Chinese Journal of Electronics
vol.XX, no.XX, pp.1– 11, XXXX
https://doi.org/XXX.XXX.XXX.XXX

Research Article

Exploring Neural Radiance Fields for Thermal View Synthesis Solely with Thermal Inputs

Haixuan Ding[1], Jialiang Tang[1], Sheng Wan[2], and Chen Gong[1]

1. *School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China*
2. *College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 211800, China*

Corresponding authors: Sheng Wan and Chen Gong; Email: wansheng315@163.com; chen.gong@njust.edu.cn.
Received December 09, 2024; Accepted April 30, 2025; Published XXX XX, 2025.

Abstract — Novel View Synthesis (NVS) for thermal scenes aims to generate thermal images from unseen viewpoints, which shows great potential in various applications, such as nighttime autonomous driving, industrial inspection, and agricultural monitoring. Recently, Neural Radiance Fields (NeRF) have emerged as a powerful approach for NVS in thermal scenes, which typically require paired RGB and thermal images to produce realistic thermal images from new views. However, practical limitations, such as insufficient lighting, the prohibitive cost of RGB image acquisition, or the lack of RGB cameras, make it challenging or even impossible to obtain high-quality RGB images, which prevents the existing NeRF methods from generating realistic thermal images. To address this problem, we devise a simple yet effective **NeRF** framework based on **T**hermal **R**adiation **P**rediction, which is termed 'NeRF-TRP', for NVS in thermal scenes. Unlike the existing NeRF techniques that rely on paired RGB and thermal images, NeRF-TRP exclusively utilizes thermal images as input. By leveraging the principle of thermal imaging, NeRF-TRP predicts the thermal radiation emitted by objects to render thermal images from novel views. Meanwhile, motivated by the thermal equilibrium observed in thermal scenes, we design a patch-based regularization to enhance the realism of the generated thermal images. Extensive experiments on thermal images demonstrate that NeRF-TRP not only produces more accurate thermal image synthesis, but also reveals superior efficiency in both training and rendering when compared with various representative baseline approaches.
Keywords — Novel view synthesis, Neural radiance fields, Thermal imaging.

I. Introduction

Novel View Synthesis (NVS) is a fundamental task in computer vision and graphics, which involves generating novel perspectives of an object or scene from a limited set of reference images. It plays a crucial role in enhancing the spatial awareness and interaction abilities of intelligent sensing systems. Recently, NVS of thermal images has attracted increasing research attention due to the unique all-weather imaging capabilities of thermal images, which remains unaffected by optical illumination and imperfect weather constraints. This makes thermal NVS particularly suitable for synthesizing novel views in challenging environments with strong environmental interference, highlighting its potential for a wide range of practical applications, such as nighttime autonomous driving, industrial inspection, and agricultural monitoring [1].

To date, various methods have been developed to tackle NVS tasks, including view interpolation [2–4], multi-view geometry [5, 6], and depth-based rendering [7, 8]. Among these, Neural Radiance Fields (NeRF) [9] have shown remarkable capability in generating highly realistic and detailed renderings from previously unseen viewpoints. NeRF employs a Multi-Layer Perceptron (MLP) to encode a 3D scene as a radiance field. Specifically, the MLP maps a series of discrete 5D coordinates, including the three-dimensional position (x, y, z), the viewing direction defined by the azimuth angle θ, and the pitch angle ϕ, to continuous representations of volume density and color for the given scene. These continuous representations naturally capture intricate variations in light and shadow, enabling accurate rendering for new views. To produce satisfactory thermal images from new views, most existing NeRF models [10, 11] utilize paired RGB and thermal images to learn scene geometry, with RGB images providing rich texture information. Nevertheless, real-world limitations, such as imperfect weather conditions or insufficient lighting, make it difficult to capture high-quality RGB images. Moreover, in the scenarios where the cost of capturing RGB images is prohibitive or RGB cameras are unavailable, obtaining RGB images may become impossible. Consequently, it is imperative to develop a new NeRF approach for

Chinese Journal of Electronics

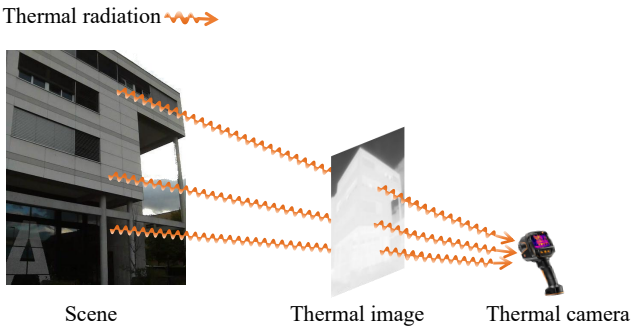


Figure 1 The diagram of thermal imaging. Here, the thermal image is generated from the emitted thermal radiation of objects, which reveals temperature differences in the scene based on variations in thermal energy.

NVS in thermal scenes that operates independently of RGB data.

To address the challenges of NVS in thermal scenes, we propose a simple yet effective **NeRF** framework based on **T**hermal **R**adiation **P**rediction termed 'NeRF-TRP', which requires only thermal images as input. Our design is inspired by the principle of thermal imaging, as illustrated in Figure 1. The formation of thermal images is intrinsically governed by thermal radiation emitted by objects, which is quantified by the product of radiant energy density and volume density as described in radiometry [12]. Building on this principle, our NeRF-TRP framework aims to predict radiant energy density and volume density for each given 3D spatial point. Thermal radiation values traced along the viewing trajectory are then aggregated through a volume rendering process to generate the final thermal images. This thermal-specific design enables NeRF-TRP to produce realistic thermal images with fine details. Furthermore, motivated by the thermal equilibrium achieved via continuous heat transfer among objects, we propose a patch-based regularization technique to enhance the quality of the generated thermal images. In summary, our contributions are as follows:

- We propose a new NeRF framework by leveraging the mechanism of thermal imaging, which enables the generation of realistic thermal images from unseen viewpoints using only a sparse set of thermal observations.
- The thermal volume rendering technique and the patch-based regularization are developed to accurately encode the thermal radiation distributions for objects, which enables our method to produce detailed and smooth thermal images.
- Experiments demonstrate that our approach achieves state-of-the-art performance in thermal image quality, training time, and rendering efficiency.

II. Related Work

This section provides an overview of key works relevant to our approach, including traditional NVS methods and the progress in NeRF.

1. Traditional Novel View Synthesis Methods

NVS is a fundamental problem in computer vision and graphics, aiming to generate new viewpoints of a scene from a limited set of input images. In general, the traditional NVS methods primarily rely on geometric and photometric principles, which can be divided into view interpolation, multi-view geometry, and depth-based rendering.

View interpolation methods produce new views directly from input images without constructing detailed 3D models. To achieve this, view morphing [2] interpolates between input views using epipolar geometry to ensure smooth transitions between views. Similarly, light field rendering [3, 4] captures dense angular examples of a scene and uses ray interpolation to synthesize new viewpoints. However, these methods usually require dense image sampling and tend to underperform in practical scenarios with a small number of images.

Multi-view geometry solutions focus on recovering explicit 3D structure and camera parameters. For example, Structure from Motion (SfM) [5, 13] and Multi-View Stereo (MVS) [6, 14–16] are widely used for reconstructing meshes and 3D point clouds [17, 18] from a collection of images. Once the geometry is recovered, traditional rendering techniques, such as texture mapping, are employed to generate novel views. While these methods can effectively generate images for new views, they depend heavily on accurate feature matching, camera calibration, and dense viewpoints.

Depth-based rendering approaches [7, 8, 19–21] leverage depth maps or layered depth images to synthesize novel views. By warping pixels from the input images to the target view using depth information, these methods enable efficient view synthesis. However, the synthetic images from novel views often suffer from low quality due to issues such as occlusion handling, depth discontinuities, and artifacts in poorly textured regions.

2. Neural Radiance Fields

Recently, NeRF [9] has shown highly realistic rendering capability and revolutionized the field of NVS. By encoding input points in 3D space as continuous volume density and color information, NeRF can produce photo-realistic images from new perspectives. Its success has garnered considerable attention, inspiring several follow-up works. Mip-NeRF [22] and Mip-NeRF 360 [23] introduce a cone-based positional encoding to address aliasing issues in neural rendering. Plenoxels [24] directly store the density and color value of each point in a voxel grid, accelerating the synthesis process. TensorRF [25] employs a tensor-based representation to improve memory efficiency. Instant NGP [26] leverages multi-resolution hash encoding to efficiently represent

Chinese Journal of Electronics

3D scenes, enabling real-time neural rendering with high-quality details.

The above NeRF methods can effectively generate high-quality new perspective images in RGB scenes. However, many practical applications, such as pharmacy [27], agriculture [28, 29], and advanced driver assistance system (ADAS) [30, 31], require the ability to synthesize realistic thermal images from new views. Existing NeRF methods often fail in these scenarios due to the fundamental differences between RGB and thermal imaging. Several recent studies have attempted to synthesize thermal images. For instance, X-NeRF [32] and ThermalNeRF [33] create a multi-spectral scene representation by combining infrared and visible light images to generate new views. ThermoNeRF [10] and ThermalMix [11] propose to utilize paired RGB and thermal images to learn scene geometry information for rendering thermal images. In general, these methods can generate realistic thermal images with the assistance of RGB images, but their effectiveness is limited in real-world scenarios where capturing RGB data could be challenging due to poor lighting or imperfect weather conditions. Moreover, RGB data may be unavailable owing to the limitations in capturing equipment or environment. In this work, we propose NeRF-TRP, a novel approach specifically designed for thermal view synthesis. Unlike prior methods, NeRF-TRP operates solely on thermal images, leveraging the unique characteristics of thermal data to effectively synthesize thermal images from unseen viewpoints without relying on RGB data.

III. Preliminaries

Given a specific scene, the vanilla NeRF can synthesize photo-realistic images from arbitrary unseen viewing directions. To achieve this, NeRF employs a MLP to parametrize the radiance field of the target scene with a sparse set of images captured from various viewpoints. During training, NeRF casts rays from camera centers through each pixel of the images. Each ray $\mathbf{r}$ is associated with a viewing direction $\mathbf{d} \in \mathbb{R}^3$ and N points $\{\mathbf{x}_i \mid i = 1, ..., N, \mathbf{x}_i \in \mathbb{R}^3\}$ sampled progressively along the path of the ray, where $\mathbf{x}_i$ corresponds to the three-dimensional spatial coordinate of a sampled point. Subsequently, the MLP maps the spatial coordinate of each point and the viewing direction to the corresponding color $\mathbf{c} \in \mathbb{R}^3$ and volume density $\sigma \in \mathbb{R}$. In the vanilla NeRF, the density value is predicted solely based on the 3D location, while the color value is influenced by both the 3D location and the viewing direction. This design makes the color prediction view-dependent, allowing it to account for non-Lambertian surface effects.

Specifically, given a camera ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$, originating from the camera center $\mathbf{o} \in \mathbb{R}^3$ and propagating a distance $t \in \mathbb{R}$ along the direction $\mathbf{d}$. The expected color $C(\mathbf{r})$ rendered from the ray within the range of near bound t_n and far bound t_f is calculated according to [34]:

$$C(\mathbf{r}) = \int_{t_n}^{t_f} T(t)\, \sigma(\mathbf{r}(t))\, \mathbf{c}(\mathbf{r}(t), \mathbf{d})\, dt, \tag{1}$$

where $\sigma(\mathbf{r}(t))$ and $\mathbf{c}(\mathbf{r}(t), \mathbf{d})$ represent the volume density and color value of a point in 3D space. The expression $T(t) = \exp(-\int_{t_n}^{t} \sigma(\mathbf{r}(s))\, ds)$ denotes the accumulated transmittance along the ray, indicating the probability that the ray travels from t_n to t without hitting any other particle. To compute this integral, a series of points are progressively sampled along the ray's path. The value of the integral is then approximated using numerical quadrature, with the discrete form as follows:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i\, (1 - \exp(-\sigma_i \delta_i))\, \mathbf{c}_i, \tag{2}$$

where $T_i = \exp\left(-\sum_{j=1}^{i-1} \sigma_j \delta_j\right)$, and δ_i represents the distance between adjacent sampled points. The symbols σ_i and $\mathbf{c}_i$ denote the volume density and color value of the sampled point $\mathbf{x}_i$, respectively.

During training, the NeRF model is optimized by minimizing a photometric loss, *i.e.*, the squared error between the predicted pixel color $\hat{C}(\mathbf{r})$ and the corresponding ground truth color $C(\mathbf{r})$:

$$\mathcal{L} = \sum_{\mathbf{r} \in \mathcal{R}} \left\| C(\mathbf{r}) - \hat{C}(\mathbf{r}) \right\|^2, \tag{3}$$

where $\mathcal{R}$ represents the set of rays in each batch.

IV. The Proposed Method

In this section, we propose a novel framework named NeRF-TRP for synthesizing thermal images from unobserved views using only a sparse set of thermal images. An overview of NeRF-TRP is shown in Figure 2. In Section 4.1, we introduce the thermal volume rendering method specifically designed for rendering thermal images. In Section 4.2, we elucidate the application of multi-resolution hash encoding to accelerate both the training and rendering processes. In Section 4.3, we present a patch-based regularization to improve the quality of the generated thermal images.

1. Thermal Volume Rendering

Unlike the original NeRF designed for RGB scenes, we aim to develop a thermal scene representation for synthesizing thermal views. Based on the principles of thermal imaging and radiometry [12], thermal cameras detect the heat energy emitted by objects as the thermal image. The thermal radiation, which represents the heat radiated by an object, is quantified as the product of radiant energy density (the amount of radiant energy emitted per unit volume) and volume density. According to Stefan-Boltzmann's law [35], the

Chinese Journal of Electronics

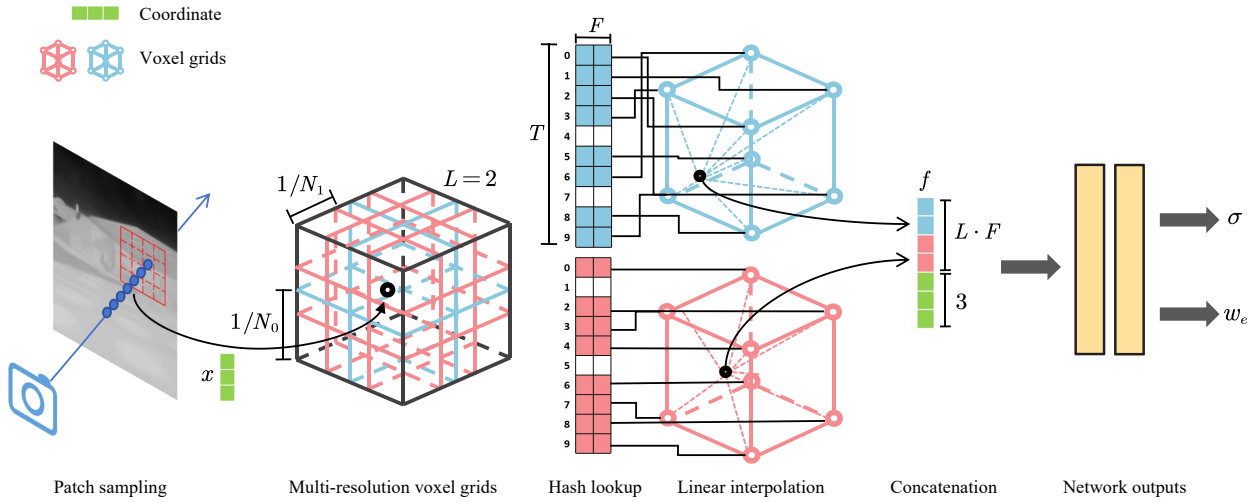


Figure 2 **An overview of our method.** Given a set of thermal images, we first adopt a random patch sampling approach to extract small patches and construct rays emitted from these patches. Subsequently, based on the coordinate $\mathbf{x}$ sampled along these rays, we retrieve feature vectors from the multi-resolution hash table and produce the feature $\mathbf{f}$. Finally, we employ a lightweight MLP to predict the volume density σ and radiant energy density w_e, which are used to render the image pixel.

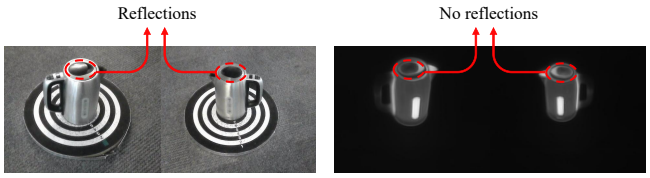
intensity of thermal radiation emitted by a material is determined solely by its temperature and is independent of the viewing angle, therefore excluding non-Lambertian effects commonly observed in RGB images. As depicted in Figure 3(a), the RGB images display view-dependent light reflections, whereas thermal images, depicted in Figure 3(b), do not exhibit such effects. Therefore, as illustrated in Figure 2, we leverage spatial positioning to predict the radiant energy density $w_e \in \mathbb{R}$ and volume density $\sigma \in \mathbb{R}$ of each point in 3D space. This position-based approach ensures that NeRF-TRP accurately represents thermal properties while simplifying the model by excluding viewing direction as input, thereby reducing network complexity and accelerating both training and rendering. Similar to Equation (2), the expected color $\hat{C}(\mathbf{r})$ of a pixel can be approximated as:

$$\hat{C}(\mathbf{r}) = \sum_{i=1}^{N} T_i \left(1 - \exp\left(-\sigma_i \delta_i\right)\right) w_{e_i}, \tag{4}$$

where w_{e_i} represents the radiant energy density of the sampled point $\mathbf{x}_i$.

2. Multi-Resolution Hash Encoding

In Section 4.1, we shed light on the importance of predicting the volume density and radiant energy density of each 3D point to enable high-quality thermal NVS. Here, we describe the method employed to efficiently predict these values. The original NeRF utilizes a large MLP to predict the properties (*e.g.*, color and volume density) of each sampled point in order to generate high-fidelity renderings, resulting in a time-intensive optimization process that requires millions of MLP queries per iteration. To accelerate the optimization process, we follow [26] and map the sampled point features into a multi-resolution hash-grid table. This multi-resolution hash-grid serves as an efficient data structure that divides space into small cubes at multiple resolutions. Instead of explicitly storing features for each vertex in a dense grid, it utilizes a fixed-size hash table to store the feature representations, which significantly reduces the number of parameters. By adopting this structure, the features of sampled points can be obtained via fast feature interpolation, which greatly accelerates the optimization process.



(a) RGB images (b) Thermal images

Figure 3 Illustration of the differences between RGB and thermal images. (a) RGB images exhibit non-Lambertian effects, *i.e.*, view-dependent light reflections, while (b) thermal images are unaffected by such effects.

Specifically, the process of multi-resolution hash encoding is depicted in Figure 2, where the voxel grids in blue and red have different resolutions (for simplicity, two resolutions are used in the illustration). Given a three-dimensional spatial coordinate $\mathbf{x}$, we initially search the neighboring voxel vertices $\mathbf{V} = \left\{\mathbf{p}_i \mid i = 1, ..., 8, \mathbf{p}_i \in \mathbb{R}^3\right\}$ within each grid of the multi-resolution structure, where $\mathbf{p}_i$ denotes the three-dimensional spatial coordinate of a voxel vertex. Using the hash function h, we then fetch the features of the voxel ver-

Chinese Journal of Electronics

tices from the hash table and perform linear interpolation based on the relative position of $\mathbf{x}$ within each resolution grid. Finally, the resulting features from all levels along with the original input coordinate $\mathbf{x}$ are concatenated to produce the multi-resolution hash features $\mathbf{f}$. By leveraging multi-resolution hash encoding, the initial input coordinate $\mathbf{x}$ is transformed into the feature representation $\mathbf{f}$ that integrates information across various scales of the scene, capturing both fine-grained details and global structure. On this basis, the accuracy of point property prediction can be improved, which ultimately enhances the quality of the synthesized thermal images. For a vertex coordinate $\mathbf{p} = (p_x, p_y, p_z)$, we employ the following spatial hash function to extract features:

$$h(\mathbf{p}) = \left(\bigoplus_{i=x,y,z} p_i \pi_i \right) \bmod T, \tag{5}$$

where $\bigoplus_{i=x,y,z}$ denotes the bit-wise XOR operation performed over the dimensions x, y, and z, and π_i denotes the large prime number used to minimize the number of hash collisions. The number of parameters in the multi-resolution hash grid is bounded by $L \cdot T \cdot F$, where L represents the number of resolution levels, T and F denote the hash table size and feature dimension of each resolution, respectively. The resolution at each level is determined between the coarsest and finest resolutions following a geometric progression.

By adopting the multi-resolution hash grid feature $\mathbf{f}$ as input, a lightweight MLP can be utilized to predict the properties of the point in the scene, including volume density σ and radiant energy density w_e. The prediction process is then defined as follows:

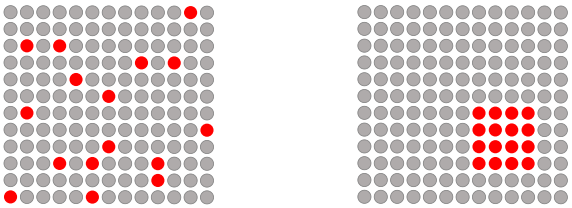
$$[\sigma,\ w_e] = \mathrm{MLP}_\theta(\phi_H(\mathbf{x})), \tag{6}$$

where $\phi_H(\cdot)$ denotes the multi-resolution hash encoding, $\mathbf{x}$ represents the three-dimensional spatial coordinate of the point, and θ corresponds to the trainable parameters of the MLP.

3. Patch-Based Regularization

In the thermal scene, objects naturally reach a state of thermal equilibrium through continuous heat transfer. Thermal cameras capture this heat distribution, and the resulting images tend to appear relatively smooth. This is different from RGB images, which often display significant fluctuations due to variations in light and surface properties. Motivated by this, to enhance the realism of the generated thermal images, we introduce Total Variation (TV) loss as a regularization technique, which is defined as follows:

$$\mathcal{L}_{TV} = \sum_{i=1}^{S_{patch}-1} \sum_{j=1}^{S_{patch}-1} \left((c_{i+1,j} - c_{i,j})^2 + (c_{i,j+1} - c_{i,j})^2 \right), \tag{7}$$



(a) Random sample (b) Patch sample

Figure 4 Comparison of different ray sampling methods during the training process, with red dots denoting the sampled pixels used to construct rays. (a) Randomly selected pixels, where rays are independent. (b) Randomly sampled patches, where rays are adjacent.

where S_{patch} is the size of the sampled patch and $c_{i,j}$ represents the rendered pixel at position (i, j). As shown in Figure 4(a), previous works utilize stochastic ray sampling during training. However, since TV loss is computed by evaluating the differences in pixel values between each pixel and its adjacent ones, a more structured approach is necessary. To address this, as shown in Figure 4(b), we adopt a patch-based sampling strategy that randomly extracts small patches from the input images.

In addition to the photometric loss $\mathcal{L}_{recon}$ and the TV loss $\mathcal{L}_{TV}$, we incorporate the proposal loss $\mathcal{L}_{prop}$ and the distortion loss $\mathcal{L}_{dist}$, following the approach in [23], to optimize the proposal sampler and reduce distortions. Finally, the total loss function of our method is calculated as follows:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \mathcal{L}_{TV} + \mathcal{L}_{dist} + \mathcal{L}_{prop}. \tag{8}$$

V. Experiments

This section evaluates the effectiveness of the proposed NeRF-TRP in addressing thermal NVS. We provide a comprehensive comparison with existing state-of-the-art NeRF methods and examine the model performance from both quantitative and qualitative perspectives. First, we introduce the experimental setups in Section 5.1. Afterwards, we present the experimental results in Section 5.2 and analyze the efficiency of the proposed NeRF-TRP in Section 5.3. Finally, we conduct ablation study to shed light on the contributions of designed components in Section 5.4.

1. Experimental Setups

1) Dataset

We conduct intensive experiments on a wide range of thermal scenes sourced from the ThermoScenes dataset [10]. Specifically, the ThermoScenes dataset comprises four outdoor and six indoor scenes, all captured using a FLIR One Pro LT thermal camera. Here, the camera poses are estimated using the COLMAP structure-from-motion package [5]. Following the standard dataset division outlined in [9], 1/8 of the data was allocated to the test set, with the remaining data used for training. All images are captured at a resolution of 480 × 640 pixels.

Chinese Journal of Electronics

Table 1 Quantitative results on the ThermoScenes dataset, with **bold** indicating the best results. Here, metric marked with ↑ (↓) represents that the higher (lower) the metric, the better the quality of the generated images.

Metric	Method	Heated Water Cup	Heated Water Kettle	Freezing Ice Cup	Melting Ice Cup	Building (Spring)	Building (Winter)	Double Robot	Raspberrypi	Exhibition Building	Trees	Avg
PSNR↑	NeRF [9]	17.32	19.65	25.92	19.07	22.18	26.55	25.53	27.46	30.86	25.04	23.96
	TensoRF [25]	22.32	33.45	27.72	19.61	21.47	27.29	20.16	24.26	32.51	21.93	25.07
	Instant NGP [26]	16.54	29.89	15.24	34.70	27.29	27.35	28.22	36.15	32.27	31.49	27.91
	Nerfacto [36]	31.95	17.93	18.91	18.30	14.82	19.46	16.12	15.34	26.06	23.41	20.23
	ThermoNeRF [10]	32.38	33.68	33.12	34.32	25.97	29.90	31.28	33.63	35.05	30.44	31.98
	NeRF-TRP (ours)	**34.72**	**33.95**	**33.46**	**38.07**	**27.38**	**30.37**	**31.50**	**36.17**	**36.49**	**36.17**	**33.83**
SSIM↑	NeRF [9]	0.155	0.042	0.979	0.947	0.918	0.872	0.909	0.950	0.948	0.912	0.763
	TensoRF [25]	0.807	0.958	0.981	0.935	0.904	0.865	0.842	0.914	0.959	0.898	0.906
	Instant NGP [26]	0.743	0.941	0.901	0.983	0.933	0.875	0.925	0.968	0.957	0.957	0.918
	Nerfacto [36]	0.888	0.618	0.956	0.981	0.852	0.900	0.682	0.630	0.961	0.966	0.843
	ThermoNeRF [10]	0.925	0.926	0.984	0.985	0.919	0.901	0.951	0.960	0.966	0.942	0.946
	NeRF-TRP (ours)	**0.930**	**0.961**	**0.987**	**0.991**	**0.936**	**0.923**	**0.953**	**0.968**	**0.974**	**0.974**	**0.960**
LPIPS↓	NeRF [9]	0.122	0.172	0.038	0.120	0.194	0.302	0.194	0.125	0.142	0.180	0.159
	TensoRF [25]	0.157	0.096	0.037	0.160	0.213	0.278	0.271	0.145	0.088	0.199	0.164
	Instant NGP [26]	0.265	0.096	0.153	0.034	0.175	0.332	0.147	0.068	0.120	0.084	0.148
	Nerfacto [36]	0.101	0.404	0.074	0.049	0.369	0.235	0.252	0.362	0.054	0.068	0.197
	ThermoNeRF [10]	0.066	0.074	0.042	0.034	0.177	0.223	0.125	0.054	0.067	0.154	0.102
	NeRF-TRP (ours)	**0.024**	**0.039**	**0.024**	**0.008**	**0.148**	**0.190**	**0.111**	**0.043**	**0.047**	**0.048**	**0.068**

2) Baseline Methods

We compare our proposed NeRF-TRP against several popular baseline approaches: (1) Methods for RGB NVS, including vanilla NeRF [9] and approaches that enhance both rendering efficiency and quality, such as TensoRF [25], InstantNGP [26], and Nerfacto [36]; (2) A method for thermal NVS, namely ThermoNeRF [10], which uses paired RGB and thermal images to render thermal images.

3) Evaluation Metrics

We employ the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index Measure (SSIM) to assess the fidelity of the generated images against the ground truth, with higher PSNR and SSIM values indicating better performance. Furthermore, we utilize the Learned Perceptual Image Patch Similarity (LPIPS) [37] score to evaluate the perceptual quality of the rendered images, where lower LPIPS scores reflect better visual quality.

4) Implementation Details

For the multi-resolution hash grid, we follow the configuration of Instant NGP [26] and set π_x=1, π_y=2,654,435,761, π_z=805,459,861, the number of resolution levels L to 16, the hash table size T to 2^{19}, and the feature vector dimension F to 2 at each level, which results in a total parameter count of 2^{24}. The lowest and highest resolutions are set to 16 and 2048, respectively (see Section 4.2). This configuration strikes a balance between model capacity and computational efficiency. Since we only use the positional information of sampled points to predict values, we employ a lightweight MLP consisting of two hidden layers, each containing 64 channels. The model is trained for 20,000 iterations with a batch size of 4,096 rays, each containing 48 sampled points, on a single NVIDIA RTX 4090 GPU. We use the Adam optimizer, starting with a learning rate of 10^{-2} that decays exponentially to 10^{-3}. Under this configuration, the training typically converges within 10 to 15 minutes. Moreover, we set the sampled patch size (S_{patch}) to 4×4 based on the thermal image resolution. For a fair comparison, all baseline methods are trained using their default configurations.

2. Experimental Results

1) Quantitative Results

Table 1 reports the quantitative comparisons between our proposed NeRF-TRP and various baseline methods. We observe that methods designed for rendering RGB images fail to yield satisfactory results for thermal scenes. Differently, our method is specifically tailored for thermal imaging and thus generally achieves better performance than the compared baseline methods, which demonstrates its ability to effectively capture the unique characteristics of thermal scenes. In addition, it is notable that our method achieves a 1.85 dB improvement in PSNR compared with ThermoNeRF. This is due to that ThermoNeRF relies on RGB content and visible light imaging mechanism to capture scene geometry, which might be unsuitable for thermal view synthesis, whereas our approach is specifically designed based on the principle of thermal imaging. These experimental results highlight the effectiveness of our NeRF-TRP in generating high-quality thermal images from unseen views, particularly in real-world scenarios where RGB data is unavailable.

2) Qualitative Results

We also conduct qualitative experiments to further verify the effectiveness of our proposed NeRF-TRP. As visualized in Figure 5, our method performs well in recovering intricate details in both visual aesthetics and geometry, such as the clarity of building windows and tree branches. In contrast, the baseline methods designed for rendering RGB images yield poor results due to the inherited differences between RGB and thermal imaging mechanisms. This often results in inaccurate geometric structures and noticeable artifact blurring. Among these methods, Nerfacto exhibits relatively better ge-

Chinese Journal of Electronics

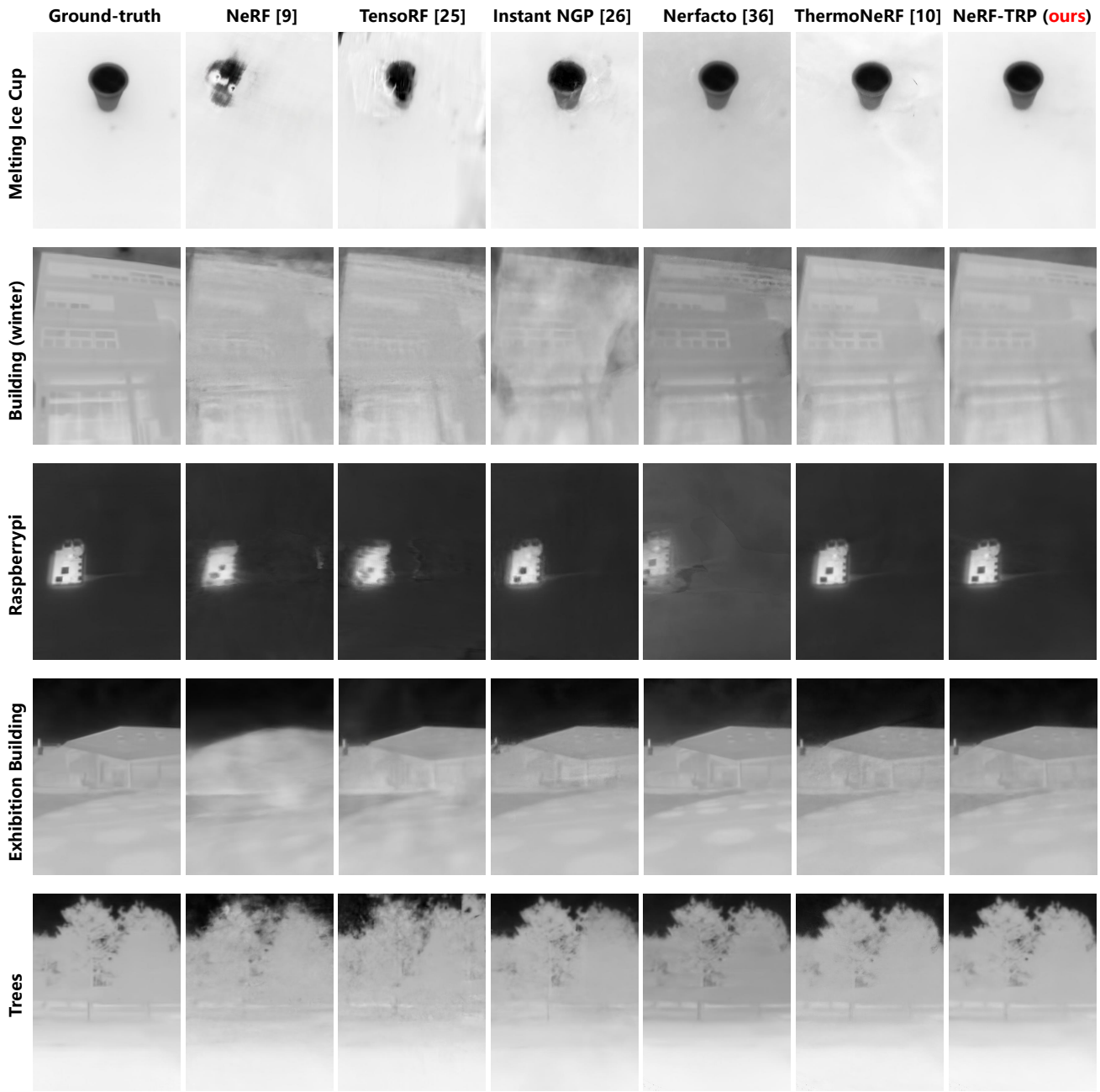


Figure 5 Qualitative results on the ThermoScenes dataset.

ometry recovery than the others but suffers from suboptimal color representation in the rendered images. On the other hand, ThermoNeRF, which utilizes paired RGB and thermal images, produces improved results but relies heavily on the availability of RGB data. This reliance poses challenges in real-world scenarios with extreme weather or poor lighting, where capturing high-quality RGB images is difficult. Moreover, obtaining RGB images may be entirely impossible due to the high cost of image acquisition or the lack of suitable cameras. Notably, our proposed method is capable of generating high-quality thermal images using only thermal data, making it a robust solution for scenarios where RGB data is difficult to obtain or even unavailable.

To further evaluate the quality of synthesized images generated by different methods, we present the difference images between the synthesized images and ground-truth image in Figure 6. Here, the MAE (Mean Absolute Error) values are displayed in the bottom-right corner of each image for quan-

Chinese Journal of Electronics

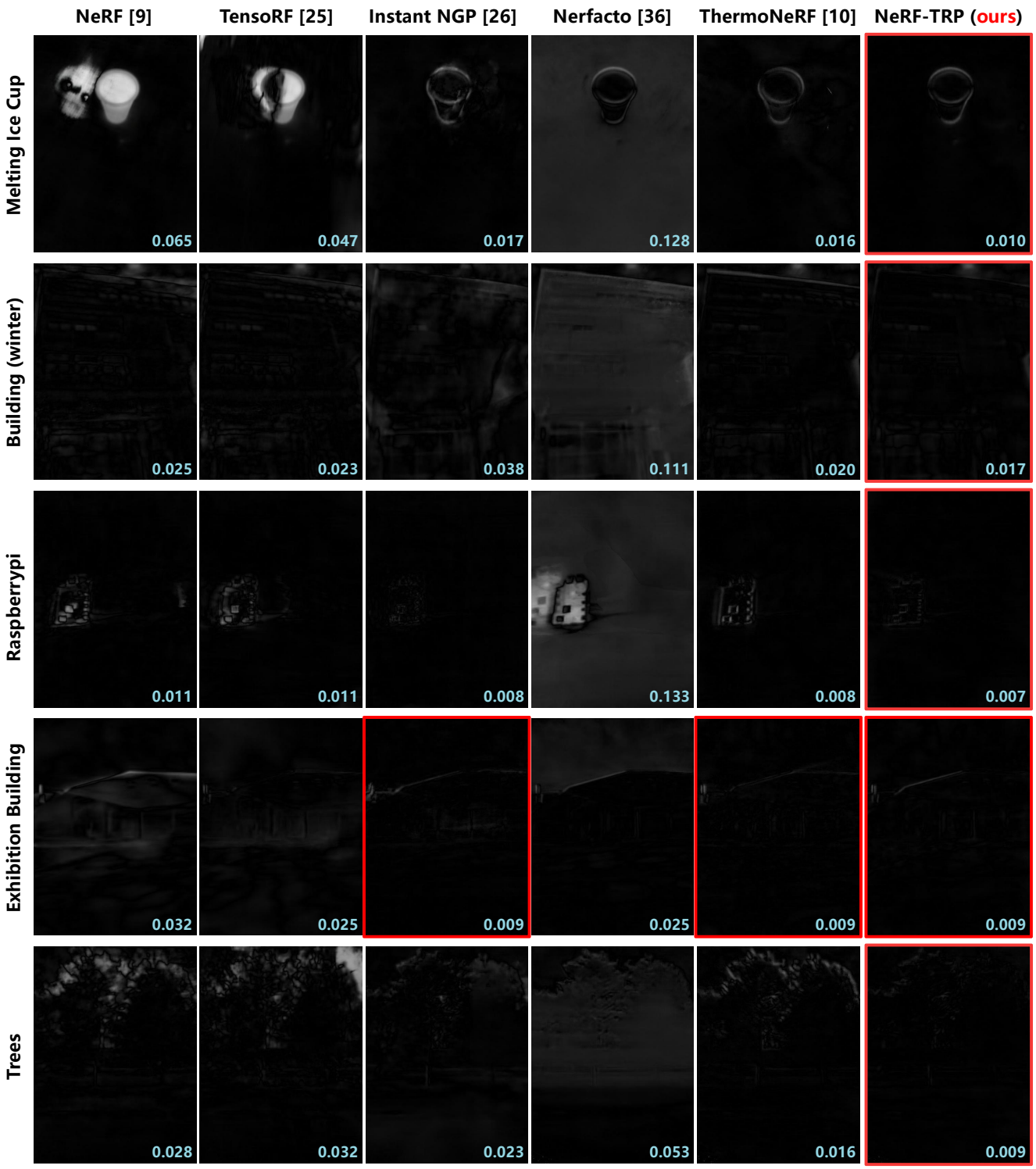


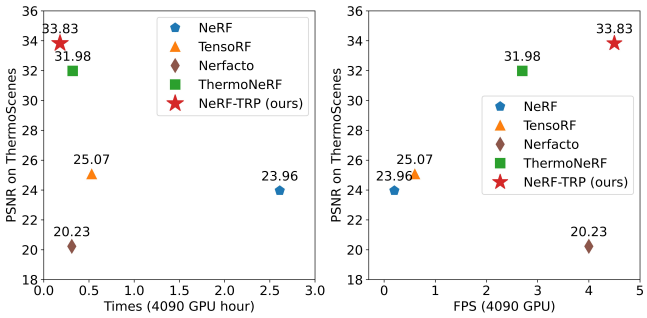
Figure 6 Difference images corresponding to the qualitative results in Figure 5. The MAE values are displayed in the bottom-right corner of each image. Here, the image with the lowest MAE value in each row is highlighted by a red box.

titative evaluation. These results demonstrate that the proposed NeRF-TRP achieves the best synthesis quality among all compared approaches.

Chinese Journal of Electronics

3. Efficiency Analysis

We compare the efficiency of the proposed NeRF-TRP with multiple baseline methods, focusing on two critical metrics, namely training time (measured as total training time in hours) and rendering efficiency (evaluated in Frames Per Second (FPS)). The results depicted in Figure 7 demonstrate that NeRF-TRP attains the highest image quality (*i.e.*, PSNR), the shortest training time, and the fastest rendering FPS among all the compared methods. The high efficiency of our method stems from the employment of multi-resolution hash encoding, which replaces computationally expensive MLP operations with efficient feature interpolation for point feature extraction. Additionally, unlike the baseline methods that incorporate the view direction as an additional input, ThermoNeRF relies solely on the positional information of sampled points for property prediction. This allows ThermoNeRF to employ a more lightweight MLP when compared with other methods, which further accelerates both training and rendering processes. These improvements make our proposed method well-suited for the applications which require both high-quality output and fast processing.



(a) PSNR *vs.* training time (b) PSNR *vs.* rending FPS

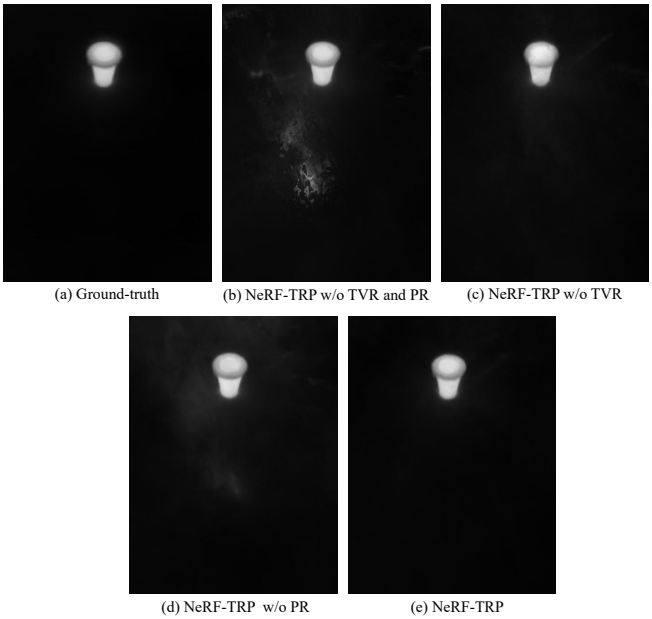
Figure 7 Training time, rendering efficiency, and image quality of different methods. (a) PSNR vs. training time (measured in 4090 GPU hours). (b) PSNR vs. rendering efficiency (evaluated in FPS).

4. Ablation Study

To validate the contributions of the devised Thermal Volume Rendering (TVR) and Patch-based Regularization (PR), we conduct ablation experiments on the ThermoScenes dataset. In these studies, we use multi-resolution hash encoding to obtain the sampled point features. The quantitative results of these experiments are presented in Table 2, and the corresponding visual comparison is shown in Figure 8. We observe that the model excluding both the TVR and PR components produces the worst performance when compared with the other models. By incorporating PR, the performance can be improved, with noise and artifacts in the generated images significantly reduced. However, geometric representation remains suboptimal (*e.g.*, the lip of the cup). Meanwhile, the utilization of TVR further enhances the results. Finally, our proposed method combining TVR and PR achieves the best overall performance among all the compared models, thereby highlighting the effectiveness of these designed components in thermal image synthesis.

Table 2 Ablation study results averaged over ten thermal scenes from the ThermoScenes dataset, with ↑ (↓) indicating higher (lower) values are better.

Model Variant	PSNR↑	SSIM↑	LPIPS↓
NeRF-TRP w/o TVR and PR	20.65	0.839	0.204
NeRF-TRP w/o TVR	28.10	0.909	0.073
NeRF-TRP w/o PR	33.31	0.954	0.078
NeRF-TRP	**33.83**	**0.960**	**0.068**



(a) Ground-truth (b) NeRF-TRP w/o TVR and PR (c) NeRF-TRP w/o TVR

(d) NeRF-TRP w/o PR (e) NeRF-TRP

Figure 8 Ablation visual comparison on the Heated Water Cup scene.

VI. Conclusions

In this paper, we propose NeRF-TRP, a neural radiance field model for synthesizing thermal images from unseen viewpoints. Unlike previous NeRF approaches that rely on paired RGB and thermal images to produce realistic thermal images, NeRF-TRP achieves this solely with thermal images as input, making it a robust solution for scenarios where RGB data is difficult or impossible to obtain. Specifically, based on the principle of thermal camera imaging, NeRF-TRP predicts thermal radiation to render images, enabling accurate representations of thermal scenes. Meanwhile, we propose a patch-based regularization inspired by the thermal equilibrium phenomenon to ensure the smooth synthesis of thermal images. Experimental results demonstrate that NeRF-TRP outperforms the compared methods, achieving state-of-the-art performance in thermal image quality, training time, and rendering efficiency.

Chinese Journal of Electronics

Although NeRF-TRP demonstrates strong capability in addressing thermal NVS, it still has several potential limitations. Similar to most NeRF-based methods, NeRF-TRP suffers from performance degradation when the input views are sparse. Additionally, NeRF-TRP requires accurate camera poses of images for training. However, unlike RGB images, thermal images often have sparse features and limited textures, making it challenging to estimate accurate camera poses. Addressing these limitations will be a focus of our future work.

Acknowledgements

This research is supported by NSF for Distinguished Young Scholar of Jiangsu Province (No: BK20220080).

References

[1] R. Gade and T. B. Moeslund, "Thermal cameras and applications: A survey", *Machine Vision and Applications*, vol.25, no.1, pp.245–262, doi:10.1007/s00138-013-0570-5, 2014.

[2] S. M. Seitz and C. R. Dyer, "View morphing", in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, New York, NY, USA, pp.21–30, doi:10.1145/237170.237196, 1996.

[3] M. Levoy and P. Hanrahan, "Light field rendering", in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, New York, NY, USA, pp.31–42, doi:10.1145/237170.237199, 1996.

[4] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph", in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, New York, NY, USA, pp.43–54, doi:10.1145/237170.237200, 1996.

[5] J. L. Schonberger and J.-M. Frahm, "Structure-from-motion revisited", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Las Vegas, NV, USA, pp.4104–4113, doi:10.1109/CVPR.2016.445, 2016.

[6] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multiview stereopsis", *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol.32, no.8, pp.1362–1376, doi:10.1109/TPAMI.2009.161, 2010.

[7] J. Shade, S. Gortler, L.-w. He, and R. Szeliski, "Layered depth images", in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, New York, NY, USA, pp.231–242, doi:10.1145/280814.280882, 1998.

[8] G. Chaurasia, S. Duchene, O. Sorkine-Hornung, and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation", *ACM Transactions on Graphics (TOG)*, vol.32, no.3, pp.1–12, doi:10.1145/2487228.2487238, 2013.

[9] B. Mildenhall, P. Srinivasan, M. Tancik, J. Barron, *et al.*, "NeRF: Representing scenes as neural radiance fields for view synthesis", in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Glasgow, UK, pp.405–421, doi:10.1007/978-3-030-58452-8_24, 2020.

[10] M. Hassan, F. Forest, O. Fink, and M. Mielle, *"ThermoNeRF: Multimodal neural radiance fields for thermal novel view synthesis", arXiv preprint*, arXiv:2403.12154, doi:10.48550/arXiv.2403.12154, 2024.

[11] M. Özer, M. Weiherer, M. Hundhausen, and B. Egger, *"Exploring multi-modal neural scene representations with applications on thermal imaging", arXiv preprint*, arXiv:2403.11865, doi:10.48550/arXiv.2403.11865, 2024.

[12] W. R. McCluney, *Introduction to Radiometry and Photometry*, Artech House, 2014.

[13] S. Vijayanarasimhan, S. Ricco, C. Schmid, R. Sukthankar, and K. Fragkiadaki, *"SfM-Net: Learning of structure and motion from video", arXiv preprint*, arXiv:1704.07804, doi:10.48550/arXiv.1704.07804, 2017.

[14] J. L. Schönberger, E. Zheng, J.-M. Frahm, and M. Pollefeys, "Pixelwise view selection for unstructured multi-view stereo", in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Amsterdam, The Netherlands, pp.501–518, doi:10.1007/978-3-319-46487-9_31, 2016.

[15] J. Zheng, B. Jiang, W. Peng, and Q. Zhang, "Multi-Scale binocular stereo matching based on semantic association", *Chinese Journal of Electronics (CJE)*, vol.33, no.4, pp.1010–1022, doi:10.23919/cje.2022.00.338, 2024.

[16] Y. Yao, Z. Luo, S. Li, T. Fang, and L. Quan, "MVSNet: Depth inference for unstructured multi-view stereo", in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Munich, Germany, pp.767–783, doi:10.1007/978-3-030-01237-3_47, 2018.

[17] Y. Yue, Y. Cai, and D. Wang, "GridNet-3D: A novel real-time 3d object detection algorithm based on point cloud", *Chinese Journal of Electronics (CJE)*, vol.30, no.5, pp.931–939, doi:10.1049/cje.2021.07.004, 2021.

[18] J. Wang, Y. Zhang, B. Zhang, J. Xia, and W. Wang, "IPFA-Net: Important points feature aggregating net for point cloud classification and segmentation", *Chinese Journal of Electronics (CJE)*, vol.34, no.1, pp.1–15, doi:10.23919/cje.2023.00.065, 2025.

[19] G. Riegler and V. Koltun, "Free view synthesis", in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Glasgow, UK, pp.623–640, doi:10.1007/978-3-030-58529-7_37, 2020.

[20] S. Wang, C. Cui, and X. Niu, "A novel DIBR 3D image watermarking algorithm resist to geometrical attacks", *Chinese Journal of Electronics (CJE)*, vol.26, no.6, pp.1184–1193, doi:10.1049/cje.2017.09.025, 2017.

[21] X. Deng, X. Cao, and Q. Dai, "Depth and residual images based rendering", *Chinese Journal of Electronics (CJE)*, vol.25, no.1, pp.131–138, doi:10.1049/cje.2016.01.020, 2016.

[22] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman, *et al.*, "Mip-NeRF: A multiscale representation for anti-aliasing neural radiance fields", in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, IEEE, Montreal, QC, Canada, pp.5835–5844, doi:10.1109/ICCV48922.2021.00580, 2021.

[23] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, "Mip-NeRF 360: Unbounded anti-aliased neural radiance fields", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, New Orleans, LA, USA, pp.5460–5469, doi:10.1109/CVPR52688.2022.00539, 2022.

[24] S. Fridovich-Keil, A. Yu, M. Tancik, Q. Chen, *et al.*, "Plenoxels: Radiance fields without neural networks", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, New Orleans, LA, USA, pp.5491–5500, doi:10.1109/CVPR52688.2022.00542, 2022.

[25] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "TensoRF: Tensorial radiance fields", in *Proceedings of the European Conference on Computer Vision (ECCV)*, Springer, Tel Aviv, Israel, pp.333–350, doi:10.1007/978-3-031-19824-3_20, 2022.

[26] T. Müller, A. Evans, C. Schied, and A. Keller, "Instant neural graphics primitives with a multiresolution hash encoding", *ACM Transactions on Graphics (TOG)*, vol.41, no.4, pp.1–15, doi:10.1145/3528223.3530127, 2022.

[27] E. Ring and K. Ammer, "Infrared thermal imaging in medicine", *Physiological Measurement*, vol.33, no.3, pp.R33–R46, doi:10.1088/0967-3334/33/3/R33, 2012.

[28] H. Yun, S. Lo, C. H. Diepenbrock, B. N. Bailey, and J. M. Ear-

les, "VisTA-SR: Improving the accuracy and resolution of low-cost thermal imaging cameras for agriculture", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Seattle, WA, USA, pp.5470–5479, doi: 10.1109/CVPRW63382.2024.00556, 2024.

[29] R. Ishimwe, K. Abutaleb, and F. Ahmed, "Applications of thermal imaging in agriculture—a review", *Advances in Remote Sensing (ARS)*, vol.3, no.3, pp.128–140, doi:10.4236/ars.2014.33011, 2014.

[30] M. Ding, W.-H. Chen, and Y.-F. Cao, "Thermal infrared single-pedestrian tracking for advanced driver assistance system", *IEEE Transactions on Intelligent Vehicles (TIV)*, vol.8, no.1, pp.814–824, doi:10.1109/TIV.2022.3140344, 2023.

[31] M. A. Farooq, P. Corcoran, C. Rotariu, and W. Shariff, "Object detection in thermal spectrum for advanced driver-assistance systems (ADAS)", *IEEE Access*, vol.9, pp.156465–156481, doi:10.1109/ACCESS.2021.3129150, 2021.

[32] M. Poggi, P. Z. Ramirez, F. Tosi, S. Salti, *et al.*, "Cross-spectral neural radiance fields", in *Proceedings of the International Conference on 3D Vision (3DV)*, IEEE, Prague, Czech Republic, pp.606–616, doi: 10.1109/3DV57658.2022.00071, 2022.

[33] Y. Y. Lin, X.-Y. Pan, S. Fridovich-Keil, and G. Wetzstein, "ThermalNeRF: Thermal radiance fields", in *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*, IEEE, Lausanne, Switzerland, pp.1–12, doi:10.1109/ICCP61108.2024.10644336, 2024.

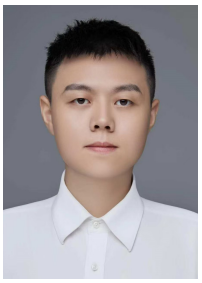
[34] J. T. Kajiya and B. P. Von Herzen, "Ray tracing volume densities", in *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, ACM, New York, NY, USA, pp.165–174, doi:10.1145/800031.808594, 1984.

[35] L. Reggiani and E. Alfinito, *"Stefan-boltzmann law revisited", arXiv preprint*, arXiv:2112.12090, doi:10.48550/arXiv.2112.12090, 2021.

[36] M. Tancik, E. Weber, E. Ng, R. Li, *et al.*, "Nerfstudio: A modular framework for neural radiance field development", in *Proceedings of the ACM Special Interest Group on Computer Graphics (SIGGRAPH)*, ACM, Los Angeles, CA, USA, pp.1–12, doi:10.1145/3588432.3591516, 2023.

[37] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, Salt Lake City, UT, USA, pp.586–595, doi:10.1109/CVPR.2018.00068, 2018.

Haixuan Ding is currently pursuing the M.S. degree in the School of Computer Science and Engineering, Nanjing University of Science and Technology, China. His research interests include novel view synthesis and neural rendering. (Email: haixuanding@njust.edu.cn)



Jialiang Tang is a Ph.D. candidate student of Nanjing University of Science and Technology. His research interests mainly include machine learning, model compression, and time series analysis. He has published more than 10 technical papers at prominent journals and conferences such as IEEE T-IP, ICCV, ECCV, AAAI, etc. He serves as the reviewer for IEEE T-CSVT, Neural Networks, NeurIPs, AAAI, AISTATS, ACM MM, etc. (Email: tangjialiang@njust.edu.cn)

Sheng Wan received his Ph.D. degree in Computer Science and Technology from the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China, in 2023. He is currently an Associate Professor with the School of Artificial Intelligence, Nanjing Agricultural University, Nanjing, China. His research interests include graph machine learning, weakly supervised learning, and hyperspectral image processing. He has published more than ten papers in top-tier conferences such as NeurIPS and AAAI, and prominent journals such as IEEE T-NNLS and IEEE T-GRS. (Email: wansheng315@163.com)

Chen Gong received his dual doctoral degree from Shanghai Jiao Tong University (SJTU) and University of Technology Sydney (UTS), respectively. Currently, he is a full professor of Nanjing University of Science and Technology. His research interests mainly include machine learning, data mining, and learning-based vision problems. He has published more than 130 technical papers at prominent journals and conferences such as IEEE T-PAMI, JMLR, IJCV, IEEE T-NNLS, IEEE T-IP, ICML, NeurIPS, ICLR, CVPR, ICCV, ECCV, AAAI, IJCAI, ICDM, etc. He serves as the associate editor for IEEE T-CSVT, Neural Networks, and NePL, and also the Area Chair or Senior PC member of several top-tier conferences such as AAAI, IJCAI, ICML, ICLR, ECML-PKDD, AISTATS, ICDM, ACM MM, etc. He won the ICDM Best Student Paper Runner-Up Award, the second prize of Natural Science Award of the Chinese Institute of Electronics, "Excellent Doctorial Dissertation Award" of Chinese Association for Artificial Intelligence, "Wu Wen-Jun AI Excellent Youth Scholar Award", and the Scientific Fund for Distinguished Young Scholars of Jiangsu Province. He was also selected as the "Global Top Chinese Young Scholars in AI" released by Baidu, and "World's Top 2% Scientists" released by Stanford University. (Email: chen.gong@njust.edu.cn)