

CAwa-NeRF: Instant Learning of Compression-Aware NeRF Features

Omnia Mahmoud, Théo Ladune and Matthieu Gendrin

Orange Innovation, France

E-mail: omnia.mohamed, theo.ladune, matthieu.gendrin@orange.com

Abstract—Modeling 3D scenes by volumetric feature grids is one of the promising directions of neural approximations to improve Neural Radiance Fields (NeRF). Instant-NGP (INGP) introduced multi-resolution hash encoding from a lookup table of trainable feature grids which enabled learning high-quality neural graphics primitives in a matter of seconds. However, this improvement came at the cost of higher storage size. In this paper, we address this challenge by introducing instant learning of compression-aware NeRF features (CAwa-NeRF), that allows exporting the zip compressed feature grids at the end of the model training with a negligible extra time overhead without changing neither the storage architecture nor the parameters used in the original INGP paper. Nonetheless, the proposed method is not limited to INGP but could also be adapted to any model. By means of extensive simulations, our proposed instant learning pipeline can achieve impressive results on different kinds of static scenes such as single object masked background scenes and real-life scenes captured in our studio. In particular, for single object masked background scenes CAwa-NeRF compresses the feature grids down to 6% (1.2 MB) of the original size without any loss in the PSNR (33 dB) or down to 2.4% (0.53 MB) with a slight virtual loss (32.31 dB).

Index Terms—Compression, Deep-Learning, Hash-Encoding, Multi-Resolution, Neural Radiance Fields, Rate-Distortion

I. INTRODUCTION

Novel view synthesis has recently caught the attention of researchers due to its huge potential use in a wide range of Virtual Reality (VR) and Augmented Reality (AR) applications. It seeks to achieve photo-realistic rendering for a 3D scene at non-captured viewpoints, given a set of images captured from multiple views with assigned camera poses. The exposure of Neural Radiance Fields (NeRF) [17] has strongly influenced 3D scenes modeling and novel-views synthesis by the use of learned Deep Neural Networks (DNN) to map each sampled 3D position at a given viewing direction, to its corresponding volume density and view-dependent color. The learning process in NeRF relied on feeding huge number of ray samples to a large-scale Multi-Layer Perceptrons (MLPs) that implicitly represents the scene. Despite the significant achievements of NeRF, it has a massive computational overhead during both training and rendering the scene. Subsequent developments to analyze and improve NeRF [30], [4], [28], [11], [20], [21], [19] have proved their ability to reconstruct numerous 3D scenes using images and corresponding camera poses. However, there is still the burden of the computational complexity and time.

In view of addressing this issue, recent advancements introduced explicit scene representation methods that uses auxiliary voxel grids to geometrically divide the scene into encoded local features, usually referred as Feature Grids. While the actual radiance field representations (density and color) are computed through the use of small-scale MLPs leading to less computational cost. These voxel-based methods have been following various data structures of explicit representations.

DVGO [23] used dense voxel grids to represent the scene, while [24] added some CUDA bindings as well as the distortion loss introduced in mip-NeRF [5] to improve their quality and training speed. In [13], the authors introduced Plenoxels which made use of sparse voxel grids with spherical harmonics to explicitly represent the scene without the use of any MLPs which optimized the training time two orders of magnitudes faster than NeRF. In [7], they presented TensoRF which is an approach to model radiance fields as a 4D tensor. TensoRF represented the scene as a 3D voxel grid with per voxel multi-channel features. It used a set of vectors and matrices to describe scene appearance and geometry along their corresponding axes which boosted the training time. Moreover, K-Planes [9] further improved the learning quality and speed for a pure PyTorch-based implementations.

On the other hand, in [16], [29] they explicitly represented the scene using 3D octree-based architectures to speed-up the training time over NeRF while preserving the same quality. Nonetheless, the authors in [18] came up with a multi-level hash encoding architecture that maps multi-resolution voxel grids into multi-level hash tables of feature grids. The use of hash tables reduced the storage size compared to other voxel grids-based methods and significantly boosted the training time of very high quality models in a matter of seconds. However, there is still the challenge of the required high storage which limits the ability of including NeRFs in some of the applications with storage constraints, particularly streaming, network, and cloud applications.

In this paper, we introduce a compression-aware counter part implementation of INGP, where we focus on learning low entropy features throughout the training. We achieve this by adopting an entropy aware training scheme with a rate-distortion trade-off, inspired by the work done in COOL-CHIC [14]. To sum up, the paper provides the following contributions:

- The introduction of an entropy-aware 3D features learning technique that could be applied to any NeRF explicit

scene representation method regardless of the architecture of the stored features.

- The per-batch measure of the entropy by directly approximating to a known mathematical probability distribution.
- The selection of two different continuous probability distributions (Laplace and Cauchy) to measure the global features entropy and evaluate the rate distortion performance for each distribution separately.
- The rate-loss trade-off adaptation (Adaptive or Hybrid) for performance enhancement according to the type of the chosen probability distribution.

Consequently, CAwa-NeRF achieves outstanding lossless results without sacrificing neither quality nor the super fast training performance of INGP for different types of static scenes compared to the state-of-the-art work listed in Section II. Precisely, for the popular lego scene the feature grids size is scaled down from 21.9 MB to 1.2 MB without loss in the PSNR (even slightly improved from 32.96 dB to 33.14 dB) or down to 0.6 MB with only 0.59 dB loss (32.37 dB). Additionally, with the suitable rate distortion trade-off further lossy compression results could be achieved.

II. COMPRESSING NEURAL FIELDS RELATED WORKS

In light of the huge storage constraints imposed by the accelerated training of explicit voxel grid representations, compression has emerged as a critical point. Several work attempted to reduce the required storage of the learned features. For instance, PlenOctrees [29] and Plenoxels [13] filtered the voxels storing only a sparse set of the most important voxels, however the size remained notably quite large compared to other models. In [26], they learned a multi-resolution tree of feature vectors which can be truncated at any depth to achieve an adaptive bit rate with PSNR penalty. In [27], the authors learned a hybrid tensor rank decomposition of the scene without neural networks to compress TensoRF [7]. However, their method achieved a near-optimal compression that sacrifices visual loss and training speed. Other work [31] compressed voxel grids of Plenoxels [13] and DVGO [23] by applying, frequency domain transformation, pruning, and quantization which reduced the model size by two orders of magnitude with nearly the same PSNR and training speed.

PlenOctrees [29], PeRFception [12], and Re:NeRF [8] all apply bit quantization of the trained local features. Where, Re:NeRF [8] and VQRF [15] have proposed post-processing compression optimizations pipelines that can achieve significant compression results with the expense of a time consuming overhead limited by the performance of the pre-trained model.

Moreover, other work [25], [22] successfully reduced the size following binary-based features learning. Such that, in [25] they proposed vector-quantized auto-decoder (VQ-AD), which adapted soft-indexing on-train quantization to achieve variable bit rates for the feature grids with drawbacks of training footprints of memory and time. On the other hand, in [22], they proposed binary radiance-field (BiRF) which is a binary-based storage-efficient radiance field representations

with longer training time compared to INGP with hybrid 2D-3D feature grids explicit representations instead of 3D only.

Finally, SHACIRA [10] managed to compress the hash-grid used for implicit representation-based models by the use of an MLP to learn the distribution of the quantized features to minimize its entropy. However, adding more MLPs to the learning model quiet affects the instant learning speed achieved by INGP.

III. PREREQUISITES

Neural radiance fields (NeRFs) learn an optimized continuous 5D function that maps a 3D point $\mathbf{x} = (x, y, z)$ sampled along the ray $\mathbf{r}(t)$ and the 2D viewing direction $\mathbf{d} = (\theta, \phi)$ to emitted view-dependent color $\mathbf{c} = (r, g, b)$, and volume density σ which models the differential probability of ray termination at that point \mathbf{x} [17]. The implicit 5D function consists of a large-scale MLP such that,

$$(\mathbf{c}, \sigma) = \text{MLP}_\Phi(\mathbf{x}, \mathbf{d}), \quad (1)$$

where Φ represents the learnable MLP weights. Afterwards for volumetric rendering, the colors \mathbf{c}_i and the densities σ_i of all the sampled points along the ray $\mathbf{r}(t) = \mathbf{o} + t\mathbf{d}$ starting at the ray origin \mathbf{o} , are accumulated to obtain the ray color

$$\hat{\mathcal{C}}(\mathbf{r}) = \sum_{i=1}^N T_i \alpha_i \mathbf{c}_i, \quad (2)$$

where T_i and α_i represent accumulated transmittance and ray termination probability respectively for the i -th sampled point and defined as follows,

$$T_i = \prod_{j=1}^i -1(1 - \alpha_j), \quad \alpha_i = 1 - \exp(-\sigma_i \delta_i), \quad (3)$$

where δ_i denotes the distance between adjacent sampled points.

To accelerate the learning, INGP relies on an occupancy grid to skip empty spaces for efficient ray sampling. To further accelerate the learning and rendering processes, a volumetric multi-resolution feature grids representation of the scene was proposed, such that local features were encoded in an explicit data structure of multi-resolution (L levels, ranges from N_{\min} to N_{\max} as in [18]) hash tables of size T and F features per entry. Each hash table l has $\theta_l \in \Theta$ set of features, where $\theta_{l,\tau,f} \in \theta_l$ denotes one feature entry. For each sampled point, the selected features $\mathbf{f} \in \Theta$ mapped by the hash encoded corners points of the sample location inside the multi-resolution voxel grids, are then linearly interpolated per level to compute \mathbf{y} as follows,

$$\mathbf{y} = \text{interp}(\mathbf{f}), \quad \text{where } \mathbf{f} = \text{enc}(\mathbf{x}; \Theta) \quad (4)$$

which is inputted to two small-scale density and colors MLPs Φ such that,

$$\sigma = \text{MLP}_{\Phi_1}(\mathbf{y}), \quad \mathbf{c} = \text{MLP}_{\Phi_2}(\sigma, \mathbf{d}) \quad (5)$$

Since the size of the implicit representation MLPs became relatively small, the set of learned explicit hash tables features Θ significantly affected the learning parameters and

constrained the required storage size. Therefore we need to constrain the entropy of the learning parameters referred as feature grids throughout the paper.

IV. METHODOLOGY

This section presents CAwa-NeRF, the proposed method for instant learning of compression-aware feature grids (latents representations) for any explicit scene representation model. Unlike other state-of-the-art compression work, CAwa-NeRF is simple, effective, could be applied to any scene representation model with stored features. However, this paper focuses on the pipeline introduced in INGP [18]. Fig 1 shows the overall proposed scheme of learning entropy constrained compression-aware feature grids. We first introduce the proposed compression-aware updates stated in Fig. 1. Furthermore, we illustrate exporting the zip compressed feature grid from the learned compression-aware feature grids as depicted in Fig. 2.

A. Compression-Aware Learning

To ensure a compression-aware learning of feature grids, the model is prepared for two separate concurrent tasks. The first task is quantization-aware learning and the second task is entropy-aware learning.

1) *Quantization-Aware Learning:* The process of quantization itself is not a differentiation friendly operation for gradients calculations during the learning process. However, following the footprints of other work [14], [6], the noise introduced by quantization could be added during the learning process to the values that needs to be quantized after training. Consequently, we prepare the system for quantization by adding uniform quantization noise on the interpolated features \mathbf{y} such that,

$$\hat{\mathbf{y}} = \mathbf{y} + \mathbf{n}, \text{ where } \mathbf{n} \sim \mathcal{U}\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right], \quad (6)$$

where Δ is the quantization step. Consequently, σ in equation (5) changes to,

$$\sigma = \text{MLP}_{\Phi_1}(\hat{\mathbf{y}}) \quad (7)$$

It should be noted that adding uniform noise on the input of the MLP does not only immunizing the system against quantization noise, but also against any sort of noise which might comes from either hash table collisions or the randomized ray batch selection process.

2) *Entropy-Aware Learning:* To achieve a compression-aware feature grid learning, the approximate feature grids rate term $R(\Theta)$ is measured. The objective of the learning process becomes minimizing the RGB rendering loss L_{rgb} as well as minimizing the amount of information needed to explicitly represent the scene, measured by the rate term $R(\Theta)$. Therefore, this is modeled by minimizing the following global loss $\mathcal{L}_{\text{global}}$ per batch of rays ($\mathbf{r} \in \mathcal{R}$),

$$\mathcal{L}_{\text{global}} = \mathcal{L}_{\text{rgb}} + \lambda R(\Theta), \quad (8)$$

where λ is the rate distortion loss trade-off that controls the compression to achieve variable bit rate. The RGB loss L_{rgb} is calculated as,

$$\mathcal{L}_{\text{rgb}} = \sum_{\mathbf{r} \in \mathcal{R}} \|\hat{C}(\mathbf{r}) - C(\mathbf{r})\|_2^2, \quad (9)$$

and the rate loss $R(\Theta)$ is calculated as follows,

$$R(\Theta) = \frac{1}{LTF} \sum_{\theta_{l,\tau,f} \in \Theta} -\log_2 \left[P(\hat{\theta}_{l,\tau,f} + \frac{\Delta}{2}) - P(\hat{\theta}_{l,\tau,f} - \frac{\Delta}{2}) \right] \quad (10)$$

Accordingly, to calculate the rate loss term $R(\Theta)$, we directly approximate the distribution of the feature grids to a probability distribution with the cumulative density function (CDF) of $P(\theta)$. Two parameters μ , and b are learned for the distribution. Afterwards, the rate is approximated by the log probability of the quantized features $\theta_{l,\tau,f}$ over one quantization interval. The quantized feature $\hat{\theta}_{l,\tau,f}$ is calculated from the learned feature $\theta_{l,\tau,f}$ Similar to equation (6).

To validate our simple approach of rate measurement, two different continuous probability distributions with a CDF $P(\theta)$ are examined. The first distribution is the Laplace distribution, which has a defined variance and is known by its compressing ability among compression related work [14]. The two learned variables of the Laplace distribution are μ and b which represent the mean and the standard-deviation of the distribution respectively.

On the other hand, the other probability distribution is the Cauchy distribution. Unlike the Laplace distribution, the Cauchy distribution does not have a defined variance. Instead, the two parameters μ , and b represent the median and the median absolute deviation (MAD) respectively. Despite that the Laplace distribution has higher tendency to reduce the entropy, the Cauchy distribution has higher weights on the distribution tails which benefits the learned model quality.

For either distribution, the parameters μ and b are initialized at the beginning of the training with μ_{init} and b_{init} respectively then learned throughout the training.

B. Exporting the compressed Feature-Grid

After the training stage, the feature grid values are thus learned with minimized entropy. Yet, they are stored in fixed floating point representation which does not provide the maximum benefit from the entropy-aware learning. Therefore, in the following we explain how we export the compressed version of the learned feature grids as depicted in Fig. 2. First we quantize the features according to the type of the distribution used for measuring the rate (whether the distribution is centered or not). Such that, if the Laplace distribution was the one used for rate measure we quantize using the non-centered (mid-rise) quantizer,

$$Q_{\text{mid-rise}}(\theta_{l,\tau,f}) = \Delta \left[\left\lceil \frac{\theta_{l,\tau,f}}{\Delta} \right\rceil - \frac{1}{2} \right] \quad (11)$$

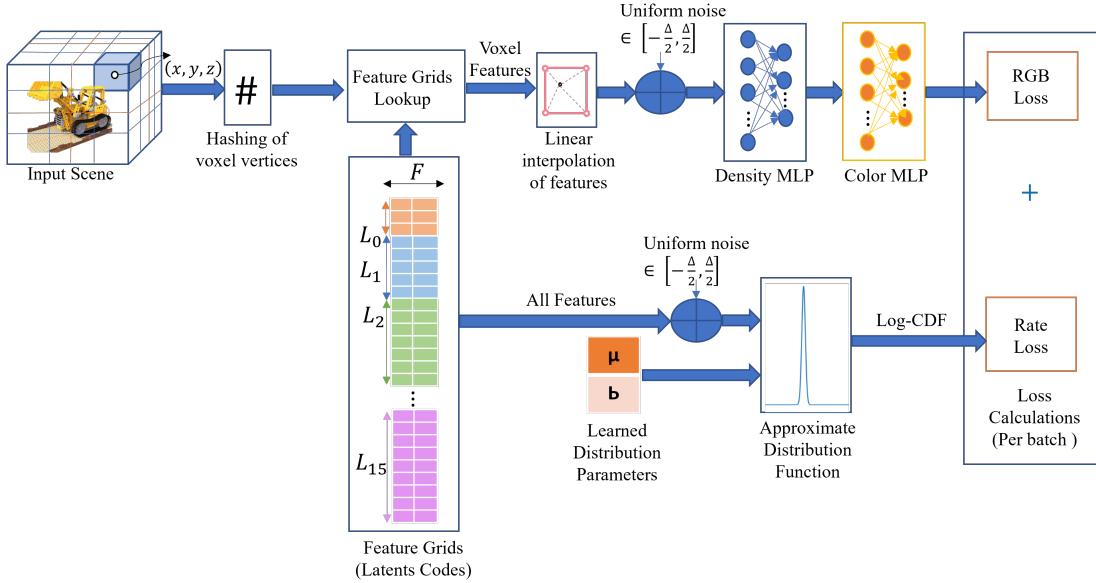


Fig. 1: Illustration of the proposed compression-aware features learning method on INGP pipeline.

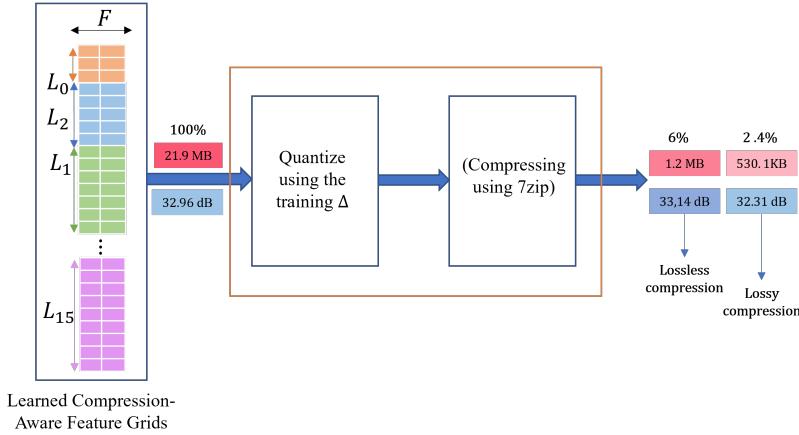


Fig. 2: Exporting the compressed features from the learned compression-aware features.

However, if the Cauchy distribution was used for rate measure, we quantize using the following centered (mid-tread) quantizer,

$$Q_{\text{mid-tread}}(\theta_{l,\tau,f}) = \Delta \left\lfloor \frac{\theta_{l,\tau,f}}{\Delta} + \frac{1}{2} \right\rfloor, \quad (12)$$

Afterwards, we apply any lossless compression tool such as 7zip [1] on the quantized feature grids.

It should be noted that the two exporting steps takes only a very few seconds while saving the model. Unlike other time consuming post-processing optimizations proposed in the related work [8], [15], the time consumed to save our compressed feature grids is negligible.

V. EXPERIMENTAL RESULTS

This section demonstrates the conducted experiments and the achieved results of CAwa-NeRF. All of our experiments are conducted using nerfstudio [2], which provides a simple

TABLE I: Parameter settings used in CAwa-NeRF experiments and results.

Parameter	Symbol	Value
Number of levels	L	16
Hash table size	T	2^{19}
Number of features per level	F	2
Coarsest (min) resolution	N_{\min}	16
Finest (max) resolution	N_{\max}	2048
Number of training iterations	-	30000
Quantization step	Δ	0.15
Initial distribution parameter 1	μ_{init}	0
Initial distribution parameter 2	b_{init}	0.01
Fixed rate loss trade-off	λ	$\{1e^{-4}, \dots, 5e^{-3}\}$
Adaptive rate loss trade-off coefficient	$\bar{\lambda}$	$\{0.1, \dots, 10\}$
Hybrid rate loss trade-off threshold (Lego)	Λ	0.0009
Hybrid rate loss trade-off threshold (Breakdance)	Λ	0.005

API for a PyTorch-based end-to-end NeRF simulator with tiny-cuda-nn [3] bindings. Table I summarizes all the simulation parameters used to obtain the results, unless stated otherwise

within the text. The common parameters between our model and INGP are set according to INGP paper. We initialize the distribution parameters μ and b for to match the small variance of the features at the beginning of the training. Since the features are initialized randomly in the range of $[-10^{-4}, 10^{-4}]$, they initially should have $\mu_{\text{init}} = 0$. With the fact that we add random noise to the features values before the rate loss calculations to model quantization, we initialize $b = 0.01$ to match the very narrow features initialization range. For the quantization step Δ , we chose the maximum quantization step that does not affect the PSNR performance when quantizing the INGP trained feature grids (CAwa-NeRF when $\lambda = 0$).

For all of the provided evaluations, we use the following two scenes to demonstrate the effectiveness of our proposed compression method CAwa-NeRF. The two scenes are chosen to generalize the validity of CAwa-NeRF for both of masked background single object scenes and real-life scenes with different levels of details whether

1. Synthetic (Lego): Consists of 100 images of 800×800 pixel² views taken from either the upper hemisphere or entire sphere around the lego object. We put the last three images aside for evaluating the model and train with the rest.

2. Real (Breakdance): Consists of 34 images of 960×540 pixel² captured with 33 cameras installed in our Orange studio (shown in Fig. 8) for a break-dancer shown in Fig. 9. We put the last three images aside for evaluating the trained model and train with the other 31 images.

Initially, CAwa-NeRF Rate distortion performance is evaluated over a wide range of fixed rate loss trade-off $\lambda \in \{-10^{-4}, \dots, 5 \times -10^{-3}\}$ for both of the considered distributions (Laplace and Cauchy). Afterwards, we propose an adaptive/hybrid rate loss trade-off and verify its effectiveness compared to fixed rate loss trade-off. Following that, we analyze the performance of the proposed method CAwa-NeRF for both of the aforementioned scenes. Last but not least, we evaluate the performance of CAwa-NeRF compared to the most relative state-of-the-art work for the synthetic lego scene.

A. Fixed rate loss trade-off

Starting by the Lego scene, we evaluate and compare CAwa-NeRF with both of the Laplace and Cauchy distributions for the range of fixed rate loss trade-off λ mentioned in Table I while the other simulation parameters are set as stated in the table. Fig. 3 depicts the rate distortion performance for fixed rate loss trade-off for the Laplace and Cauchy distributions in dark green (....) and dark pink (---) respectively. For the Laplace distribution (....), with a minimal rate loss trade-off the feature grid size is scaled down from 6.4 MB to 1.1MB with a slight loss in the PSNR performance. However, to further decrease the size, the PSNR performance aggressively degrades such that the PSNR versus size relation is no more semi-linear.

Unlike Laplace, the Cauchy distribution (---) with the same fixed rate loss trade-off λ achieves a remarkable performance that outperforms Laplace with higher rate loss trade-off values

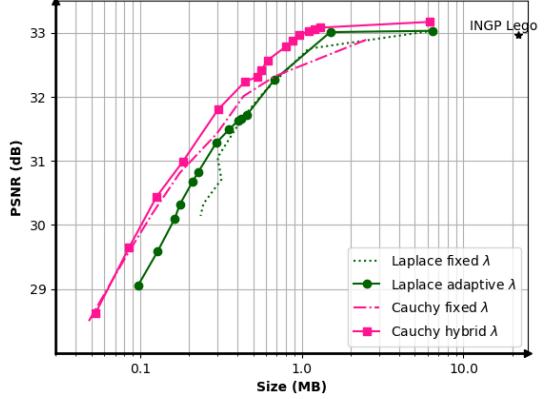


Fig. 3: Rate Distortion (PSNR vs. Feature-Grids Size) results for CAwa-NeRF for fixed and updated (adaptive/hybrid) rate loss trade-off λ for both of the studied distributions (Laplace, Cauchy).

(smaller sizes). However, fails to achieve high PSNR qualities for minimal compression similar with small λ values.

Therefore in the following subsection, the rate loss trade-off is updated to an adaptive or hybrid rate loss according to the distribution type to obtain the best rate-distortion results for both of the distributions.

B. Updated adaptive/hybrid rate loss trade-off

For the Laplace distribution (....), to overcome the aggressive loss on the PSNR especially for higher compression constrains, the rate versus PSNR trade-off λ is updated to an adaptive value ($\mathcal{L}_{\text{rgb}} \bar{\lambda}$). Such that, it changes throughout the training according to the current achieved PSNR quality. Therefore, the loss in equation (8) becomes,

$$\mathcal{L}_{\text{global}} = \mathcal{L}_{\text{rgb}}(1 + \bar{\lambda} R(\Theta)), \quad (13)$$

where, $\bar{\lambda}$ denotes the adaptive rate loss trade-off coefficient. This adaptive rate loss trade-off modifications is done to linearize the relation between the compressed feature grids size and the model quality measured by the PSNR in the semi-log plot, to insure that the entropy of the Laplace distribution is reduced hand-in-hand while improving the model quality.

On the other hand, the Cauchy distribution naturally has more weight in the distribution tails which favours the PSNR over entropy minimization. Therefor with fixed rate loss (---), it manages to provide substantial results for higher λ values (small file sizes), yet has a problem in providing higher PSNR values with small λ values (larger file sizes). Therefore, λ in equation (8) is set as in equation (14) to be only adaptive at the beginning of the training, to ensure minimal entropy from the beginning of the training and continue prioritizing the PSNR afterwards, yet still provide entropy minimized features.

$$\lambda = \begin{cases} \lfloor \lambda \rfloor \mathcal{L}_{\text{rgb}} & \text{if } \lambda < \Lambda < \mathcal{L}_{\text{rgb}} \\ \mathcal{L}_{\text{rgb}} & \text{if } \Lambda \leq \lambda < \mathcal{L}_{\text{rgb}}, \\ \lambda & \text{otherwise} \end{cases}, \quad (14)$$

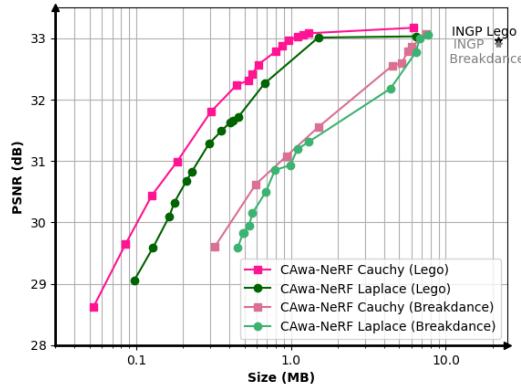


Fig. 4: Rate Distortion (PSNR vs. Feature-Grids Size) results of CAwa-NeRF for both of Laplace and Cauchy distributions with their updated rate-loss trade-off λ for the synthetic scene (Lego) and the real scene (Breakdancer).

where Λ is a threshold for the hybrid rate loss trade-off with the Cauchy distribution only.

To sum up, the Laplace distribution has higher tendency to provide minimal entropy even with the cost of aggressively distorting the quality, therefore an adaptive rate loss trade-off is always needed to maintain the best performance out of the Laplace distribution. While the Cauchy distribution has higher tendency to maintain the best model quality since it has higher weights at the tails of the distribution, therefore an adaptive rate loss trade-off is only needed at the beginning of the training afterwards it should be fixed to provide small sizes with high PSNR at the same time.

Fig. 3 depicts the clear improvements of CAwa-NeRF performance for both of, the Laplace distribution with adaptive trade-off λ (—●—) with coefficients $\bar{\lambda} \in \{0.1, 0.5, 1, 1.1, 1.2, \dots, 10\}$, as well as the Cauchy distribution with hybrid trade-off λ (—■—). As clearly shown, there is a significant improvement for both of the distribution compared to their fixed λ . The Laplace distribution with the adaptive rate loss trade-off can reduce the feature grids size from 21.9 MB (INGP) down to 1.3 MB without loss in model quality. Moreover, there is a slight improvement (33.1 dB) over the PSNR performance of the uncompressed INGP model (32.96 dB). Additionally, thanks to the adaptive rate loss trade off in equation (13), we can further compress the model without aggressively distorting the PSNR quality. Moreover, the Cauchy distribution outperforms the Laplace distribution, but with the cost of choosing the suitable threshold Λ . Such that, a compressed features file with size 1.2 MB is achieved with a PSNR improvement to 33.14 dB. On the other hand, for smaller file sizes the Cauchy distribution always outperforms Laplace even with fixed trade-off λ . It is true that theoretically the Laplace distribution is more compression friendly, however, the higher tendency of the Cauchy distribution to achieve better model qualities eventually serves compression.

Now we evaluate CAwa-NeRF performance with the two distributions and their updated trade-offs for captured real

life scenes, represented by the breakdance scene. Fig. 4 shows the achieved rate distortion results for the Breakdance scene in comparison to the synthetic lego scene, where the uncompressed models for both scenes are plotted in black (\star) and gray (\star) for the lego and breakdance scenes respectively. It could be generally noticed that, there is a clear PSNR gap between the compressed lego and breakdance scenes. However this could be explained, since the lego scene has masked background while the breakdance has more information in the background therefore more distortion loss produced by compression.

In fact for both of the distributions, the real life scene (Breakdance) is compressed from 21.9 MB (INGP) down to 6.4 MB without loss in quality, or down to 1.3 MB with 31.4 dB PSNR for Laplace or down to 1.1 MB with the same PSNR for Cauchy. It is clear that, the Cauchy distribution still outperforms Laplace for the breakdance scenes as well. However, the trade-off threshold Λ is set to 0.005. Since the breakdance scene contains more details, the threshold Λ needs to be increased compared to the lego in order to relax the rate of entropy minimization and prioritize the PSNR more.

The impact of CAwa-NeRF on the compressed features distributions versus uncompressed (INGP) is visually justified in Fig. 5. It depicts a histogram with the occurrence of the features for the quantization step $\Delta = 0.15$. We compare the shape of the uncompressed feature grids distributions of both of the synthetic Lego scene and the real Breakdance scene, with their compressed counter parts using both of the Laplace and Cauchy distributions at $\bar{\lambda} = 1$ and $\lambda = 0.0008$ respectively for almost the same PSNR with either distribution. Generally we can notice a significantly dense distribution for the real scene (breakdance) compared to the synthetic one (lego), which is present since the real scene contains 3D geometric representations of the multiple objects in the whole scene while the synthetic scene has a single object with a masked background. In particular, for the same rate loss trade-off the compressed breakdance scene has almost double the synthetic scene size for the same model quality.

It is highlighted that, the further compressed the features size, the wider the learned features values range and the narrower the distribution variance around the center unlike the uncompressed case (INGP). Which verifies that throughout the learning process, the actual features distribution is entropy minimized by constraining the entropy directly using either Laplace or Cauchy distributions. It is also depicted that, approximating the rate using the Cauchy distribution yields denser features at the center and less dense at the distribution tails which guarantees smaller sizes over using Laplace distribution for the same PSNR result.

To this end, we have been overfitting the model for a fixed number of training iterations as given in Table I. Fig. 6 demonstrates the effect of the number of training iterations on the learned model quality (in Fig. 6a) and the corresponding features size (in Fig. 6b) for both of the studied distributions and scenes at the same rate loss trade-off. It is generally noticed that, the higher the number of iterations, the lower the

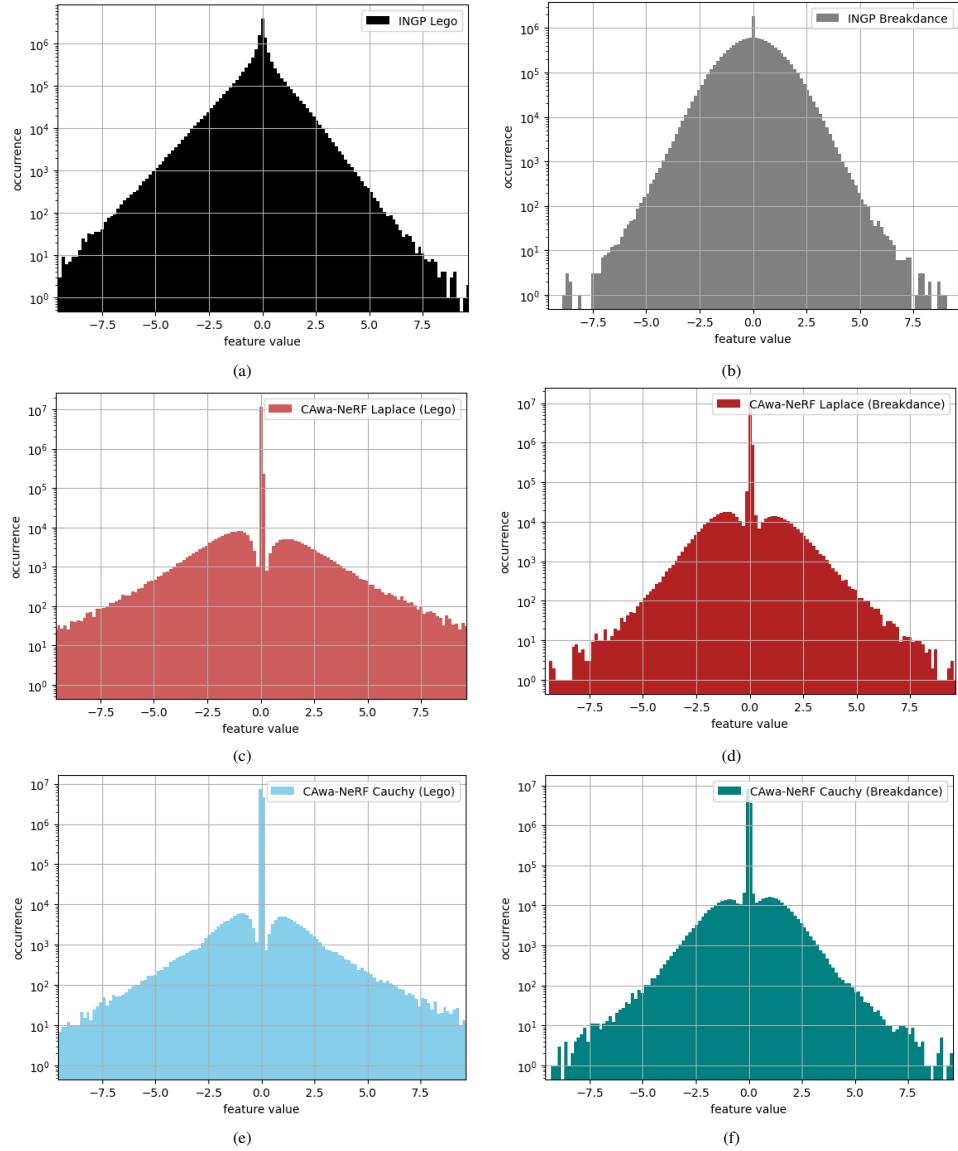


Fig. 5: Histograms of the learned multi-resolution feature grids: (a) uncompressed INGP Lego. (b) uncompressed INGP Breakdance. (c) CAwa-NeRF Laplace (Lego), $\bar{\lambda} = 1$ (PSNR=31.71dB, Size=0.45MB). (d) CAwa-NeRF Laplace (Breakdance), $\bar{\lambda} = 1$ (PSNR=31.31dB, Size=1.1MB). (e) CAwa-NeRF Cauchy (Lego), $\lambda = 0.001$ (PSNR=31.74dB, Size=0.31MB). (f) CAwa-NeRF Cauchy (Breakdance), $\lambda = 0.0008$ (PSNR=31.32dB, Size=1.1MB).

file size and the higher the PSNR quality till both saturates. It is also clearly noticeable that the single object scene (lego) saturates faster than the real scene (breakdance). Additionally, the Cauchy distribution always yields the best PSNR and size performance especially for the lego scene. For the real scene, the Laplace distribution converges faster since the Cauchy distribution doesn't have a defined variance so its more sensitive to changes in the PSNR due to compressing the details in the real scene.

To further analyze and highlight the difference between CAwa-NeRF performance for the two studied distributions with their updated rate loss trade-offs, Fig. 7a depicts the model PSNR (Lego(●), Breakdance(●)) and the com-

pressed features file size (Lego(●), Breakdance(●)) versus the adaptive rate loss trade-off coefficient $\bar{\lambda}$ for the Laplace distribution. Fig. 7b shows the PSNR (Lego(■), Breakdance(■)) and the compressed features file size (Lego(■), Breakdance(■)) versus the hybrid rate loss trade-off λ with the Cauchy distribution.

It is generally noticeable that, the higher the rate loss trade-off, the exponentially decaying the size as well as the PSNR. Moreover, the exponential decaying rate of the compressed feature grids size is larger than the PSNR decaying rate such that, for small $\bar{\lambda}$ values there is a noticeable reduction in the size without a severe loss in the PSNR. This explains that the uncompressed model (INGP) considered that all of

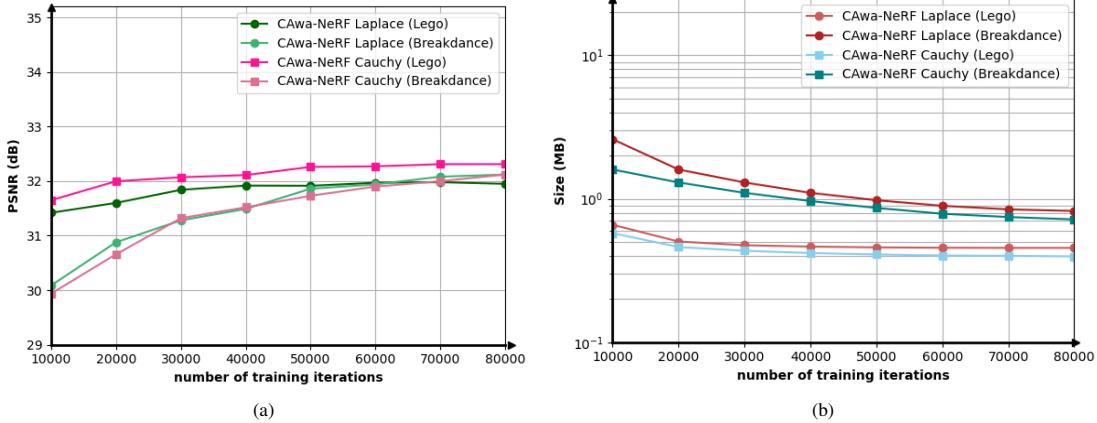


Fig. 6: CAwa-NeRF ($\bar{\lambda} = 1$) performance for different number of training iterations. (a) The model PSNR plotted in dark and light green for the synthetic scene (lego) and the real scene (breakdancer) respectively. (b) Feature grids size plotted in light and dark red for the synthetic scene (lego) and the real scene (breakdancer) respectively.

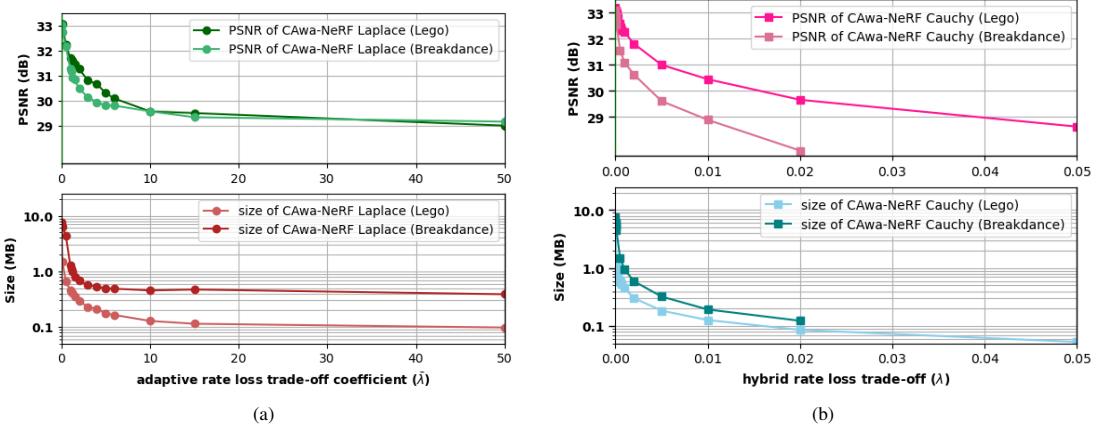


Fig. 7: CAwa-NeRF PSNR and compressed features size versus rate loss trade-off λ for both of the studied scenes and distributions.

the features have an equal share in the rendering quality while they do not. For the lego scene, we can notice that the Cauchy distribution provides a less decaying PSNR curve over Laplace. For the breakdance scene, the Cauchy distribution does not significantly improve the PSNR decaying rate, and yet it is still required to select the best threshold Λ to achieve the slightly better performance over the Laplace distribution.

VI. COMPARISON WITH RELATED WORK

In this section we discuss our achieved CAwa-NeRF compression results compared to the related work mentioned in Section II. To sum up, the previous conducted work to compress NeRF models of explicit scene features representations could be divided into three types. The first one focuses on applying some post-processing steps such as but not limited to, voxel pruning, quantization, and fine tuning on the learned features. These models cannot be directly compared to CAwa-NeRF since we focus on learning compressed features instantly during the training. The second type adapts binary-based learning methods. Such as the proposed auto encoder

in [25], where each features entry is learned directly in n bits representation by proposing a soft indexing method for quantization (VQ-AD) to overcome the non-differentiable nature of vector quantization. Another binary-based method is (BiRF) introduced in [22] where they binarize each feature entry for the hash table and they add extra hash table for 2D features with size T_{2D} as well as the standard 3D features hash table with size T_{3D} . Finally, entropy minimization based methods which includes SHACIRA [10], where the authors relied into integer quantization as well as minimizing the features entropy learned using an MLP. Table II summarizes our rate distortion achieved results for the synthetic scene compared to both of VQ-AD and BiRF. It is clearly noticeable that CAwa-NeRF is a strong competitor against both of VQ-AD and BiRF especially that it could be integrated to any model not only the hash table based INGP one without a significant increase in the training time. Finally, it still outperforms SHACIRA especially with the ability of reducing the size without any loss in the model PSNR with a much simpler and effective

TABLE II: Comparison between compression related work and CAwa-NeRF in terms of the storage needed for the feature representations versus PSNR for the synthetic NeRF scene (lego). INGP is used as a base-line for the results.

Model Name	PSNR (dB)	Size (MB)
INGP	32.96	21.9
VQ-AD (6 bits) [25]	30.76	0.49
VQ-AD (4 bits) [25]	30.09	0.33
BiRF ($T_{3D} = 2^{21}$) [22]	31.61	1.94
BiRF ($T_{3D} = 2^{19}$) [22]	31.53	0.72
BiRF ($T_{3D} = 2^{17}$) [22]	30.66	0.27
SHACIRA [10]	32.21	1.39
CAwa-NeRF Laplace ($\bar{\lambda} = 0.2$)	32.95	1.3
CAwa-NeRF Laplace ($\bar{\lambda} = 0.5$)	32.37	0.63
CAwa-NeRF Laplace ($\bar{\lambda} = 1$)	31.71	0.45
CAwa-NeRF Laplace ($\bar{\lambda} = 4$)	30.82	0.22
CAwa-NeRF Cauchy ($\lambda = 0.00008, \Lambda = 0.0009$)	33.24	1.3
CAwa-NeRF Cauchy ($\lambda = 0.0007, \Lambda = 0.0009$)	32.31	0.53
CAwa-NeRF Cauchy ($\lambda = 0.002, \Lambda = 0.0009$)	31.74	0.31
CAwa-NeRF Cauchy ($\lambda = 0.005, \Lambda = 0.0009$)	31.01	0.19

method of entropy constraining. It is worth mentioning that CAwa-NeRF takes only around twice the time needed for training the uncompressed (INGP) model for both of the studied distributions (19 minutes instead of 10 minutes, for 30K iterations).

VII. CONCLUSION AND FUTURE WORK

This paper has proposed an instant learning technique for explicit feature representation NeRF models called compression-aware NeRF (CAwa-NeRF). This is achieved by directly approximating the learned features rate to one of the known continuous probability distributions throughout the model training. Consequently, we have been able to learn minimum entropy and maximum quality feature grids simultaneously with a rate distortion trade-off. We verified that having an adaptive/hybrid rate distortion trade-off dependent on, the model quality throughout the training and the type of the chosen probability distribution, allows further compressing the features with minimal loss in the model quality measured by the PSNR. Moreover, CAwa-NeRF have been tested for different types of scenes and achieved significant performance compared to other conducted work. CAwa-NeRF differentiates itself from other related work in that, it could be integrated to any 3D explicit scene representation model without any changes in the architecture.

Additionally, CAwa-NeRF is not time consuming unlike previous post-processing based or binary based compression related work, such that it converges to the maximum compression results with minimum entropy in just double the time needed for INGP which is still significantly fast compared to previous work. It is also worth mentioning that the achieved compressed sizes are measured with half precision PyTorch numbers in 16-bit representations per feature entry.

Despite the impressive compression results of CAwa-NeRF on the explicit feature grids representations, we further need to work on compressing the implicit MLPs weights even with

their relatively small storage footprint compared to the uncompresses feature grids. We should also examine the compression results for different bit representations per feature other than the fixed 16 bits representation in the paper. Moreover, further work to enhance CAwa-NeRF results by examining other probability models than the Laplace distribution is encouraged. Last but not least, examining how to extent CAwa-NeRF to compressing dynamic scene representations is one of our main interests.

ACKNOWLEDGEMENTS

We would like to thank our Volumetric video team members Christophe Daguet, Sylvain Kedvadec, Cedric Chedaleux, and Stephane Denis for, the numerous fruitful discussions for this project, capturing and calibrating our own testing datasets on the studio, and providing technical references needed to facilitate information collection.

REFERENCES

- [1] 7-Zip lossless compression file archiver. <http://7-zip.org/>. Accessed: 2023-07-30.
- [2] nerfstudio: a simple api of a simplified end-to-end process of creating, training, and visualizing nerfs. <http://docs.nerf.studio/en/latest/index.html>. Accessed: 2023-07-30.
- [3] Tiny cuda neural networks. <http://github.com/NVlabs/tiny-cuda-nn/blob/master/README.md>. Accessed: 2023-07-30.
- [4] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022.
- [5] Jonathan T Barron, Ben Mildenhall, Dor Verbin, Pratul P Srinivasan, and Peter Hedman. Mip-nerf 360: Unbounded anti-aliased neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5470–5479, 2022.
- [6] Thomas Bird, Johannes Ballé, Saurabh Singh, and Philip A Chou. 3d scene compression through entropy penalized neural representation functions. In *2021 Picture Coding Symposium (PCS)*, pages 1–5. IEEE, 2021.
- [7] Anpei Chen, Zexiang Xu, Andreas Geiger, Jingyi Yu, and Hao Su. Tensorf: Tensorial radiance fields. In *European Conference on Computer Vision*, pages 333–350. Springer, 2022.
- [8] Chenxi Lola Deng and Enzo Tartaglione. Compressing explicit voxel grid representations: fast nerfs become also small. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1236–1245, 2023.
- [9] Sara Fridovich-Keil, Giacomo Meanti, Frederik Rahbæk Warburg, Benjamin Recht, and Angjoo Kanazawa. K-planes: Explicit radiance fields in space, time, and appearance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12479–12488, 2023.
- [10] Sharath Girish, Abhinav Shrivastava, and Kamal Gupta. Shacira: Scalable hash-grid compression for implicit neural representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 17513–17524, 2023.
- [11] Yuan-Chen Guo, Di Kang, Linchao Bao, Yu He, and Song-Hai Zhang. Nerfren: Neural radiance fields with reflections. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18409–18418, 2022.
- [12] Yoonwoo Jeong, Seungjoo Shin, Junha Lee, Chris Choy, Anima Anandkumar, Minsu Cho, and Jaesik Park. Perception: Perception using radiance fields. *Advances in Neural Information Processing Systems*, 35:26105–26121, 2022.
- [13] Ratish Jha, Sakshi Seth, and Likitha Karnati. Plenoxels: Radiance fields without neural networks. 2021.
- [14] Théo Ladune, Pierrick Philippe, Félix Henry, and Gordon Clare. Cool-chic: Coordinate-based low complexity hierarchical image codec. *arXiv preprint arXiv:2212.05458*, 2022.

- [15] Lingzhi Li, Zhen Shen, Zhongshu Wang, Li Shen, and Liefeng Bo. Compressing volumetric radiance fields to 1 mb. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4222–4231, 2023.
- [16] Lingjie Liu, Jiatao Gu, Kyaw Zaw Lin, Tat-Seng Chua, and Christian Theobalt. Neural sparse voxel fields. *Advances in Neural Information Processing Systems*, 33:15651–15663, 2020.
- [17] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.
- [18] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4), 7 2022.
- [19] Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5480–5490, 2022.
- [20] Keunhong Park, Utkarsh Sinha, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Steven M Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5865–5874, 2021.
- [21] Keunhong Park, Utkarsh Sinha, Peter Hedman, Jonathan T Barron, Sofien Bouaziz, Dan B Goldman, Ricardo Martin-Brualla, and Steven M Seitz. Hypernerf: A higher-dimensional representation for topologically varying neural radiance fields. *arXiv preprint arXiv:2106.13228*, 2021.
- [22] Seungjoo Shin and Jaesik Park. Binary radiance fields. *arXiv preprint arXiv:2306.07581*, 2023.
- [23] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5459–5469, 2022.
- [24] Cheng Sun, Min Sun, and Hwann-Tzong Chen. Improved direct voxel grid optimization for radiance fields reconstruction. *arXiv preprint arXiv:2206.05085*, 2022.
- [25] Towaki Takikawa, Alex Evans, Jonathan Tremblay, Thomas Müller, Morgan McGuire, Alec Jacobson, and Sanja Fidler. Variable bitrate neural fields. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–9, 2022.
- [26] Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. Neural geometric level of detail: Real-time rendering with implicit 3d shapes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11358–11367, 2021.
- [27] Jiaxiang Tang, Xiaokang Chen, Jingbo Wang, and Gang Zeng. Compressible-composable nerf via rank-residual decomposition. *Advances in Neural Information Processing Systems*, 35:14798–14809, 2022.
- [28] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T Barron, and Pratul P Srinivasan. Ref-nerf: Structured view-dependent appearance for neural radiance fields. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5481–5490. IEEE, 2022.
- [29] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenocubes for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5752–5761, 2021.
- [30] Kai Zhang, Gernot Riegler, Noah Snavely, and Vladlen Koltun. Nerf++: Analyzing and improving neural radiance fields. *arXiv preprint arXiv:2010.07492*, 2020.
- [31] Tianli Zhao, Jiayuan Chen, Cong Leng, and Jian Cheng. Tinynerf: Towards 100 x compression of voxel radiance fields. *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(3):3588–3596, 06 2023.

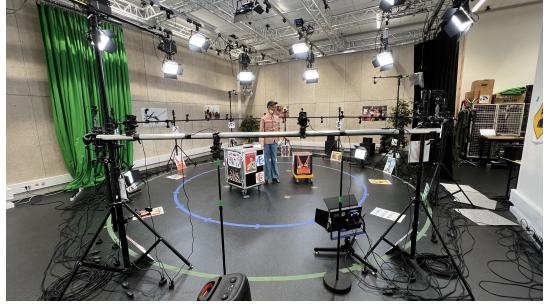


Fig. 8: The Camera Setup used to capture the images of the real scene (Breakdance).



Fig. 9: One camera shot for the Breakdance scene used in the conducted experiments.