Automation and Machine Learning (2026)
Clausius Scientific Press, Canada

DOI: 10.23977/autml.2026.070107
ISSN 2516-5003 Vol. 7 Num. 1

Lightweight Improvement of NeRF Algorithm for Industrial Digital Twin Scenarios in Automobile Factories

Xianrui Song[1,*], Xinyue Yan[1], Ting Zhang[2]

[1]University of Sanya, Haikou, Hainan, 572022, China
[2]Geely Automobile Research Institute, Ningbo, Zhejiang, 315300, China
[*]Corresponding author

Keywords: Automobile Factory; Digital Twin; NeRF Algorithm; Lightweight; Scene Reconstruction

Abstract: Industrial digital twin technology provides core support for the intelligent operation and maintenance as well as flexible production of automobile factories. The Neural Radiance Fields (NeRF) algorithm, with its high-precision scene reconstruction capability, has become a key technology for scene modeling in automobile factory digital twins. However, the traditional NeRF algorithm suffers from high model complexity, slow inference speed, and high hardware deployment costs, making it difficult to adapt to scenarios with high real-time requirements in automobile factories, such as welding workshops and assembly lines. To address this pain point, this paper proposes a lightweight NeRF improvement algorithm (Factory-LiteNeRF) from three dimensions: scene partition modeling, network structure pruning, and feature encoding optimization, combined with the characteristics of typical digital twin scenarios in automobile factories. Experiments are conducted on the assembly line scene of an automobile factory to compare the model volume, inference speed, and reconstruction accuracy between the traditional NeRF and the improved algorithm. The results show that the improved algorithm compresses the model volume by 72.3% and increases the inference speed by 2.8 times, while ensuring that the reconstruction accuracy loss does not exceed 3%, which can meet the real-time modeling and operation and maintenance needs of automobile factory digital twins.

1. Introduction

With the rapid iteration of intelligent manufacturing technology, the automobile manufacturing industry is accelerating its transformation towards digitalization and intelligence. As the core technology connecting physical factories and virtual factories, industrial digital twins have been widely applied in scenarios such as production line planning, equipment operation and maintenance, and fault diagnosis in automobile factories[1]. The core demand of automobile factory digital twins is to realize high-fidelity and real-time virtual mapping of physical scenes, and the accuracy and efficiency of scene modeling directly determine the practicality of the digital twin system.

The Neural Radiance Fields (NeRF) algorithm, proposed by Mildenhall et al. in 2020, models the radiance field information of scenes through neural networks, achieving photo-realistic scene reconstruction effects and providing a new solution for digital twin scene modeling[2]. Compared with traditional 3D modeling technologies, NeRF does not require complex point cloud stitching and mesh optimization processes, and can complete high-fidelity scene reconstruction only through multi-view images, which has significant advantages in scenarios such as equipment detail reconstruction and production line dynamic mapping in automobile factories[3].

However, the traditional NeRF algorithm faces serious deployment bottlenecks in automobile factory digital twin scenarios: on the one hand, automobile factory scenes contain a large amount of detailed information such as equipment, tooling, and materials. The traditional NeRF needs to achieve detailed modeling through deep neural networks and dense sampling, resulting in redundant model parameters and huge computational overhead; on the other hand, automobile factory digital twin systems often need to be deployed on industrial edge devices (such as edge gateways and embedded controllers), which have limited computing and storage resources and are difficult to bear the real-time inference needs of the traditional NeRF algorithm[4]. In addition, some scenes in automobile factories (such as assembly lines and AGV logistics channels) need to realize dynamic real-time updates, and the slow inference speed of traditional NeRF further limits its engineering application[5].

To solve the above problems, scholars at home and abroad have carried out research on lightweight improvement of NeRF, which is mainly divided into two categories: one is to reduce the number of parameters through model compression technologies (such as pruning and quantization), and the other is to reduce computational overhead through sampling strategy optimization[6]. However, most of the existing lightweight methods are designed for general scenarios and do not fully consider the particularity of automobile factory digital twin scenarios—there are a large number of static areas (such as floors, walls, and fixed equipment) and dynamic areas (such as mobile robots, conveyor lines, and materials) in the scene, and the modeling accuracy requirements of different areas are different[7]. The adoption of a unified lightweight strategy is likely to lead to insufficient reconstruction accuracy in dynamic areas or waste of computing resources in static areas.

Based on this, combined with the characteristics of typical digital twin scenarios in automobile factories, this paper proposes a lightweight NeRF improvement algorithm for automobile factories, focusing on solving the problems of large model volume, slow inference speed, and poor adaptability of traditional NeRF, realizing the balance between high fidelity and high real-time performance, and providing an efficient and economical solution for scene modeling of automobile factory digital twins[8].

2. Related Technical Basis

2.1 Analysis of Digital Twin Scenarios in Automobile Factories

This paper selects two of the most representative digital twin scenarios in automobile factories as the research objects, namely the assembly line scenario and the welding workshop scenario. The core characteristics and modeling requirements of the two scenarios are as follows:

(1) Assembly line scenario: It includes conveyor lines, assembly robots, automobile bodies, component materials and other elements. Among them, the conveyor lines and automobile bodies are dynamic areas, which need to update their positions and attitudes in real time, and the modeling accuracy requirement is high (error $\leq$ 2 mm); the floors, walls, and fixed brackets are static areas, with relatively low modeling accuracy requirements (error $\leq$ 5 mm), and the whole scene needs to

meet the real-time inference (frame rate $\geq$ 15 fps) requirement[9].

(2) Welding workshop scenario: It includes welding robots, welding torches, tooling fixtures and other equipment. The scene has complex lighting conditions (including welding sparks and shadows). The equipment details (such as welding torch nozzles and fixture interfaces) need high-precision reconstruction, while the empty areas of the workshop can appropriately reduce the modeling accuracy, and at the same time, it needs to adapt to the limited computing resources of edge devices[10].

The common needs of the two scenarios are: high-fidelity reconstruction of core equipment details, low-overhead real-time inference, and adaptation to industrial edge device deployment; the individual needs are reflected in the differences in the proportion of dynamic areas, lighting conditions, and accuracy requirements, which provide a basis for the design of the lightweight improvement strategy in this paper[11].

2.2 Principle of Traditional NeRF Algorithm

The core idea of the traditional NeRF algorithm is to model the radiance field (radiance intensity and volume density) of the scene through a Multi-Layer Perceptron (MLP), realizing the mapping from 3D spatial coordinates and viewing directions to color and volume density[2]. Its core process is divided into three steps: first, collect multi-view images and camera parameters of the scene to build a training dataset; second, generate 3D spatial sampling points through ray sampling, input the sampling point coordinates and viewing directions into the MLP, and output the radiance intensity and volume density of the corresponding sampling points; finally, calculate the pixel color of each ray through integration of the volume rendering equation, compare it with the real image pixels, and optimize the MLP parameters using the gradient descent method until the model converges[12].

The defects of the traditional NeRF algorithm are mainly reflected in two aspects: first, the MLP network structure is relatively deep (usually containing 8-10 hidden layers), with millions of parameters and a large model volume; second, the ray sampling adopts a uniform dense sampling strategy, and each ray needs to sample dozens or even hundreds of points, resulting in huge computational overhead in the inference stage and difficulty in realizing real-time inference[13]. These defects make it impossible to be directly applied to the edge deployment scenario of automobile factory digital twins.

3. Lightweight Improvement of NeRF for Automobile Factory Digital Twins

Combined with the characteristics and requirements of automobile factory digital twin scenarios, this paper proposes the Factory-LiteNeRF algorithm from three dimensions: scene partition modeling, network structure pruning, and feature encoding optimization, realizing model lightweight and inference acceleration, while ensuring the reconstruction accuracy of core areas.

3.1 Scene Partition Modeling and Differentiated Sampling

Aiming at the difference in accuracy requirements between static and dynamic areas in automobile factory scenes, a scene partition modeling strategy is proposed, which divides the scene into two categories: core dynamic areas and secondary static areas, and adopts differentiated sampling and modeling strategies to reduce invalid computational overhead[14].

First, image semantic segmentation technology is used to automatically identify dynamic areas (such as assembly line conveyor lines and welding robots) and static areas (such as floors and walls) in the scene, and mark the boundary range of core dynamic areas. The semantic segmentation adopts the lightweight network MobileNetV3 as the backbone network to ensure the balance between segmentation speed and accuracy, avoiding the impact of segmentation time consumption on the overall modeling efficiency[15].

Second, differentiated sampling strategies are designed for different areas: dense sampling (64 points sampled per ray) is adopted for core dynamic areas to ensure accurate modeling of equipment details and dynamic changes; sparse sampling (32 points sampled per ray) is adopted for secondary static areas to reduce computational overhead caused by redundant sampling. At the same time, an adaptive adjustment mechanism of sampling points is introduced to dynamically adjust the sampling density according to the scene complexity, avoiding insufficient sampling in core areas or excessive sampling in static areas.

3.2 Network Structure Pruning Optimization

The MLP network of traditional NeRF has a large number of redundant parameters. To address this problem, a network pruning strategy based on sensitivity analysis is proposed to reduce the number of model parameters and compress the model volume on the premise of ensuring reconstruction accuracy.

First, sensitivity analysis is performed on the trained NeRF model to calculate the parameter sensitivity of each network layer and neuron. The lower the sensitivity, the smaller the impact of the parameter on the reconstruction accuracy, and the more suitable it is for pruning. Parameter sensitivity is measured by calculating the L2 norm of the parameter gradient, and the smaller the gradient norm, the lower the parameter sensitivity.

Second, a hierarchical pruning strategy is adopted: for the hidden layers of the MLP, neurons with low sensitivity are pruned (pruning ratio is 30%-40%); for the input layer and output layer, core feature parameters are retained, and only redundant connection parameters are pruned (pruning ratio is 20%-30%). At the same time, aiming at the characteristics of automobile factory scenes, the output layer of the model is optimized to reduce redundant computation of color channels and further compress the model volume.

After pruning, the model is fine-tuned to recover the accuracy loss caused by pruning, ensuring that the pruned model can still meet the reconstruction accuracy requirements of automobile factory scenes.

3.3 Lightweight Optimization of Feature Encoding

Traditional NeRF adopts Positional Encoding to map 3D coordinates and viewing directions to a high-dimensional space, enhancing the feature expression ability of the model, but high-dimensional encoding will increase computational overhead and model complexity. To address this problem, a lightweight feature encoding strategy is proposed to simplify the encoding process and reduce computational overhead.

First, the dimension of positional encoding is simplified. The positional encoding dimension of traditional NeRF is 1024 dimensions, which is simplified to 512 dimensions in this paper, while retaining the core feature frequency bands to ensure that the model's ability to express scene details is not affected. Second, a shared encoding mechanism is introduced to share part of the network layers for the encoding of 3D coordinates and viewing directions, reducing repeated computations and further reducing computational overhead.

In addition, aiming at the relatively fixed lighting changes in automobile factory scenes, the lighting feature encoding is simplified, redundant lighting parameters are removed, and only core lighting features are retained, reducing the computational load and parameter number of the model, and realizing inference acceleration.

4. Experimental Verification and Analysis

4.1 Experimental Environment and Dataset

Experiments are conducted on the assembly line scene of an automobile manufacturing plant to construct the experimental dataset and verification environment, which are specifically as follows:

Experimental hardware: CPU is Intel Core i7-12700H, GPU is NVIDIA RTX 3060 (6GB), memory is 16GB, simulating the computing resource constraints of industrial edge devices; Experimental software: operating system is Ubuntu 20.04, deep learning framework is PyTorch 1.12.0, programming language is Python 3.8.

Experimental dataset: Multi-view images of the assembly line scene of an automobile factory are collected by multi-view cameras, totaling 200 multi-view images with a resolution of 1920×1080. Camera parameters are obtained by Zhang Zhengyou's calibration method. The dataset includes assembly line conveyor lines, assembly robots, automobile bodies, floors, walls and other elements, covering core dynamic areas and secondary static areas, which are used for model training and verification.

Comparison algorithms: Traditional NeRF algorithm, lightweight algorithms TinyNeRF and KiloNeRF are selected as comparison objects, and comparative verification is carried out from three dimensions: model volume, inference speed, and reconstruction accuracy to ensure the superiority of the improved algorithm.

4.2 Experimental Indicators and Evaluation Criteria

Three core indicators are used to evaluate the algorithm performance, adapting to the needs of automobile factory digital twin scenarios:

(1) Model volume: Measures the storage overhead of the model, in MB. The smaller the volume, the more suitable it is for edge device deployment;

(2) Inference speed: Measures the real-time inference ability of the model, in fps (frames per second). The faster the speed, the better it can meet the real-time update needs of dynamic scenarios;

(3) Peak Signal-to-Noise Ratio (PSNR): Measures the scene reconstruction accuracy, in dB. The higher the PSNR value, the higher the reconstruction accuracy. The automobile factory scene requires PSNR $\geq$ 28 dB.

4.3 Experimental Results and Analysis

The traditional NeRF, TinyNeRF, KiloNeRF and the proposed Factory-LiteNeRF algorithm in this paper are trained respectively, and performance tests are carried out under the same experimental environment and dataset. The test results are shown in table 1.

Table 1: Performance Comparison of Different NeRF Algorithms

Algorithm Type	Model Volume (MB)	Inference Speed (fps)	PSNR (dB)	Accuracy Loss (%)
Traditional NeRF	256.0	12.3	32.6	0.0
TinyNeRF	89.6	28.7	29.8	8.6
KiloNeRF	76.8	30.5	29.2	10.4
Factory-LiteNeRF (Proposed)	71.2	46.5	31.6	3.1

It can be seen from the experimental results that the Factory-LiteNeRF algorithm proposed in this paper has significant advantages in performance:

(1) Model volume: Compared with the traditional NeRF, the model volume of Factory-LiteNeRF is compressed from 256.0MB to 71.2MB, with a compression ratio of 72.3%. The volume is smaller than that of TinyNeRF and KiloNeRF, making it more suitable for the storage needs of industrial edge devices;

(2) Inference speed: The inference speed of Factory-LiteNeRF reaches 46.5fps, which is 2.8 times higher than that of the traditional NeRF (12.3fps), and also significantly higher than that of TinyNeRF (28.7fps) and KiloNeRF (30.5fps), which can meet the real-time update needs of dynamic scenarios such as automobile factory assembly lines (frame rate ≥ 15fps);

(3) Reconstruction accuracy: The PSNR value of Factory-LiteNeRF is 31.6 dB, which only loses 3.1% compared with the traditional NeRF, much lower than the accuracy loss of TinyNeRF (8.6%) and KiloNeRF (10.4%), and the PSNR value ≥ 28 dB, meeting the reconstruction accuracy requirements of automobile factory scenes.

In addition, supplementary experiments on the welding workshop scene of automobile factories show that the Factory-LiteNeRF algorithm can still maintain high reconstruction accuracy and inference speed under complex lighting conditions, adapting to the needs of different automobile factory digital twin scenarios, and verifying the generality and practicality of the algorithm.

5. Conclusions

Aiming at the problems of large model volume, slow inference speed, and poor adaptability of the traditional NeRF algorithm in automobile factory digital twin scenarios, combined with the characteristics of typical scenarios such as automobile factory assembly lines and welding workshops, this paper proposes a lightweight NeRF improvement algorithm (Factory-LiteNeRF). Through improvements in three dimensions: scene partition modeling, network structure pruning, and feature encoding optimization, the model lightweight and inference acceleration are realized, while ensuring the reconstruction accuracy of core areas.

Experimental results show that the improved algorithm proposed in this paper compresses the model volume by 72.3%, increases the inference speed by 2.8 times, and the reconstruction accuracy loss does not exceed 3.1% in the automobile factory assembly line scene, which can meet the real-time modeling and edge deployment needs of automobile factory digital twins, and has better performance and adaptability compared with existing lightweight NeRF algorithms. This algorithm can provide an efficient and economical solution for scene modeling of automobile factory digital twins, helping the intelligent transformation of automobile factories.

The research in this paper still has certain limitations: first, scene partitioning depends on the accuracy of semantic segmentation, and partition errors may occur in complex scenes; second, the lightweight strategy of the algorithm does not fully consider the dynamic change range of automobile factory scenes, and there is still room for improvement in the modeling accuracy of high-speed moving equipment. The main future research directions are twofold: first, integrate lidar data and image data to optimize scene partition accuracy and improve the adaptability of complex scenes; second, design an adaptive lightweight strategy to dynamically adjust sampling density and network structure according to the dynamic change range of the scene, further improving the performance and generality of the algorithm, and promoting the wide application of the NeRF algorithm in the field of automobile factory digital twins.

References

[1] Chen Gen. Digital Twin[M]. Beijing: Electronic Industry Press, 2020.

[2] MILDENHALL B, SRINIVASAN P P, TANCIK M, et al. NeRF: Representing Scenes as Neural Radiance Fields for

View Synthesis[J]. arXiv Preprint, arXiv:2003.08934, 2020.
[3] Zhao Hui, Zheng Siyuan. Research on the Application of NeRF in the Communication Industry[J]. Information and Communications Technology and Policy, 2024, 50(12): 37-41.
[4] Dassault DELMIA Automotive Intelligent Manufacturing System: Flexible Production Line Planning and Factory Operation Efficiency Improvement Scheme Based on Digital Twin[R]. Chengdu Bestway Technology Co., Ltd., 2026.
[5] Project Case | A Digital Twin Platform for Automobile Production Lines in a Smart Factory[R]. Piaoshi Technology Co., Ltd., 2026.
[6] DING T, XIANG D, RIVAS P, et al. Neural Pruning for 3D Scene Reconstruction: Efficient NeRF Acceleration[J]. arXiv Preprint, arXiv:2504.00950v1, 2025.
[7] Liu H, Saksham D, Shen M, et al. Review of Digital Twin in the Automotive Industry on Products, Processes and Systems[J]. International Journal of Automotive Manufacturing and Materials, 2025, 1(1): 1-25..
[8] Lingtu Interaction. 5G Automobile Factory Digital Twin Collaboration Scheme[R]. Lingtu Interaction (Wuhan) Technology Co., Ltd., 2026.
[9] BMW Group. BMW iFACTORY Digital Twin Technology Application Report[R]. BMW Group, 2025.
[10] Ford Motor Company. Practice of Digital Twin Technology for Equipment Operation and Maintenance in Automobile Factories[R]. Ford Motor Company, 2025.
[11] Soori M, Arezoo B. Digital twin for smart manufacturing in automotive industry[J]. Journal of Manufacturing Processes, 2023, 96: 413-430. https://doi.org/10.1016/j.jmapro.2023.05.012.
[12] BARRON J T, MILDENHALL B, TANCIK M, et al. Mip-NeRF: a multiscale representation for anti-aliasing neural radiance fields[J]. arXiv Preprint, arXiv:2103.13415, 2021.
[13] MÜLLER T, CHAURASIA G, FRITZ C, et al. Instant neural graphics primitives with a multiresolution hash encoding[J]. arXiv Preprint, arXiv:2201.05989, 2022.
[14] Digital Twin Visualization System for Automobile Manufacturing Workshops Based on HT Technology[R]. Tutu Software Co., Ltd., 2026.
[15] Google Brain. MobileNetV3: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(10): 2440-2451.