Appendix

A. Implementation

For all experiments, we evaluate ReasonX on two representative intrinsic decomposition architectures: (i) PRISM [5], a rectified-flow diffusion transformer for RGBX prediction, and (ii) Marigold IID Lighting v1.1 [7], a diffusion-based model for joint albedo-irradiance estimation. We fine-tune both models using our Intrinsic-GRPO framework, with the MLLM judge providing relative intrinsic comparison rewards. Unless otherwise stated, all fine-tuning follows the setup described in the main paper: a training set of 10k COCO [9] images, T=15 denoising steps (T=50 steps during inference), AdamW optimizer with a learning rate of 10^{-5}, group size G=8, and SDE noise scale a=0.7. PRISM is conditioned using an empty text prompt in addition to the RGB image condition at all steps during training and post-training / PRISM-**X** inference.

For evaluation, we compare our ReasonX variants against their base models and a diverse set of strong supervised and synthetic-data methods. Additionally, we evaluate against OmniGen2 [12], where we follow the fine-tuning strategy described in the main paper to adapt its unified generation framework for multi-task intrinsic prediction. We evaluate all baselines using their publicly released models, official code bases, and default / highest performing settings.

B. MLLM Judge

B.1. Training Details

We use InternVL2.5-4B [4] as our base model for our MLLM judge and fine-tune it on (point-pair annotated RGB image, ground truth judgment) pairs. For training and evaluation, we use the same training / test split of the synthetic HyperSim [11] and InteriorVerse [14] datasets as our base model PRISM, use an image resolution of 512×512px and fine-tune the base MLLM for 5 epochs. For point-pair sampling, we set a minimum and maximum distance threshold of 20 and 350 between the points, and discard point pairs with less than 2% difference between them.

B.2. Absolute vs Relative Intrinsic Judgments

To assess whether relative comparisons are advantageous over absolute predictions, and whether such relative comparisons can be derived from absolute intrinsic predictions, we conduct an ablation study in which the base MLLM is fine-tuned to predict discretized absolute intrinsic values. For each modality, we partition the continuous ground-truth range into five ordered bins (e.g. for depth: very near, near, mid, far, very far; for albedo: very dark to very bright). We fine-tune separate MLLMs for each modality to perform 5-way classification on single annotated points, using the same synthetic training and test split as our relative judge. At test time, we evaluate both absolute accuracy and the quality of *derived* relative judgments by comparing the predicted bins for a pair of points and inferring their ordinal relationship. We exclude point pairs where both points belong to the same bin for simplicity.

Table 1. Performance of MLLM fine-tuned to perform absolute predictions (5-bin). We report accuracy and F1 score metrics for both the absolute predictions and relative judgments derived from them.

Modality	**Absolute**		**Relative (Derived)**	
	Accuracy	**F1 Score**	**Accuracy**	**Macro F1**
Depth	0.634	0.401	0.475	0.451
Normal	0.800	0.265	0.702	0.690
Albedo	0.449	0.426	0.483	0.488
Irradiance	0.541	0.437	0.418	0.387

As shown in Table 1, absolute predictions prove substantially more challenging for the MLLM, with accuracies ranging from 44.9% to 80.0% depending on the modality. Moreover, relative judgments obtained by differencing these absolute predictions perform significantly worse than those from our directly trained relative judge. For example, derived depth comparisons achieve only 47.5% accuracy compared to 96.2% with our relative model. These results validate our relative judgment approach as both more effective and better suited to the comparative nature of the task.

B.3. Judge Generalization and Sampling Ablations

To further validate the reliability of our MLLM judge, we evaluate it on both held-out synthetic splits (HyperSim + InteriorVerse test sets, denoted *Syn*) and real-world benchmarks (ETH3D for depth and IIW for albedo). We additionally ablate two design choices: the point-pair sampling strategy and the breadth of modality-specific training. We present all results in Tab. 2.

Sampling strategy. Our default point-pair sampler is guided by base model predicted intrinsics: we enforce minimum and maximum pixel distances of 20 and 350, and discard ambiguous pairs with less than 2% intrinsic difference (see Sec. 2.1). To ablate this choice, we train and evaluate a judge variant using *uniform random* point-pair sampling (*Random sampling*), which yields substantially lower accuracy and F1 across all modalities on the synthetic test set, confirming the importance of informed sampling.

Limited modality training. We train a *Limited* judge variant using only albedo and normal comparison data. This variant matches our full judge on albedo and normals, and partially transfers to unseen depth and irradiance modalities, indicating cross-modal correlations. However, the lim-

ited transfer underscores the benefit of modality-specific training for achieving high accuracy across all channels.

Evaluation on real benchmarks. On ETH3D (depth) and IIW (albedo), our judge achieves 91% and 81% accuracy respectively, substantially outperforming analytic judgments derived from PRISM predictions (72% and 68%), as shown in Table 2. This confirms that the judge learns modality-aware perceptual signals that generalize beyond the synthetic training domain, rather than relying on simple heuristics or memorization.

Table 2. Per modality judge accuracy and F1 scores on synthetic (Syn) and real benchmarks. *Random sampling* denotes a judge trained with uniform point-pair sampling. *Limited* denotes a judge trained only on albedo and normal comparisons; gray entries indicate unseen modalities. On real data, we compare our judge against analytic judgments derived from PRISM predictions.

Method	Data	Depth	Normal	Albedo	Irradiance
Our judge	Syn	0.96/0.96	0.94/0.93	0.89/0.89	0.88/0.88
Random sampling	Syn	0.79/0.77	0.78/0.76	0.81/0.79	0.75/0.73
Limited	Syn	0.62/0.62	0.94/0.93	0.90/0.90	0.60/0.44
Our judge	ETH3D	0.91/0.91	–	–	–
PRISM (analytic)	ETH3D	0.72/0.66	–	–	–
Our judge	IIW	–	–	0.81/0.80	–
PRISM (analytic)	IIW	–	–	0.68/0.64	–

B.4. Effect of ReasonX on Alignment Rewards

A central goal of ReasonX is to improve the agreement between (i) the MLLM judge's relative intrinsic judgments and (ii) the analytic relations implied by the model's predicted intrinsics. To illustrate this effect, we visualize the alignment before and after GRPO finetuning (Fig. 3). For each example, we show the judge MLLM's responses to point-pair annotated RGB image queries per modality-specific question alongside the corresponding analytic comparisons computed from the predicted intrinsics. In the top rows, ReasonX produces noticeably stronger alignment: the improved intrinsic predictions yield analytic relations that match the judge's comparative assessments more consistently. The bottom rows highlight challenging cases, such as highly textured materials (e.g. marble, wood), intricate patterns, mirrors and reflective surfaces, and scenes with dim or spatially complex lighting) where both the intrinsic predictions and the judge's assessments may be unreliable. These examples illustrate the inherent ambiguity of such scenes and mark natural failure modes shared by both analytic and perceptual signals. Overall, the pre/post ReasonX comparison supports that ReasonX training strengthens cross-modal consistency and enhances the agreement between analytic intrinsic relations and MLLM-based comparative reasoning.

C. Additional Results

We provide further quantitative and qualitative evaluations to complement the results in the main paper. First, we assess the generalization of ReasonX variants on the unseen synthetic ARAP dataset [1], following the evaluation protocol of CID [3]. As shown in Table 4, our PRISM-**X** and Marigold-**X** variants demonstrate improvements over the base PRISM and Marigold models as well as comparisons to strong baselines. Across all metrics, ReasonX yields consistent gains and achieves state-of-the-art performance.

We then evaluate our ReasonX variants on the synthetic HyperSim and InteriorVerse test sets against several strong baselines, and present our results in Tables 5 and 6. On both datasets, ReasonX consistently improves the quality of albedo, normals, and irradiance predictions. Finally, we include extensive qualitative comparisons for both PRISM-**X** and Marigold-**X** on challenging in-the-wild and out-of-distribution images (Figs. 1 and 2), illustrating the improved decomposition quality and cross-modal consistency achieved by our framework.

Table 3. Average MLLM judgment - predicted intrinsics alignment rewards pre- and post-optimization of our base models on our COCO test set.

Method	Albedo	Irradiance	Normals	Depth
Marigold (base variant) [6]	0.735	0.588	-	-
Marigold-**X**	0.801	0.694	-	-
PRISM [5]	0.701	0.597	0.618	0.591
PRISM-**X**	0.789	0.708	0.701	0.771

Table 4. Albedo decomposition quality evaluation on the synthetic ARAP Dataset [1]. Our ReasonX variants achieve 4.7–5.7% RMSE improvement over their base models.

Method	LMSE ↓	RMSE ↓	SSIM ↑
Zhu et al. [14]	0.029	0.184	0.729
Careaga and Aksoy [2]	0.035	0.162	0.751
Kocsis et al. [8]	0.030	0.160	0.738
RGB↔X [13]	0.025	0.177	0.738
Careaga and Aksoy [3]	0.021	0.149	0.796
OmniGen2 [12]	0.046	0.225	0.668
Marigold [7]	0.022	0.158	0.732
Marigold-**X**	0.021	0.149	0.781
Improvement	*+4.5%*	*+5.7%*	*+6.7%*
PRISM [5]	0.022	0.149	0.798
PRISM-**X**	**0.020**	**0.142**	**0.820**
Improvement	*+9.1%*	*+4.7%*	*+2.8%*

C.1. Cross-Modal Alignment

We perform a cyclic RGB reconstruction experiment on a subset of the COCO test set [9] consisting of 5000 images to evaluate the effect of ReasonX on the alignment between the estimated modalities. Given a source RGB image, we reconstruct the diffuse appearance of the target

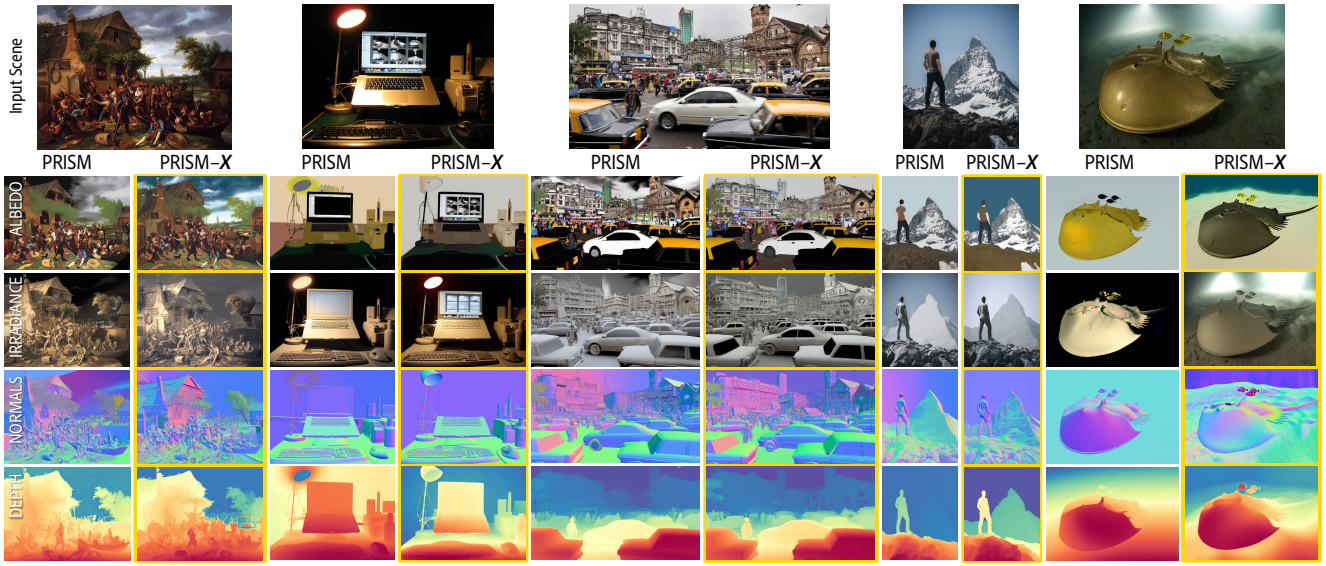

Figure 1. Additional intrinsic decomposition samples on challenging out-of-distribution images. Our PRISM-**X** significantly improves its base model PRISM across all intrinsic channels with respect to decomposition quality and in-the-wild generalization performance.

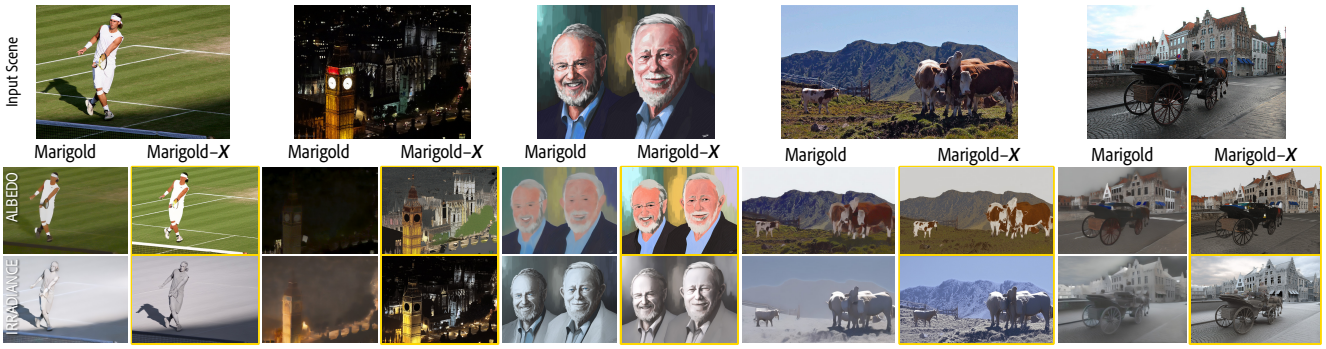

Figure 2. Intrinsic decomposition on challenging out-of-distribution images with the base Marigold (IID Lightning v1.1) model and our Marigold-**X**. Our ReasonX variant significantly improves its base model Marigold, especially in cases where the true albedo / base color is difficult to interpret.

Table 5. Quantitative evaluation of our method's intrinsic image decomposition performance on the HyperSim test set against various baselines.

Method	Albedo		Normal		Irradiance	
	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓
Zhu et al. [14]	11.7	0.54	16.5	0.45	–	–
Careaga and Aksoy [2]	13.5	0.34	–	–	14.5	0.22
Kocsis et al. [8]	12.1	0.41	–	–	–	–
RGB↔X [13]	17.4	0.18	19.8	0.18	14.1	0.22
Careaga and Aksoy [3]	17.1	0.21	–	–	17.2	0.21
OmniGen2 [12]	17.6	0.19	19.5	0.21	16.5	0.19
Marigold-Normals v1.1 [7]	–	–	19.8	0.20	–	–
Marigold [7]	18.2	0.22	–	–	17.6	0.26
Marigold-**X**	19.0	0.20	–	–	18.2	0.21
PRISM [5]	19.3	0.18	19.9	0.18	18.5	0.20
PRISM-**X**	**19.9**	**0.17**	**20.3**	**0.17**	**18.9**	**0.18**

Table 6. Quantitative evaluation of our method's intrinsic image decomposition performance on the InteriorVerse test set against various baselines.

Method	Albedo		Normal	
	PSNR ↑	LPIPS ↓	PSNR ↑	LPIPS ↓
Zhu et al. [14]	13.6	0.24	17.1	0.26
Careaga and Aksoy [2]	17.4	0.20	–	–
Kocsis et al. [8]	12.2	0.30	20.1	0.21
RGB↔X [13]	16.6	0.17	20.2	0.19
Careaga and Aksoy [3]	17.7	0.27	–	–
OmniGen2 [12]	13.9	0.21	19.0	0.18
Marigold-Normals v1.1[7]	–	–	20.2	0.16
Marigold [7]	19.5	0.19	–	–
Marigold-**X**	19.8	0.16	–	–
PRISM [5]	19.9	0.14	21.2	0.15
PRISM-**X**	**20.7**	**0.12**	**21.5**	**0.14**

image simply by multiplying the estimated albedo and irradiance maps. For our Marigold base model and ReasonX variant, we additionally apply a shift using the Marigold predicted residual ($RGB = A * I + R$). We then compare the reconstructed image to the source image and report the reconstruction metrics in Table 7. As shown in Table 7, both ReasonX variants exhibit improved cyclic reconstruction quality over their base models, indicating that our ReasonX framework not only improves individual modalities

Table 7. Cyclic RGB (RGB-to-X-to-RGB) reconstruction error on the real COCO test set. We perform intrinsic decomposition with our ReasonX variants and reconstruct RGB images using the estimation albedo and irradiance. Our ReasonX variants show improved reconstruction over their base models, indicating better cross-alignment of modalities.

Method	RMSE ↓	PSNR ↑	SSIM ↑	LPIPS ↓
Marigold (base variant) [6]	0.184	14.7	0.564	0.35
Marigold-**X**	0.178	15.0	0.581	0.32
Improvement	*+3.3%*	*+2.0%*	*+3.0%*	*+8.6%*
PRISM [5]	0.192	14.3	0.584	0.30
PRISM-**X**	0.184	14.7	0.610	0.28
Improvement	*+4.2%*	*+2.8%*	*+6.7%*	*–3.1%*

but also improves their cross-modal alignment. We note that these improvements are strongly correlated with more stable and accurate albedo estimation.

D. Ablation Studies

D.1. Analytical Depth-Normal Consistency Reward

To assess the contribution of the MLLM judge, we perform an ablation study in which the MLLM-based comparative reward used in our GRPO loop is replaced with a purely analytic geometric reward based on the consistency of the predicted depth and normals. We note that this ablation leverages the fact that PRISM predicts multiple intrinsic modalities jointly, making this type of consistency reward generally unavailable to intrinsic decomposition models. Given PRISM predicted depth $\hat{D}$ and normals $\hat{N}$, we estimate a depth map $\tilde{D}$ implied by the predicted normals using a least-squares Poisson integration. The reward penalizes disagreement between the two depth maps: $r_{\text{DN}} = -\|\hat{D} - \tilde{D}\|_1$.

This consistency signal enforces geometric coherence but provides no explicit supervision on albedo or irradiance, and does not incorporate perceptual cues. We finetune ReasonX using this analytic reward under the same GRPO setup as our main experiments and present the results in Table 8. While the depth-normal reward improves geometric consistency, it yields smaller gains than the full ReasonX framework and does not transfer improvements to other modalities (e.g., albedo). These findings highlight that purely analytic self-consistency is insufficient for robust improvement on real images, whereas MLLM-derived relative feedback offers a richer and more transferable signal.

D.2. Effect of KL Regularization

We use a KL regularization term during GRPO to prevent reward hacking and overly aggressive updates that drift from the pretrained model distribution [10]. To isolate its contribution, we consider a variant of ReasonX in which the KL term is removed from the GRPO objective. As such a setup leads to rapid model drift, we adopt an interleaved training strategy where we perform GRPO on real images without KL regularization, followed by standard flow-matching updates on the original synthetic training set of PRISM. This alternating procedure stabilizes optimization while allowing us to assess the effect of removing KL regularization. We report results on synthetic intrinsic benchmarks and depth estimation datasets in Tables 9 and 10 respectively.

Table 8. Depth and normals ablation using a depth-normal consistency (DN) reward. DN improves over the base PRISM model but is consistently weaker than ReasonX.

Method	NYU-v2		ETH3D		DIODE	
	AbsRel↓	δ_1↑	AbsRel↓	δ_1↑	Mean↓	$<11.5°$↑
PRISM	0.061	0.922	0.142	0.836	14.6	42.6
PRISM + DN reward	0.060	0.930	0.089	0.905	14.8	**43.0**
PRISM-**X**	**0.053**	**0.958**	**0.077**	**0.950**	**14.5**	41.3

Table 9. Ablation study: Impact of KL regularization on albedo, normals and irradiance decomposition across synthetic datasets.

Method	Albedo		Normal		Irradiance	
	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓	PSNR↑	LPIPS↓
HyperSim						
PRISM-**X**	19.9	0.17	20.3	0.17	18.9	0.18
w/o KL reg.	19.5	0.18	20.0	0.18	18.6	0.19
InteriorVerse						
PRISM-**X**	20.7	0.12	21.5	0.14	-	-
w/o KL reg.	20.1	0.14	21.2	0.15	-	-

Table 10. Ablation study: Impact of KL regularization on depth estimation.

Method	NYU-v2		ETH3D	
	AbsRel↓	δ_1↑	AbsRel↓	δ_1↑
PRISM-**X**	0.053	0.958	0.077	0.950
w/o KL reg.	0.059	0.930	0.135	0.864

As shown in Tables 9 and 10, despite the interleaved stabilization strategy, removing the KL term yields consistent degradation across albedo, normals, irradiance, and depth. The drop is most pronounced on ETH3D, a mixed indoor-outdoor dataset, which indicates reduced generalization capabilities compared to our PRISM-**X** model. These results highlight that KL regularization is required for stable optimization under comparative rewards, ensuring that improvements arise from meaningful alignment with the judge rather than from exploiting the reward structure or drifting away from the pretrained model manifold.

D.3. Effect of Number of Point Pairs

We conduct a controlled ablation on the number of point pairs used per training step by fine-tuning PRISM on a 1k-image subset for 5 epochs using depth comparisons only. We train variants with 10, 20, and 40 pairs per step and

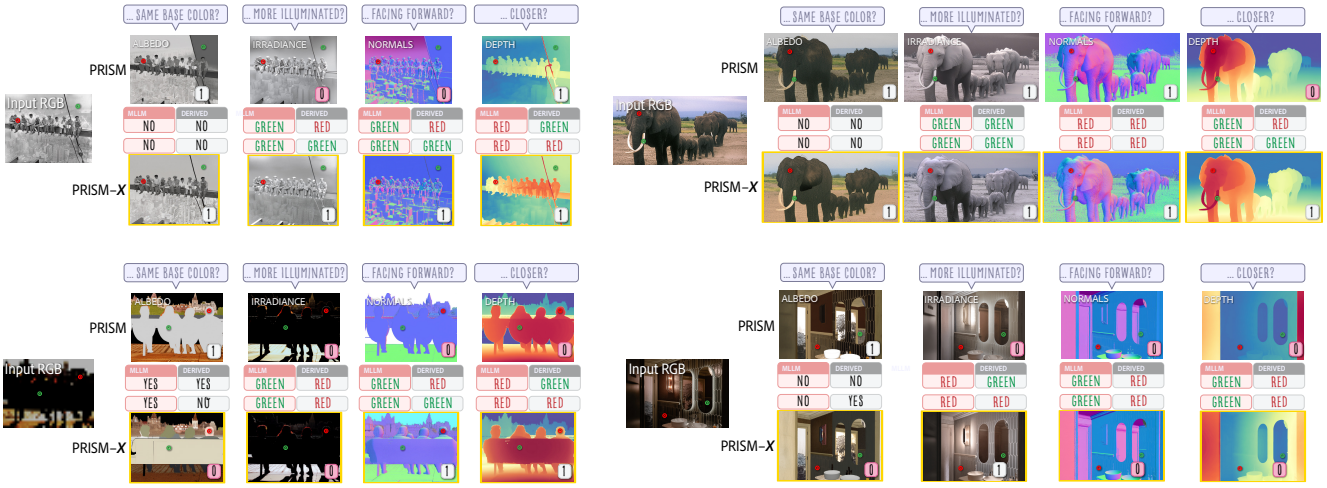

Figure 3. Alignment of MLLM judgments with analytically derived judgments from predicted intrinsics pre and post ReasonX training of base model PRISM. Top rows shows samples with accurate MLLM judgments where PRISM-**X** generated intrinsics result in improved alignment. Bottom row show failure cases where both the MLLM and analytically derived judgments might be unreliable due to dim lighting, intricate patterns and mirrored / reflective surfaces.

evaluate on NYU-v2 and ETH3D. As shown in Table 11, increasing the number of point pairs consistently improves performance across both datasets, with substantial gains from 10 to 20 pairs and further improvements at 40 pairs. The 40-pair variant remains slightly below the full PRISM-**X** model due to the reduced training set and depth-only supervision.

Table 11. Effect of the number of point-pair samples per training step. We fine-tune PRISM on a 1k-image subset for 5 epochs with depth comparisons only and evaluate on NYU-v2 and ETH3D.

Method	**NYU-v2**		**ETH3D**	
	AbsRel↓	δ_1↑	AbsRel↓	δ_1↑
PRISM	0.061	0.922	0.142	0.836
PRISM-**X** (10 pairs)	0.060	0.927	0.132	0.858
PRISM-**X** (20 pairs)	0.056	0.945	0.109	0.920
PRISM-**X** (40 pairs)	0.054	0.952	0.094	0.935
PRISM-**X** (full, 40 pairs)	**0.053**	**0.958**	**0.077**	**0.950**

References

[1] Nicolas Bonneel, Balazs Kovacs, Sylvain Paris, and Kavita Bala. Intrinsic Decompositions for Image Editing. *Computer Graphics Forum (Eurographics State of The Art Report)*, 2017. 2

[2] Chris Careaga and Yağız Aksoy. Intrinsic image decomposition via ordinal shading. *ACM Transactions on Graphics (ToG)*, 2023. 2, 3

[3] Chris Careaga and Yağız Aksoy. Colorful diffuse intrinsic image decomposition in the wild. *ACM Transactions on Graphics (ToG)*, 43(6), 2024. 2, 3

[4] Zhe Chen, Weiyun Wang, Yue Cao, Yangzhou Liu, Zhangwei Gao, Erfei Cui, Jinguo Zhu, Shenglong Ye, Hao Tian, Zhaoyang Liu, et al. Expanding performance boundaries of open-source multimodal models with model, data, and test-time scaling. *arXiv preprint arXiv:2412.05271*, abs/2412.05271, 2024. 1

[5] Alara Dirik, Tuanfeng Wang, Duygu Ceylan, Stefanos Zafeiriou, and Anna Frühstück. PRISM: A unified framework for photorealistic reconstruction and intrinsic scene modeling. *arXiv preprint arXiv:2504.14219*, 2025. 1, 2, 3, 4

[6] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 2, 4

[7] Bingxin Ke, Kevin Qu, Tianfu Wang, Nando Metzger, Shenflgyu Huang, Bo Li, Anton Obukhov, and Konrad Schindler. Marigold: Affordable adaptation of diffusion-based image generators for image analysis. *IEEE transactions on pattern analysis and machine intelligence*, PP, 2025. 1, 2, 3

[8] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for indoor single-view material estimation. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5198–5208, 2024. 2, 3

[9] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, 2014. 1, 2

[10] Jie Liu, Gongye Liu, Jiajun Liang, Yangguang Li, Jiaheng Liu, Xintao Wang, Pengfei Wan, Di Zhang, and Wanli Ouyang. Flow-GRPO: Training flow matching models via online rl. *arXiv preprint arXiv:2505.05470*, 2025. 4

[11] Mike Roberts, Jason Ramapuram, Anurag Ranjan, Atulit Kumar, Miguel Angel Bautista, Nathan Paczan, Russ Webb, and Joshua M Susskind. Hypersim: A photorealistic synthetic dataset for holistic indoor scene understanding. In

Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pages 10912–10922, 2021. 1

[12] Chenyuan Wu, Pengfei Zheng, Ruiran Yan, Shitao Xiao, Xin Luo, Yueze Wang, Wanli Li, Xiyan Jiang, Yexin Liu, Junjie Zhou, et al. OmniGen2: Exploration to advanced multimodal generation. *arXiv preprint arXiv:2506.18871*, 2025. 1, 2, 3

[13] 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. In *ACM SIGGRAPH Conference Proceedings*, 2024. 2, 3

[14] Jingsen Zhu, Fujun Luan, Yuchi Huo, Zihao Lin, Zhihua Zhong, Dianbing Xi, Rui Wang, Hujun Bao, Jiaxiang Zheng, and Rui Tang. Learning-based inverse rendering of complex indoor scenes with differentiable monte carlo raytracing. In *ACM SIGGRAPH Asia Conference Proceedings*. ACM, 2022. 1, 2, 3