

POP-GS: Next Best View in 3D-Gaussian Splatting with P-Optimality

Joey Wilson
University of Michigan
wilsoniv@umich.edu

Martin Labrie
Amazon Lab 126
labrieml@amazon.com

Min Sun
Amazon Lab 126
aliensunmin@gmail.com

Marcelino Almeida
Amazon Lab 126
mmalmeid@amazon.com

Maani Ghaffari
University of Michigan
maanigj@umich.edu

Cheng-Hao Kuo
Amazon Lab 126
chkuo@amazon.com

Sachit Mahajan
Amazon Lab 126
msachit@amazon.com

Omid Ghasemalizadeh
Amazon Lab 126
ghasemal@amazon.com

Arnab Sen
Amazon Lab 126
senarnie@amazon.com

Abstract

In this paper, we present a novel algorithm for quantifying uncertainty and information gained within 3D Gaussian Splatting (3D-GS) through P-Optimality. While 3D-GS has proven to be a useful world model with high-quality rasterizations, it does not natively quantify uncertainty or information, posing a challenge for real-world applications such as 3D-GS SLAM. We propose to quantify information gain in 3D-GS by reformulating the problem through the lens of optimal experimental design, which is a classical solution widely used in literature. By restructuring information quantification of 3D-GS through optimal experimental design, we arrive at multiple solutions, of which T-Optimality and D-Optimality perform the best quantitatively and qualitatively as measured on two popular datasets. Additionally, we propose a block diagonal covariance approximation which provides a measure of correlation at the expense of a greater computation cost.

1. Introduction

In order to operate in a novel environment, robots must be capable of quantifying uncertainty in their surroundings due to occlusions and unobserved regions, as well as quantifying information gained by exploring new areas. While many prior methods for map representations were built on a metric space with natural uncertainty quantification, such representations result in discretization errors of the environment [38]. Recently, 3D-Gaussian Splatting (3D-GS) was proposed as a method for high-quality novel-view synthesis [14], which captivated the attention of the robotics community due to its explicit world model representation, with the

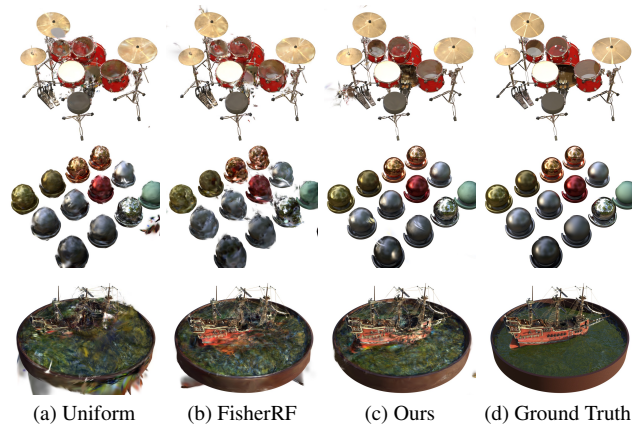


Figure 1. We propose a novel method for calculating information gain from image in 3D Gaussian Splatting. In the above images, each method is provided a set of one hundred candidate views on scenes from the Blender dataset, and selects ten views to train a 3D-GS model on. Compared to state-of-the-art, our method more accurately estimate the information value of images to train a 3D-GS model. In this figure we demonstrate block D-Optimality, however our derivation also provides multiple solutions discussed later.

potential to function as a map [6, 23, 37, 42].

Given a set of views of a scene, 3D-GS learns to render novel views from any angle through gradient descent. 3D ellipsoids are iteratively fine-tuned, which can then be rasterized to a 2D image at novel views. Several works have shown that the 3D ellipsoids can be extended to include features beyond just color, such as the category of objects the ellipsoid represents, resulting in an improved level of semantic scene understanding [30, 41]. However, 3D-GS does not natively quantify uncertainty, which leads to issues when determining whether an image has been seen before,

or when quantifying the information gained by perceiving a new view. While several works have studied this problem for neural radiance fields (NeRF) [8, 25, 26], the explicit representation of 3D-GS presents a unique challenge.

Since the conception of 3D-GS, several works have sought to quantify uncertainty directly upon a trained 3D-GS model [9, 13, 18]. Of these works, one solution which has appeared effective is relating the per-pixel gradient of the 3D-GS parameters to the information or contribution of the parameters. Notably, FisherRF [13] derived a solution for information quantification of views in 3D-GS through a diagonal approximation of Fisher Information [16]. FisherRF has since been applied to mobile manipulation [33] and active perception [21], however it does not leverage prior literature in effective functions for active perception [15, 32] or consider any correlation between parameters.

Classically, the problem of quantifying information gain from views is known as active perception in robotics literature, and has been studied extensively in probabilistic robotics applications [7, 22]. More generally, active perception can be solved through application of optimal experimental design [19], which defines solutions for identifying the most informative design parameters for an experiment in order to learn about a set of unknown parameters [10, 11, 17, 20]. One particular solution to experimental design is the use of P-Optimality, which specifies a class of solutions depending on the covariance matrix and the choice of P. P-Optimality has been successfully applied in robotics to keyframe selection, optimal loop closure, and active perception, however has not been explored in 3D-GS [3, 28, 29].

Building upon the recent work of FisherRF, which shows that a diagonal approximation of the Hessian matrix can be effective when calculating information gain in 3D-GS, we derive a general solution for the covariance matrix in 3D-GS and apply optimal experimental design techniques. Since the covariance matrix is too large to store in memory, we propose a simple diagonal approximation following FisherRF [13], as well as a block diagonal approximation which captures cross-correlation of parameters within the same ellipsoid [9]. From our general derivation, we construct and compare different p-optimality solutions, and find that D-Optimality and T-Optimality lead to significant improvements. We also find that our block diagonal approximation improves information gain quantification compared to the baselines.

To summarize, our contributions are:

1. Derive a general solution for the covariance matrix of 3D-GS, which allows application of optimal experimental design techniques.
2. Propose several novel methods for quantifying the information value of images in 3D-GS.
3. Block-diagonal approximation of 3D-GS covariance matrix which captures correlation between parameters of

the same ellipsoid.

4. Quantitative and qualitative comparison of optimal experimental design solutions on 3D-GS to each-other, and to current state-of-the-art solutions.

2. Related Work

In this section, we explore literature related to 3D Gaussian Splatting (3D-GS) and optimal experimental design for active perception. In this paper, we propose to leverage the theory of p-optimality from optimal experimental design literature to quantify information gain within 3D-GS.

2.1. 3D Gaussian Splatting

3D Gaussian Splatting (3D-GS) is a new method for novel view synthesis which models scenes through 3D ellipsoids [14], different from previous methods which model scenes implicitly [25, 34]. Each scene is modeled through thousands or millions of 3D ellipsoids where each ellipsoid contains an opacity and color modeled by spherical harmonics to capture lighting effects. 3D ellipsoids are trained through gradient descent, and at inference time are rasterized to create 2D images at candidate views through a process known as “splatting” [6].

Due to the explicit representation of scenes as 3D ellipsoids, 3D-GS has attracted a significant amount of attention from the computer vision and robotics research communities [23, 39, 42]. 3D-GS has the potential to substitute as a more expressive world model representation, and many works have explored adding additional features such as from vision-language (VL) networks to create higher levels of scene understanding [30, 41]. However, despite the name, 3D-GS does not provide a measure of uncertainty which limits applications in safety-critical or resource-constrained environments. Recently, several works have investigated uncertainty quantification and have found a promising research direction of relating Fisher Information to the explicit 3D-GS parameters [9, 13]. In particular, FisherRF developed a formulation for calculating information gain which treats the 3D-GS model as a black box, not requiring any additional training. However, FisherRF does not leverage any inter-parameter correlation when calculating the information of values to images, and does not utilize the rich literature of optimal experimental design and active perception which derive solutions for calculating information gain.

2.2. Active Perception

Active perception is a well-studied problem in robotics literature which seeks to identify the optimal path to improve a map. Since capturing new views in the real world requires robot traversal, a significant amount of research has focused on optimal solutions to determining which view-points are most valuable. One successful approach is the application

of P-optimality [15], which defines a class of optimal experimental design solutions based on functionals of the covariance matrix which vary with the choice of an integer p . P-Optimality (P-Opt.) has been widely and successfully applied to SLAM prior to the conception of 3D-GS in keyframe selection, optimal loop closure, and active perception [3, 28, 29].

Early research in optimal decision-taking for Simultaneous Localization and Mapping problems used T-optimality due to efficient computation as the trace of the covariance matrix [24, 32], eliminating the need to compute the eigenvalues. On the other hand, recent research has focused on D-optimality as a reliable metric for an Optimal Experimental Design [4, 27–29]. As advocated in [3], D-optimality seems to be a more rewarding metric towards the fulfillment of an active mapping session. The recent success of D-optimality can also be explained by its monotonicity property in active mapping scenarios [31], which guarantees that uncertainty increases monotonically as a robot moves through the scene. In this work, we propose to expand upon the formulation from FisherRF to incorporate parameter correlation and develop a more general solution which allows for application of classical optimal experimental design techniques.

3. Method

In this section, we introduce our method for quantifying uncertainty and information gain in 3D-Gaussian Splatting. First, we introduce preliminaries on the 3D Gaussian Splatting representation. Next, we describe our approximation of the covariance matrix for each ellipsoid, which provides a measure of uncertainty on the parameters. Finally, we detail our method for efficiently calculating the informational value of a candidate image.

3.1. Preliminaries: Gaussian Splatting

3D Gaussian splatting represents a scene through volumetric rendering of optimized 3D ellipsoids. The geometry of each 3D ellipsoid is parameterized by a center μ , scale S , and rotation R , while the color contribution of each ellipsoid is defined by opacity α and color c . Together, the rotation and scale define the shape of the 3D ellipsoid:

$$\Sigma = RSS^T R^T. \quad (1)$$

To render images, 3D ellipsoids are first splatted into 2D projections from a provided viewpoint, resulting in a 2D shape Σ' and location μ' . Next, the contribution α'_n of each 2D ellipsoid n to pixel x' is calculated through a kernel as:

$$\alpha'_n = \alpha_n \times \exp\left(-\frac{1}{2}(x' - \mu'_n)^T \Sigma'^{-1}_n (x' - \mu'_n)\right). \quad (2)$$

Finally, 2D ellipsoids are blended into pixels through a process known as alpha compositing, which computes the color of each pixel from a depthwise sorted list of Gaussians \mathcal{N} :

$$C = \sum_{n=1}^{\mathcal{N}} c_n \alpha'_n \prod_{j=1}^{n-1} (1 - \alpha'_j). \quad (3)$$

Parameters are optimized by comparing the rendered image from a viewpoint with the ground truth image and performing gradient descent over a weighted loss function [6, 14, 36]:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_1 + \lambda\mathcal{L}_{\text{D-SSIM}}, \quad (4)$$

where λ is a weighting function on the \mathcal{L}_1 and structural similarity loss. Once the 3D-GS model is fitted to the training data, it can render views from any perspective, however it lacks any information on when the rendering may fail. In order to quantify the amount of information the fitted Gaussian Splatting model has on each parameter, we note that the \mathcal{L}_1 loss is proportional to the partial derivative of the rendered pixel’s color with respect to the parameters of the 3D-GS model, summed over all pixels in the training set. Based on this insight, we construct our covariance matrix to capture this information.

3.2. Information Gain through Optimal Experimental Design

In order to quantify information gained through adding an image, we first require a measure of uncertainty, which is derived in this section. Approaching the problem from a maximum likelihood perspective, the maximum likelihood formulation aims to determine the solution variable θ that minimizes the pixel error $e = c - h(\theta)$ across all pixels in all frames, where $h(\cdot)$ is the 3D-GS model described in Section 3.1. Assuming that all measured pixel errors are normal zero mean, independent and identically distributed (IID), i.e., $e \sim \mathcal{N}(\mathbf{0}, \sigma_e \cdot \mathbb{I})$, then the maximum likelihood formulation seeks an optimal solution variable $\theta \in \mathbb{R}^{l \times l}$ that maximizes the likelihood function:

$$p(c|\theta) = \exp\left(-\frac{1}{2} \frac{e^T e}{\sigma_e^2}\right). \quad (5)$$

Due to the monotonicity of the log function, maximizing the function above is equivalent to minimizing the log-likelihood function:

$$-\sigma_e^2 \cdot \log p(c|\theta) = \frac{1}{2} e^T e \quad (6)$$

Assuming that we have an estimate of the solution variables θ_* , then we can expand the system’s model in the vicinity of θ_* using Taylor expansion as: $h(\theta) \approx h(\theta_*) +$

Ellipsoid 1	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,1}} \frac{dc_i}{d\theta_{1,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,1}} \frac{dc_i}{d\theta_{1,2}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,1}} \frac{dc_i}{d\theta_{2,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,1}} \frac{dc_i}{d\theta_{2,2}}$
	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,2}} \frac{dc_i}{d\theta_{1,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,2}} \frac{dc_i}{d\theta_{1,2}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,2}} \frac{dc_i}{d\theta_{2,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{1,2}} \frac{dc_i}{d\theta_{2,2}}$
Ellipsoid 2	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,1}} \frac{dc_i}{d\theta_{1,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,1}} \frac{dc_i}{d\theta_{1,2}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,1}} \frac{dc_i}{d\theta_{2,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,1}} \frac{dc_i}{d\theta_{2,2}}$
	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,2}} \frac{dc_i}{d\theta_{1,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,2}} \frac{dc_i}{d\theta_{1,2}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,2}} \frac{dc_i}{d\theta_{2,1}}$	$\sum_{i \in N} \frac{dc_i}{d\theta_{2,2}} \frac{dc_i}{d\theta_{2,2}}$
	Ellipsoid 1		Ellipsoid 2	

Figure 2. The Hessian matrix captures the information content of each parameter in the trained 3D-GS model, approximated as per-pixel gradients over the training set of images. Since a trained 3D-GS model may contain millions of parameters, we approximate the Hessian matrix through a block or main diagonal.

$J\Delta\theta$, where $\Delta\theta \triangleq \theta - \theta_*$ and $J \triangleq \frac{\partial h}{\partial \theta}|_{\theta=\theta_*}$. We can rewrite the residual function as:

$$\begin{aligned} e &\approx c - h(\theta_*) - J\Delta\theta \\ &= e_* - J\Delta\theta, \end{aligned} \quad (7)$$

where the optimal residual is defined as $e_* \triangleq c - h(\theta_*)$. Substituting this in Eq. 6, we have that:

$$\frac{1}{2}e^T e \approx \frac{1}{2}e_*^T e_* - e_*^T J\Delta\theta + \frac{1}{2}\Delta\theta^T J^T J\Delta\theta. \quad (8)$$

In order to satisfy the first order conditions for optimality in Eq. 8, it is necessary for its first order partial derivative w.r.t. the solution variables to be zero [1]. This leads to the Gauss-Newton iterative optimization equation where $\Delta\theta$ is updated as:

$$\Delta\theta = (J^T J)^{-1} J^T e_*. \quad (9)$$

Assuming that the optimal solution θ_* is unbiased, then it follows that the expected residual $\mathbb{E}[e_*] = \mathbf{0}$ and $\mathbb{E}[e_* e_*^T] = \sigma_e \cdot \mathbb{I}$, leading to $\mathbb{E}[\Delta\theta] = \mathbf{0}$ and:

$$\begin{aligned} \mathbb{E}[\Delta\theta \Delta\theta^T] &= (J^T J)^{-1} J^T \mathbb{E}[e_* e_*^T] J (J^T J)^{-1} \\ &= \sigma_e^2 (J^T J)^{-1} J^T J^T J (J^T J)^{-1} \\ &= \sigma_e^2 (J^T J)^{-1}. \end{aligned} \quad (10)$$

Without loss of generality, this work assumes¹ $\sigma_e = 1$. Defining $H \triangleq J^T J \in \mathbb{R}^{l \times l}$, the matrix H is known as an approximation of the Hessian, or the information matrix

¹Given that we've assumed that the pixel error is zero-mean IID, the value of σ_e does not matter for this application.

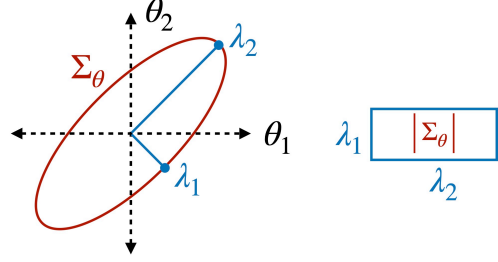


Figure 3. Optimal experimental design defines functionals of the eigenvalues of the covariance matrices, each with geometric intuitions. D-Optimality approximates the volume of the covariance matrix, as shown in this figure.

for this nonlinear optimization problem. Therefore, the covariance matrix associated with the solution variables θ is given by the inverse of the Hessian (information) matrix.

3.2.1. Uncertainty Decrease due to an Added Image

In this paper, we assume that we have a set of n images to choose from (along with their respective original poses) to determine which of the images will lead to maximal uncertainty reduction among all n candidates. Therefore, our Next-Best-View formulation attempts at maximally decreasing uncertainty of the covariance matrix Σ_i that is obtained by adding the i -th image to the model, $i \in \{1, \dots, n\}$. Note that the contents of the i -th image are not necessary for our formulation, only the pose at which we plan to take the i -th image from. This is an important feature of our solution, as it attempts to evaluate the amount of uncertainty reduction (or information increase) that can be achieved by taking an image from a new perspective without actually having the image available.

In this section, we assume that we already have an initial guess of the map θ_* and that its associated prior Jacobian J_- and Hessian $H_- = J_-^T J_-$ have been computed. As we add one new prospective candidate image i taken from a pose p_i , it is possible to compute the Jacobian associated with the new image using prior map parameters θ_* and the image's pose p_i as $J_i = \frac{\partial h}{\partial \theta}|_{\theta_*, p_i}$. Defining H_i as the Hessian of the problem as we add the i -th image, it can be computed as:

$$H_i = H_- + J_i^T J_i. \quad (11)$$

For exponential likelihoods, there is an asymptotic inverse relationship in the maximum likelihood estimator between the Hessian and covariance matrices [35]. Therefore, our goal is to maximize the information $\mathcal{I}(\cdot)$ obtained from the i -th image, which can be found by minimizing the uncertainty $U(\cdot)$:

$$\begin{aligned} \arg \max_i \mathcal{I}(H_i) &= \arg \min_i U(\Sigma_i) \\ &= \arg \min_i U(H_i^{-1}). \end{aligned} \quad (12)$$

Table 1. Properties of P-Optimality for different values of p . Note: λ_k represent the eigenvalues of the covariance matrix Σ_i .

	T-optimality	A-optimality	D-optimality	E-optimality
p	1	-1	0	$\mp\infty$
Equivalent Formulae	$\frac{1}{l}\text{tr}(\Sigma_i) = \frac{1}{l}\sum_k \lambda_k$	$(\frac{1}{l}\text{tr}(\mathbf{H}_i))^{-1} = (\frac{1}{l}\sum_k \lambda_k^{-1})^{-1}$	$\sqrt[l]{ \Sigma_i } = \exp\left(\frac{1}{l}\sum_k \log \lambda_k\right)$	\min_{λ_k} \max_{λ_k}
Meaning	Average Variance	Harmonic Mean Variance	Volume of covariance hyper-ellipsoid	Single extreme eigenvalue

In this work, we rely on the Theory of Optimal Experimental Design [15, 29], which defines the P-Optimality uncertainty metric as:

$$U_p(\Sigma_i) = \left(\frac{1}{l}\text{trace}(\Sigma_i^p)\right)^{\frac{1}{p}}, \quad (13)$$

where p is an integer. Depending on the chosen value for p , the uncertainty function can have some special properties [27], as detailed in Table 1.

3.3. Approximating the Covariance

In practice fitted 3D-GS models may contain millions of parameters, which is intractable due to the cubic computational and quadratic memory complexity of computing the eigenvalues. Therefore, we propose two approximations of the Hessian matrix to save memory and computation.

Simple Diagonal: First, following the work of FisherRF we propose a simple diagonal approximation of the covariance matrix. FisherRF derives the approximation from a Laplace approximation [5], where the covariance matrix is approximated as the main diagonal plus a small regularizing constant λ_θ :

$$\Sigma_\theta \approx \text{diag}(\Sigma_i) + \lambda_\theta, \quad (14)$$

This formulation allows for an efficient and direct comparison of our method of computing information gain with FisherRF, without consideration of correlation. Intuitively, the constant λ_θ represents prior information for our method.

Block Diagonal: In order to capture some of the correlation between 3D-GS parameters, we propose to approximate the Hessian matrix as a block diagonal matrix where each block diagonal element contains the parameters of a single ellipsoid. Please note that a block diagonal approximation has also been proposed within 3D-GS, however the approximation was applied to the task of pruning 3D-GS models to remove redundant ellipsoids [9]. Our insight is that the parameters of the same ellipsoid are most likely to be correlated, and a block diagonal matrix can be processed in parallel on a GPU for efficient computation.

When constructing the block diagonal matrix, we note that computing partial derivatives w.r.t. each pixel leads to singularity issues since the partial derivative of each color is the value α'_n for the ellipsoid. To avoid singularity issues,

we therefore compute the Hessian matrix by separately calculating partial derivatives for each channel of every pixel, instead of for every pixel for all channels. The block diagonal approximation is shown in Fig. 2.

3.4. Batch Selection

In practical applications it may be valuable to measure information gain over a set of candidate images, such as along a trajectory or in keyframe selection. Following FisherRF, we implement a simple approach which iteratively adds or removes the most optimal candidate image, updates the Hessian, and repeats the process without additional training. While simple, this batch information implementation does capture the redundancy of views, as the change in the Hessian is reflected when a candidate image is added.

4. Results

In this section, we study the effectiveness of our method on quantifying information gained from obtaining new images. Following the experimental setup of FisherRF, we quantitatively and qualitatively evaluate our method against baselines on the task of single view selection, where a 3D-GS model is trained by iteratively selecting the most informative proposal view and fitting the model to a set of candidate view. Next, we compare our method against the same baselines on batch view selection. Third, we compare the ability of each method to quantify uncertainty in novel views by studying the correlation of information gain with reconstruction metrics. Last, we perform ablation studies on the parameters of the 3D-GS model to identify the most important parameters for calculating information gain.

Baselines: We compare against the recently published FisherRF [13], which calculates information gain as:

$$\mathcal{I}(\mathbf{H}_i) = \text{tr}\left((\mathbf{J}_i^T \mathbf{J}_i) \mathbf{H}_i^{-1}\right) \quad (15)$$

with matrices modeled by the simple diagonal approximation. While FisherRF compared against a random and ActiveNeRF baseline [26], we omit these baselines as they performed worse than FisherRF in their experiments, and instead focus on comparing our results with FisherRF. We also add a uniform sampling baseline, which is slightly different from the implementation of random in FisherRF that

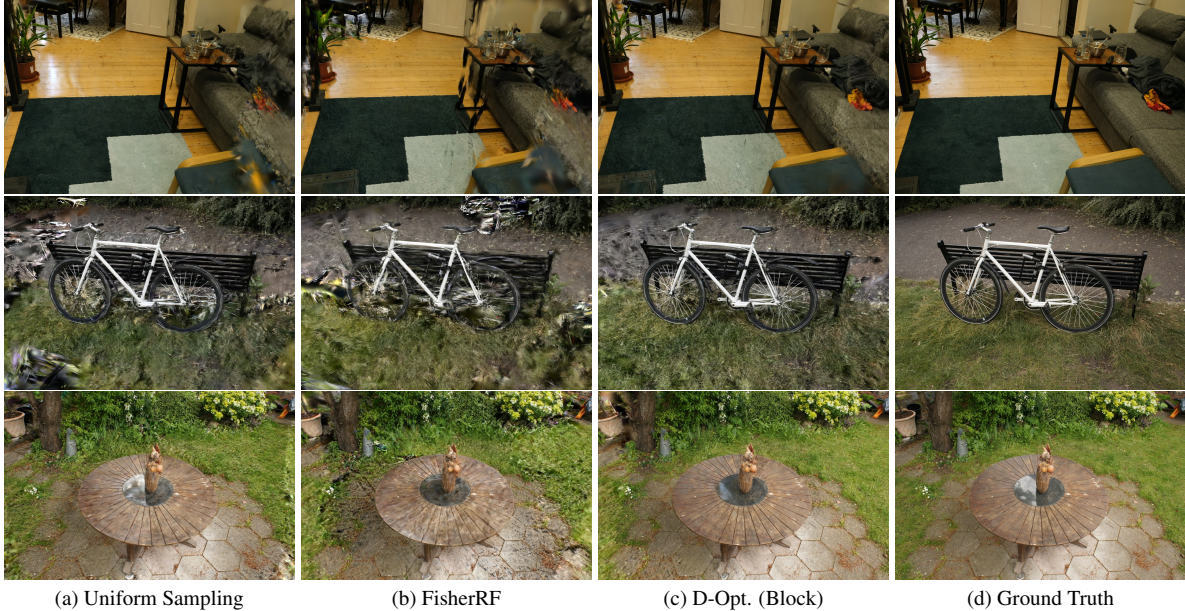


Figure 4. Comparison of view selection methods on the Mip-Nerf360 dataset with 10 views. The columns are built by different methods and in order are: uniform sampling, FisherRF, Block D-GS, and the ground truth image.

we found was incorrectly implemented. Instead, our uniform sampling baseline samples views uniformly from the training set, which results in a high coverage of the views in the test set, especially if the training and testing views are similarly distributed. As one of the merits of our approach is a more general solution with optimal experimental design, we implement A, D, E, and T Optimality baselines for comparison. All baselines are compared over the peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) [36], and LPIPS metrics [40].

Dataset: All methods are compared on two common radiance field datasets. First is the Mip-NeRF360 dataset [2], which is a real-world high-resolution dataset commonly used in novel view synthesis literature as well as by FisherRF. Mip-NeRF360 contains nine scenes, with five outdoor scenes and four indoor, which we average performance over to obtain the final results. Following FisherRF and the original 3D-GS paper, we train all models at resolutions of 1060×1600 pixels. Additionally, the prior information constant is set to a value of $\lambda_\theta = 10^{-6}$ for all models. The Mip-NeRF360 dataset contains complex scenes, where the benefit of strong view selection models is clear. However, the dataset also contains some noisy images which are distributed at random, and can impact the results.

Therefore, following FisherRF, we also evaluate all models on the Blender dataset [25], which contains eight high-fidelity objects modeled synthetically. While the scenes in this experiment are less complex, this dataset allows us to study information gain quantification without the stochasticity introduced by real-world noisy images.

Table 2. Results on Single View Selection with 10 Views on the Mip-Nerf360 Dataset.

Method	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
Uniform Sampling	17.29	0.508	0.432
FisherRF	16.81	0.493	0.445
A-Opt. (Simple)	15.55	0.452	0.480
E-Opt. (Simple)	15.33	0.436	0.488
T-Opt. (Simple)	17.91	0.520	0.420
D-Opt. (Simple)	<u>17.95</u>	<u>0.535</u>	<u>0.411</u>
D-Opt. (Block)	18.15	0.548	0.401

4.1. Single View Selection

First, we compare all methods on single view selection, where one candidate view is selected at a time. We follow the experimental setup of FisherRF with minimal modifications, including evaluating models on the ability to select both ten views and twenty views. Note that accurate information quantification is more apparent with ten views due to the limited amount of training information.

Ten Views: For the ten view setup, each method begins with 2 training views, and is trained for $100v$ iterations, where v is the number of training views. The method then selects a single candidate view to add to the training set, and repeats the process until $v = 10$, at which point the model trains until a cumulative total of 10,000 training steps. Qualitative examples on the Mip-NeRF360 dataset are shown in Fig. 4, and qualitative examples on the Blender dataset are shown in Fig. 1.

First, we compare models on the Mip-NeRF360 dataset, whose performance metrics can be found in Table 2. In this

Table 3. Results on Single View Selection with 10 Views on the Blender Dataset.

Method	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
Uniform Sampling	23.32	0.885	0.101
FisherRF	24.59	0.897	0.091
A-Opt. (Simple)	22.39	0.876	0.116
E-Opt. (Simple)	21.40	0.862	0.129
T-Opt. (Simple)	25.40	0.908	0.080
D-Opt. (Simple)	25.52	0.909	0.078
D-Opt. (Block)	<u>25.41</u>	<u>0.908</u>	<u>0.078</u>

experiment, FisherRF is slightly outperformed by the uniform sampling method. We would like to note that uniform sampling is actually an effective method for this experimental setup, and will achieve a near optimal performance if the test and train set are similarly distributed. By reframing the approach to information gain within 3D-GS as optimal experimental design, we find that both T-Optimality and D-Optimality approaches improve significantly over the FisherRF and uniform sampling baselines. Additionally, the block diagonal approximation significantly improves the structural quality measured by SSIM and LPIPS metrics.

Next, we repeat the same set of experiments on the synthetic dataset, shown in Table 3. Here, we find that FisherRF performs significantly better than the uniform sampling method, which may be due to the less complex synthetic scenes, where FisherRF is able to more accurately quantify information on a single object. Similar to before, we find that both T-Optimality and D-Optimality methods outperform the baselines by a wide margin. Across all three metrics, simple and block D-Optimality perform similarly, which we expect is due to the saturation of performance as the approaches are near optimal.

Twenty Views: The twenty view experimental setup follows a very similar approach, of iteratively adding views and training for $100v$ iterations before selecting the next view. However, this setup begins with $v = 4$ training views, adds images until $v = 20$, and trains until a cumulative total of 21,000 training steps. Experimental results on the MIP dataset can be found in Table 4. Similar to the ten view experiment, we find that uniform sampling slightly outper-

Table 4. Results on Single View Selection with 20 Views on the Mip-Nerf360 Dataset.

Method	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
Uniform Sampling	20.86	0.616	0.408
FisherRF	20.89	0.608	0.416
A-Opt. (Simple)	18.62	0.558	0.452
E-Opt. (Simple)	19.57	0.580	0.433
T-Opt. (Simple)	21.07	0.615	0.409
D-Opt. (Simple)	<u>21.09</u>	<u>0.624</u>	<u>0.406</u>
D-Opt. (Block)	21.32	0.636	0.397

forms FisherRF while T and D Optimality achieve the highest performance with an improvement from the block diagonal approximation.

4.2. Batch View Selection

Next, we compare all methods on batch view selection, where information gain is evaluated over *several* views simultaneously before training. This problem is more applicable to real-world scenarios such as information gained over a robot trajectory, or identification of the best set of views of an object.

Iterative: In the first experiment on batch view selection we follow the experimental set-up of FisherRF. The procedure is similar to single view selection, however models begin with 4 training views, are trained for $150v$ iterations between view selections, and select 4 views at a time until 20 views are obtained. All models are trained for a cumulative total of 10,000 training steps. Results on the Mip-Nerf360 dataset are shown in Table 5, and similar to previous experiments demonstrate superior performance of D and T optimality. Additionally, the block diagonal approximation improves structural quality measured by SSIM and LPIPS metrics. Note that due to the iterative view selection and training, this experiment may not reward batch view diversity, motivating our next experiment.

Keyframe Selection: To further study batch view selection in a setting more similar to SLAM applications, we compare each approach on keyframe selection. All methods are provided the same pre-trained 3D-GS model and select ten keyframes without replacement from a set of views. The selected views are then used to re-train a new 3D-GS model, which is evaluated and compared as a measure of batch information quantification. Results on the Blender dataset are summarized in Table 6, demonstrating low performance of FisherRF, A Optimality and E Optimality which select similar views. Instead, T and D Optimality select informative and different views resulting in a large performance gap.

4.3. Correlation with Render Quality

Intuitively, we expect information gain and rasterization quality at view points to be inversely related. For instance, if a 3D-GS model has only been trained on the front side of

Table 5. Results on Batch View Selection on Mip-Nerf360 dataset.

Method	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
Uniform Sampling	20.42	0.613	0.389
FisherRF	20.50	0.603	0.399
A-Opt. (Simple)	18.14	0.553	0.428
E-Opt. (Simple)	17.88	0.535	0.440
T-Opt. (Simple)	20.73	0.611	0.391
D-Opt. (Simple)	20.86	<u>0.624</u>	<u>0.383</u>
D-Opt. (Block)	<u>20.79</u>	0.631	0.378

Table 6. Results on Keyframe Selection on Blender Dataset.

Method	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
Uniform Sampling	23.47	0.888	0.109
FisherRF	18.37	0.829	0.184
A-Opt. (Simple)	17.05	0.811	0.226
E-Opt. (Simple)	16.55	0.786	0.255
T-Opt. (Simple)	24.90	0.903	0.096
D-Opt. (Simple)	24.26	0.899	0.101
D-Opt. (Block)	<u>24.53</u>	<u>0.902</u>	<u>0.099</u>

a chair, view points from the back side would have poor renderings while providing high information gain to the 3D-GS model. Therefore, as another test of our proposed method of quantifying information gain, we create a sparsification plot [12] to study the *inverse* relationship between image information gain and image render quality.

The sparsification plot in Fig. 5 is created by first training a 3D-GS model on ten randomly selected images for 2,000 iterations. Each method, using the same random seed and trained model, sorts candidate views by expected information gain. At decile increments, the cumulative average reconstruction quality of views is calculated and plotted for each method. Reading from left to right, the plot indicates the average reconstruction quality of the most informative views. Due to the inverse nature between image uncertainty and information gain, we would expect the most informative candidate views (left) to have *low* reconstruction quality.

We can see from Fig. 5 that D-Opt. (Block) has a monotonic relationship with the reconstruction quality, whereas FisherRF has difficulty with some objects. To better understand the relative performance, we also include plots of the Uniform Sampling and Oracle methods, where the Uniform Sampling method shows no correlation between the selected images and the reconstruction quality as expected. The Oracle baseline sorts the candidate views by the actual PSNR value, representing a perfect baseline, with similar performance to D-Opt. (Block).

4.4. Ablation Study

Last, we conclude with ablation studies on the most important parameters for information gain quantification. While

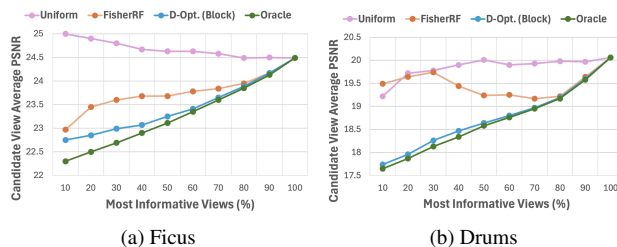


Figure 5. Correlation of expected information gain with PSNR of candidate views on two objects in Blender dataset.

Table 7. Ablation study on D-Opt. (Block) parameters on Blender Dataset with 10 images. The parameters are *sh*: spherical harmonics, α : opacity, μ : location, R: rotation, and S: scale.

Parameters Removed	PSNR (\uparrow)	SSIM (\uparrow)	LPIPS (\downarrow)
{ <i>sh</i> }	25.52	0.9089	0.0784
{ α }	25.56	0.9087	0.0778
{ μ , R, S}	25.36	0.9074	0.0793
\emptyset	25.41	0.9084	0.0776

we evaluated our methods with all parameters, reducing the number of parameters can improve computational efficiency. We compare the performance of D-Opt. (Block) with different parameter combinations in Table 7 on the task of single view selection with 10 views on the Blender dataset. Peculiarly, we find that the geometric parameters are important to quantifying information with few images while the spherical harmonics do not result in a significant difference, supporting the approach of PUP 3D-GS [9]. We expect that the opacity decreases performance since our approach does not capture the cross-correlation of ellipsoids. Additionally, we suspect that the spherical harmonics may be more useful with a well fitted scene which already has learned the geometric structure. Nonetheless, these results indicate that information may be evaluated with a minimal set of the geometric parameters, which can increase inference speed. Concretely, on the ship scene from the Blender dataset the memory and latency are as follows: 2.62 GB at 0.15 Hz for full block diagonal, 75.3 MB at 1.71 Hz for block diagonal without spherical harmonics, and 44.4 MB at 12.16 Hz for the simple approximation. Note that the cost of the simple approximation is the same as that of FisherRF.

5. Conclusion

In this paper, we introduced a novel method for calculating the information gain from images in 3D Gaussian Splatting which builds on prior literature of P-Optimality. Information quantification for 3D-GS is an important problem for evaluating uncertainty in novel environments, selecting key frames for SLAM algorithms, and next best view applications. Our novel formulation leads to a general solution with a simple and block diagonal information matrix approximation, with computational and performance trade-offs. We demonstrate quantitatively that formulating the information quantification with T and D Optimality improves performance compared to the state of the art, supporting results from prior literature. While our method achieves significant results quantitatively and qualitatively, the simple and block diagonal approximations discard correlation *between* ellipsoids. For future work we would like to investigate re-formulating the problem to include inter-ellipsoid information, such as from the structural similarity loss or the opacity parameter.

References

- [1] Yaakov Bar-Shalom, X Rong Li, and Thiagalingam Kirubarajan. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, 2004. 4
- [2] Jonathan T. Barron, Ben Mildenhall, Dor Verbin, Pratul P. Srinivasan, and Peter Hedman. Mip-NeRF 360: Unbounded Anti-Aliased Neural Radiance Fields. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 5460–5469, 2022. 6
- [3] Henry Carrillo, Ian Reid, and José A. Castellanos. On the comparison of uncertainty criteria for active SLAM. In *Proc. IEEE Int. Conf. Robot. and Automation*, pages 2080–2087, 2012. 2, 3
- [4] Yongbo Chen, Shoudong Huang, and Robert Fitch. Active SLAM for mobile robots with area coverage and obstacle avoidance. *IEEE/ASME Trans. Mechatronics*, 25(3):1182–1192, 2020. 3
- [5] Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux – effortless Bayesian deep learning. In *Proc. Advances Neural Inform. Process. Syst. Conf.*, 2021. 5
- [6] Ben Fei, Jingyi Xu, Rui Zhang, Qingyuan Zhou, Weidong Yang, and Ying He. 3D Gaussian Splatting as New Era: A Survey. *IEEE Trans. Graph.*, pages 1–20, 2024. 1, 2, 3
- [7] Maani Ghaffari Jadidi, Jaime Valls Miro, and Gamini Dissanayake. Gaussian processes autonomous mapping and exploration for range-sensing mobile robots. *Auton. Robot.*, 42(2):273–290, 2018. 2
- [8] Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, and Andrea Tagliasacchi. Bayes’ Rays: Uncertainty Quantification for Neural Radiance Fields. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 20061–20070, 2024. 2
- [9] Alex Hanson, Allen Tu, Vasu Singla, Mayuka Jayawardhana, Matthias Zwicker, and Tom Goldstein. PUP 3D-GS: Principled Uncertainty Pruning for 3D Gaussian Splatting. *arXiv*, abs/2406.10219, 2024. 2, 5, 8
- [10] Xun Huan and Youssef M. Marzouk. Simulation-based optimal Bayesian experimental design for nonlinear systems. *J. of Comput. Phys.*, 232(1):288–317, 2013. 2
- [11] Xun Huan and Youssef M. Marzouk. Gradient-based stochastic optimization methods in Bayesian experimental design. *Int. J. for Uncertainty Quant.*, 4(6):479–510, 2014. 2
- [12] Eddy Ilg, Özgün Çiçek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty Estimates and Multi-hypotheses Networks for Optical Flow. In *Proc. European Conf. Comput. Vis.*, pages 677–693, 2018. 8
- [13] Wen Jiang, Boshu Lei, and Kostas Daniilidis. FisherRF: Active View Selection and Uncertainty Quantification for Radiance Fields using Fisher Information. In *Proc. European Conf. Comput. Vis.*, pages 422–440, 2024. 2, 5
- [14] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *IEEE Trans. Graph.*, 42(4), 2023. 1, 2, 3
- [15] Jack Kiefer. General equivalence theory for optimum designs (approximate theory). *The Annals of Statistics*, pages 849–879, 1974. 2, 3, 5
- [16] Andreas Kirsch and Yarin Gal. Unifying Approaches in Active Learning and Active Sampling via Fisher Information and Information-Theoretic Quantities. *J. Mach. Learning Res.*, 2022. 2
- [17] S. Kullback and R. A. Leibler. On Information and Sufficiency. *The Ann. Math. Stat.*, 22(1):79–86, 1951. 2
- [18] Ruiqi Li and Yiu ming Cheung. Variational Multi-scale Representation for Estimating Uncertainty in 3D Gaussian Splatting. In *Proc. Advances Neural Inform. Process. Syst. Conf.*, 2024. 2
- [19] D. V. Lindley. On a Measure of the Information Provided by an Experiment. *The Ann. Math. Stat.*, 27(4):986–1005, 1956. 2
- [20] D. V. Lindley. On a Measure of the Information Provided by an Experiment. *The Ann. Math. Stat.*, 27(4):986–1005, 1956. 2
- [21] Guangyi Liu, Wen Jiang, Boshu Lei, Vivek Pandey, Kostas Daniilidis, and Nader Motee. Beyond Uncertainty: Risk-Aware Active View Acquisition for Safe Robot Navigation and 3D Scene Understanding with FisherRF. *arXiv*, abs/2403.11396, 2024. 2
- [22] Ruben Martinez-Cantin, Nando de Freitas, Eric Brochu, José A. Castellanos, and A. Doucet. A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Auton. Robot.*, 27:93–103, 2009. 2
- [23] Hidenobu Matsuki, Riku Murai, Paul H.J. Kelly, and Andrew J. Davison. Gaussian Splatting SLAM. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 18039–18048, 2024. 1, 2
- [24] Lyudmila Mihaylova, Tine Lefebvre, Herman Bruyninckx, Klaas Gadeyne, and Joris De Schutter. A comparison of decision making criteria and optimization methods for active robotic sensing. In *Numerical Methods and Applications*, pages 316–324. Springer, 2003. 3

- [25] 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. In *Proc. European Conf. Comput. Vis.*, pages 405–421, 2020. [2](#), [6](#)
- [26] Xuran Pan, Zihang Lai, Shiji Song, and Gao Huang. ActiveNeRF: Learning Where to See with Uncertainty Estimation. In *Proc. European Conf. Comput. Vis.*, pages 230–246, 2022. [2](#), [5](#)
- [27] Julio A Placed and José A Castellanos. A General Relationship between Optimality Criteria and Connectivity Indices for Active Graph-SLAM. *IEEE Robot. Autom. Letter.*, 8(2):816–823, 2022. [3](#), [5](#)
- [28] Julio A Placed, Juan J Gómez Rodríguez, Juan D Tardós, and José A Castellanos. Explorb-slam: Active visual slam exploiting the pose-graph topology. In *Iberian Robotics conference*, pages 199–210. Springer, 2022. [2](#), [3](#)
- [29] Julio A. Placed, Jared Strader, Henry Carrillo, Nikolay Atanasov, Vadim Indelman, Luca Carlone, and José A. Castellanos. A Survey on Active Simultaneous Localization and Mapping: State of the Art and New Frontiers. *IEEE Trans. Robot.*, 39:1686–1705, 2022. [2](#), [3](#), [5](#)
- [30] Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, and Hanspeter Pfister. LangSplat: 3D Language Gaussian Splatting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 20051–20060, 2024. [1](#), [2](#)
- [31] María L. Rodríguez-Arévalo, José Neira, and José A. Castellanos. On the Importance of Uncertainty Representation in Active SLAM. *IEEE Trans. Robot.*, 34(3):829–834, 2018. [3](#)
- [32] Robert Sim and Nicholas Roy. Global a-optimal robot exploration in slam. In *Proc. IEEE Int. Conf. Robot. and Automation*, pages 661–666. IEEE, 2005. [2](#), [3](#)
- [33] Matthew Strong, Boshu Lei, Aiden Swann, Wen Jiang, Kostas Daniilidis, and Monroe Kennedy III au2. Next Best Sense: Guiding Vision and Touch with FisherRF for 3D Gaussian Splatting. *arXiv*, abs/2410.04680, 2024. [2](#)
- [34] Matthew Tancik, Vincent Casser, Xinchun Yan, Sabeek Pradhan, Ben Mildenhall, Pratul P. Srinivasan, Jonathan T. Barron, and Henrik Kretschmar. Block-NeRF: Scalable Large Scene Neural View Synthesis. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 8248–8258, 2022. [2](#)
- [35] A. W. van der Vaart. *Asymptotic Statistics*, chapter 4, page 35–40. Cambridge University Press, 1998. [4](#)
- [36] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. [3](#), [6](#)
- [37] Joey Wilson, Marcelino Almeida, Min Sun, Sachit Mahajan, Maani Ghaffari, Parker Ewen, Omid Ghahmizadeh, Cheng-Hao Kuo, and Arnie Sen. Modeling Uncertainty in 3D Gaussian Splatting through Continuous Semantic Splatting. *ArXiv*, abs/2411.02547, 2024. [1](#)
- [38] Joey Wilson, Yuewei Fu, Joshua Friesen, Parker Ewen, Andrew Capodiceci, Paramsothy Jayakumar, Kira Barton, and Maani Ghaffari. ConvBKI: Real-Time Probabilistic Semantic Mapping Network With Quantifiable Uncertainty. *IEEE Trans. Robot.*, 40:4648–4667, 2024. [1](#)
- [39] Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 19595–19604, 2024. [2](#)
- [40] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 586–595, 2018. [6](#)
- [41] Shijie Zhou, Haoran Chang, Sicheng Jiang, Zhiwen Fan, Zehao Zhu, Dejia Xu, Pradyumna Chari, Suyu You, Zhangyang Wang, and Achuta Kadambi. Feature 3dgs: Supercharging 3d gaussian splatting to enable distilled feature fields. In *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, pages 21676–21685, 2024. [1](#), [2](#)
- [42] Liyuan Zhu, Yue Li, Erik Sandström, Shengyu Huang, Konrad Schindler, and Iro Armeni. LoopSplat: Loop Closure by Registering 3D Gaussian Splats. *arXiv*, abs/2408.10154, 2024. [1](#), [2](#)