SViM3D: Stable Video Material Diffusion for Single Image 3D Generation

Supplementary Material

Contents

Overview

In the supplement to Stable Video Materials 3D (SViM3D), a foundational multi-view material model with camera control, we first expand the background on video diffusion models and neural fields, add information on the optimization and finally present more results, in-depth analysis and applications of our method. Please also consider watching the **supplemental video** that gives an overview of this work and contains further visual results.

A. Additional Background

A.1. Video Diffusion Denoising

The conditioning image is concatenated to the noisy latent state input z_t at noise timestep t. The CLIP-embedding [17] matrix of the conditioning image is provided to the cross-attention layers of each transformer block as its key and value. The camera poses, represented as angles e_i and a_i as well as the noise timestep t are encoded into sinusoidal position embeddings. The camera pose embeddings are linearly transformed and added to the noise timestep embedding. The result is added to each residual block's output features after being run through another linear layer to match the feature dimension as in SV3D [19].

A.2. Coordinate-based MLPs and NeRF

[15] NeRFs [15] use a dense neural network to model a continuous function that takes 3D location $\boldsymbol{x} \in \mathbb{R}^3$ and view direction $\boldsymbol{d} \in \mathbb{R}^3$ and outputs a view-dependent output color $\boldsymbol{c} \in \mathbb{R}^3$ and volume density $\sigma \in \mathbb{R}$. A camera ray $r(t) = \boldsymbol{o} + t\boldsymbol{d}$ is cast into the volume, with ray origin $\boldsymbol{o} \in \mathbb{R}^3$ and view direction $\boldsymbol{d}$. The final color is then approximated via numerical quadrature of the integral: $\hat{\boldsymbol{c}}(\boldsymbol{r}) = \int_{t_n}^{t_f} T(t)\sigma(t)\boldsymbol{c}(t)\,dt$ with $T(t) = \exp(-\int_{t_n}^{t} \sigma(t)\,dt)$, using the near and far bounds of the ray t_n and t_f respectively [15].

Figure A1. **Multiple samples.** Demonstrating the stochastic sampling process by taking three samples with the same condition image. For views that are less constrained by the conditioning diverse examples can be generated depending on the initial noise. Note, that the roughness and metallic parameters (blue and green here) are consistent with the RGB predictions, though.

B. Optimization

B.1. UNet training details

We pre-compute latents and CLIP-embeddings [17] for all training data. The RGB color rendering is composed on a solid random color or white, the basecolor AOV stays always on white. The other outputs keep their black backgrounds. We follow the EDM framework and use the diffusion loss for fine-tuning described by Blattmann *et al.* [2]. We employ Flash Attention v2 [4, 5] to keep the memory footprint low such that a batch size of two is still possible for 21 frames on similar hardware to the SV3D [19] training.

Guidance. Compared to a conventional video generation with a reference frame as the starting point we have circular orbits both starting and ending close to the reference view. To reduce over-sharpening caused by classifier-free-guidance (CFG) [9] we also adapt a triangular CFG scaling similar to the one proposed in [20] where the guidance scale is adapted based on the distance to the reference view.

B.2. Geometry regularization.

We adopt several geometric priors to regularize the reconstructed shape. Firstly we supervise the normal using the predicted normal maps. Especially during the beginning of the NeRF optimization this supervision loss is strictly enforced eliminating the need for any additional monocular prior. Since our normal maps generally contain more detail than can be represented by the mesh representation, starting from the second half of phase 1, we additionally optimize a bump map represented by a small auxiliary field conditioned on the coordinate embeddings from DMTet. A bilateral smoothness loss is also added to the normals in phase 1 and increased during phase 2. Similarly, we utilize the smooth depth loss from RegNeRF [16]. While the supervision loss with the pseudo-GT (pGT) and the photometric rendering loss are high in the beginning of the NeRF reconstruction (Phase 1) we slowly increase the weight of the LPIPS [26] over the course of the reconstruction ultimately dominating the reconstruction at the end of Phase 1. Our homography correction scheme is also added in Phase 1 after an initial warmup phase of 400 steps. In Phase 2 the LPIPS loss is slowly reduced a little and bilateral smoothness regularizers increased in weight to clean up remaining noise.

B.3. View dependent masking

We normalize the masks by the maximum value over all views and apply a smoothstep function f_s followed by a gamma correction to smoothly clip to the range of 0 to 1 and to steer the mask contrast.

B.4. Homography correction

To make the optimization more robust to outlier views where the image is warped wrongly due to homogeneous image regions or complex edge features, we introduce a masking scheme in Phase 2. Based on the loss difference in the albedo map, it is decided if the current view is warped or not. If a view is consistently masked, then H_i is reinitialized and further refined.

C. Further results

In the following section we provide additional results including evaluation on additional datasets and qualitative comparisons related to the reconstruction pipeline.

C.1. Overview of baseline methods

Intrinsic Image Diffusion (IID) [13] is one of the first works to explore diffusion models for PBR material estimation. Their model outputs albedo, roughness and metallic parameters for a single frame. Originally trained on interior scenes, it has also been applied to general 3D reconstruction [7].

Table C2. **Baseline Methods.** Features of existing methods used in our evaluation compared to SViM3D.

Method	RGB NVS	Multi-view	Joint PBR	Spatially-varying PBR	Normals	Textured mesh
SV3D [20]	✓	✓	✗	✗	✗	✓
SF3D [3]	✗	✗	✓	✗	✓	✓
IID [13]	✗	✗	✓	✓	✗	✗
RGB↔X [24]	✗	✗	✗	✓	✓	✗
SM [14]	✗	✗	✓	✓	✗	✓
SViM3D	✓	✓	✓	✓	✓	✓

Table C3. **Baseline Methods Relighting.** Features of existing methods for image based relighting compared to SViM3D.

Method	LDR output	HDR output	Global Illum	NVS	Multi-view	Material Editing	Interactive speed
IC Light [25]	✓	✗	✓	✗	✗	✗	✗
Neural Gaffer [11]	✓	✗	✓	✗	✗	✗	✗
SViM3D	✓	✓	✗	✓	✓	✓	✓

MaterialFusion [14] proposes a 2D material denoising diffusion prior based on StableDiffusion 2.1 [18] with the same output as above but trained on object centric data. They employ an SDS based optimization to achieve 3D asset generation. Finally, RGB↔X [24] released a latent image diffusion model that can generate PBR data as part of their material- and lighting-aware neural rendering pipeline. Their material model can generate either albedo, roughness, metallic or diffuse irradiance maps conditioned on a single image and a text prompt to select the task. Significantly faster is SF3D [3] which is based on a transformer decoder architecture like LRM [10, 22]. Since the 3D reconstruction code for SV3D [19] is not publicly available at the time of writing we decide to compare against SF3D instead. As evident in Fig C3 SF3D's material model is limited as it does not allow for spatially-varying roughness and metallic values. This poses a severe limitation for real-world objects composed from multiple materials. Our spatially-varying parametrization yields shading results closer to the GT. Tab. C2 gives a high-level overview of the features available in the compared methods. SViM3D is the only one offering RGB view synthesis and material synthesis as a multi-view task with joint spatially-varying PBR and normal prediction as well as 3D reconstruction of a textured mesh.

C.2. Additional multi-view material results

In Fig C2, Fig. C5 and Fig C9 we show additional raw outputs of our diffusion model given reference images from multiple datasets. SViM3D generates plausible material maps for a variety of object classes and surface materials. The high metallic value in Fig C9 is questionable in a physical sense but apparently helps the model to represent the specific shine of the dinosaur figure which might correspond to the way an artist might work in this case. In Fig. C16 we compare the generated material maps to the ground truth AOVs from synthetic data. Despite the ambiguity the model is able to predict plausible solutions also reflected in the RMSE values in Tab. 1. In addition to our newly introduced Poly Haven [8] object dataset we also evaluate our model on a test split of

Figure C2. **Multi-view material prediction.** Additional examples from the Poly Haven [8] test dataset. SViM3D successfully converts a single image to a sequence of novel views with spatially-varying PBR material parameters and surface normals. These can directly be used to relight the novel views as shown in the two bottom rows.

(a) GT rendering (b) SF3D (c) SViM3D

Figure C3. **Material parametrization.** Compared to SF3D [3], a recent method for single image to 3D generation, our material model is able to replicate spatially-varying roughness and metallic parameters which help to represent real-world objects realistically.

Table C1. **View consistency.** Multi-view consistency evaluated using MEt3R [1] on the Poly Haven test data.

Method	MEt3R score↑
SV3D RGB↔X [24]	0.54
SV3D + IID [13]	0.51
SV3D + SM [14]	0.54
SViM3D	0.57

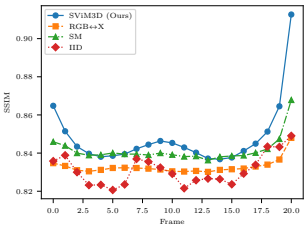

Figure C4. **Multi-view error distribution.** We compare the SSIM results of the Basecolor prediction across frames over the Poly Haven test set.

the recently introduced BlenderVault dataset [14] in Tab. C5. The results are consistent with our evaluation on Poly Haven verifying the plausabiliy of our test results.

C.3. Quantitative evaluation across views

Fig. C4 compares the mean error across all generated views between all evaluated models from Tav. 2. Our method consistently yields the best results over all views, although it varies depending on the camera view. The observation that the side views are the most challenging generations might be explained by the occurrence of more extreme angle configurations in the context of the surface shading. Traditionally, grazing angles and samples close to object boundaries can lead to inconsistencies in 3D reconstruction [27] and generation might suffer from similar effects. Additionally, Tab. C1 shows the results of MeT3R [1], a view consistency metric based on the recently introduced DUST3R [21] for calibration free 3D point cloud reconstruction. The metric also reflects the improved multi-view consistency in SViM3D com-

Figure C5. **Multi-view PBR materials.** Given the input image SViM3D generates multi-view consistent novel views with corresponding basecolor, roughness, metallic and normal maps. These can directly be used to generate views under novel illumination. We show 5 samples from a generated orbit and two new illumination settings as examples. The objects are sourced from our Poly Haven [8] test dataset. Please find additional results in the supplementary material.

Figure C6. **More 3D reconstruction results.** Objects sourced from Poly Haven [8] and GSO [6], rendered in Blender.

Figure C7. **2.5D Relighting.** Using the output of SViM3D and an environment map we can directly relight an object. We can use the same illumination representation and deferred shading as in the differentiable rendering pipeline.

pared to the SV3D [19] baselines.

C.4. 3D reconstruction

Fig. C8 illustrates our single image to 3D reconstruction pipeline using an example image from our test set. Starting with the multi-view novel view synthesis with material parameters and surface normals, the output is lifted to a 3D representation, first a NeRF [15], then a polygon mesh. It is worth noting that the material parameters are well preserved thanks to our pseudo GT supervision. Finally, the mesh can be rendered under novel illumination, again. We show additional 3D reconstruction results in Fig. C6. Fig. C15 features two generations conditioned on a smartphone capture illustrating in-the-wild performance.

C.5. Multiple samples

Fig. A1 compares three samples of denoising process given the same condition image. It is visible that there is some diversity in the predictions while they still all represent physically plausible solutions in the context of the conditioning given the underconstrained task. The diversity of the deviations increases the further the camera moves away from the condition frame, of course.

C.6. 3D Geometry

We evaluate the quality of the reconstructed geometry using Chamfer distance and Intersection over Union (IoU) against ground truth point clouds provided by the Google Scanned Objects (GSO) [6] dataset and report the results in Tab. C4. We select a random subset of 80 real-world objects for the comparison against SF3D [3]. Compared to the feed-forward architecture of SF3D can our reconstruction method fail in rare cases where some views do not align for some reason. This is reflected in the slightly lower scores. In cases where reconstruction succeeds the quality is visually very close, often keeping a bit finer detail in the case of SViM3D at the expensive of some additional noise (see also Fig C3).

RGB only view synthesis Using the SV3D [19] baseline without PBR material prediction yields lower quality results also for the RGB color generation as reported in Tab. 2. We argue that enforcing reasoning over illumination as part of the material estimation also helps the generation of consistent lighting in the RGB views.

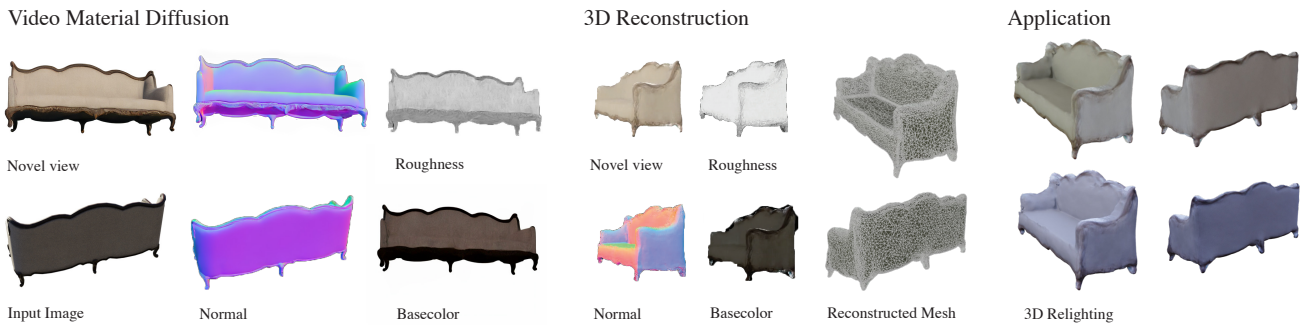

Figure C8. **3D reconstruction example.** SViM3D 's pipeline starts with a single image at the bottom left. First novel views and the corresponding material parameters and surface normals are generated. Following, an intermediate 3D representation is optimized given the multi-view material prior. Finally, a 3D mesh can be extracted and integrated into downstream applications. Here we show an example from our Poly Haven [8] test dataset.

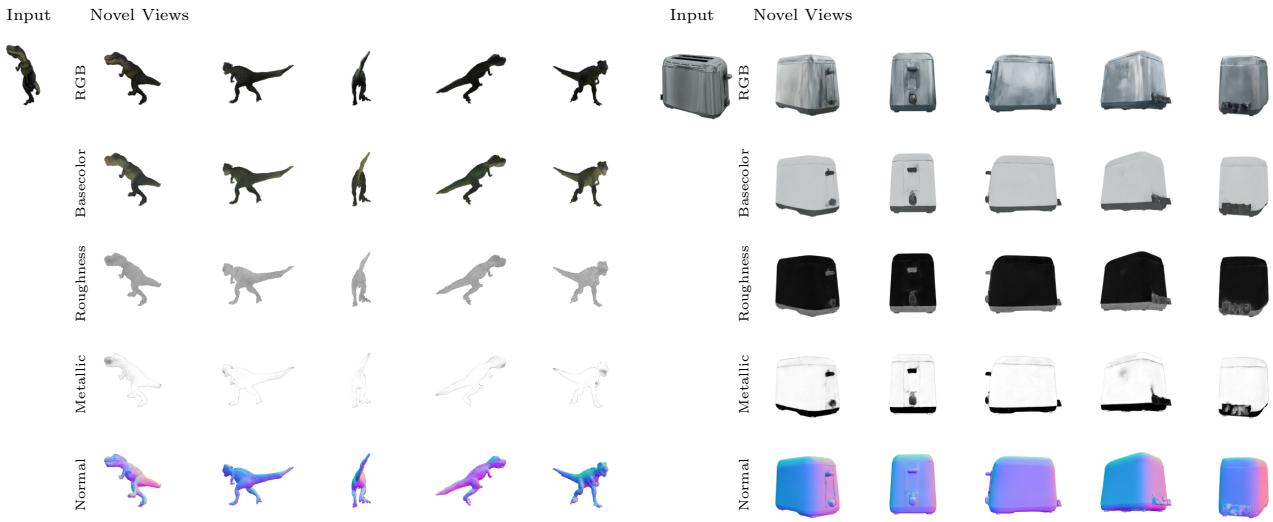

Figure C9. **Multi-view material examples from GSO.** Two objects from the GSO [6] dataset representing common real-world houshold items. SViM3D generalizes well to this domain as long as the scene is object centric.

Table C4. **3D reconstruction.** We evaluate the model against SF3D [3] on a subset of the Google Scanned Objects (GSO) [6] featuring real-world household items. The mesh quality is reported as Chamfer distance and IoU compared to the scanned GT point-clouds.

Method	3D Geometry	
	Chamfer↓	IoU↑
SF3D [3]	0.031	0.52
SViM3D	0.034	0.48

C.7. Relighting

In Fig. C7 we give additional insights into our 2.5D relighting approach. We show a metallic and plastic surface lit by

Table C5. **Multi-view NVS with material parameters on BlenderVault dataset.** Given a single RGB image a multi-view orbit around the scene center is generated with corresponding PBR materials and normals. We compare RGB NVS and albedo / basecolor generation as stand-in for PBR materials against rendered GT on a subset (100 objects) of the BlenderVault dataset [14]. We also compare against the MaterialFusion [14] baseline on their single view prediction task.

Method	PSNR↑	SSIM↑	LPIPS↓	FID↓	CLIPS↑	CMMD↓
RGB radiance 21 images						
SViM3D (ours)	20.22	0.86	0.081	24.95	0.86	1.14
Basecolor / Albedo 21 images						
SViM3D (ours)	19.80	0.86	0.08	40.0	0.81	1.08
Basecolor single image (ref view)						
SM [14] (*from paper*)	24.70	0.91	-	-	-	-
SViM3D (ours)	27.35	0.92	0.05	46.0	0.83	1.08

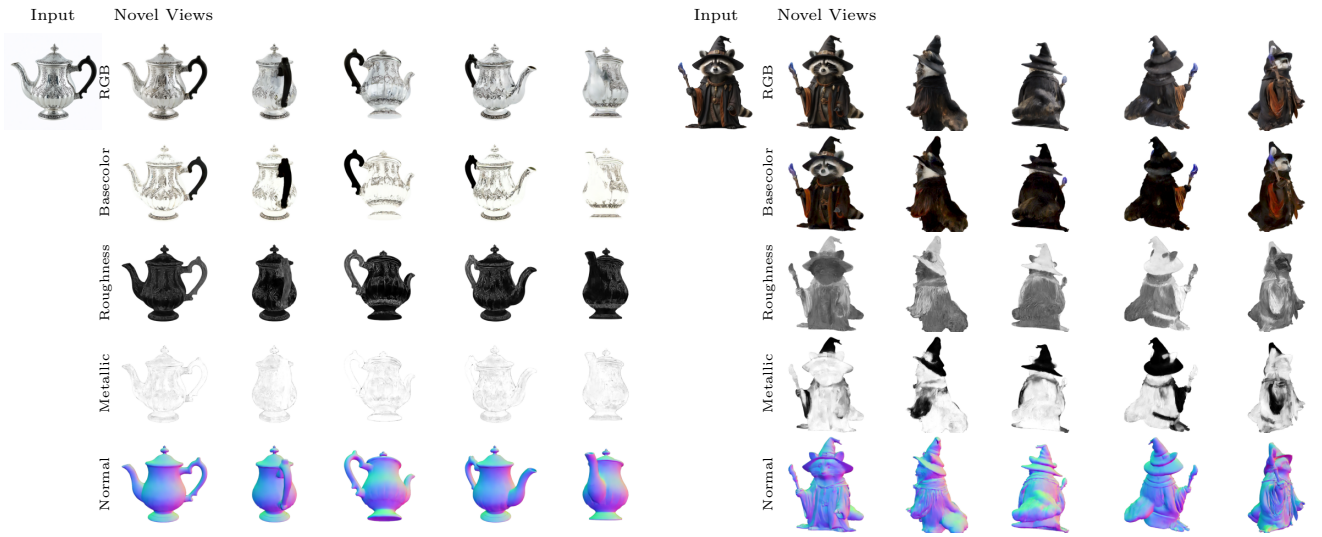

Figure C10. **Multi-view material examples from generated images..** Multi-view generations conditioned on generated images from text-to-image models, a wizard raccoon and a silver teapot. SViM3D is capable of estimating plausible and view consistent results. The wizard raccoon is an out-of-distribution example due to the lack of stylized character models in the training data.

Figure C11. **Material editing.** The explicit material parameters of SViM3D's output can be edited in a physically-plausible way and the result visualized using our rendering framework. In this example the material roughness is varied between almost zero and close to one while the original value is close to the version second to left.

different rotations of the spherical environment map. Using all the generated material channels and the normal directions we can achieve dynamic direct illumination at real-time speed. We also present the intermediate illumination representation used in our deferred shading pipeline. Our pipeline also enables material editing as further analyzed in Fig. C11. Fig C13 shows examples for different illumination directions and camera views. To achieve indirect illumination, a full 3D reconstruction can be completed.

Relighting comparison We present additional results from our 2.5D relighting pipeline in Fig. C12. As baselines we use IC-Light [25], Neural Gaffer [12] and DiLightNet [23], three diffusion based methods for image-based relighting recently introduced. In Tab. C3 we give an overview of the feature sets of all relighting methods. Neural Gaffer supports environment map inputs as conditioning which is fed as low and high dynamic range representation. IC-Light provides image editing based on a background image. And DiLightNet adds radiance hints to the conditioning via environment maps. In our comparison we preprocess the environment maps to serve the methods, respectively. We compare the results against the GT obtained from our 2.5D rendering pipeline here, using the synthetic PBR material maps. SViM3D is the only model capable of joint novel view synthesis and relighting. This is reflected in better multi-view consistency and fewer artifacts like the residual highlight in the example of Neural-Gaffer. IC-Light generally generated high-contrast output which is difficult to edit in real-world use cases.

3D relighting application As shown in Fig. C8 as well as Fig. 1 the 3D reconstructed models can be easily integrated into new environments thanks to the PBR materials. Using a path tracer global illumination effects can then be achieved, too. Please find additional dynamic relighting and scene integration examples in the supplemental video.

Analysis of ambiguous materials We constructed a small dataset of pathological test cases for the ambiguity between metallic and glossy plastic surfaces. In over 90% of the cases a low roughness value with near zero metalness is predicted. The predictions of higher values often are for objects that would usually have metal in their material. See Fig. C14 for a visual example. These findings can be explained by dataset bias.

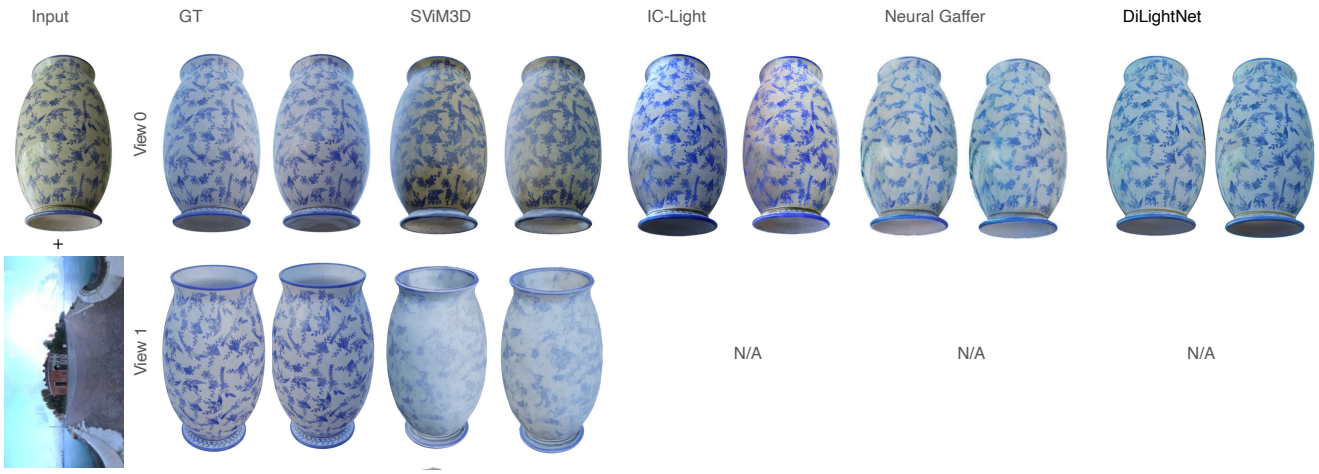

Figure C12. **Relighting comparison.** We compare image-based relighting results on an example object from the Poly Haven [8] dataset between the synthetic ground truth (GT), IC-Light [25], Neural-Gaffer [12], DiLightNet [23] and SViM3D (ours).

Figure C13. **Relighting.** Using the output of SViM3D and an environment map (HDRI) we can directly relight any view on the camera trajectory using our 2.5D approach.

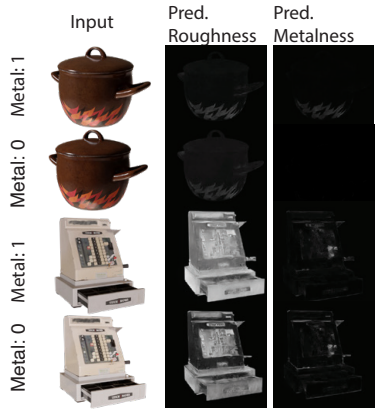

Figure C14. **Glossiness vs. Metalness ambiguity.** Examples from our generated test cases and the corresponding model predictions.

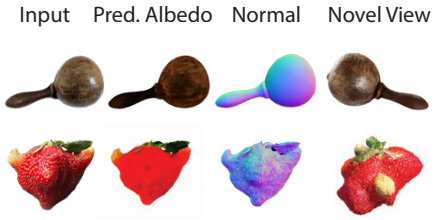

Figure C15. **Real-world results.** Example generations from casual smartphone captures of a shaker instrument and a strawberry.

References

[1] Mohammad Asim, Christopher Wewer, Thomas Wimmer, Bernt Schiele, and Jan Eric Lenssen. Met3r: Measuring multi-view consistency in generated images, 2024. 3

[2] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, Varun Jampani, and Robin Rombach. Stable Video Diffusion: Scaling Latent Video Diffusion Models to Large Datasets, 2023. 1

[3] Mark Boss, Zixuan Huang, Aaryaman Vasishta, and Varun Jampani. SF3D: Stable Fast 3D Mesh Reconstruction with UV-unwrapping and Illumination Disentanglement. *arXiv preprint*, 2024. 2, 3, 4, 5

[4] Tri Dao. FlashAttention-2: Faster Attention with Better Parallelism and Work Partitioning, 2023. 1

[5] Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and memory-efficient exact attention with IO-awareness. In *Advances in Neural*

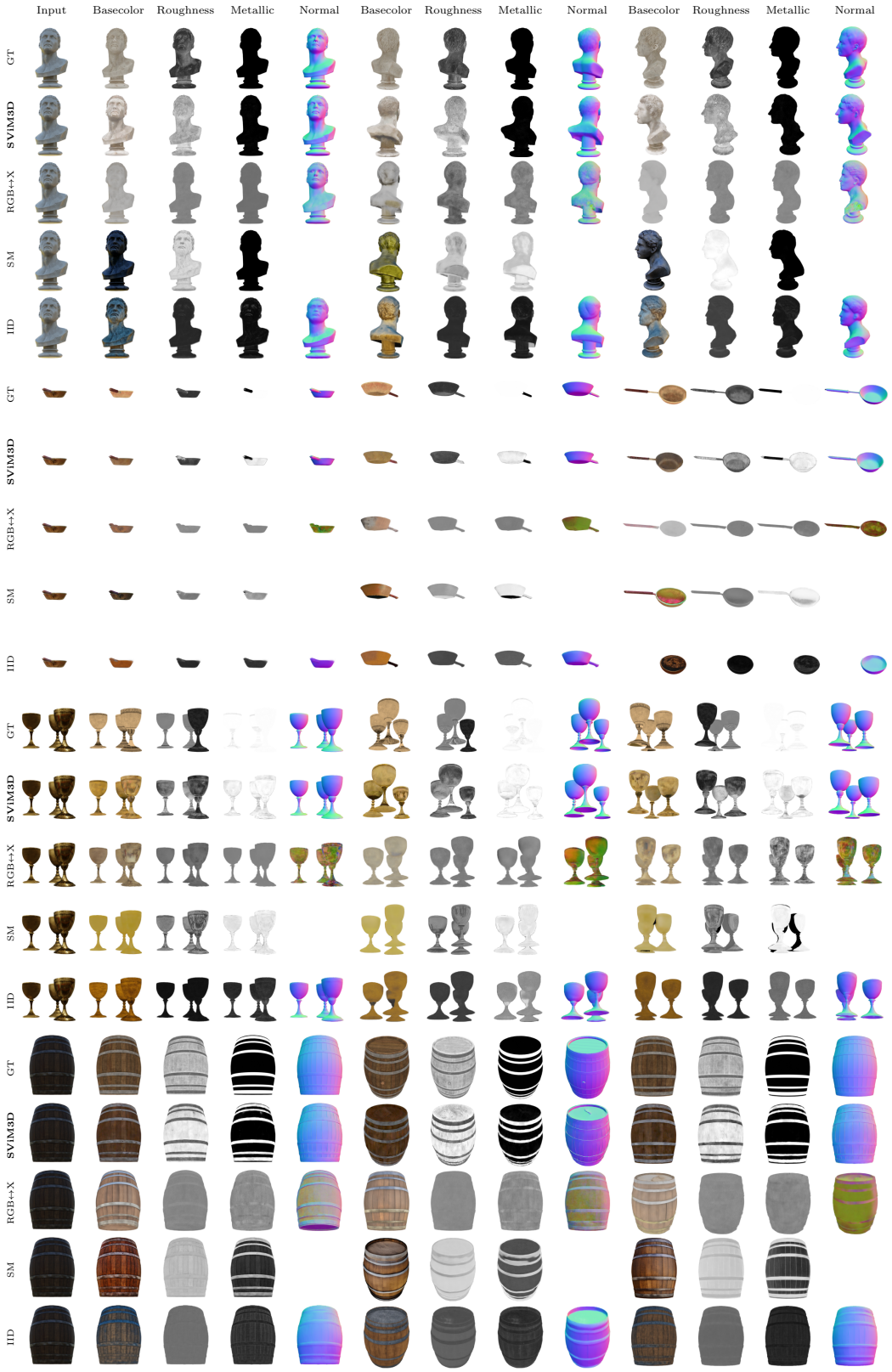

Figure C16. **Comparison of multi-view material generation on Poly Haven objects.** We compare generated materials of RGB↔X [24], StableMaterial (SM) of MaterialFusion [14] and Intrinsic Image Diffusion (IID) [13] based on SV3D [19] generations and SViM3D for three views around the object against GT renders.

Information Processing Systems, 2022. 1

[6] Laura Downs, Anthony Francis, Nate Koenig, Brandon Kinman, Ryan Hickman, Krista Reymann, Thomas B. McHugh, and Vincent Vanhoucke. Google scanned objects: A high-quality dataset of 3d scanned household items. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 2553–2560, 2022. 4, 5

[7] Kang Du, Zhihao Liang, and Zeyu Wang. GS-ID: Illumination Decomposition on Gaussian Splatting via Diffusion Prior and Parametric Light Source Optimization, 2024. 2

[8] Poly Haven. Poly Haven • Poly Haven — polyhaven.com. https://polyhaven.com/, 2024. [Accessed 22-08-2024]. 2, 3, 4, 5, 7

[9] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. In *NeurIPS 2021 Workshop on Deep Generative Models and Downstream Applications*, 2021. 1

[10] Yicong Hong, Kai Zhang, Jiuxiang Gu, Sai Bi, Yang Zhou, Difan Liu, Feng Liu, Kalyan Sunkavalli, Trung Bui, and Hao Tan. LRM: Large reconstruction model for single image to 3D. *arXiv preprint arXiv:2311.04400*, 2023. 2

[11] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural gaffer: Relighting any object via diffusion. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2

[12] Haian Jin, Yuan Li, Fujun Luan, Yuanbo Xiangli, Sai Bi, Kai Zhang, Zexiang Xu, Jin Sun, and Noah Snavely. Neural Gaffer: Relighting Any Object via Diffusion, 2024. arXiv:2406.07520 [cs]. 6, 7

[13] Peter Kocsis, Vincent Sitzmann, and Matthias Niessner. Intrinsic image diffusion for indoor single-view material estimation. *CVPR*, 2024. 2, 3, 8

[14] Yehonathan Litman, Or Patashnik, Kangle Deng, Aviral Agrawal, Rushikesh Zawar, Fernando De la Torre, and Shubham Tulsiani. MaterialFusion: Enhancing Inverse Rendering with Material Diffusion Priors, 2024. 2, 3, 5, 8

[15] Ben Mildenhall, Pratul Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *ECCV*, 2020. 1, 4

[16] Michael Niemeyer, Jonathan T. Barron, Ben Mildenhall, Mehdi S. M. Sajjadi, Andreas Geiger, and Noha Radwan. Regnerf: Regularizing neural radiance fields for view synthesis from sparse inputs. *CVPR*, 2022. 2

[17] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning transferable visual models from natural language supervision. In *Proceedings of the 38th International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021. 1

[18] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Bjorn Ommer. High-Resolution Image Synthesis with Latent Diffusion Models. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10674–10685, New Orleans, LA, USA, 2022. IEEE. 2

[19] Vikram Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz, Dmitrii Tochilkin, Christian Laforte, Robin Rombach, and Varun Jampani. SV3D: Novel multi-view synthesis and 3D generation from a single image using latent video diffusion. In *European Conference on Computer Vision*, 2024. 1, 2, 4, 8

[20] Vikram Voleti, Chun-Han Yao, Mark Boss, Adam Letts, David Pankratz, Dmitrii Tochilkin, Christian Laforte, Robin Rombach, and Varun Jampani. SV3D: Novel multi-view synthesis and 3D generation from a single image using latent video diffusion. In *European Conference on Computer Vision (ECCV)*, 2024. 1, 2

[21] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20697–20709, 2024. 3

[22] Xinyue Wei, Kai Zhang, Sai Bi, Hao Tan, Fujun Luan, Valentin Deschaintre, Kalyan Sunkavalli, Hao Su, and Zexiang Xu. MeshLRM: Large reconstruction model for high-quality mesh. *arXiv preprint arXiv:2404.12385*, 2024. 2

[23] Chong Zeng, Yue Dong, Pieter Peers, Youkang Kong, Hongzhi Wu, and Xin Tong. Dilightnet: Fine-grained lighting control for diffusion-based image generation. In *ACM SIGGRAPH 2024 Conference Papers*, 2024. 6, 7

[24] Zheng Zeng, Valentin Deschaintre, Iliyan Georgiev, Yannick Hold-Geoffroy, Yiwei Hu, Fujun Luan, Ling-Qi Yan, and Miloš Hašan. RGB¡-¿X: Image decomposition and synthesis using material- and lighting-aware diffusion models. *ArXiv*, 2024. 2, 3, 8

[25] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Scaling in-the-wild training for diffusion-based illumination harmonization and editing by imposing consistent light transport. In *The Thirteenth International Conference on Learning Representations*, 2025. 2, 6, 7

[26] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 2

[27] Michael Zollhöfer, Patrick Stotko, Andreas Görlitz, Christian Theobalt, Matthias Nießner, Reinhard Klein, and Andreas Kolb. State of the Art on 3D Reconstruction with RGB-D Cameras. *Computer Graphics Forum*, 37(2):625–652, 2018. 3