IDArb: Intrinsic Decomposition for Arbitrary Number of Input Views and Illuminations

Zhibing Li, Tong Wu, Jing Tan, Mengchen Zhang, Jiaqi Wang, Dahua Lin

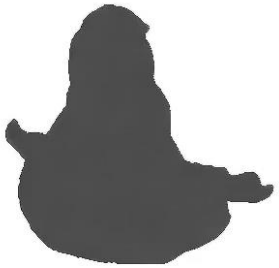
Input images

Albedo

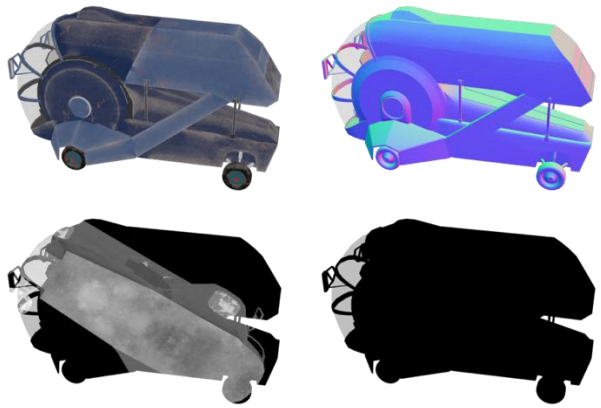
Normal

Metallic

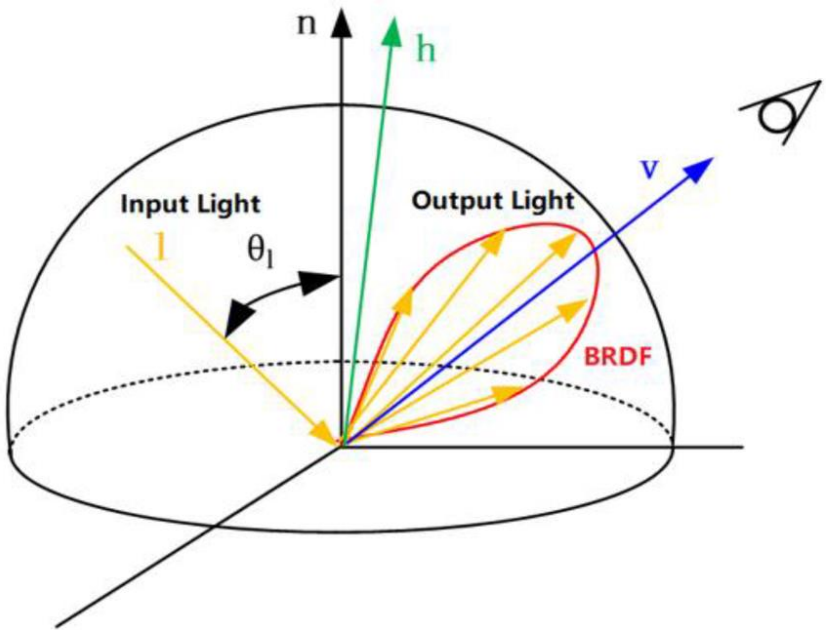
Roughness



Rendering $L_o(\omega_o; x) = \int_\Omega L_i(\omega_i; x) f_r(\omega_o, \omega_i; x)(\omega_i \cdot n) d\omega_i$



Intrinsic components



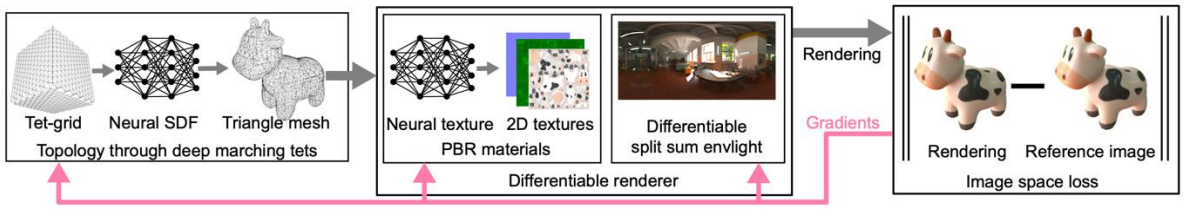
RGB colour

Inverse Rendering/ Intrinsic Decomposition

The colour we perceive results from a complex interaction between the incident light, the material properties, and the surface geometry.

How Existing Methods Address this Problem

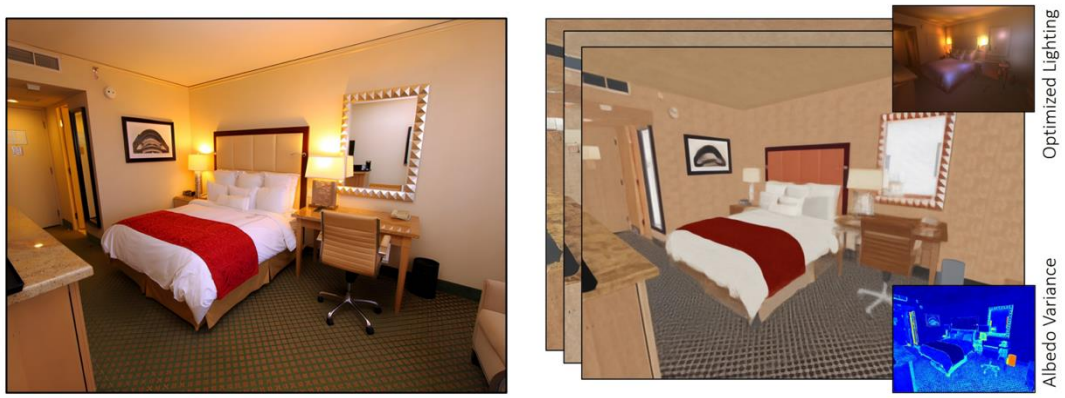
Optimization-Based Methods



(NVDiffRec, Jacob et al, 2022)

- Jointly reconstruct shape, materials, and lighting from multi-view images.
- Require dense multi-view inputs
- Face inherent ambiguity between lighting and materials

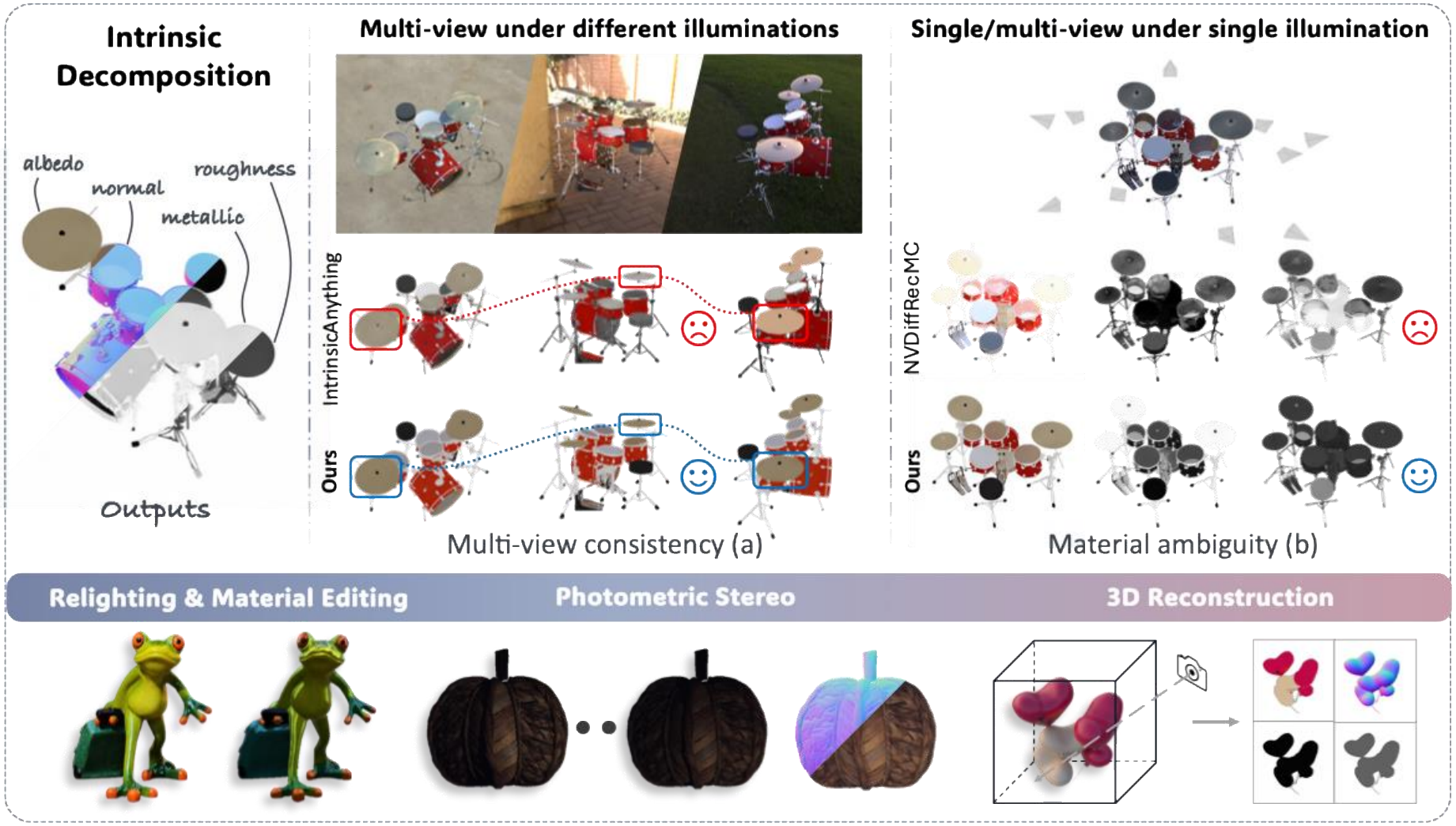
Learning-Based Methods



(IID, Kocsis et al, 2024)

- Leverage large datasets to learn priors
- Usually take a single image as input
- Decompose intrinsic components in a feed-forward way

Ours: Efficient, Versatile and Accurate Intrinsic Decomposition



- ✓ Efficient
  - Feed-forward method
- ✓ Versatile
  - Arbitrary input views
  - Different illuminations
- ✓ Accurate
  - SOTA performance
- ✓ Down-stream tasks

Multi-view & Multi-light Dataset

Our dataset contains **5.7 million** images, including rendered RGBs and their corresponding intrinsic components.

ABO

✓ Multiple illuminations

✗ Universal objects

G-Objaverse

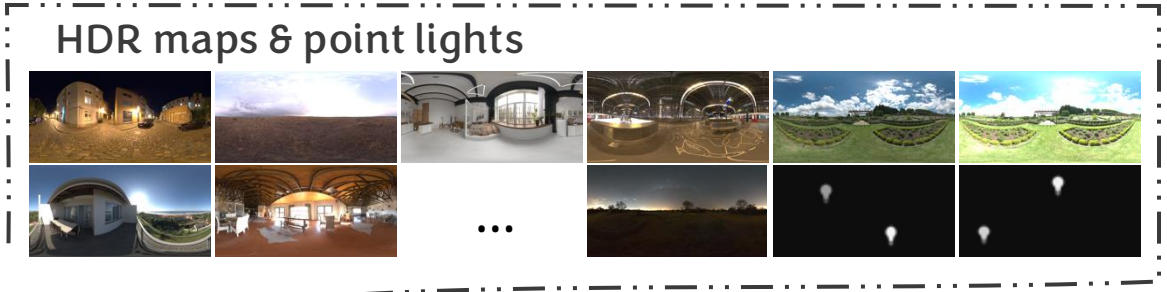
✗ Multiple illuminations

✓ Universal objects

Arb-Objaverse

✓ Multiple illuminations

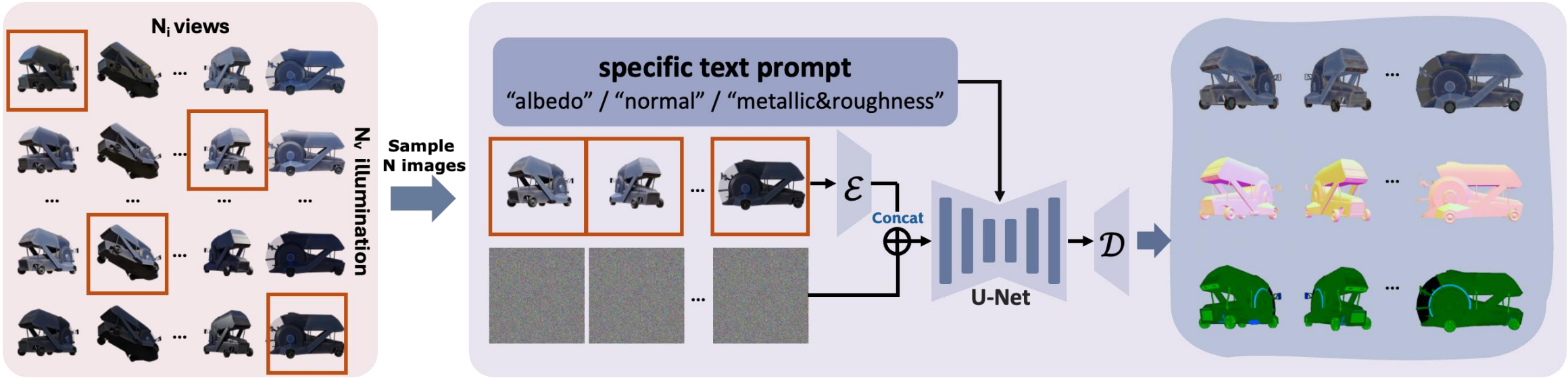
✓ Universal objects



Albedo/ Normal

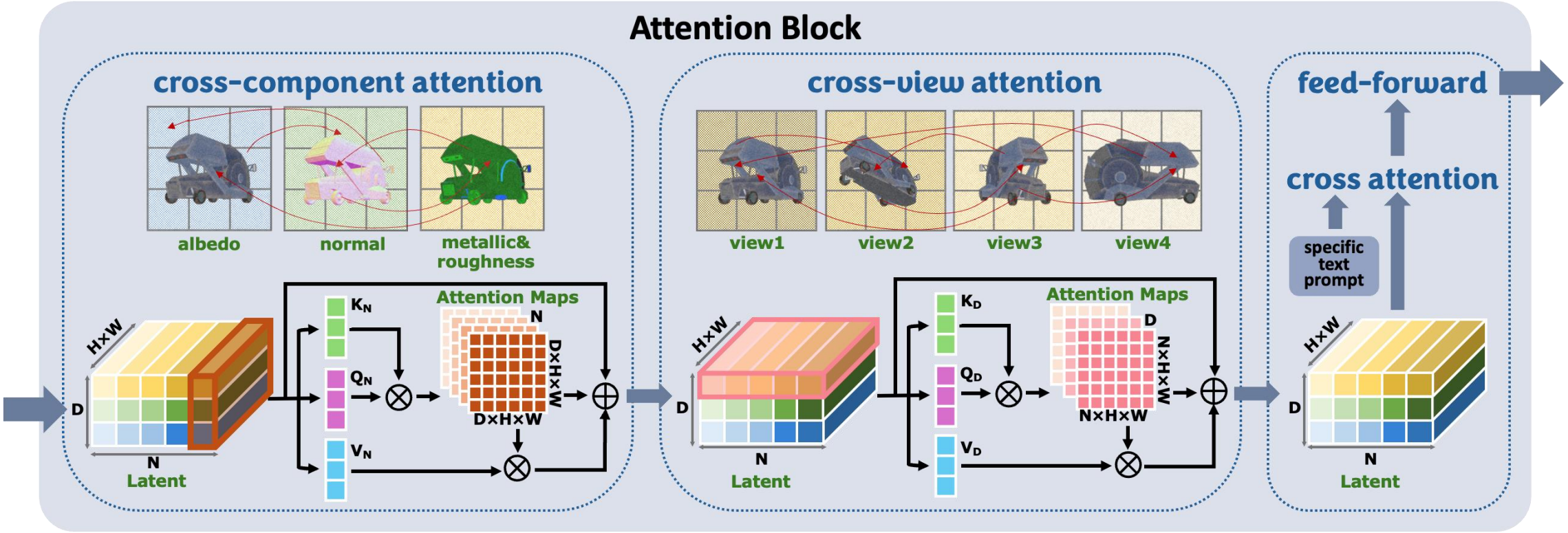
Metallic/ Roughness

Diffusion for Intrinsic Decomposition



- Frames intrinsic decomposition as a generative problem.
- Leverages prior knowledge from Stable Diffusion.

Cross-View-Cross-Component Module



- Ensures view-consistent and plausible estimation
- Eliminates the need for camera pose information

Comparison on **Real-World** Data

Quantitative Evaluations

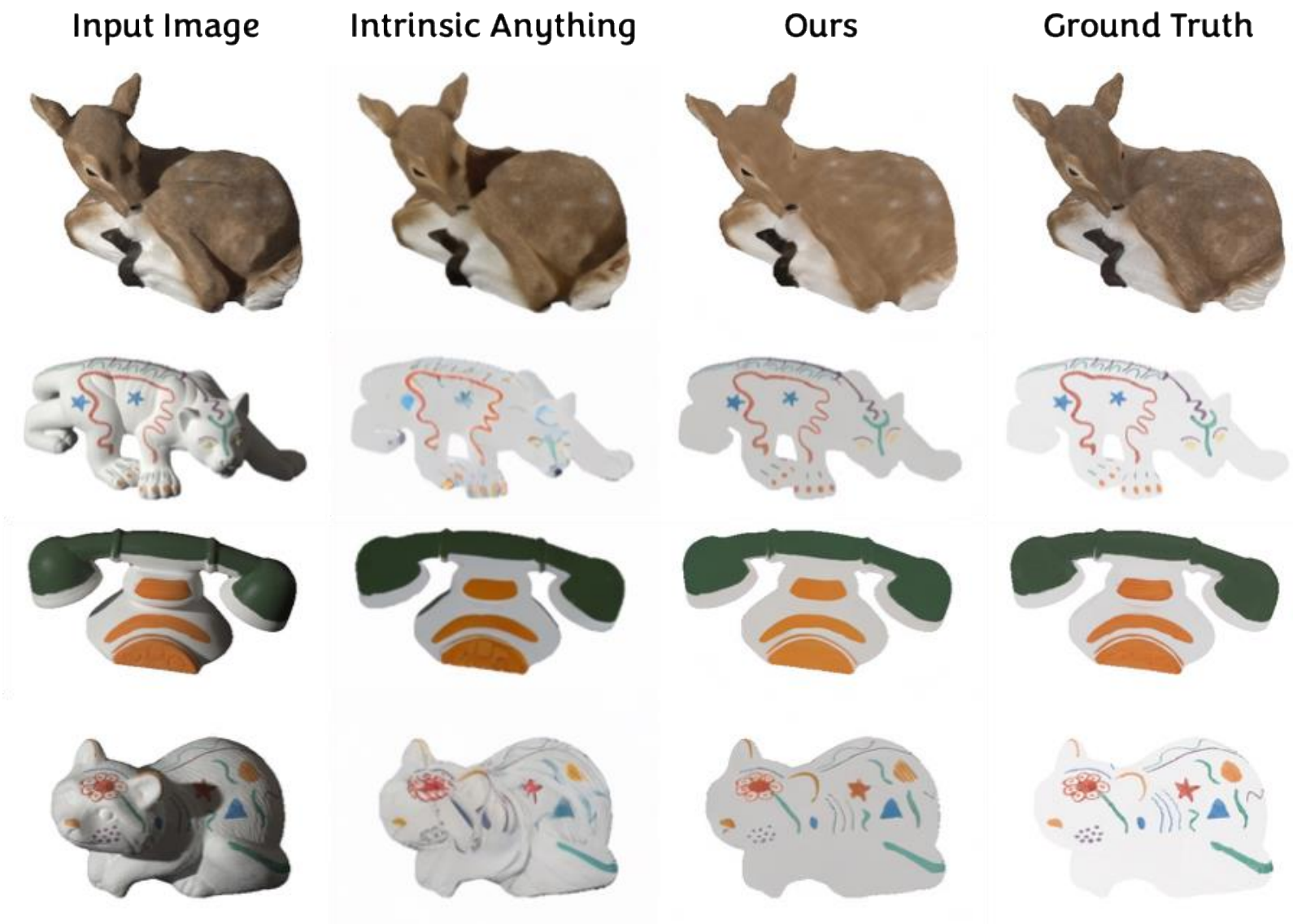
Quantitative comparisons on **MIT-Intrinsic**

	SSIM↑	PSNR↑	LPIPS↓
Ours	0.876	**27.98**	**0.117**
IntrinsicAnything	**0.896**	25.66	0.150

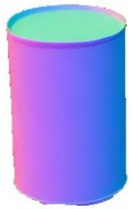
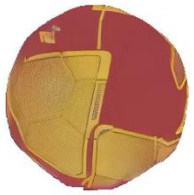
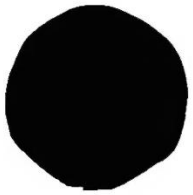
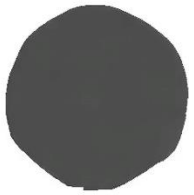
Quantitative comparisons on **Stanford-ORB**

	Normal	Albedo			Re-rendering			
	Cosine Distance↓	SSIM↑	PSNR↑	LPIPS ↓	PSNR-H↑	PSNR-L↑	SSIM↑	LPIPS ↓
Ours(single)	0.041	0.978	<u>41.30</u>	<u>0.039</u>	24.11	31.28	0.969	0.024
Ours(multi)	**0.029**	<u>0.978</u>	**41.46**	**0.038**	**24.36**	**31.43**	**0.970**	**0.024**
StableNormal	<u>0.038</u>							
IntrinsicNeRF		**0.981**	39.31	0.048				

Results on MIT-Intrinsic



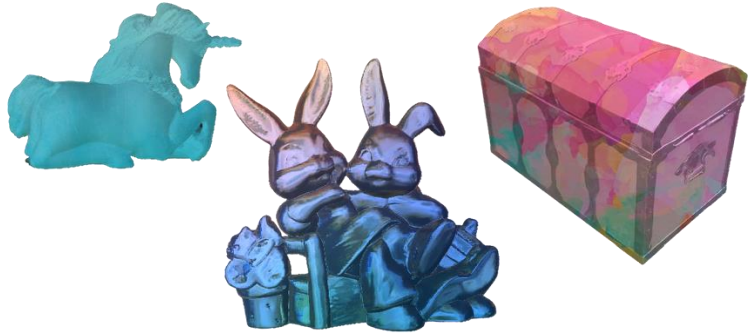
Results on Stanford-ORB

Input images	IntrinsicAnything	Albedo	Normal	Metallic	Roughness
					
					

Application: Relighting and Editing

(a) Captured real data

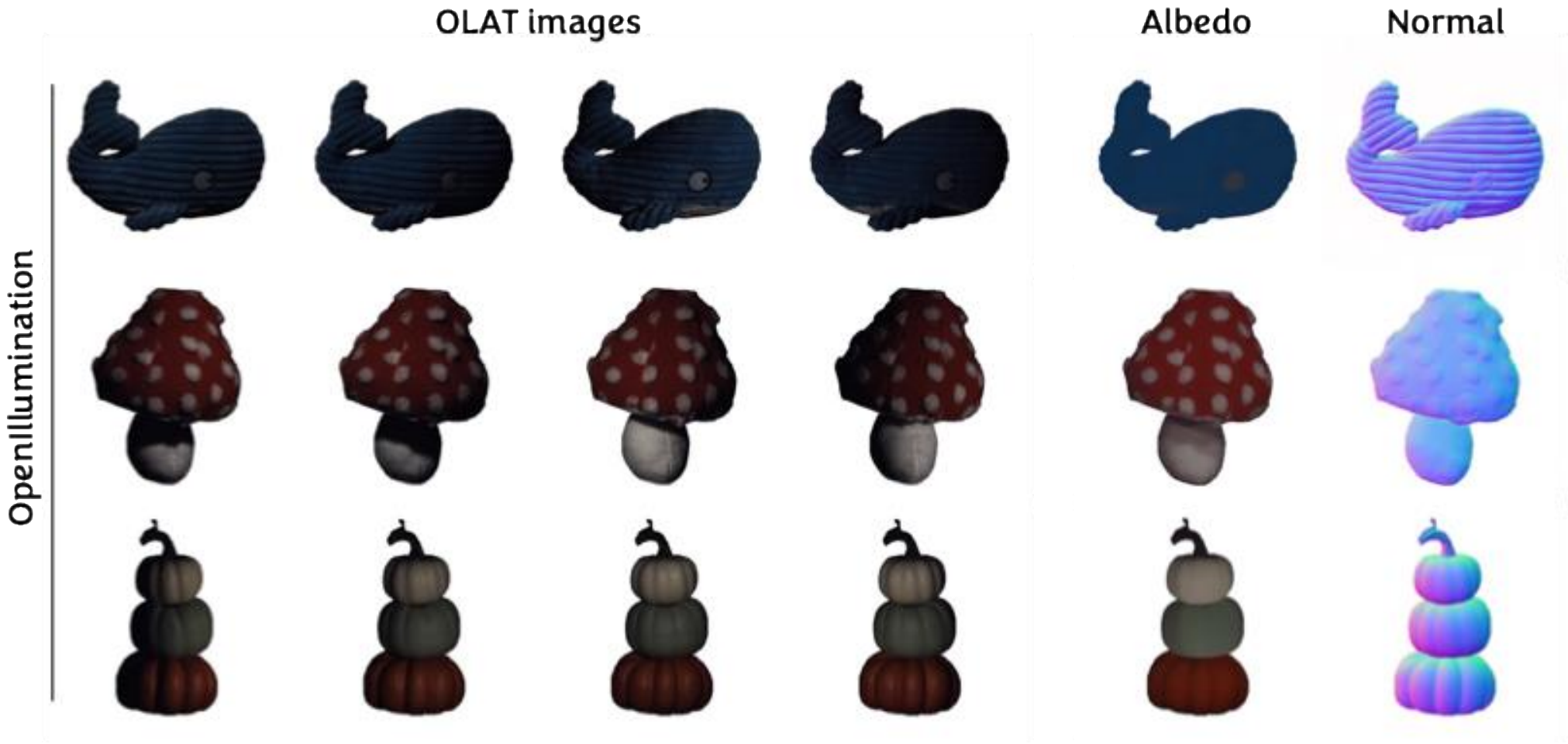
(b) Relighting



(c) Material editing

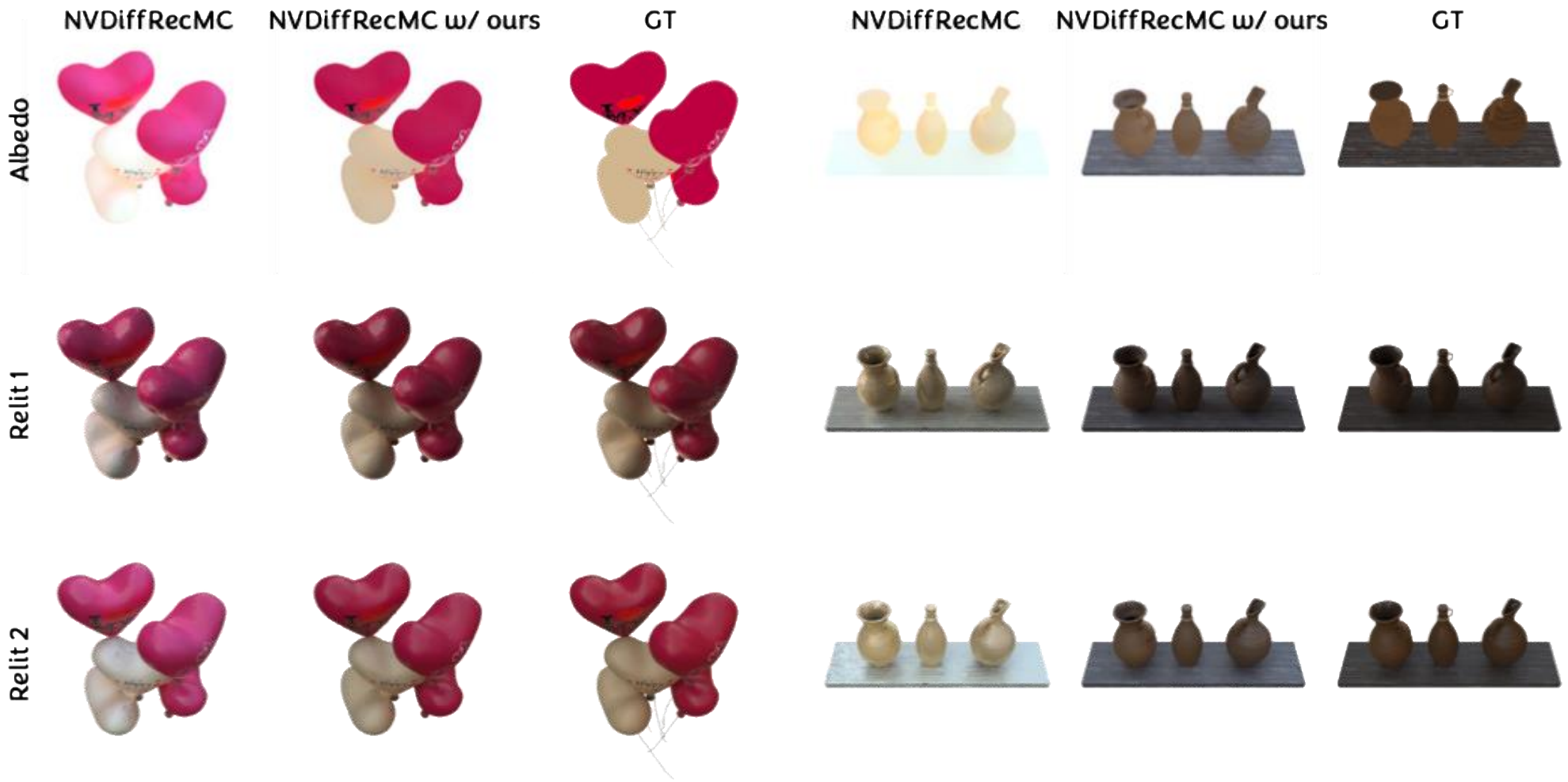
Application: Photometric Stereo

Our method delivers reasonable estimations under the **One-Light-At-a-Time** condition.



Application: 3D Reconstruction

Our method can be used as a prior to enhance 3D reconstruction.



Conclusion

Intrinsic decomposition for arbitrary number of images under varying light via diffusion model.

Project Page: https://lizb6626.github.io/IDArb/

Code: https://github.com/Lizb6626/IDArb/