



HAL
open science

N-Dimensional Gaussians for Fitting of High Dimensional Functions

Stavros Diolatzis, Tobias Zirr, Alexander Kuznetsov, Georgios Kopanas,
Anton Kaplanyan

► **To cite this version:**

Stavros Diolatzis, Tobias Zirr, Alexander Kuznetsov, Georgios Kopanas, Anton Kaplanyan. N-Dimensional Gaussians for Fitting of High Dimensional Functions. SIGGRAPH 2024 Special Interest Group on Computer Graphics and Interactive Techniques Conference, Jul 2024, Denver (Colorado), United States. pp.1-11, 10.1145/3641519.3657502 . hal-04723115

HAL Id: hal-04723115

<https://hal.science/hal-04723115v1>

Submitted on 6 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

N-Dimensional Gaussians for Fitting of High Dimensional Functions

Stavros Diolatzis
stavros.diolatzis@intel.com
Intel Labs
France

Tobias Zirr
tobias.zirr@intel.com
Intel Labs
Germany

Alexander Kuznetsov
alexander.kuznetsov@intel.com
Intel Labs
USA

Georgios Kopanas
georgios.kopanas@inria.fr
Inria, Université Côte d'Azur
France

Anton Kaplanyan
anton.kaplanyan@intel.com
Intel Labs
USA

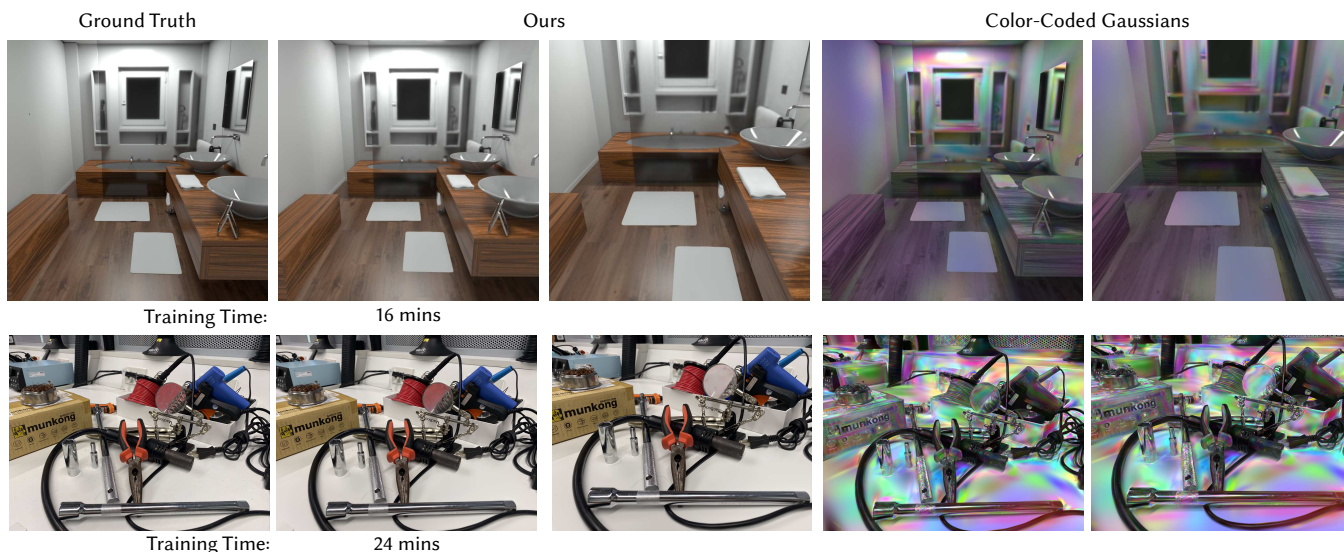


Figure 1: Our method optimizes N-Dimensional Gaussians to approximate high dimensional anisotropic functions in a few minutes. Our parameterization, culling and optimization-controlled refinement allows us to quickly estimate Gaussian parameters to represent surface (top) and volume radiance fields (bottom). Top: A synthetic scene with given geometry and material information creates a 10 dimensional input space. While the Gaussians live on surfaces they can represent rough and specular reflections by moving faithfully with viewpoint. Bottom: Similarly, 6D Gaussians of position and direction correctly reconstruct the appearance through the magnifying glass while being consistent with camera movement. We recommend watching the supplementary video to observe the movement of the complex effects.

ABSTRACT

In the wake of many new ML-inspired approaches for reconstructing and representing high-quality 3D content, recent hybrid and explicitly learned representations exhibit promising performance and quality characteristics. However, their scaling to higher dimensions is challenging, e.g. when accounting for dynamic content with respect to additional parameters such as material properties,

illumination, or time. In this paper, we tackle these challenges for an explicit representations based on Gaussian mixture models. With our solutions, we arrive at efficient fitting of compact N-dimensional Gaussian mixtures and enable efficient evaluation at render time: For fast fitting and evaluation, we introduce a high-dimensional culling scheme that efficiently bounds N-D Gaussians, inspired by Locality Sensitive Hashing. For adaptive refinement yet compact representation, we introduce a loss-adaptive density control scheme that incrementally guides the use of additional capacity towards missing details. With these tools we can for the first time represent complex appearance that depends on many input dimensions beyond position or viewing angle within a compact, explicit representation optimized in minutes and rendered in milliseconds.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).
SIGGRAPH Conference Papers '24, July 27-August 1, 2024, Denver, CO, USA
© 2024 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0525-0/24/07.
<https://doi.org/10.1145/3641519.3657502>

CCS CONCEPTS

• **Computing methodologies** → *Reflectance modeling*; **Computer graphics**.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:

Stavros Diolatzis, Tobias Zirr, Alexander Kuznetsov, Georgios Kopanas, and Anton Kaplanyan. 2024. N-Dimensional Gaussians for Fitting of High Dimensional Functions. In *Special Interest Group on Computer Graphics and Interactive Techniques Conference Conference Papers '24 (SIGGRAPH Conference Papers '24), July 27-August 1, 2024, Denver, CO, USA*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3641519.3657502>

1 INTRODUCTION

Computer graphics has recently seen an influx of new optimization-based approaches for reconstructing and representing high-quality graphics content such as complex geometry, lightfields, or high-fidelity avatars from multiview images. While inspired by machine learning, where neural networks are often used to handle high-dimensional data, many of these techniques are optimized for lower-dimensional domains with well-understood frequency content (e.g., spatial or separable spatio-angular domains). In fact, purely ML-derived approaches, such as implicit neural representations, where all information is implicitly stored in network weights, are quickly outperformed by hybrid or explicit representations. In these representations learned parameters are directly associated with certain input dimensions of the overall queried distribution. Thus, explicit disentanglement of certain query inputs such as spatial coordinates can be used to reduce the data that needs to be accessed per inference or query. We observe that this trend towards more explicit representations has led to a limiting of their usefulness for practical computer graphics problems that require handling higher input dimensionality, e.g. to handle high-quality lighting under changing conditions, paired with diverse material properties.

In this paper, we recover this ability to scale to higher numbers of input dimensions, for an explicit representation based on N -dimensional Gaussian mixtures (GMMs). Efficient splatting of GMMs for lower-dimensional radiance fields [Kerbl et al. 2023] recently revived interest in such representations. We evaluate two relevant applications, shading synthetic scenes with high appearance variability, and capturing scenes from images with strong view-dependent effects, showcased in Figure 1. We tackle two particular challenges in the construction of compact explicit N -dimensional Gaussian mixtures: (1) recover training and evaluation efficiency for larger numbers of mixture components, each with high-dimensional anisotropic distributions; and (2) adaptive refinement of the mixture representation in the face of unknown domain-specific dependencies or inter-dimensional correlations. We introduce:

- An unconstrained N -dimensional adaptive Gaussian mixture representation,
- Fast fitting and evaluation based on a high-dimensional culling scheme for N -D Gaussians inspired by Locality Sensitive Hashing, and
- A domain-independent, loss-adaptive refinement scheme to arrive at compact and high-quality mixtures.

Our code implementation is available at <https://github.com/intel/ngd-fitting>.

2 RELATED WORK

We review previous work related to high-dimensional fitting, neural graphics, and rendering of learned representations. We focus on the most relevant building blocks and refer to the survey by Tewari et al. [2022] for a more extensive overview.

2.1 High-Dimensional Anisotropic Distribution Fitting

When fitting distributions varying with respect to high numbers of input parameters, correlations arising from mixed input types (e.g., spatial, angular, path-space, temporal) are hard to predict. The problem of compactly representing high-dimensional anisotropic distributions is closely related to (unsupervised) clustering [Figueiredo and Jain 2002], where the meaning of distance and similarity is reduced with increasing number of dimensions. Locality Sensitive Hashing (LSH) techniques are an inspiration for our approach to practically bound and cull our mixture components. The idea of projection onto random unit vectors [Andoni et al. 2015] for localization and bounding proved particularly useful for us, as e.g. also used to cut down on attention computations for irrelevant tokens in Diffusion Models [Kitaev et al. 2020]. To arrive at compact adaptive representations, many applications use domain knowledge about the structure of their high-dimensional data distributions, in order to derive explicit clustering criteria, such as for high-energy paths in path space for path guiding [Simon et al. 2018], or probes in radiance caching [Krivánek et al. 2006]. In our general scenario, we build on careful iterative refinement strategies such as commonly used for optimizing implicit neural representations [Tewari et al. 2022] and explicit representations [Zhao et al. 2020] alike.

2.2 Gaussian Mixtures in Rendering

Gaussians or Gaussian Mixture Models (GMMs) have been used extensively in Computer Graphics to represent signals due to their simplicity and convenient mathematical properties. Zhou et al. [2008] used isotropic Gaussians to speed up volumetric path tracing by caching volumetric radiance fields. Jakob et al. [2011] fit a hierarchical mixture of anisotropic Gaussians to approximate million of photons with significantly fewer parameters. GMMs have also been utilized more recently for learning guiding distributions during path tracing [Vorba et al. 2014] to reduce noise. Such GMMs can be optimized robustly using expectation maximization even with noisy training samples. The convenient property of GMMs is that they can be used for product importance sampling when the both the known and learned components are available as Gaussians, resulting in a Gaussian product distribution [Herholz et al. 2016] that predicts reflected light with high accuracy.

2.3 Implicit Neural Rendering

Neural networks were demonstrated to fit well to radiance fields of real [Mildenhall et al. 2021] and virtual scenes [2013] when handed scene coordinates in the right format. Networks have also been used to estimate the zero-level set of signed distance functions

for neural surface reconstruction [Wang et al. 2021]. After training, radiance or distance information is stored implicitly within the network’s weights and biases. For every query point, an inference pass is required to retrieve the predicted value. Full-image generation via neural networks [Eslami et al. 2018] conditioned on observations from a small set of viewpoints and decomposition of image-space scene information into smaller, less entangled representations [Granskog et al. 2020] for alternative renditions have also been studied for neural rendering. Diolatzis et al. [2022] demonstrate a generator network for variable radiance fields based on explicit encoding of scene states. They achieve adaptive training data sampling by a guiding scheme that identifies elusive effects such as caustics or transparency. Small fully-fused neural networks that can be trained and evaluated in real time were popularized by Mueller et al. [Müller et al. 2021] for dynamic radiance caching and surface representations. NeRF [Mildenhall et al. 2021] demonstrated that neural radiance field paired with volume ray casting can be fit to real-world view points and generate novel views with high quality. In this application the encoding [Tancik et al. 2020] and parameterization of the inputs can have significant impact on quality. Ref-NeRF [Verbin et al. 2022] replaced outgoing radiance with reflected radiance for improved reflections and view dependent effects.

2.4 Hybrid & Explicit Representations

The findings of neural radiance fields were subsequently reproduced with more efficient hybrid and explicit representations, where data structures tailored to the spatio-angular domain are optimized to store some or all of the information needed to reproduce training views and synthesize novel views. Octrees were applied to signed distance fields [Takikawa et al. 2021] and radiance fields [Yu et al. 2021], omitting any neural components while retaining the iterative training scheme for volume ray casting samples. Dense grids [Sun et al. 2022], and more memory-efficient hash grids [Müller et al. 2022] jointly optimized with small networks were shown effective and led to hybrid representations outperforming pure neural representations. In generative adversarial networks for objects such as faces, triplane encodings [Chan et al. 2022] proved effective for many types of geometry despite their 2D nature. As Radial Basis Functions [Carr et al. 2001; Dinh et al. 2002] have been shown in the past to be an efficient representation for 3D reconstruction, Neural Radial Basis functions were proposed in [Chen et al. 2023]. They offered increased spatial adaptivity and outperformed grid-based approaches but extension to N-D would require a unified initialization and refinement scheme. Our approach is inspired by the recent 3D Gaussian Splatting (3DGS) method [Kerbl et al. 2023] which demonstrates both impressive quality and performance for novel view synthesis of real world scenes using lower-dimensional GMMs, particularly due to their effective empty space skipping by way of direct projection to the image plane, paired with an efficient GPU rasterization implementation. Such 3D Gaussian mixtures also can be turned into meshes as in [Guédon and Lepetit 2023] for improved editability.

3 METHOD

In the following we describe the key aspects of our method, which allow us to efficiently arrive at a compact explicit representation of density functions that are defined on high-dimensional input domains. For our optimized target representation, we use Gaussian mixture models which fit well to highly anisotropic distributions of high-dimensional data. Specifically, we optimize the shape and position of n N-Dimensional Gaussian components jointly with learnable color values c and an application-dependent parameter a for brightness or opacity.

A key ingredient to efficient training is our culling step that omits any computations relating to Gaussians that are irrelevant to any current query point, by way of a Locality Sensitive Hashing-inspired approach that bounds their projections onto random unit vectors. For adaptive refinement of our optimized representation, previously proposed schemes with explicit splitting and merging heuristics become difficult to tune to the sparse distributions of high-dimensional reference data. We introduce an optimizer-controlled refinement scheme, which smoothly fades in additional dependent Gaussians to enhance quality and add details where necessary.

3.1 N-Dimensional Gaussian Parameterization

We parameterize each of our N -Dimensional Gaussian components through its mean $\mathbf{m} : \mathbb{R}^N$ and its full covariance $\mathbf{V} : \mathbb{R}^{N^2}$:

$$G_{\mathbf{V}}(\mathbf{x} - \mathbf{m}) = e^{-\frac{1}{2}(\mathbf{x} - \mathbf{m})^T \mathbf{V}^{-1}(\mathbf{x} - \mathbf{m})} \quad (1)$$

The scale and rotation parameterization used in previous work [Kerbl et al. 2023] is problematic when we operate in N dimensions, as rotations become increasingly difficult to describe. Instead, we opt for a Cholesky decomposition of \mathbf{V} :

$$\mathbf{V} = \mathbf{L}\mathbf{L}^T \quad (2)$$

where \mathbf{L} is a lower triangular matrix which we optimize directly. We use an exponential activation function for the diagonal elements $\mathbf{L}[i, i](x) = e^x$ to ensure positive-definiteness and stabilize training. For the lower triangular elements we restrict them to be in the -1 to 1 range using a sigmoid $\mathbf{L}[i, j](x) = 2 \cdot \text{sigmoid}(x) - 1 : i < j$. Note that up to unitary transformations, the matrix \mathbf{L} uniquely transforms a standard normal distribution $G_{\mathbf{I}}$ to a fitted Gaussian component $G_{\mathbf{V}=\mathbf{L}\mathbf{L}^T}$, which makes the decomposition well-suited for hierarchical transformations of Gaussians as used in our smooth optimization-controlled refinement.

3.2 N-Dimensional Gaussians Culling

In order to achieve acceptable performance during training and inference, we must avoid evaluating all the N -D Gaussians for each query point. If all $N-3$ inputs except world position are constant for a single image, culling is straight-forward as global conditioning to the current values $\mathbf{x}_c = x_4, \dots, x_N$ followed by standard spatial culling as in Kerbl et al. [2023]. In the general case that we aim at, however, inputs may vary between pixels (e.g. material-dependent fitting such as for surface roughness or refraction indices), requiring effective high-dimensional culling before costly projection and evaluation per pixel.

Our approach to culling is inspired by Locality Sensitive Hashing, which is sometimes used for high-dimensional approximate nearest

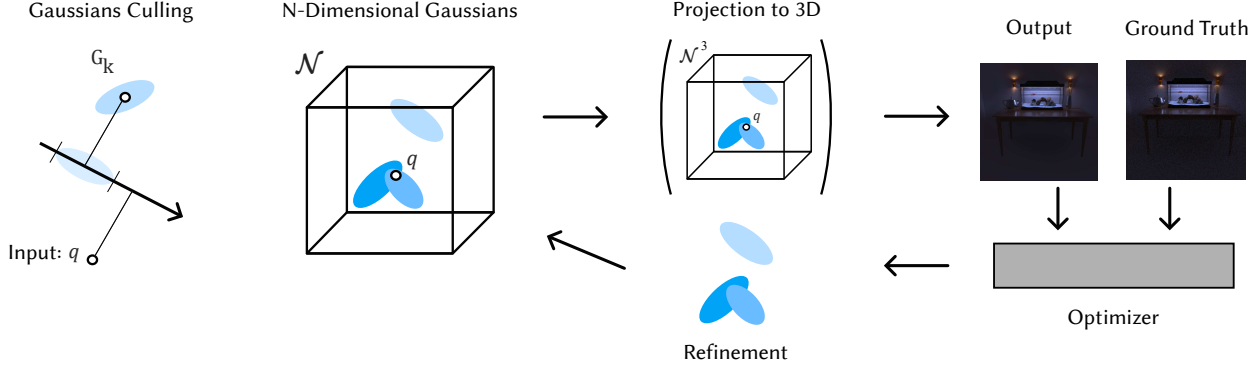


Figure 2: Our optimization receives a number of query points q of N dimensionality as input. For these given points we estimate which Gaussians can be discarded safely through our N -Dimensional culling. With the remaining ones we evaluate for each q our Gaussian mixture either in N dimensions for surface radiance fields or by first projecting the Gaussians to 3D. Our optimization converges to high quality while it also controlling the introduction of new Gaussians via our Optimization-Controlled Refinement.

neighbor search. The intuition behind LSH is that projection onto random vectors within the high-dimensional data space provide valuable information for estimating and bounding proximity. We similarly project our Gaussian components onto random unit vectors to quickly discard components for which the query point falls outside their projected footprint. Given a query point $q : \mathbb{R}^N$ and an N -D Gaussian, we project both its mean m and covariance V onto k randomly sampled vectors $r : \mathbb{R}^N$. The projections of the query point and the Gaussian mean are:

$$q_r = q^T r, \quad m_r = m^T r \quad (3)$$

To take the Gaussian’s anisotropy into account we also project the covariance V on the vector r :

$$\sigma_r^2 = r^T V r \quad (4)$$

With this projection we can safely cull any Gaussian if:

$$|q_r - m_r| < 3\sigma_r \quad (5)$$

i.e. if the query projection is outside the project 3σ -confidence interval. Doing this for all pixels would be costly, instead we do it for each tile of 16 by 16 pixels, bounding the query point by their spatial extent. In this way our culling is conservative, i.e. it will never discard a Gaussian that should have been evaluated.

3.3 Optimization-Controlled Refinement

In the following, we tackle the problem of refining our explicit Gaussian mixture representation in the face of unknown domain-specific splitting or merging heuristics (as previously proposed for e.g. spatio-angular domains [Kerbl et al. 2023; Krivánek et al. 2006; Ruppert et al. 2020]). Note that heuristics based on auxiliary statistical estimates become increasingly unreliable with increasing number of dimensions due to increased scarcity of training samples w.r.t. any respective combinations of parameter values. In order to arrive at a general, purely optimizer-controlled solution, we instead regularly introduce small amounts of additional capacity to our representation in the form of nested Gaussian components with a dependency structure that constrains their use: During training,

by default each initial or newly materialized Gaussian component contains an additional child Gaussian with negligible randomized contribution, and spanning the full parent Gaussian. In order to avoid redundant fitting of parent mixture components, the child Gaussian G_{V_c} is defined in the reference frame of its parent Gaussian G_{V_p} by relative parameters U, m_u , thus receiving all parent updates by default:

$$V_c = LU(LU)^T, \quad V_p = LL^T, \quad (6)$$

$$m_c = Lm_u + m_p. \quad (7)$$

We visualize this relationship in Figure 3. Thus, divergence of parent and child Gaussians is caused only in the need of introducing new details which cannot be represented just by the parent Gaussian. If such a case arises, meaning the opacity/brightness of the child Gaussian is increased by the optimizer crossing a threshold of non-negligible contribution, the child Gaussian is materialized. This results in both parent and child becoming independent entities and each is assigned a new default-initialized child Gaussian. Note that since we also parameterize child covariance in terms of a local lower triangular matrix U , the materialized child Gaussian is again

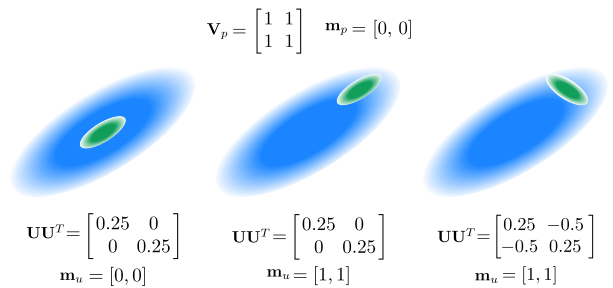


Figure 3: We visualize the relationship between parent (blue) and child (green) Gaussians for different covariance matrices and means.

parameterized by a lower triangular matrix \mathbf{LU} as the product of two such matrices preserves this property.

Practical Observations. In this scheme, enforcing a dependency between the parent and child Gaussians is an important factor. Without this relationship the child Gaussians tend to move outside the parent Gaussian and become used redundantly. To avoid this, the covariance of the parent should impact the shape and mean of the child from the start to guide the optimization into using it only to enhance the parent Gaussian.

Note that we avoid explicit subdivision in our refinement. In typical hierarchical subdivision schemes, finding the direction of the best split axis becomes increasingly complex. In our scheme, the iterative optimization is responsible for driving any added capacity of the representation into the right direction. In Figure 4 we show an example of how the refinement introduces details using dependent Gaussians.

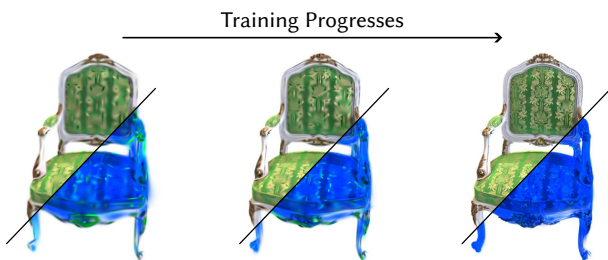


Figure 4: Our refinement scheme allows the optimizer to choose where to introduce new Gaussians. In the figure, at different parts of the training procedure, we show in blue the already existing Gaussians and in green the ones that the optimizer has chosen to utilize. Notice the dependent Gaussians adding details to the chair texture.

Practical Initialization and Refinement Scheduling. We fit our representation in phases between refinements of about 300 iterations each, such that all Gaussians stabilize towards the end of each phase before any existing child Gaussians are materialized and new children added. In the initial phase, we expect all Gaussians to move and change a lot, therefore we omit any nested child Gaussians until the initial state is sufficiently converged. Afterwards, the first set of child Gaussians are introduced and quickly start introducing some of the finer-grained effects and details. In order to determine which child Gaussians should be materialized after each phase, we check their brightness resp. opacity (depending on the application) to see if they are being utilized, in our experiments we use a threshold $t = 0.1$ for opacity and $t = 0.01$ for brightness. Different values of this threshold control how easily we want new Gaussians to be introduced versus forcing the optimizer to make do with fewer. Note that when we introduce new child Gaussians, the overall values represented by the mixture do not change significantly, since we avoid any explicit splitting, subdivision, or other perturbation as used in previous work [Kerbl et al. 2023]. This avoids spiking in the loss and domain-specific tuning of heuristics. The contribution

threshold t is directly related to the expected output and thus an intuitive hyperparameter overall.

4 APPLICATIONS

Our method offers a powerful representation for fitting high dimensional functions with strong anisotropy dependencies between dimensions. Such dependencies occur between position and direction for reflections or between position and time in time varying appearance. We apply our method to two different problems, global illumination of variable synthetic scenes and novel view synthesis with an emphasis on capturing view-dependent effects. We focus on these scenarios as they typically offer sufficiently dense capture data to allow for the optimization of high-dimensional functions. We leave generalization with sparser captures for future work.

4.1 Global Illumination with Variability

Neural networks have been utilized to shade synthetic scenes conditioned on fast to render G-Buffers [Diolatzis et al. 2022; Granskog et al. 2020]. When we apply our N -Dimensional Gaussians as a replacement for the generator networks used in these methods we achieve high quality within just a few minutes of training.

In this scenario, geometry and material information are given through the G-Buffers, and each channel of these buffers provides a dimension for our Gaussians. We evaluate all Gaussians on the surfaces of the synthetic scene. More specifically the dimensionality of the Gaussians equals to the G-Buffers’ dimensions plus any additional variable dimensions. The geometry and material information include 10 dimensions: world position xyz , view direction xyz_{dir} , albedo rgb and roughness r . Any variable element of the scene at a given image has a specific state which is normalized using its variable range. For example, if an object can move from $x = -2$ to $x = 2$, when it is at $x = 0$ this variable dimension will have a value $v = 0.5$. We concatenate these variable dimensions to the ones from the G-Buffers.

Since the geometry is given through G-Buffers the Gaussians have no opacity parameter, the parameter α controls instead the brightness of their color c . In other words for a given point on a surface the color computed from k Gaussians is:

$$c(\mathbf{x}) = G_{\mathbf{v}_1}(\mathbf{x} - \mathbf{m}_1)\alpha_1c_1 + \dots + G_{\mathbf{v}_k}(\mathbf{x} - \mathbf{m}_k)\alpha_kc_k \quad (8)$$

Since we train our method on high dynamic range renderings, we use an exponential activation for the brightness α and a sigmoid for the color c . Our loss function is a relative L2 loss [Lehtinen et al. 2018], since the ground truth Monte Carlo renderings of our scenes have some residual noise. We use Mitsuba 3 [Jakob et al. 2022] and its Python bindings to render our scene ground truths and G-Buffers. To reduce the time generating high quality datasets for each scene we train on 256 by 256 images and during rendering use higher resolution 800 by 800 G-Buffers.

To achieve high performance during training and inference, we analytically compute the gradients for our trainable parameters and implement their calculations in Taichi [Hu et al. 2019] kernels.

The scenes we showcase have different amounts of variability. In the Bathroom scene we can change the camera and inspect the rough reflections of the geometry even though the Gaussians are evaluated on the surface of the transparent bathroom door. In the

Living Room scene we can change the position of the light source while changing viewpoint. Finally, the Aquarium scene offers some complex multi bounce effects through the aquarium and the water, all which are simulated simply with anisotropy. In this scene we can also control the position of the fish. We showcase these results in Figure 1, Figure 7 and in our video.

4.2 Volumetric Radiance Fields

Another application where neural networks have been used extensively is novel view synthesis, with methods like NeRF training a network to store the volumetric radiance field of a scene. One common bias in both implicit and explicit methods for novel view synthesis is the reduced representation power in the view direction component compared to the spatial one. Except for specialized methods, this leads to high-frequency reflections being reconstructed poorly or as duplicate geometry. Such approaches fail in the case of complex anisotropic reflections and leading to low-quality reconstructions.

In contrast, a 6D spatio-angular representation constructed using our method can have as much adaptivity in the angular as in the spatial domain, and anisotropic relationships between the two domains are modeled explicitly. We demonstrate our method on such challenging scenes as the ones in the dataset provided by [Wizadwongsa et al. 2021]. We show this application to demonstrate the versatility and robustness of our method. To apply our method to this scenario, we only modify the rendering step, i.e. we apply a projection from N-D to 3D to our Gaussians (details of the projection are included in supplemental) and then use the splatting process from 3DGS. Despite the quite different application of rendering real world scenes our N-D parameterization and dependent refinement scheme works well, adapting to details and giving high quality reconstruction in both diffuse and view dependent effects. We showcase the results in Figure 1, Figure 6 and in the video.

5 COMPARISONS & ABLATIONS

To evaluate the effectiveness and robustness of our method we compare with implicit, hybrid and explicit methods in our two different applications.

Specifically for the shading of variable scenes we train a Pixel Generator network (a MLP network with skip connections) that has been used in previous methods [Diolatzis et al. 2022; Granskog et al. 2020], and we use TinyCudaNN [Müller 2021] to train a hash grid and a small MLP decoder.

For novel view synthesis on scenes with hard to render anisotropic reflections we compare against 3D Gaussian Splatting for comparable training time.

Comparison for Synthetic Scenes. We train a deep neural network (8 layers of 512 features) with the Pixel Generator architecture used and an implementation of the hash grid encoding for comparison.

For the hash encoding we parameterize xyz using the hash grid, spherical harmonics for the view direction and an identity encoding for the rest of the dimensions, all of which is decoded by a 64 feature 2 layers MLP. Despite our method not using any explicit domain knowledge of the inputs to use different encodings, it achieves superior quality as showcased in Figure 7. In contrast, the shallow

MLP decoder is unable to handle the collisions of the hash grid while taking into account the extra 7 input dimensions.

The Pixel Generator is quite a big network and one that needs close to tens of hours of training to converge. For similar training times it learns a blurry representation compared to our explicit optimization (Figure 7).

We give our quantitative evaluations in Table 1 for all 3 methods and scenes. In Table 4 we also report the iteration timings for each method. Note that the Hash Grid baseline utilizes Python bindings to optimized C++ code compared to our method which uses purely Python.

Table 1: In this table we report Mean Absolute Percentage Error (MAPE) for each method in our synthetic scenes dataset.

	BATHROOM	AQUARIUM	LIVING ROOM
Hash Grid	0.161	0.180	0.128
Pixel Generator	0.415	0.168	0.178
Ours	0.127	0.153	0.126

Comparison for Specular Real World Scenes. In this application we initialize the spatial component and color of our Gaussians, similarly to 3DGS, using Structure-from-Motion (SfM) [Snavely et al. 2006] when available. Since these points can create erroneous Gaussians, after initialization we discard any that have very low opacity ($a < 0.1$). Compared to 3DGS, our method doesn't require regular discarding or resetting of opacity.

We compare against 3DGS as it is conceptually close to our method, Instant NGP as a state of the art hybrid method, and to NeX as an implicit method with a focus on specular effects. We train 3DGS and INGP for similar training times using the respective publicly available code. As NeX requires tens of hours to train we use the numbers reported in their paper and train our method in the same (lower) resolution.

As we show in Figure 6 for scenes with highly anisotropic reflections which cannot be faked by floating geometry the spherical harmonics encoding used in 3DGS is not enough. Instant NGP fares better in reproducing some specular effects, but the capacity of the small network cannot handle all the collisions efficiently, leading to blurring and fogginess. Our method learns correctly to optimize the anisotropy between world position and viewing direction which results in reflections actually moving in world space as we move the camera. In our video please appreciate that many of the reflections are parameterized by a few Gaussians (demonstrated through the color-coded Gaussians) resulting in fewer ones necessary to achieve higher quality. These observations are also reflected in PSNR in Table 2. Since our method utilizes fewer Gaussians, it also has overall shorter evaluation times. In Table 3 we show the inference times for our method and 3DGS using the Python implementation of the splatting renderer provided by the 3DGS authors. Note that 3DGS also provides a C++ renderer, part of [Bonopera et al. 2020], which provides much better performance than the Python version but our 6D parameterization could be integrated within the same framework to profit from similar performance gains.

Table 2: Quantitative evaluation of our method against 3DGS, Instant NGP and NeX for our two scenes. We show training resolution, training times and PSNR for all methods. Next metrics are provided from their paper.

Scene	CD		TOOLS	
Resolution	1600 x 900		1600 x 1200	
	PSNR		PSNR	
Ours	28.39		27.58	
3DGS	25.76		26.32	
INGP	27.23		25.70	
Resolution	1008 x 567		1008 x 756	
	PSNR	Training	PSNR	Training
Ours	30.06	21 mins	29.07	21 mins
NeX	31.43	>18 hours	28.16	>18 hours

Table 3: We report the number of Gaussians at the end of training for our method and 3DGS as well as the inference time for both methods using the Python splatting renderer of [Kerbl et al. 2023].

Scene	CD		TOOLS	
	Inference (ms)	# Gaussians	Inference (ms)	# Gaussians
Ours	24.5	$3.2 \cdot 10^5$	41.23	$2.7 \cdot 10^5$
3DGS	62.92	$10.1 \cdot 10^5$	70.57	$13.6 \cdot 10^5$

Compared to NeX we demonstrate similar PSNR, on average, within half an hour of training compared to the tens of hours required by NeX. Also note that training NeX in the same higher resolution as the previous results required more memory than the 24 GB of available GPU memory in our tests.

LSH Ablation. Our tile-based high dimensional culling using projections allows us to train and render fast our mixtures. We showcase these advantages against brute force evaluation in Table 4 for the synthetic scenes application. During training we use lower resolution images and in that case the 16 by 16 tiles cover a larger area resulting in fewer Gaussians culled. Still avoiding the back propagation step for these Gaussians saves a considerable amount of time for each training iteration.

At inference time each tile covers a small area and as a result a big percentage of the Gaussians can be discarded resulting in three times faster evaluation.

To test the impact of the chosen confidence interval on performance and image quality we experiment with different values including our choice of 3σ and report the results in Figure 5. Lower confidence values lead to more aggressive culling with lower average evaluations (more Gaussians discarded) and faster inference times. But in Figure 5 we observe that for values lower than the chosen 3σ threshold, blocky artifacts start to appear (highlighted by

Table 4: Iteration time (during training and inference) of our method (with and without LSH culling), the Pixel Generator and the Hash Grid baselines for the BATHROOM scene.

	Training		Inference	
	Time (s)	Avg. Evals	Time (ms)	Avg. Evals
Brute Force	0.194	7856.0	317	7856.0
Ours	0.091	6208.2	131	1984.89
Pixel Generator	0.397	-	210	-
Hash Grid	0.010	-	5	-

red arrows) due to the aggressive culling. This means that for these values the culling discards Gaussians which normally contributed to the image leading to artifacts.

LSH Threshold	Brute Force	6σ	3σ (Ours)	2σ	1.5σ
Inference (ms)	245	192	131	81	65
Avg. Evaluations	7856.0	4319.74	1984.89	872.16	477.31

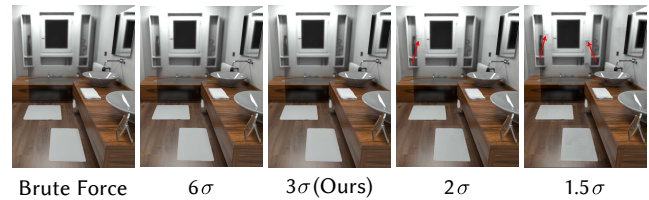


Figure 5: Demonstration of the impact to image quality and inference time for different values of LSH threshold. With red arrows we point out the artifacts that appear when the threshold is set to lower values than our choice of 3σ .

6 LIMITATIONS AND FUTURE WORK

Naturally, any representation fitting to reference data is prone to overfitting and aliasing problems, to which our method is no exception. It is well-known that this challenge increases with the number of variables affecting the results [Figueiredo and Jain 2002, Sect. 5.1.3], calling for according increases in training sample size. In our experiments fitting to real-world view points, we avoid overfitting our directionally adaptive representation to sparse view point sets using typical directional regularization [Niemeyer et al. 2022]: Smoothness of directionally varying components is enforced in-between captured view points by randomly perturbing training directions to cover the full space. Even with this regularization real-world 360 datasets are more challenging in their data distribution and capture quality, and it could require specific tuning/heuristics which we leave for future work.

Adapting our refinement scheme in the future to other covariance parameterizations, such as the one used in 3DGS could help reduce the amount of hyperparameters required. However such an adaptation is non-trivial due to additional conversions required in

the materialization (composite covariance to quaternions + scaling) and would require extra care to not degrade performance overall.

The computational scaling with respect to number of input dimensions differs for implicit neural and explicit Gaussian representations. In the plain Gaussian mixture parameterization used throughout this paper, storage cost per Gaussian component grows quadratically with the number of dimensions. More sparse parameterizations would be possible (e.g. trading full inter-dimensional adaptivity for fewer principal component axes), but are subject to future work. We note that while in contrast, neural representations may implicitly expose sub-quadratic scaling behavior, no such representations suitable for interactive rendering are currently easily extensible to high-dimensional domains.

The choice of representations and their parameterization has observable impact on results. Like in previous work [Kerbl et al. 2023], the representation of discontinuities by Gaussian mixture results in somewhat scale-dependent sharpness, whereas other representations such as implicit neural models using ReLU activations are naturally well-suited to place truly discontinuous boundaries. Studying alternative distribution primitives as well as alternative distribution parameters for the best control, convergence, and compactness of resulting representations, could be interesting future work. Capturing coherent motion of primarily visible objects is another interesting open issue for follow-up work. While in principle, our Gaussian components can be moved easily in the spatial domain, we note that such a manual animation would be in contrast to captured time-dependent signals: Even time-dependent evaluation of our representation is effectively a slice through a static high-dimensional Gaussian mixture, therefore our method does not yet implicitly solve correspondence-finding of coherently transformed objects. We expect interesting challenges for future high-dimensional explicit representations to reach this secondary objective.

7 CONCLUSION

We demonstrated feasibility of efficiently constructing compact explicit representations for the reconstruction of high-dimensional data in interactive computer graphics applications. In particular, we achieve similar or higher quality in our results when compared to previous methods, without leveraging domain knowledge that is restricted to lower-dimensional domains. We hope that our insights can inspire both explicit and neural implicit representations going forward to expand their scope to adaptivity in higher-dimensional input domains, either by way of improved fitting of explicit primitives or by way of derived parametric encodings.

The locality of parameter impact in explicit representations is a strong property not only for culling and efficiency purposes, but also in the sense of distributed storage capacity, providing good opportunities for an implicit adaptive scaling to larger input domains. In contrast, implicit neural representations commonly assign a certain fixed capacity to certain parts of the domain. Ultimately, however, both approaches will still require efficient large-scale culling mechanisms (e.g. using hierarchical schemes) accounting for high variability with respect to high-dimensional inputs, which is another interesting area of future work.

ACKNOWLEDGMENTS

We would like to thank the anonymous referees for their valuable comments and helpful suggestions. We also thank Laurent Belcour and Sebastian Herholz for their valuable input and suggestions. G. Kopanas was supported by ERC Advanced grant FUNGRAPH No. 788065 (<http://fungraph.inria.fr>).

REFERENCES

- Alexandr Andoni, Piotr Indyk, Thijs Laarhoven, Ilya Razenshteyn, and Ludwig Schmidt. 2015. Practical and optimal LSH for angular distance. *Advances in neural information processing systems* 28 (2015).
- Sebastien Bonopera, Peter Hedman, Jerome Esnault, Siddhant Prakash, Simon Rodriguez, Theo Thonat, Mehdi Benadel, Gaurav Chaurasia, Julien Philip, and George Drettakis. 2020. sibr: A System for Image Based Rendering. <https://sibr.gitlabpages.inria.fr/>
- Jonathan C Carr, Richard K Beatson, Jon B Cherrie, Tim J Mitchell, W Richard Fright, Bruce C McCallum, and Tim R Evans. 2001. Reconstruction and representation of 3D objects with radial basis functions. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*. 67–76.
- Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. 2022. Efficient geometry-aware 3D generative adversarial networks. In *Proc. IEEE CVPR*. 16123–16133.
- Zhang Chen, Zhong Li, Liangchen Song, Lele Chen, Jingyi Yu, Junsong Yuan, and Yi Xu. 2023. NeuRBF: A Neural Fields Representation with Adaptive Radial Basis Functions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*. 4182–4194.
- Huong Quynh Dinh, Greg Turk, and Greg Slabaugh. 2002. Reconstructing surfaces by volumetric regularization using radial basis functions. *IEEE transactions on pattern analysis and machine intelligence* 24, 10 (2002), 1358–1371.
- Stavros Diolatzis, Julien Philip, and George Drettakis. 2022. Active exploration for neural global illumination of variable scenes. *ACM Trans. Graph.* 41, 5 (2022), 1–18.
- SM Ali Eslami, Danilo Jimenez Rezende, Frederic Besse, Fabio Viola, Ari S Morcos, Marta Garnelo, Avraham Ruderman, Andrei A Rusu, Ivo Danihelka, Karol Gregor, et al. 2018. Neural scene representation and rendering. *Science* 360, 6394 (2018), 1204–1210.
- M.A.T. Figueiredo and A.K. Jain. 2002. Unsupervised learning of finite mixture models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24, 3 (2002), 381–396. <https://doi.org/10.1109/34.990138>
- Jonathan Granskog, Fabrice Rousselle, Marios Papas, and Jan Novák. 2020. Compositional neural scene representations for shading inference. *ACM Trans. Graph.* 39, 4 (2020), 135–1.
- Antoine Guédon and Vincent Lepetit. 2023. Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering. *arXiv preprint arXiv:2311.12775* (2023).
- Sebastian Herholz, Oskar Elek, Jiří Vorba, Hendrik Lensch, and Jaroslav Křivánek. 2016. Product importance sampling for light transport path guiding. In *Comp. Graph. Forum*, Vol. 35. Wiley Online Library, 67–77.
- Yuanming Hu, Tzu-Mao Li, Luke Anderson, Jonathan Ragan-Kelley, and Frédo Durand. 2019. Taichi: a language for high-performance computation on spatially sparse data structures. *ACM Trans. Graph.* 38, 6 (2019), 1–16.
- Wenzel Jakob, Christian Regg, and Wojciech Jarosz. 2011. Progressive expectation-maximization for hierarchical volumetric photon mapping. In *Comp. Graph. Forum*, Vol. 30. 1287–1297.
- Wenzel Jakob, Sébastien Speierer, Nicolas Roussel, and Delio Vicini. 2022. DrJit: A Just-In-Time Compiler for Differentiable Rendering. *Transactions on Graphics (Proceedings of SIGGRAPH)* 41, 4 (July 2022). <https://doi.org/10.1145/3528223.3530099>
- Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Trans. Graph.* 42, 4 (2023).
- Nikita Kitaev, Lukasz Kaiser, and Anselm Levskaya. 2020. Reformer: The efficient transformer. *arXiv preprint arXiv:2001.04451* (2020).
- Jaroslav Křivánek, Kadi Bouatouch, Sumanta N Pattanaik, and Jiri Zara. 2006. Making Radiance and Irradiance Caching Practical: Adaptive Caching and Neighbor Clamping. *Rendering Techniques* 2006 (2006), 127–138.
- Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. 2018. Noise2Noise: Learning image restoration without clean data. *arXiv preprint arXiv:1803.04189* (2018).
- Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. 2021. Nerf: Representing scenes as neural radiance fields for view synthesis. *Comm. ACM* 65, 1 (2021), 99–106.
- Thomas Müller. 2021. *tiny-cuda-nn*. <https://github.com/NVlabs/tiny-cuda-nn>

- Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. 2022. Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.* 41, 4 (2022), 1–15.
- Thomas Müller, Fabrice Rousselle, Jan Novák, and Alexander Keller. 2021. Real-time neural radiance caching for path tracing. *ACM Trans. Graph.* 40, 4 (2021), 1–16.
- Michael Niemeyer, Jonathan T Barron, Ben Mildenhall, Mehdi SM Sajjadi, Andreas Geiger, and Noha Radwan. 2022. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5480–5490.
- Peiran Ren, Jiaping Wang, Minmin Gong, Stephen Lin, Xin Tong, and Baining Guo. 2013. Global illumination with radiance regression functions. *ACM Trans. Graph.* 32, 4 (2013), 130–1.
- Lukas Ruppert, Sebastian Herholz, and Hendrik P. A. Lensch. 2020. Robust fitting of parallax-aware mixtures for path guiding. *ACM Trans. Graph.* 39, 4, Article 147 (aug 2020), 15 pages. <https://doi.org/10.1145/3386569.3392421>
- Florian Simon, Alisa Jung, Johannes Hanika, and Carsten Dachsbacher. 2018. Selective guided sampling with complete light transport paths. *Transactions on Graphics (Proceedings of SIGGRAPH Asia)* 37, 6 (Dec. 2018). <https://doi.org/10.1145/3272127.3275030>
- Noah Snavely, Steven M Seitz, and Richard Szeliski. 2006. Photo tourism: exploring photo collections in 3D. In *ACM siggraph 2006 papers*. 835–846.
- Cheng Sun, Min Sun, and Hwann-Tzong Chen. 2022. Direct voxel grid optimization: Super-fast convergence for radiance fields reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 5459–5469.
- Towaki Takikawa, Joey Litalien, Kangxue Yin, Karsten Kreis, Charles Loop, Derek Nowrouzezahrai, Alec Jacobson, Morgan McGuire, and Sanja Fidler. 2021. 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*. 11358–11367.
- Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. 2020. Fourier features let networks learn high frequency functions in low dimensional domains. *Advances in neural information processing systems* 33 (2020), 7537–7547.
- Ayush Tewari, Justus Thies, Ben Mildenhall, Pratul Srinivasan, Edgar Treitsch, Wang Yifan, Christoph Lassner, Vincent Sitzmann, Ricardo Martin-Brualla, Stephen Lombardi, et al. 2022. Advances in neural rendering. In *Comp. Graph. Forum*, Vol. 41. 703–735.
- Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, and Pratul P. Srinivasan. 2022. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. *CVPR* (2022).
- Jiří Vorba, Ondřej Karlík, Martin Šik, Tobias Ritschel, and Jaroslav Křivánek. 2014. On-line learning of parametric mixture models for light transport simulation. *ACM Trans. Graph.* 33, 4 (2014), 1–11.
- Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. 2021. NeuS: Learning Neural Implicit Surfaces by Volume Rendering for Multi-view Reconstruction. *Advances in Neural Information Processing Systems* 34 (2021), 27171–27183.
- Suttisak Wizadwongsa, Pakkapon Phongthawee, Jiraphon Yenphraphai, and Supasorn Suwajanakorn. 2021. NeX: Real-time View Synthesis with Neural Basis Expansion. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. 2021. Plenotrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 5752–5761.
- Shuang Zhao, Wenzel Jakob, and Tzu-Mao Li. 2020. Physics-based differentiable rendering: a comprehensive introduction. *ACM SIGGRAPH 2020 Courses* 14 (2020), 1–14.
- Kun Zhou, Zhong Ren, Stephen Lin, Hujun Bao, Baining Guo, and Heung-Yeung Shum. 2008. Real-time smoke rendering using compensated ray marching. In *ACM Trans. Graph. (SIGGRAPH)*. 1–12.

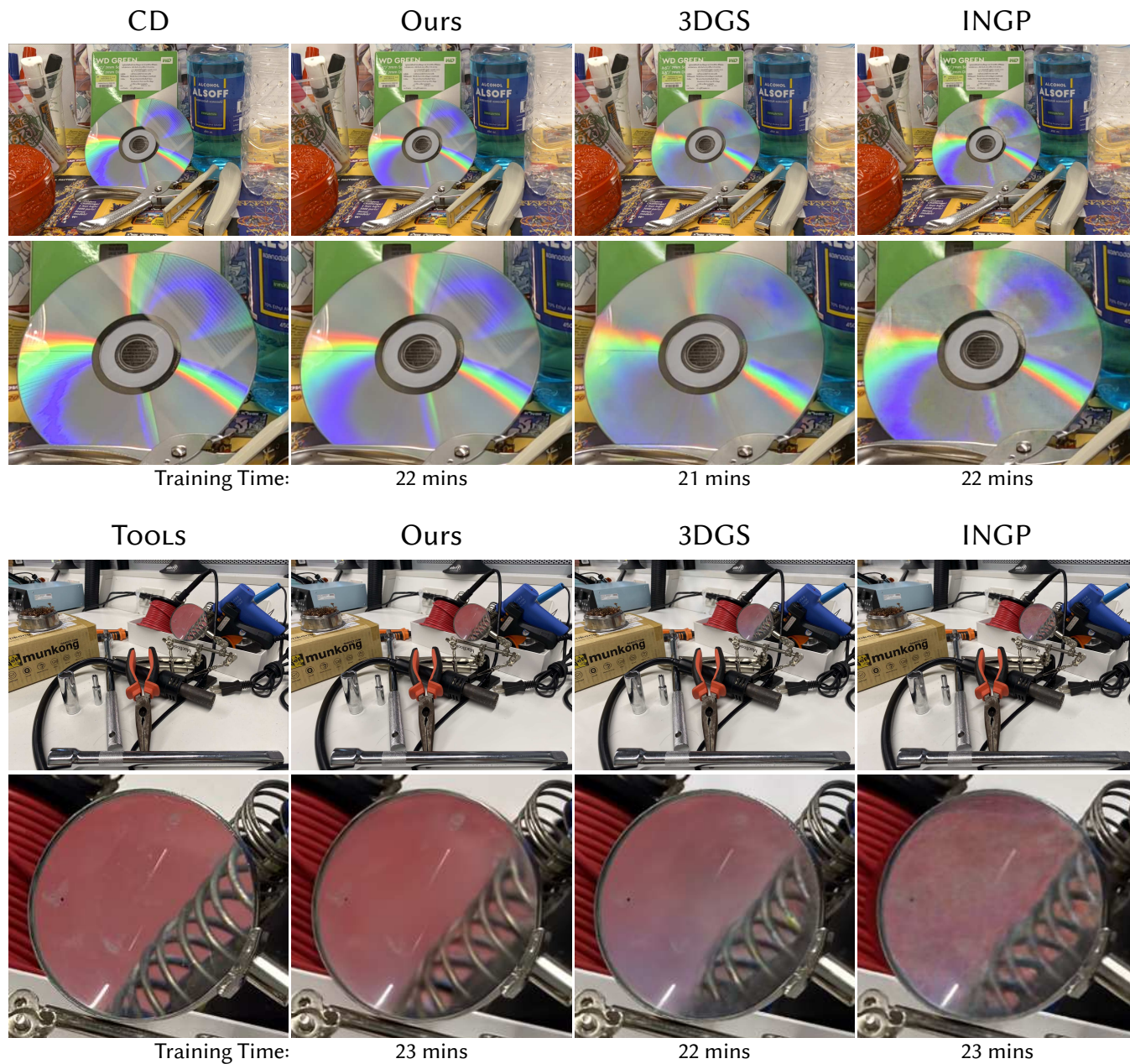


Figure 6: Qualitative results of our method compared 3DGS and Instant-NGP for two different scenes with complex view dependent effects.

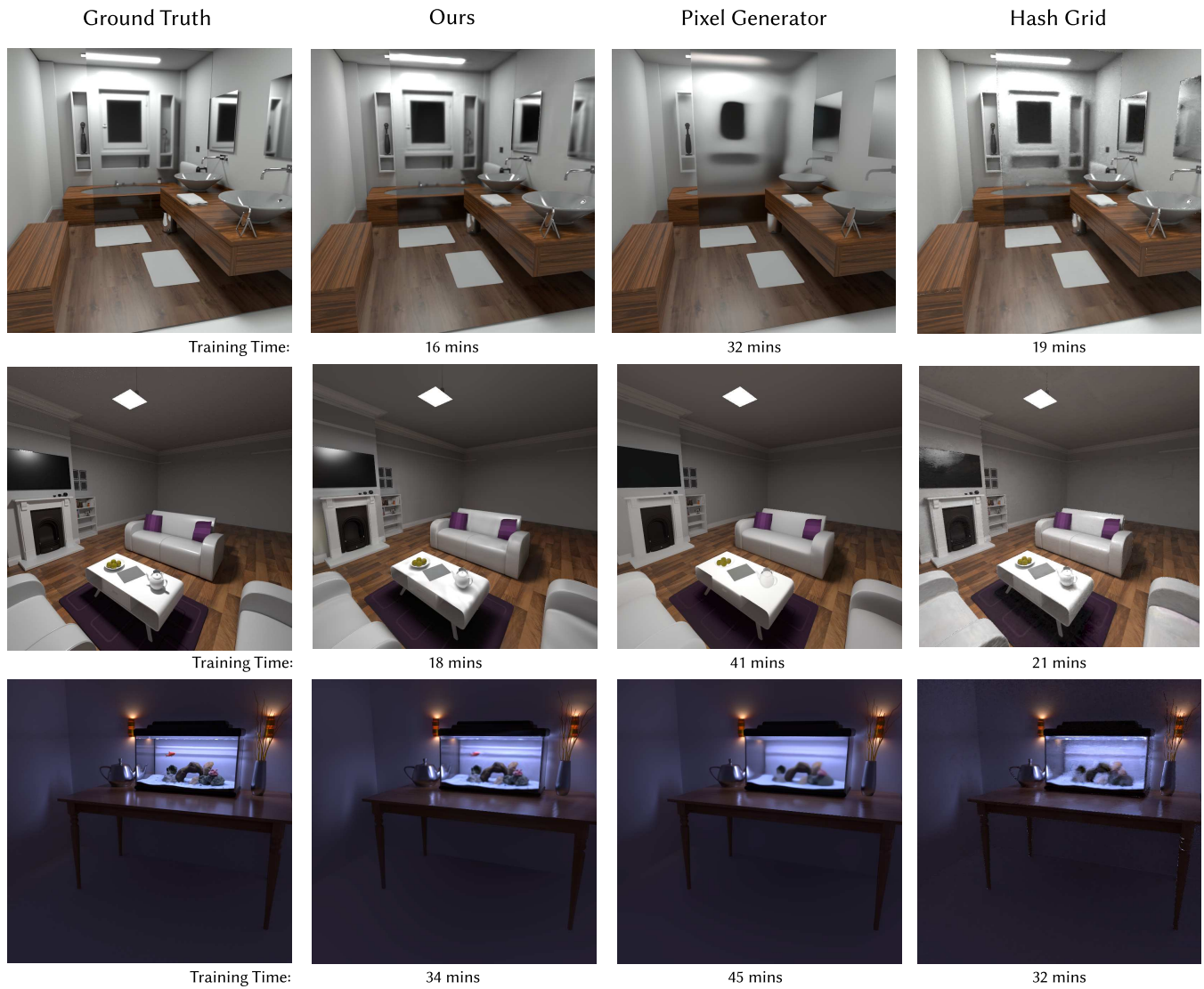


Figure 7: We compare with an implicit (Pixel Generator) and hybrid (Hash Grid) method for our application of shading variable synthetic scenes. Our method achieves high quality in a few minutes of training while enabling fast rendering and interactivity.