuoc, C. Li, L. Theis, C. Richardt, Y. L. Yang, HoloGAN: Unsupervised learning of 3D representations from natural images, in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, (2019), 7588–7597.
47. W. Wang, H. L. Liu, K. C. Tan, A surrogate-assisted differential evolution algorithm for high-dimensional expensive optimization problems, *IEEE Trans. Cybern.*, **53** (2023), 2685–2697. https://doi.org/10.1109/TSMC.2023.3281822

AIMS Press