Proceedings of the 4th International Conference on Machine Learning and Automation
DOI: 10.54254/2755-2721/2026.37870

Research on Neural Radiance Field Technology: Principles, Implementation, and Evolution from NeRF to NeRF²

Yanpu Cai

School of Mechanical & Electrical Engineering, Heilongjiang University, Harbin, China
1584081386@qq.com

Abstract. Neural radiance fields (NeRFs) represent 3D scenes as continuous implicit functions and use differentiable volume rendering to convert 2D observations into 3D reconstruction and novel-view synthesis. The technology has expanded from static scenes to dynamic rendering, real-time neural graphics, and 3D generation, and into wireless communications as NeRF² for radio-frequency propagation modeling. Under a unified view of optical implicit fields and electromagnetic radiance fields, this paper reviews the frameworks, core methods, representative results, and limitations of visual NeRF and radio-frequency NeRF², and proposes a task-matched evaluation framework. It covers positional encoding, sampling, volume rendering, dynamic modeling, and 3D AIGC in the visual domain, and complex-valued signals, attenuation and reradiation, spatial-spectrum prediction, and few-shot data augmentation in the radio-frequency domain. Rather than rank the families by task-independent quality, the paper compares them through output objects, physical constraints, data requirements, and application objectives, supporting cross-modal implicit-field research in visual reconstruction, wireless-channel prediction, and 6G digital twins.

Keywords: NeRF, NeRF², volume rendering, wireless channel modeling, digital twins

1. Introduction

NeRF's main contribution is to represent a 3D scene as a continuous function that maps spatial position and viewing direction to density and view-dependent color, rather than as fixed grids or meshes [1]. Differentiable volume rendering and multi-view image supervision enable novel-view synthesis without explicit 3D annotations. By optimizing geometry, appearance, and rendering together, NeRF became foundational in 3D vision, robotics, virtual reality, and digital content generation [2].

Wireless systems must model propagation paths, attenuation, reflection, diffraction, and multipath in complex environments. Ray tracing is physically interpretable but can be expensive for large, dynamic, or multi-configuration scenes [3]. 6G digital twins also require joint modeling of environment, channel, and network state for coverage prediction, channel estimation, and deployment optimization [4], motivating transfer of neural implicit fields from visible light to radio frequency.

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.37870

Research has followed two paths: visual methods improve representation capacity, sampling efficiency, dynamic-scene modeling, and generative control; wireless methods introduce continuous fields with complex-valued signals and propagation priors because visual NeRF cannot directly represent electromagnetic propagation. Both use continuous field functions, neural fitting, and differentiable rendering, but their outputs differ: visual NeRF predicts images, depth, or geometry, whereas radio-frequency NeRF² predicts signal strength, spatial spectra, arrival angles, and channel states.

Based on a systematic review, this paper reorganizes the trajectory of visual NeRF, summarizes radio-frequency modeling pathways, and proposes an integrated evaluation framework. Visual development is structured as static representation, efficiency improvement, dynamic modeling, and 3D AIGC; NeRF², NeWRF, R-NeRF, and VoxelRF are reviewed in parallel. The framework covers visual quality, radio-frequency prediction, computational and data efficiency, physical consistency, and scene generalization. Literature checking, numerical reporting, and figure and table preparation follow a traceable scientific-writing workflow [5].

The primary innovations and contributions of this work are as follows:

• Visual NeRF research is reorganized as static representation, efficiency improvement, dynamic modeling, and 3D AIGC, strengthening the review's structure.

• This review further contributes a systematic survey of representative radio-frequency methods (including NeRF², NeWRF, R-NeRF, and VoxelRF), distilling their technical characteristics and development trends.

• A multidimensional framework covers visual quality, radio-frequency prediction, computational and data efficiency, physical consistency, and scene generalization for horizontal comparison and future work.

2. Visual neural radiance fields (NeRF)

2.1. Overall framework

Standard NeRF represents a static scene as a five-dimensional continuous function. Given a 3D position and 2D viewing direction, it predicts density and view-dependent color [1]. Equation (1) defines the field, Eq. (2) a camera ray, and Eqs. (3)-(4) the volume-rendered pixel color from samples along that ray.

$$F_{\Theta}(x, d) = (\sigma, c) \tag{1}$$

where: F_{Θ} is the neural radiance field function; Θ are the learnable weights and biases of the network; $x = (x, y, z)$ is the three-dimensional position, where x, y, z are the Cartesian coordinate components; d is the unit viewing direction; σ is the volume density, measured as the reciprocal of the chosen length unit; c is the view-dependent RGB color vector.

$$r(t) = o + td \tag{2}$$

where: $r(t)$ is the position on the ray at distance parameter t; o is the camera center; d is the unit direction vector; t is the distance measured from the camera center along this direction. Position and distance use the same length unit.

$$C(r) = \int_{t_{\mathrm{n}}}^{t_{\mathrm{f}}} T(t)\sigma(r(t))c(r(t), d)\mathrm{d}t \tag{3}$$

DOI: 10.54254/2755-2721/2026.37870

where: $C(r)$ is the rendered color of ray r; t_n, t_f are the near and far bounds, respectively; t is the distance parameter; $T(t)$ is the accumulated transmittance; $\sigma(r(t))$ is the volume density at position $r(t)$; $c(r(t), d)$ is the color at that position along unit direction d; dt is the differential distance.

$$T(t) = \exp(-\int_{t_n}^{t} \sigma(r(s))\mathrm{d}s) \tag{4}$$

where: $T(t)$ is the dimensionless accumulated transmittance from near bound t_n to distance t; s is the integration variable; $r(s)$ is the ray position; $\sigma(r(s))$ is the density at that position; exp is the natural exponential function; ds is the differential distance.

Engineering-wise, visual NeRF has four modules: input receives multi-view images, camera intrinsics and extrinsics, and sampled rays; feature encoding maps coordinates and directions into a learnable high-dimensional space; the implicit network predicts density and color; rendering integrates each ray to form training or novel-view images. Rendering errors backpropagate to close the observation-encoding-field-rendering-optimization loop.

2.2. Core methods

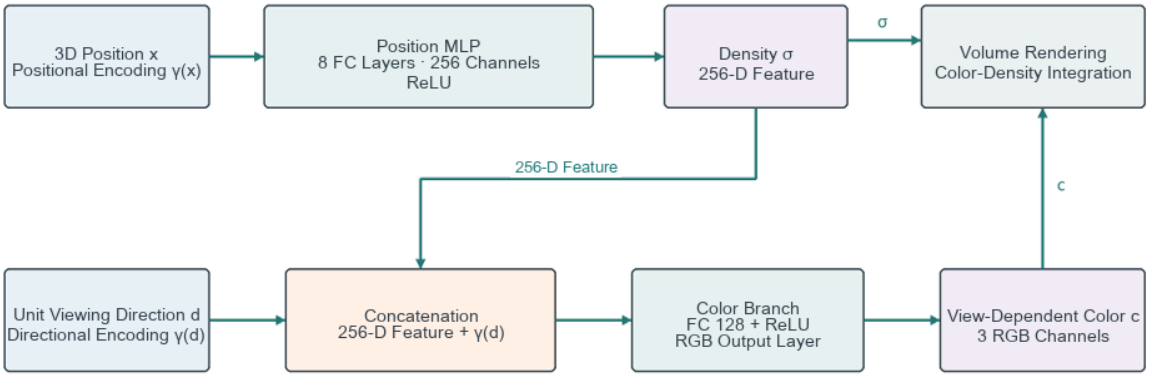


Figure 1. Schematic of the visual NeRF MLP architecture

2.2.1. Continuous implicit representation and volume rendering

NeRF uses an MLP rather than voxels, point clouds, or meshes. The backbone predicts density and features from position, then combines them with encoded viewing direction to predict color. Density is ideally direction-independent, while color varies with direction to model specular and other non-Lambertian effects. Volume rendering both maps the field to pixels and transfers pixel errors to sampled points.

2.2.2. Positional encoding and efficiency improvements

Standard MLPs have spectral bias and struggle with high-frequency texture and geometry from low-dimensional coordinates. Fourier features use sine and cosine mappings to expose higher frequencies [6]. Multiresolution hash encoding stores learnable features in multilevel tables and uses lookup/interpolation to reduce query cost, greatly shortening training in Instant-NGP [7]. Fourier encoding offers interpretable frequency control; hash encoding prioritizes speed and memory efficiency, so they serve different rapid-training and analysis needs.

DOI: 10.54254/2755-2721/2026.37870

2.2.3. Sampling and anti-aliasing

Sampling controls rendering cost. Uniform sampling wastes queries in empty space; hierarchical sampling uses a coarse network to refine high-contribution regions, while occupancy grids and empty-space skipping remove invalid samples. Mip-NeRF represents a ray as a conical frustum across continuous scales to reduce aliasing and stabilize multiscale detail. It reports 17% lower average error than NeRF, 60% lower error on challenging multiscale data, and much lower rendering cost at comparable accuracy [8].

2.2.4. Dynamic scenes and 3D generation

Static NeRF assumes positions do not change across observations. Dynamic models add time or deformation: D-NeRF maps observation space to canonical space to reconstruct rigid and non-rigid motion [9], while Gafni et al. combine a dynamic field with controllable facial deformation for monocular-video 4D avatars [10]. Generalization remains limited by motion type, camera trajectory, and training coverage.

DreamFusion uses a pretrained 2D diffusion prior and score distillation to optimize a NeRF for text-to-3D generation without large-scale 3D training data [11]. This shifts NeRF from reconstruction to conditional creation, but multiview consistency, geometric completeness, and controllability remain open.

2.3. Representative results and limitations

Under the original settings, NeRF achieved high novel-view quality on synthetic and real forward-facing scenes: PSNR 40.15, 31.01, and 26.50 dB; SSIM 0.991, 0.947, and 0.811; and LPIPS 0.023, 0.081, and 0.250 on three datasets [1]. These values show a balance of pixel error, structural fidelity, and perceptual similarity. Mip-NeRF confirms the value of multiscale representations for anti-aliasing and cross-resolution generalization [8].

Visual NeRF has three main limitations. Deep MLP queries make training and inference expensive, especially as scenes grow. Static-scene assumptions and reliance on accurate camera poses make occlusion, non-rigid motion, and sparse views difficult. Generative methods add creative control but remain dependent on the prior and optimization strategy for 3D consistency, detail, and controllability. The focus has therefore shifted from feasibility to fast, stable, and controllable reconstruction or generation.

3. Neural radio-frequency radiance fields (NeRF²) for wireless communications

3.1. Overall framework

Visual NeRF predicts optical color and density, not complex wireless amplitude and phase, multipath interference, or transmitter dependence. NeRF² models radio-frequency propagation as a continuous spatial field, learns unmeasured received results from sparse measurements, and incorporates ray tracing as a physical prior [12]. Inputs may include transmitter and receiver positions, propagation direction, and array configuration; outputs include received signals, RSSI, spatial spectra, or channel states.

NeRF² queries a field, accumulates rays, models the receiver, and trains against measurements. A voxel position enters an attenuation network; its features are combined with transmitter position and measurement direction in a radiation network that predicts complex attenuation and reradiation.

DOI: 10.54254/2755-2721/2026.37870

Electromagnetic ray tracing accumulates these terms, the antenna model integrates directions or forms a spatial spectrum, and differentiable propagation jointly optimizes the networks using complex signals or spectra [12]. This learned field represents complex multipath interactions.

3.2. Core methods

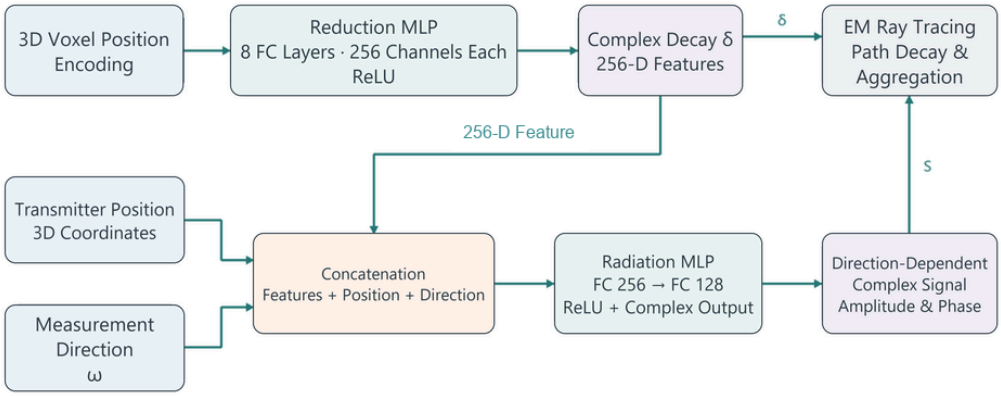


Figure 2. Schematic of the NeRF² dual-MLP architecture

3.2.1. Attenuation and reradiation fields

Following the Huygens-Fresnel view, NeRF² treats propagation points as local secondary-wave sources. Each point has attenuation and reradiation attributes: attenuation describes amplitude change and phase shift through the point; reradiation describes the complex response that sends a received signal toward other directions [12]. Along a path, attenuation coefficients multiply; training often accumulates them in the log domain to reduce underflow and vanishing gradients.

$$\delta(P_x) = \Delta a(P_x)\mathrm{e}^{\mathrm{j}\Delta\theta(P_x)} \tag{5}$$

where: $\delta(P_x)$ is the complex attenuation factor of the voxel at position P_x; x is the voxel index; $\Delta a(P_x)$ is the nonnegative amplitude transmission factor; $\Delta\theta(P_x)$ is the phase rotation in radians; e is the natural exponential base; $\mathrm{j}^2 = -1$ defines the imaginary unit. Here, δ has a different meaning from the sampling interval in visual NeRF; Δa is a multiplicative factor rather than an amplitude difference.

$$F_\Theta : (P_{\mathrm{TX}}, P_x, \boldsymbol{\omega}) \mapsto (\delta(P_x), S(P_x, \boldsymbol{\omega})) \tag{6}$$

where: F_Θ is the radio-frequency radiance field; Θ are the learnable parameters of the two MLPs; P_{TX} is the transmitter position; P_x is the position of voxel x; $\boldsymbol{\omega}$ is the unit vector of the measurement direction, parameterized by azimuth α and elevation β; δ is the position-dependent complex attenuation; S is the direction-dependent complex reradiated signal; its conditional dependence on transmitter position is omitted from the output notation; $\mapsto$ denotes the mapping.

DOI: 10.54254/2755-2721/2026.37870

3.2.2. Received-signal and spatial-spectrum modeling

An omnidirectional receiver accumulates complex signals from all directions to predict received power or RSSI for coverage and fingerprint localization. An array receiver must preserve inter-element phase differences to predict directional power and spatial spectra. Spectra retain signal strength and multipath direction, enabling AoA estimation and MIMO channel prediction [12].

$$R(\boldsymbol{\omega}) = \int_0^D H_{P(r,\boldsymbol{\omega})\to P_{\mathrm{RX}}} S(P(r,\boldsymbol{\omega}), -\boldsymbol{\omega})\mathrm{d}r \tag{7}$$

where: $R(\boldsymbol{\omega})$ is the complex received signal arriving from unit direction $\boldsymbol{\omega}$; D is the scene truncation distance of the ray; r is the radial distance; $P(r,\boldsymbol{\omega})$ is the ray position; P_{RX} is the receiver position; H is the accumulated complex attenuation from that position to the receiver; S is the reradiation contribution in the continuous model; $-\boldsymbol{\omega}$ points toward the receiver, opposite to the back-traced ray; $\mathrm{d}r$ is the path differential. In numerical quadrature, the signal contribution should be scaled consistently with the sampling step.

$$R = \int_{\Omega} \sqrt{G_{\mathrm{RX}}(\boldsymbol{\omega})} R(\boldsymbol{\omega})\mathrm{d}\Omega \tag{8}$$

where: R is the total complex received signal at a single antenna; Ω is the directional domain covered by the antenna; $G_{\mathrm{RX}}(\boldsymbol{\omega})$ is the linear power gain of the receiving antenna along unit direction $\boldsymbol{\omega}$; it must not be replaced directly by a dB value, and its square root is used to weight the complex amplitude; $R(\boldsymbol{\omega})$ is the directional received signal; $\mathrm{d}\Omega$ is the differential solid angle. Compared with the original Eq. (9), the integration measure is made explicit here; when an azimuth-elevation grid is used, the quadrature weights should include the corresponding solid-angle factor.

$$\Psi(\boldsymbol{\omega}) = \frac{1}{K^2-1} \left| \sum_{i=1}^{K} \sum_{j=1}^{K} w_{i,j}(\boldsymbol{\omega}) \mathrm{e}^{\mathrm{j}\Delta\tilde{\theta}_{i,j}} \right| \tag{9}$$

where: $\Psi(\boldsymbol{\omega})$ is the relative spatial spectrum defined in the original work; $\boldsymbol{\omega}$ is the scanned unit direction; $K > 1$ is the number of array elements along each dimension, giving K^2 elements in total; i, j are the row and column indices; $w_{i,j}$ is the complex beamforming weight; $\Delta\tilde{\theta}_{i,j}$ is the measured phase difference relative to the reference element, in radians; j is the imaginary unit, distinguished from the italic index j; e is the natural exponential base; $|\cdot|$ is the complex modulus. The modulus form is retained as in the original Eq. (11); no square should be added.

3.2.3. From NeRF² to NeWRF, R-NeRF, and VoxelRF

NeWRF targets sparse wireless measurements, using arrival-direction-guided ray casting to constrain effective paths and predict channels at unvisited positions [13]. R-NeRF models RIS propagation in two stages, transmitter to RIS then RIS to receiver, to handle RIS position and complex reflection [14]. VoxelRF makes part of the field explicit as a voxel grid, uses trilinear interpolation and two shallow MLPs, and skips empty space to reduce queries [15]. They target sparse measurement, reconfigurable surfaces, and efficiency, respectively, rather than forming a linear replacement chain.

DOI: 10.54254/2755-2721/2026.37870

3.2.4. Physics-constrained data augmentation

NeRF² also proposes turbo-learning: train a field from few real measurements, use it to synthesize propagation-consistent samples, then mix real and synthetic data for a higher-level communications task. Unlike unconstrained image augmentation, the synthetic data preserve paths, amplitude relationships, and spatial structure, reducing measurement cost and label scarcity [12].

3.3. Representative results and limitations

NeRF² reconstructs a continuous radio-frequency field from few measurements and reports usable results for spatial spectra, indoor localization, and 5G MIMO channel prediction. A microbenchmark reached about 82% median spectrum similarity. With 10% real and 90% synthetic data, one AoA experiment improved accuracy by 47.9%, with about 49.5% average improvement across 14 scenes [12]. Channel prediction improved SNR and SINR by 5.97 dB and 4.32 dB over comparison methods; Argos 5G MIMO reported 20.87 dB median predicted SNR. These are task-, dataset-, and setting-specific.

Evidence boundary. The 82%, 47.9%, 49.5%, 5.97 dB, 4.32 dB, and 20.87 dB values refer respectively to spatial-spectrum similarity, few-shot AoA, 14-scene AoA, channel prediction, and Argos 5G MIMO; none is a universal wireless performance level.

Later methods emphasize task adaptation. NeWRF combines sparse channels with arrival direction for unmeasured positions; R-NeRF evaluates different RIS positions with simulated and measured data and reports lower MAE than NeRF² at low training-data ratios [14]; VoxelRF uses voxelization and empty-space skipping to reduce training and inference cost for resource-constrained devices [15]. RF implicit-field research has thus moved from modeling a field to adapting it to propagation structure, hardware, and computation.

Table 1. Representative methods: improvement targets, core mechanisms, and evidence boundaries

Method	Application scenario	Key mechanism	Advantages	Limitations
NeRF [1]	Static visual-scene representation	Position and direction inputs; volume rendering	PSNR/SSIM/LPIPS for synthetic and real forward-facing scene outputs	High training and rendering cost; depends on camera poses
Mip-NeRF [8]	Multiscale aliasing	Conical-frustum integration and scale-aware encoding	Average error reduced by 17%; challenging data reduced by 60%	Conclusion bound to the paper's experimental settings
Instant-NGP [7]	Training and query efficiency	Multiresolution hash encoding + small MLP	Substantially shortens training time	Speed and memory usage depend on hardware and scene
D-NeRF [9]	Dynamic scenes	Time conditioning and deformation to canonical space	Supports spatiotemporal field modeling	Generalization limited by motion type and view coverage

DOI: 10.54254/2755-2721/2026.37870

Table 1. (continued)

NeRF² [12]	Radio-frequency propagation field	Complex-valued attenuation/radiation dual MLPs + ray tracing	Clear task results for spatial spectra, AoA, SNR, and SINR	Sensitive to frequency, materials, and configuration
NeWRF [13]	Sparse wireless measurements	Arrival-direction-guided ray casting	Channel prediction at unmeasured positions	Cross-scene adaptation still requires calibration
R-NeRF [14]	RIS propagation environments	Two-stage Tx-RIS and RIS-Rx modeling	Signal-field prediction under different RIS positions	Changing reflection configurations increase input complexity
VoxelRF [15]	Wireless-field query efficiency	Voxelization + shallow MLPs + empty-space skipping	Reduces training and inference burden	Explicit voxels create a resolution-storage trade-off

RF fields are sensitive to frequency, polarization, materials, and transmitter configuration, so visual geometry alone cannot guarantee cross-band or cross-scene generalization. Sparse measurements reduce acquisition cost but remain dependent on locations, arrays, and environments. Physical priors must be calibrated with data; overly strong assumptions can lose complex scattering. Dynamic environments with multiple transmitters or RISs also increase input dimension, data demand, and real-time-update difficulty.

Table 2. Task boundaries and evaluation emphases of visual NeRF and radio-frequency NeRF²

Dimension	Input	Output	Physical process	Main applications
Visual NeRF [1]	Multi-view images and camera poses	Color, density, depth, or geometry	Optical volume rendering	Novel-view synthesis, reconstruction, and 3D AIGC
Radio-frequency NeRF² [12]	Sparse signal measurements, Tx/Rx, or array configuration	RSSI, spatial spectrum, AoA, or channel	Complex-valued propagation, attenuation, reflection, and multipath	Channel prediction, coverage, RIS, and 6G digital twins

4. Evaluation framework and application scenarios

Visual NeRF and radio-frequency NeRF² differ in outputs, objectives, and data, so evaluation should use domain-specific metrics within shared engineering dimensions. PSNR, SSIM, and LPIPS are appropriate for images, or auxiliary measures for spectra treated as 2D images, but cannot replace radio-frequency propagation error. Conversely, RSSI, AoA, SNR, and SINR cannot assess visual texture realism. Chapter 4 therefore avoids a single cross-domain ranking and pairs domain metrics with common dimensions [4].

DOI: 10.54254/2755-2721/2026.37870

4.1. Visual-quality evaluation

PSNR (Peak Signal-to-Noise Ratio) is based on pixel MSE and is usually reported in dB; higher is better. It compares novel-view synthesis within the same dataset and dynamic range.

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{I_{max}^2}{\mathrm{MSE}}\right) \tag{10}$$

where: PSNR is the peak signal-to-noise ratio, in dB; I_{max} is the maximum permitted image-pixel value, usually set to one for normalized images; MSE is the mean squared error between the predicted and reference images over the same pixel range; $\log_{10}$ is the base-10 logarithm. When the error is zero, PSNR tends to positive infinity; pixel ranges and error calculations must be consistent when results are compared.

SSIM (Structural Similarity Index Measure) compares luminance, contrast, and structure between rendered and reference images, usually on a 0-1 scale. It complements PSNR when pixel error misses structural fidelity.

LPIPS (Learned Perceptual Image Patch Similarity) measures deep-feature distance from a pretrained network; lower values indicate closer perception and help reveal texture misalignment, local structural changes, and artifacts. Report all three; no single metric replaces visual judgment.

4.2. Radio-frequency modeling evaluation

Spatial-spectrum and RSSI error. Use MSE, MAE, or normalized error for predicted versus reference power distributions; RSSI, reported in dB, measures signal-strength bias in coverage and fingerprint localization.

AoA error is the angular distance between predicted and reference arrival directions. For spatial-spectrum outputs, also inspect peak direction, main-lobe width, and multimodal structure.

SNR and SINR. In channel prediction, SNR reflects similarity between predicted and true channels, while SINR includes multiuser interference. Report the dataset, frequency band, antenna scale, prediction direction, and statistical procedure so task-specific values are not presented as universal.

Table 3. Evaluation metrics and applicable tasks

Domain	Metric	Meaning	Direction	Typical task
Visual	PSNR	Logarithmic expression of pixel mean squared error	Higher is better	Novel-view synthesis
	SSIM	Luminance, contrast, and structural consistency	Higher is better	Structural fidelity
	LPIPS	Deep-feature perceptual distance	Lower is better	Texture and perceptual quality

DOI: 10.54254/2755-2721/2026.37870

Table 3. (continued)

Radio-frequency	Spatial-spectrum / RSSI MAE or MSE	Power-distribution or signal-strength error	Lower is better	Coverage, localization, and array reception
	AoA angular distance	Difference between predicted and reference arrival directions	Lower is better	Angle-of-arrival estimation
	SNR / SINR	Channel-prediction similarity and performance under interference	Higher is better	Channel prediction, MU-MIMO
General	Training/inference time and memory	Deployment cost and resource usage	Must be interpreted with the task	Engineering usability
	Cross-scene error	Performance at unseen positions, materials, or configurations	Lower is better	Generalization capability

4.3. General evaluation dimensions

Computational efficiency. Report training time, single-render or inference time, parameter count, and GPU or system memory. Hash encoding, occupancy grids, and voxelization improve visual-NeRF query efficiency; NeRF² must also report ray-tracing and complex-propagation cost.

Data efficiency. For visual models report training views, coverage, and camera-pose quality; for radio-frequency models report real measurement points, array configuration, frequency, and transmitter positions. For synthetic augmentation, report the real-data proportion and synthetic-data source.

Physical consistency. Visual models should preserve geometry, occlusion, and cross-view consistency. Radio-frequency models should also satisfy distance attenuation, occlusion, reflection, diffraction, phase continuity, and plausible energy behavior, distinguishing them from black-box regression.

Scene generalization. Test unseen camera trajectories, spatial positions, materials, RIS positions, or transmitter configurations, stating which factors change. Separate results outside the training distribution; random within-scene splits do not establish cross-scene capability.

4.4. Representative application scenarios

In 3D reconstruction and AIGC, visual NeRF supports novel-view synthesis, free-viewpoint video, digitization of cultural relics and buildings, product visualization, and text-driven 3D generation.

DOI: 10.54254/2755-2721/2026.37870

Surveys classify 3D generation as 3D-native, 2D-prior-driven, or hybrid, placing NeRF both as a scene representation and as a target of generative optimization [16].

In 6G digital twins, visual NeRF supplies urban or indoor geometry and appearance, while NeRF² supplies spatial electromagnetic properties. Together they can support coverage, beam configuration, base-station placement, interference analysis, and network operations. Deployment still requires unified coordinates, time synchronization, material parameters, and multiband calibration. 6G digital-twin research coordinates environmental twins, ray-tracing engines, and channel twins [4], with neural fields as a data-driven complement.

For wireless deployment and RIS optimization, NeWRF, R-NeRF, and VoxelRF emphasize sparse measurement, reconfigurable propagation, and fast queries. Their shared goal is to reduce measurement or repeated-simulation cost. Deployment decisions must still rely on measured calibration, error bounds, failure cases, and system-level communications validation; offline visualization is insufficient.

5. Conclusions and outlook

This paper reviewed the evolution from visual to radio-frequency implicit fields. Visual NeRF uses a continuous five-dimensional field and differentiable rendering, then develops through Fourier features, hash encoding, multiscale anti-aliasing, dynamic scenes, and text-driven 3D generation. NeRF² adds complex amplitude and phase, ray-propagation priors, attenuation, and reradiation to predict spatial spectra, RSSI, AoA, and wireless channels. NeWRF, R-NeRF, and VoxelRF show the need for task-specific designs for sparse measurements, RISs, and efficiency.

The families share a mathematical skeleton but should not be collapsed into one leaderboard. Visual models target appearance and geometry; radio-frequency models target propagation signals and physical consistency. Their metrics likewise differ. Future work should (1) jointly represent visual geometry, semantics, and radio-frequency propagation in a calibratable multimodal field; (2) support incremental updates and few-shot adaptation for dynamic environments, multiple transmitters, and reconfigurable surfaces; and (3) combine voxels, tensor decomposition, hash encoding, and compression to lower radio-frequency-field training and inference cost.

NeRF's value is not only a new rendering method but a demonstration that continuous implicit representation plus a differentiable physical process can cross modalities. Real-system adoption depends on jointly validating data efficiency, physical constraints, interpretable errors, and deployment cost.

References

[1] Mildenhall, B., Srinivasan, P. P., Tancik, M., et al. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis. In *Computer Vision - ECCV 2020*, 405-421. https://doi.org/10.1007/978-3-030-58452-8_24

[2] Gao, K., Gao, Y., He, H., Lu, D., Xu, L., & Li, J. (2026). Neural radiance fields in 3D vision: A comprehensive review. *Computational Visual Media, 12*(4), 851-905. https://doi.org/10.26599/cvm.2026.9450535

[3] Yun, Z., & Iskander, M. F. (2015). Ray tracing for radio propagation modeling: Principles and applications. *IEEE Access, 3*, 1089-1100. https://doi.org/10.1109/ACCESS.2015.2453991

[4] Yu, L., Miao, Y., Zhang, J., et al. (2025). Road to 6G digital twin networks: Multi-task adaptive ray-tracing as a key enabler. *arXiv preprint arXiv:2502.14290*. https://doi.org/10.48550/arXiv.2502.14290

[5] Kassis, T., Agarwal, V., He, Y., Patel, D., & Brueckner, A. M. (2026). Scientific agent skills: A library of procedural knowledge for research agents. *arXiv preprint arXiv:2609.00065*. https://doi.org/10.48550/arXiv.2609.00065

[6] Tancik, M., Srinivasan, P., Mildenhall, B., et al. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in Neural Information Processing Systems, 33*, 7537-7547.

DOI: 10.54254/2755-2721/2026.37870

[7] Müller, T., Evans, A., Schied, C., et al. (2022). Instant neural graphics primitives with a multiresolution hash encoding. *ACM Transactions on Graphics, 41*(4), 1-15. https://doi.org/10.1145/3528223.3530127
[8] Barron, J. T., Mildenhall, B., Tancik, M., et al. (2021). Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 5835-5844. https://doi.org/10.1109/ICCV48922.2021.00580
[9] Pumarola, A., Corona, E., Pons-Moll, G., et al. (2021). D-NeRF: Neural radiance fields for dynamic scenes. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10313-10322. https://doi.org/10.1109/CVPR46437.2021.01018
[10] Gafni, G., Thies, J., Zollhofer, M., et al. (2021). Dynamic neural radiance fields for monocular 4D facial avatar reconstruction. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8645-8654. https://doi.org/10.1109/CVPR46437.2021.00854
[11] Poole, B., Jain, A., Barron, J. T., et al. (2022). DreamFusion: Text-to-3D using 2D diffusion. *arXiv preprint arXiv:2209.14988*. https://doi.org/10.48550/arXiv.2209.14988
[12] Zhao, X., An, Z., Pan, Q., & Yang, L. (2023). NeRF²: Neural radio-frequency radiance fields. *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, 1-15. https://doi.org/10.1145/3570361.3592527
[13] Lu, H., Vattheuer, C., Mirzasoleiman, B., & Abari, O. (2024). NeWRF: A deep learning framework for wireless radiation field reconstruction and channel prediction. *arXiv preprint arXiv:2403.03241*. https://doi.org/10.48550/arXiv.2403.03241
[14] Yang, H., Jin, Z., Wu, C., et al. (2024). R-NeRF: Neural radiance fields for modeling RIS-enabled wireless environments. *IEEE Global Communications Conference*, 3859-3864. https://doi.org/10.1109/GLOBECOM52923.2024.10901706
[15] Zeng, Z., Sun, S., Tao, M., et al. (2026). VoxelRF: Voxelized radiance field for fast wireless channel modeling. *IEEE Communications Letters, 30*, 617-621. https://doi.org/10.1109/LCOMM.2025.3644410
[16] Liu, J., Huang, X., Huang, T., et al. (2024). A comprehensive survey on 3D content generation. *arXiv preprint arXiv:2402.01166*. https://doi.org/10.48550/arXiv.2402.01166