https://doi.org/10.2352/CIC.2024.32.1.12

Intrinsic-GS: Multi-view Intrinsic Image Decomposition Using Gaussian Splatting and Color-Invariant Priors

Xiaoyan Xing [1]***, Konrad Groh*** [2]***, Sezer Karaoglu*** [1]***, Theo Gevers*** [1]
1. UvA-Bosch Delta Lab 2. Bosch Center of Artificial Intelligence

Abstract

Despite significant advancements in single-view intrinsic image decomposition, a domain disparity exists due to the limited information that can be obtained from a single-view image and the ill-posed nature of the problem of intrinsic image decomposition. Multi-view images offer an alternative method to circumvent the ambiguity present in 2D intrinsic image decomposition. Building on the concepts of multi-view intrinsic images and recent neural rendering techniques, we propose Intrinsic-GS, a multi-view intrinsic image decomposition method utilizing Gaussian-splatting. To achieve this, we first augment each Gaussian ellipsoid with additional attributes (i.e., albedo, shading, and a residual term) to model the intrinsic radiance field. Next, we use several color-invariants and physics-based priors to jointly regularize the optimization of the intrinsic and composited radiance fields. Finally, we conduct experiments on both synthetic and real-world datasets, demonstrating stable intrinsic decomposition results across various (including non-Lambertian) objects and scenes.

Introduction

Intrinsic image decomposition in computer vision offers opportunities to separate scene attributes like reflectance and shading, enabling better understanding and manipulation of visual content. This can enhance applications in image editing, object recognition, and scene understanding by providing clearer and more interpretable representations of visual data. However, accurately achieving intrinsic image decomposition has significant challenges due to: (1) the difficulty in reconstructing the real-world scene from images, and (2) the challenge of disentangling the complex interplay of light, shape, and material properties inherent in real-world images. Despite the rapid development of neural rendering methods [1, 2, 3, 4], which have made it easier to represent a 3D scene from a set of images, the second challenge remains.

Several studies address these challenges by integrating inverse rendering techniques with neural rendering [5, 6, 7, 8], decomposing scenes into geometry, reflectance, and illumination components. These approaches, however, are hindered by the fundamental ambiguities and ill-posed nature of inverse rendering, which necessitate prior assumptions that restrict modeling of mutual occlusions, inter-reflections, and indirect light propagation among different objects—ultimately limiting accurate 3D surface recovery and confining their utility to specific objects. A recent advancement by [9], attempts to mitigate these issues by incorporating intrinsic decomposition, i.e., separate an image to the reflectance of the materials (albedo) and the effect of illumination (shading), into neural radiance fields [1]. This sidesteps the complexities of estimating physically based rendering parameters required by physically based rendering models. Nonetheless, it struggles with decomposing high-frequency reflectance, is less effective in outdoor settings, and the lengthy training times (approximately 8 hours for a single object's decomposition) hinder its practical application in real-world scenarios.

In this work, we propose a novel approach, namely Intrinsic-GS, which incorporates intrinsic decomposition into the neural rendering pipeline of 3D Gaussian Splatting (3D-GS) [4]. This combination provides two benefits for disentangling the complex illumination interplay in a 3D scene: 1) Intrinsic-GS circumvents the need to model all complex physically based rendering components directly, instead representing them through higher-level abstractions like albedo, shading, and view-dependent residuals, which grants enhanced flexibility for application in more complex scenes. 2) Building on 3D-GS's fast and high-fidelity 3D reconstruction ability, Intrinsic-GS can reconstruct a scene and its intrinsic components quickly and at a low computational cost, significantly reducing the cost of creating an editable scene.

To introduce intrinsic decomposition into 3D-GS, we explicitly model the intrinsic radiance fields by introducing extra attributes for each Gaussian, and jointly optimizing them with the original Gaussian's attributes (Details in Sec 3.1 and 3.2). Consequently, intrinsic-specific rasterizers will aggregate the intrinsic radiance and render it as the intrinsic components. To constrain the intrinsic components, we apply a set of physically based (e.g. color invariant) priors and/or learned priors to achieve the intrinsic decomposition in a self-supervised manner (Details in Sec 3.3 to 3.5). Building on carefully designed 2D and 3D constraints and leveraging the self-clustered nature of 3D-GS, our method does not require explicit semantic labels, which are required by [9]. Moreover, since our method does not rely on accurate geometry reconstruction, we can also bypass explicit surface reconstruction, which is a weakness of the original 3D-GS as indicated by [10, 11].

Contribution We introduce intrinsic radiance fields into the 3D Gaussian-Splatting pipeline, enabling novel-view intrinsic image synthesis from object-centric data to real-world scenes. Leveraging a set of priors, our intrinsic radiance field is trained in a self-supervised manner, obviating the need for explicit surface reconstruction. Our method is simple yet efficient, it obtains a scene level decomposition within 30 minute, $\sim$ 20x faster than the Intrinsic-NeRF [9]. Table 1 shows a general comparison of our method with the recent neural-based methods.

Related work

Our method builds on both intrinsic decomposition, i.e., separate the image into its albedo and shading, and inverse rendering,

Table 1: General comparison with the advanced neural-based inverse rendering & intrinsic decomposition methods. Training-time is evaluated on object-level dataset. N/A refers to no such claim in the paper. N/A* refers to such a claim is made in the paper but can not be observed from the data used here.

Method	TensoIR [6]	Intrinsic-NeRF [9]	GS-IR [8]	Ours
Geometry-free	×	✓	×	✓
Various Material	✓	Some	N/A	✓
Scene-level	N/A	N/A*	N/A*	✓
Training-time	1h	8h	0.5h	0.5h

i.e., reconstruct the scene into its basic component based on certain rendering model (e.g. BRDF). We therefore, summarize both of them as follows:

Intrinsic decomposition. One branch of intrinsic decomposition methods focuses on single-image analysis. Some works try to address the intrinsic decomposition problem using traditional optimization methods [12, 13, 14]. Following the success of deep learning in other vision tasks, numerous methods have since employed neural network architectures for estimating these intrinsic characteristics [15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28]. For a more detailed introduction, please refer to Garces *et al.* [29]. More recently, with the development of generative models, some methods [30, 31] have showcased to leverage the distribution priors of generative models for estimating intrinsic images. These methods generally provide increasingly superior performance on images close to the learned distribution, yet often fail against out-of-distribution data and cannot guarantee consistent intrinsic imaging across the same scene.

Multi-view image intrinsic decomposition methods [32, 33, 34, 35, 36] naturally ensure consistent intrinsic imaging across views and can achieve application like scene relighting [37]. However, these methods often require extra proxy geometry information [34, 36] and cannot obtain intrinsic images from a novel view.

Our method, on the other hand, builds on the 3D Gaussian-Splatting. Such neural rendering method naturally provides the ability to render consistent intrinsic images across views and is able to generate novel-view intrinsic images in real-time, thanks to the customized tile-based rasterizer in 3D-GS.

Neural rendering. The emerging of neural rendering methods [1, 2, 3, 4] provide an alternative way to reconstruct the 3D scene. Plenty of works combining neural rendering with inverse rendering [5, 7, 6, 8, 38] have shown realistic view synthesis and consistent estimation of the underlying properties of the scenes. Yet, such an approach often requires fine-grained surface reconstruction [38] and high-quality geometry [7, 6], limiting their application to specific objects or indoor scenes. Recently, [9] introduced intrinsic decomposition into NeRF, which implicitly represents the scene using an intrinsic neural radiance field. Yet, it suffers from long training times and is not applicable to complex scenes. More recently, [8, 39] replace the NeRF with 3D-GS and achieve a higher fidelity of inverse rendering results, while it is limited by extracting the reliable geometry information from the 3D-GS representation, which is the weakness of the 3D-GS representation as indicated by [10, 11].

Our method, on the other hand, builds on the previously studied single and multi-view intrinsic imaging, and bypasses the need for accurate reconstruction of surface and geometry and is able to apply to a wider range of scenes.

Intrinsic-GS

We aim to obtain novel-view consistent intrinsic image in a scene from multi-view images. To achieve this, we build on the recent neural rendering method, 3D Gaussian-Splatting [4], and integrate the intrinsic radiance field into 3D-GS. In the following sections, we first recap the preliminaries about intrinsic decomposition and 3D-GS, then we introduce how we model and constrain the intrinsic radiance field. Finally, we show how we jointly optimize the intrinsic radiance field along with its corresponding 3D-GS. Figure 1 illustrates the Intrinsic-GS pipeline.

Preliminaries

Intrinsic decomposition. Lambertain reflectance and gray-scale shading are widely employed in solving intrinsic decomposition methods [26, 21] to simplify the inverse problem. Such assumptions enable the decomposition of an input image I into its illumination-invariant reflectance $R(I)$ and illumination-varying shading $S(I)$ components as follows:

$$I = R(I) \odot S(I), \tag{1}$$

where $\odot$ denotes channel-wise multiplication. However, the Lambertian model often falls short in realistic scenes with non-diffuse surfaces, such as those exhibiting glossy reflections or metallic properties. To address these limitations, the intrinsic residual model introduces a view-independent residual term $Re(I)$ [40, 41, 9], capturing the view-dependent effects not accounted for by the Lambertian model:

$$I = R(I) \odot S(I) + Re(I), \tag{2}$$

This model better accommodates real-world scenarios by incorporating additional components to handle specular highlights and complex material properties.

3D Gaussian Splatting. As a recent trending in neural rendering, 3D Gaussian splatting is gaining lots of attention due to its rapidly rendering speed. It can be regarded as a point cloud representation, where each point is explicitly represented by a 3D Gaussian functions $\mathscr{G}$. It can be formulated as:

$$\mathscr{G}(\boldsymbol{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \exp\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu}), \tag{3}$$

where $\boldsymbol{\mu}$ is the spatial location of the Gaussian, the $\boldsymbol{\Sigma}$ donates the shape of the Gaussian. Each Gaussian is also accompanied by an opacity α and a view-dependent color $\mathbf{c}$ (encoded by spherical harmonics).

Given a specific viewpoint, the 3D Gaussians are projected onto an image plane, resulting in 2D Gaussians. The color $\mathbf{C}$ of the 2D image at pixel $\boldsymbol{u}$ can be calculated as:

$$\mathbf{C} = \sum_{i\in N} T_i \mathscr{G}_i(\boldsymbol{u}|\boldsymbol{\mu}',\boldsymbol{\Sigma}')\alpha_i \mathbf{c}_i, T_i = \prod_{j=1}^{i-1}(1-\mathscr{G}_i(\boldsymbol{u}|\boldsymbol{\mu}',\boldsymbol{\Sigma}')\alpha_i), \tag{4}$$

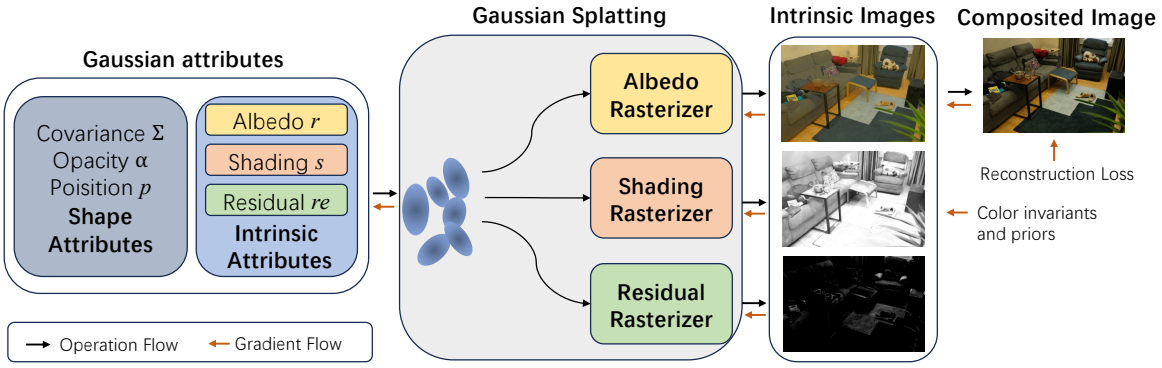

Figure 1: **The proposed Intrinsic-GS pipeline**. We first initialize Gaussian kernels with extra intrinsic-related attributes, namely, albedo, shading, and view-dependent residuals; then, we extend the original Gaussian Splatting's tile-based rendering to three individual rasterizers to render intrinsic images simultaneously. Gaussian's intrinsic-related attributes are constrained by multiple priors (includes priors for the intrinsic images and the final composited image), resulting in trained intrinsic radiance fields after convergence.

where, $\boldsymbol{\mu}', \boldsymbol{\Sigma}'$ are the projected mean vector and covariance matrix of the Gaussian, respectively (details in [4]). T_i is the accumulated transmittance, which quantifies the probability density of i-th Gaussian at pixel $\boldsymbol{u}$.

3D representation Intrinsic-GS aims to obtain intrinsic images by reconstructing the 3D scene from multi-view images. This is achieved by incorporating the reflectance model (here we employ the dichromatic model, Eq. (2)), into the GS pipeline. The input to our model is the same as that used by 3D-GS [4]. Intrinsic-GS initializes with an SfM point cloud or a random point cloud. To represent the intrinsic properties, we introduce extra attributes to each Gaussian ellipsoid and employ additional tile-based rasterizers to aggregate the corresponding property:

$$\mathbf{c}_i = \alpha_{ri}\mathbf{r}_i \odot \mathbf{s}_i + \mathbf{re}_i, \tag{5}$$

where r_i, s_i, and re_i are the reflectance (albedo), shading, and residual term at the i-th Gaussian, respectively. α_{ri} is the specific opacity for reflectance. The reflectance-specific opacity is forced to be either 0 or 1, which guarantees the reflectance attribute only models the reflectance at the opaque body.

Intrinsic field initialization

Given N points as the initialization for 3D-GS, the intrinsic field consists of attributes for modeling albedo, shading and view-dependent residual, and they are initialized as follows:

Albedo. A three-channel attribute $\mathbf{r} \in \mathrm{R}^{N\times 3}$ is employed to represent albedo. It starts with the normalized color of the original point cloud to roughly minimize the effect of shadows and highlights.

Shading. Gray shading is assumed to reduce the complexity of the inverse problem following [14, 27]. Thus, a single-channel attribute $\mathbf{s} \in \mathrm{R}^{N\times 1}$ is employed to represent shading. It is initialized by the intensity of the color of the original point cloud.

Residual term. Following the gray shading assumption, a near-natural white incident light is assumed. Hence, the residual term is represented by an attribute $\mathbf{re} \in \mathrm{R}^{N\times 3}$ with three repeated channels.

Similar to the original 3D-GS [4], the number N of 3D Gaussians is increased or decreased by duplicating or deleting 3D ellipsoids. The number of components of attributes follow the change of the number of 3D Gaussians. During training, those intrinsic components are jointly optimized with the Gaussian's shape and position attributes. In the following subsection, we introduce the related regularisation and constraints for the reflectance attributes.

Albedo constraints

Given the absence of ground-truth data for learning the albedo reconstruction, we employ multiple color-invariant and physics-based priors to constrain this process. Our approach is three-fold: at the pixel level, we use the reflectance sparsity prior to ensure that pixels with similar chromaticity share the same reflectance, following [14, 9]. At the edge level, we apply the cross color ratio prior [42] to preserve the reflectance differences across various materials. At the 3D spatial level, we utilize clustering based on the albedo properties of Gaussians.

Pixel-level Following [14, 9], given an image I, the reflectance chromaticity is approximated to $\mathscr{C}(x) = I(x)/\sum_{r,g,b}|I(x)|$, then the chromaticity similarity weight can be defined as $\omega_{cs}(x,y)$ which is associated with many priors:

$$\omega_{cs}(x,y) = exp(-\alpha_{cs}||C(x) - C(y)||_2^2), \tag{6}$$

where, x and y are the image pixel coordinates. In practice, coefficient α_{cs} is set to 60 to produce the best decomposition results. *Reflectance sparsity.* When two pixels share similar spatial locations and chromaticity, their reflectance should also be similar. We therefore, minimize the chromaticity difference according to

$$L_{rs} = \sum_{y\in\mathscr{N}(x)} \omega_{cs}||r(x) - r(y)||_2^2, \tag{7}$$

where, ω_{cs} is predefined as the chromaticity similarity [14, 9], $\mathscr{N}$ is the neighboring pixels.

Edge-level Inspired by Das *et al.* [26], we employ the cross color ratios [42] to determine albedo color differences (i.e. albedo edges).

$$L_{ccr} = ||\Delta_c(R') - \Delta_c(I)||_2^2, \tag{8}$$

where Δ_c calculates the CCR (Cross Color Ratios). R' denotes the reflectance in the 2D image plane.

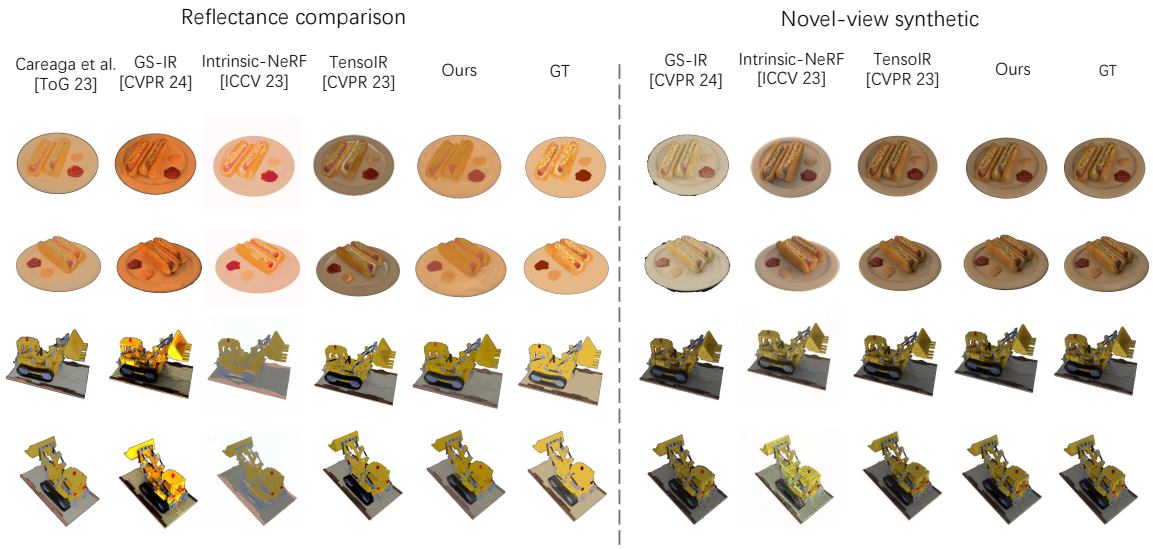

Figure 2: Visual results on NeRF-blender dataset [1]. Our method outperforms the latest multi-view intrinsic decomposition and inverse rendering methods. The state-of-the-art single view intrinsic decomposition method [27] is used as reference.

3D-spatial level Given the explicit nature of the 3D Gaussian representation, we apply clustering on albedo attributes to enhance the reflectance consistency across views.

Similar to Bell *et al.* [14], we first convert the value of albedo attributes (in *RGB*) to a 3-channel vector formulated by:

$$[r,g,b] \rightarrow [0.5 \cdot \frac{r+g+b}{3}, \frac{r}{r+g+b}, \frac{g}{r+g+b}] \tag{9}$$

Then, we apply K-Means clustering on these converted vectors to get the set of reflectance R. The number of cluster centers is empirically set to 20. Finally, the reflectance R is projected onto the 2D space denoted by R'. Details are formulated in Eq. (4).

Shading constraints

Smoothness prior Natural objects with smooth surfaces are anticipated to exhibit gradual shading variations. Additionally, this smoothness assumption is reinforced by using a pseudo-normal to ensure consistent shading across similar surfaces:

$$L_{smo} = \sum_{y \in \mathcal{N}(x)} \omega_{cs} ||r(x) - r(y)||_2^2 ||s(x) - s(y)||_2^2 \tag{10}$$

Residual prior To model non-Lambertian surfaces, a residual term re is used to capture view-dependent effects, such as specularity and glossy reflections. Given that our method focuses on real-world scenarios, where diffuse light generally predominates, we aim to recover image content primarily through reflectance and shading. Therefore, the following constraint is used:

$$L_{res} = ||re(x)||_2^2 \tag{11}$$

Jointly optimization

Intrinsic-GS is integrated into 3D-GS with additional attributes to represent intrinsic components. Using differentiable rasterizers [4], the pipeline is jointly optimized through photometric loss L_{rec} [4] and unsupervised prior constraints. The final loss is defined by:

$$L = \lambda_{rec} L_{rec} + \lambda_{rs} L_{rs} + \lambda_{ccr} L_{ccr} + \lambda_{smo} L_{smo} + \lambda_{res} L_{res}, \tag{12}$$

where λ represents the corresponding weight for each loss. We set the weights empirically.

Evaluation

In this section, we demonstrate: 1) The high-quality intrinsic decomposition achieved by Intrinsic-GS for different object-level datasets. 2) The high-fidelity and consistent intrinsic decomposition results for real-world images. 3) A quantitative ablation study evaluating the contribution of each employed prior.

Intrinsic decomposition for objects with different material

In this section, intrinsic decomposition is presented for commonly used blender object datasets [1, 43]. Intrinsic-GS is compared with recent state-of-the-art methods on intrinsic decomposition [26, 27] and neural rendering based inverse rendering [6, 8], respectively. For image-based intrinsic decomposition [26, 27], we employ their open-source models. For the neural rendering based method, we train their models on each object following their original implementations.

Figure 2 presents the reflectance decomposition results on the NeRF-synthetic dataset [1]. Our method consistently performs well on the evaluated scenes. Compared to [9], our method is capable of reconstructing the albedo in higher frequency areas. Notably, BRDF-based methods [6] and [8] exhibit leakage from the shading.

Table 2 reports the quantitative comparison on Blender objects (hotdog, Lego, ficus, and drums). Numerically, our method achieves state-of-the-art results across all metrics. Specifically, our method outperforms others by a large margin on the perceptual metric [44], showcasing its superior intrinsic decomposition capability.

To further evaluate our method's robustness against challenging scenes, such as non-Lambertian surfaces, we conduct experiments on the shiny-blender dataset [43]. Due to the unavailability of ground-truth reflectance labels, we perform a visual evaluation. As illustrated in Figure 3, our method achieves

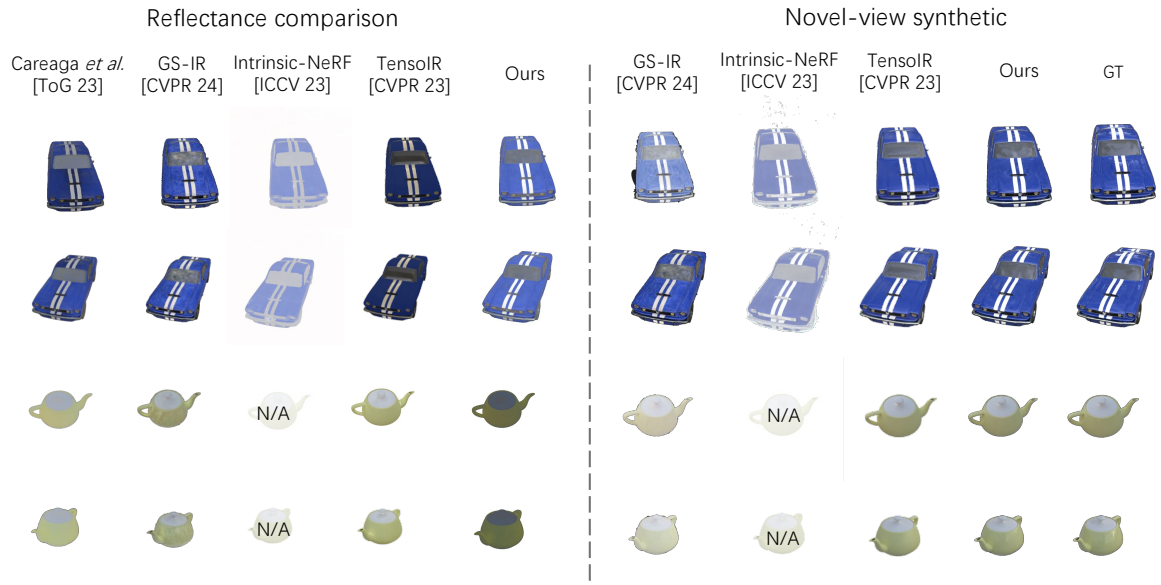

Figure 3: Qualitative results on Shiny NeRF dataset [43]. Our method performs stable on the more glossy objects, showcasing the robust design of our pipeline. The state-of-the-art single view intrinsic decomposition method [27] is used as reference.

Table 2: Numerical comparison of reflectance estimation on the NeRF synthetic dataset [1]. First 2 rows are single-view intrinsic decomposition methods. Row 3 to 5 are the neural rendering based methods.

Method	MSE ↓	SSIM ↑	PSNR ↑	LPIPS ↓
PIE-Net [26]	0.0190	0.8323	21.2113	0.1217
Careaga *et al.* [27]	0.0067	0.8877	22.2881	0.1001
TensoIR [6]	0.0078	0.9230	23.7934	0.3790
Intrinsic-NeRF [9]	0.0032	0.9350	24.4061	0.3933
GS-IR [8]	0.0042	0.9002	25.9926	0.4744
Ours	**0.0016**	**0.9436**	**26.1722**	**0.0706**

Table 3: Ablation study of the proposed constraint levels on reflectance.

Pixel	Edge	Spatial	SSIM ↑	PSNR ↑	LPIPS ↓
✓			0.9313	25.4476	0.0901
✓	✓		0.9402	26.1978	0.0875
✓	✓	✓	0.9436	26.1722	0.0706

stable results compared to other competitors. Notably, Intrinsic-NeRF [9] fails to model shiny surfaces due to the limitations of the original NeRF backbone it employs. Additionally, results from Careaga *et al.* [27], GS-IR [8], and TensoIR [6] exhibit leakages from shading or environmental illumination, which are caused by the limitations of their modeling approaches.

Intrinsic decomposition for real-world scenes

In this section, we present the results of our intrinsic decomposition on the MIP-360 dataset [2]. Again, no ground truth reflectance labels are available. We compare Intrinsic-GS with the state-of-the-art single-view intrinsic imaging methods [26, 27] and the neural rendering-based method [8], which claims the capability for open-world reflectance decomposition. An evaluation with [9] is omitted as it fails to reconstruct the scene. We do not compare with [6], since it is only applicable to object-level data.

Figure 4 showcases our method's capability in generating novel-view intrinsic images. Compared to recent advancements in single-view [27] and multi-view [8] intrinsic decomposition methods, our approach demonstrates superior intrinsic estimation in terms of view consistency and image quality. Notably, for Careaga *et al.* [27] , the input image is the ground truth, while for GS-IR [8] and Intrinsic-GS, all images are directly generated from the neural rendering pipeline, indicating that these views were not seen during training.

Ablation study

Ablation results are shown in Table 3. The three-level constraints help our pipeline achieve better reflectance estimation in various ways. Specifically, without the 3D spatial clustering, an increase in PSNR performance can be observed, indicating that the introduced 3D spatial clustering may degrade the representation of color distribution. However, 3D spatial clustering is essential for achieving reflectance sparsity. Therefore, we will still include it in our full pipeline. Additionally, without the cross-color ratio (CCR), the performance of our method decreases, highlighting the importance of calculating edge-based information.

Recoloring

Here we showcase our method's downstream application. Intrinsic-GS is able to obtain intrinsic images from any given view, and the intrinsic images can be used to achieve recoloring via directly changing the color based on the albedo attribute.

Figure 5 and Figure 6 show the recoloring results on the MIP-360 dataset [2] and the shiny-nerf dataset [43]. Despite facing complex scenarios, stable recoloring results can be obtained across the different viewing angles. This further proves our pipeline's excel ability to separate the intrinsic components.

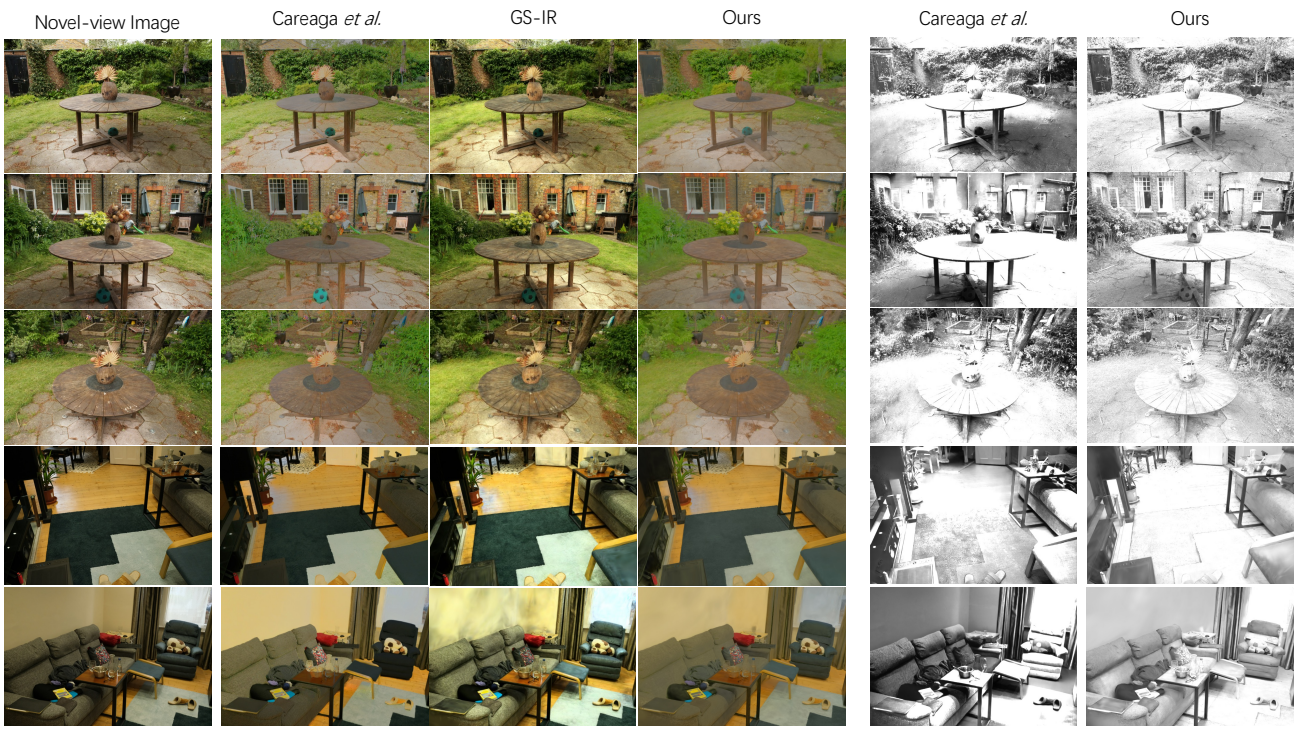

Figure 4: Real-world intrinsic decomposition visualization. Column 2 to 4 are the albedo estimation results from Careaga *et al.* [27], GS-IR [8], and ours, respectively. Column 5 and 6 are the shading estimation results from Careaga *et al.* [27] and ours, respectively. Notably, GS-IR [8] does not estimate shading.

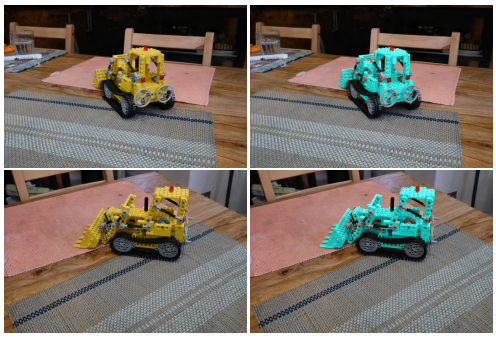

Figure 5: Our method achieves real-time novel-view recoloring on real-world scene. Left column: rendered novel-view image. Right column: the recolored novel-view image. **Zoom in to see the details.**

Figure 6: Our method achieves real-time novel-view recoloring on shiny-reflected surface. Left column: rendered novel-view image. Right 2 columns: the recolored novel-view image. **Zoom in to see the details.**

Limitations

The intrinsic images may exhibit failure scenes if 3D-GS cannot accurately represent the scene from a particular viewpoint. However, this also implies that any improvements in the rendering method would result in corresponding gains in the performance of our model. Additionally, our method assumes neutral illumination; under more colorful lighting conditions, the color of the light may leak into the albedo. Moreover, the original CCR is designed for Lambertian-like surfaces, which may introduce artifacts when applied to high-frequency reflectance. Further study into color invariants may address this issue.

Conclusion

We have proposed a pipeline, Intrinsic-GS, which integrates intrinsic decomposition into 3D Gaussian-splatting. With only multi-view images as input, our method is able to decompose a scene into reflectance, shading, and a view-dependent residual. Built on several physics-based and/or learned priors, the whole decomposition process is trained in a self-supervised way. Once trained, Intrinsic-GS can obtain consistent intrinsic images from any novel views, significantly reducing the cost for downstream applications like recoloring and material editing.

References

[1] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis," in *ECCV*, pp. 405–421, Springer, 2020.

[2] J. T. Barron, B. Mildenhall, M. Tancik, P. Hedman,

R. Martin-Brualla, and P. P. Srinivasan, "Mip-nerf: A multiscale representation for anti-aliasing neural radiance fields," in *ICCV*, 2021.
[3] A. Chen, Z. Xu, A. Geiger, J. Yu, and H. Su, "Tensorf: Tensorial radiance fields," in *ECCV*, 2022.
[4] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, "3d gaussian splatting for real-time radiance field rendering," *ACM ToG*, vol. 42, July 2023.
[5] X. Zhang, P. P. Srinivasan, B. Deng, P. Debevec, W. T. Freeman, and J. T. Barron, "Nerfactor: Neural factorization of shape and reflectance under an unknown illumination," *ACM TOG*, vol. 40, no. 6, pp. 1–18, 2021.
[6] H. Jin, I. Liu, P. Xu, X. Zhang, S. Han, S. Bi, X. Zhou, Z. Xu, and H. Su, "Tensoir: Tensorial inverse rendering," in *CVPR*, 2023.
[7] Y. Zhang, J. Sun, X. He, H. Fu, R. Jia, and X. Zhou, "Modeling indirect illumination for inverse rendering," in *CVPR*, pp. 18643–18652, 2022.
[8] Z. Liang, Q. Zhang, Y. Feng, Y. Shan, and K. Jia, "Gs-ir: 3d gaussian splatting for inverse rendering," in *CVPR*, 2024.
[9] W. Ye, S. Chen, C. Bao, H. Bao, M. Pollefeys, Z. Cui, and G. Zhang, "Intrinsicnerf: Learning intrinsic neural radiance fields for editable novel view synthesis," in *ICCV*, pp. 339–351, 2023.
[10] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, "2d gaussian splatting for geometrically accurate radiance fields," in *SIGGRAPH 2024 Conference Papers*, Association for Computing Machinery, 2024.
[11] A. Guédon and V. Lepetit, "Sugar: Surface-aligned gaussian splatting for efficient 3d mesh reconstruction and high-quality mesh rendering," in *CVPR*, 2024.
[12] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman, "Ground truth dataset and baseline evaluations for intrinsic image algorithms," in *ICCV*, 2009.
[13] J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE TPAMI*, vol. 37, no. 8, pp. 1670–1687, 2015.
[14] S. Bell, K. Bala, and N. Snavely, "Intrinsic images in the wild," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, p. 159, 2014.
[15] Z. Li and N. Snavely, "Cgintrinsics: Better intrinsic image decomposition through physically-based rendering," in *ECCV*, 2018.
[16] T. Narihira, M. Maire, and S. X. Yu, "Direct intrinsics: Learning albedo-shading decomposition by convolutional regression," in *ICCV*, 2015.
[17] J. Shi, Y. Dong, H. Su, and X. Y. Stella, "Learning non-lambertian object intrinsics across shapenet categories," in *CVPR*, 2017.
[18] T. Zhou, P. Krahenbuhl, and A. A. Efros, "Learning data-driven reflectance priors for intrinsic image decomposition," in *CVPR*, 2015.
[19] D. Zoran, P. Isola, D. Krishnan, and W. T. Freeman, "Learning ordinal relationships for mid-level vision," in *CVPR*, 2015.
[20] Y. Qian, M. Shi, J.-K. Kamarainen, and J. Matas, "Fast fourier intrinsic network," in *WACV*, 2021.
[21] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, "Revisiting deep intrinsic image decompositions," in *CVPR*, 2018.
[22] A. S. Baslamisli, H.-A. Le, and T. Gevers, "Cnn based learning using reflection and retinex models for intrinsic image decomposition," in *CVPR*, 2018.
[23] W.-C. Ma, H. Chu, B. Zhou, R. Urtasun, and A. Torralba, "Single image intrinsic decomposition without a single intrinsic image," in *ECCV*, 2018.
[24] A. S. Baslamisli, P. Das, H.-A. Le, S. Karaoglu, and T. Gevers, "Shadingnet: Image intrinsics by fine-grained shading decomposition," *IJCV*, vol. 129, no. 8, pp. 2445–2473, 2021.
[25] S. Sengupta, A. Kanazawa, C. D. Castillo, and D. W. Jacobs, "Sfsnet: Learning shape, reflectance and illuminance of facesin the wild'," in *CVPR*, pp. 6296–6305, 2018.
[26] P. Das, S. Karaoglu, and T. Gevers, "Pie-net: Photometric invariant edge guided network for intrinsic image decomposition," in *CVPR*, 2022.
[27] C. Careaga and Y. Aksoy, "Intrinsic image decomposition via ordinal shading," *ACM Transactions on Graphics*, 2023.
[28] X. Xing, K. Groh, S. Karaoglu, and T. Gevers, "Intrinsic appearance decomposition using point cloud representation," *arXiv preprint arXiv:2307.10924*, 2023.
[29] E. Garces, C. Rodriguez-Pardo, D. Casas, and J. Lopez-Moreno, "A survey on intrinsic images: Delving deep into lambert and beyond," *IJCV*, vol. 130, no. 3, pp. 836–868, 2022.
[30] P. Kocsis, V. Sitzmann, and M. Nießner, "Intrinsic image diffusion for single-view material estimation," in *CVPR*, 2024.
[31] Z. Zeng, V. Deschaintre, I. Georgiev, Y. Hold-Geoffroy, Y. Hu, F. Luan, L.-Q. Yan, and M. Hašan, "Rgb $\setminus \leftrightarrow$ x: Image decomposition and synthesis using material-and lighting-aware diffusion models," in *SIGGRAPH*, 2024.
[32] A. Meka, M. Zollhöfer, C. Richardt, and C. Theobalt, "Live intrinsic video," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1–14, 2016.
[33] R. Yi, P. Tan, and S. Lin, "Leveraging multi-view image sets for unsupervised intrinsic image decomposition and highlight separation," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 12685–12692, 2020.
[34] S. Duchêne, C. Riant, G. Chaurasia, J. Lopez-Moreno, P.-Y. Laffont, S. Popov, A. Bousseau, and G. Drettakis, "Multi-view intrinsic images of outdoors scenes with an application to relighting," *ACM Transactions on Graphics*, p. 16, 2015.
[35] P.-Y. Laffont, A. Bousseau, and G. Drettakis, "Rich intrinsic image decomposition of outdoor scenes from multiple views," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 2, pp. 210–224, 2012.
[36] Y. Yu, A. Meka, M. Elgharib, H.-P. Seidel, C. Theobalt, and W. Smith, "Self-supervised outdoor scene relighting," in *ECCV*, 2020.
[37] J. Philip, M. Gharbi, T. Zhou, A. Efros, and G. Drettakis, "Multi-view relighting using a geometry-aware network," *ACM ToG*, vol. 38, July 2019.
[38] J. Zhu, Y. Huo, Q. Ye, F. Luan, J. Li, D. Xi, L. Wang, R. Tang, W. Hua, H. Bao, *et al.*, "I2-sdf: Intrinsic indoor scene reconstruction and editing via raytracing in neural sdfs," in *CVPR*, pp. 12489–12498, 2023.
[39] Y. Jiang, J. Tu, Y. Liu, X. Gao, X. Long, W. Wang, and Y. Ma, "Gaussianshader: 3d gaussian splatting with

shading functions for reflective surfaces," *arXiv preprint arXiv:2311.17977*, 2023.

[40] B. A. Maxwell, R. M. Friedhoff, and C. A. Smith, "A bi-illuminant dichromatic reflection model for understanding images," in *CVPR*, pp. 1–8, 2008.

[41] S. Tominaga, "Dichromatic reflection models for a variety of materials," *Color Research & Application*, vol. 19, no. 4, pp. 277–285, 1994.

[42] T. Gevers and A. W. Smeulders, "Color-based object recognition," *Pattern recognition*, vol. 32, no. 3, pp. 453–464, 1999.

[43] D. Verbin, P. Hedman, B. Mildenhall, T. Zickler, J. T. Barron, and P. P. Srinivasan, "Ref-NeRF: Structured view-dependent appearance for neural radiance fields," *CVPR*, 2022.

[44] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018.