DNF-Intrinsic: Deterministic Noise-Free Diffusion for Indoor Inverse Rendering

Supplementary Material

1. Details of Generative Renderer

Our generative renderer aims to take the scene's intrinsic properties as input and generate the input image as output. It is fine-tuned from Stable Diffusion v2, with a trainable ControlNet conditioned on 9-channel intrinsic properties (albedo, metallic, roughness, normal, and depth). The model is optimized by the AdamW optimizer with a learning rate of 1e-4 and a weight decay rate of 1e-2. Unlike traditional physics-based renderers, our generative renderer does not require environmental lighting as input and can still produce realistic rendered images. Besides, as shown in Figure 1, our generative renderer enables various potential applications, including uncontrollable relighting, material editing, and object removal. Specifically, given the scene's intrinsic properties, our generative renderer can generate images with lighting conditions different from the original image, although this is uncontrollable since the lighting is represented by Gaussian noise. Additionally, we can manually adjust the albedo color for material editing or remove the target object from the intrinsic properties for object removal.

2. Details of Reconstruction Loss

Similar to Dreamfusion [8], we first compute the gradient of $\mathcal{L} = \mu_\theta(I) \cdot \text{stop\_gradient}\left[\mathcal{L}_{rec}\right]$ with respect to the parameters θ of the inverse rendering model, where $\mu_\theta(I)$ denotes the predicted intrinsic property and I is the input image. Then, the parameters can be updated using an optimizer. The pseudo-code is shown in Figure 2.

3. Details of Application

In order to achieve controllable lighting editing or object insertion, we provide a solution to optimize the environmental lighting based on the high-quality intrinsic properties predicted by our method. Specifically, similar to [6], we use pre-integrated environment lighting parameterized by Spherical Gaussians (SG) for global illumination, along with 48 SG emission profiles to represent point lights. The SG parameters are then determined by optimizing a L2 loss between the re-rendered output and the input. After fitting, the parameters of the light sources, such as color or intensity, can be adjusted independently to achieve controllable lighting editing. Meanwhile, with the environmental lighting of the scene, we can render an image with the inserted 3D object using the estimated lighting to achieve realistic object insertion.

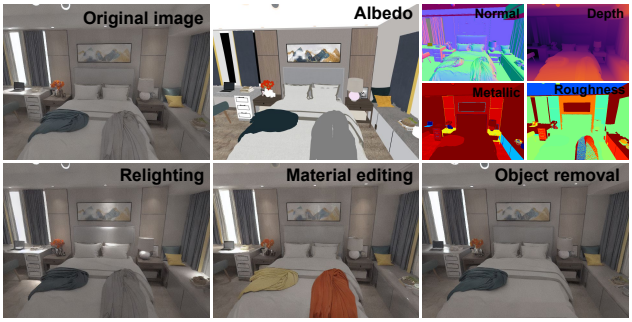


Figure 1. Examples of our generative renderer.

4. More Visual Comparison on Synthetic Data

Figures 3, 4, 5, 6, 7, 8, 9, and 10 provide more visual comparison of inverse rendering on the synthetic InteriorVerse dataset [13]. As shown, our method clearly outperforms previous methods on material and geometry estimation.

5. More Visual Comparison on Real Data

Figures 11, 12, 13, 14, 15, 16, 17, and 18 provide more visual comparison on material (albedo, metallic, and roughness) and geometry estimation (depth and normal). Comparing the results, it is clear that our method outperforms current state-of-the-art inverse rendering methods, and is able to produce comparable or even better results than specialized methods for material and geometry estimation.

6. More Application Results

We in Figure 19 provide more virtual object insertion results, while Figure 20 gives additional results on material and lighting editing. As shown, these application results are visually natural, manifesting the robustness of our predicted intrinsic properties.

References

[1] Gwangbin Bae and Andrew J Davison. Rethinking inductive biases for surface normal estimation. In *CVPR*, 2024. 8, 15

[2] Xi Chen, Sida Peng, Dongchen Yang, Yuan Liu, Bowen Pan, Chengfei Lv, and Xiaowei Zhou. Intrinsicanything: Learning diffusion priors for inverse rendering under unknown illumination. In *ECCV*, 2024. 3, 4, 5, 6, 10, 11, 12, 13

[3] Partha Das, Sezer Karaoglu, and Theo Gevers. Pie-net: Photometric invariant edge guided network for intrinsic image decomposition. In *CVPR*, 2022. 10, 11, 12, 13

[4] Xiao Fu, Wei Yin, Mu Hu, Kaixuan Wang, Yuexin Ma, Ping Tan, Shaojie Shen, Dahua Lin, and Xiaoxiao Long. Geowizard: Unleashing the diffusion priors for 3d geometry estimation from a single image. In *ECCV*, 2024. 15, 16

```
params = IR_model.init() # inverse rendering model
opt_state = optimizer.init(params)
generative_renderer = diffusion.load_ControlNet()
for nstep in iterations:
  t = random.uniform(0., 1.)
  alpha_t, sigma_t = diffusion.get_coeffs(t)
  eps = random.normal(img_shape) # sample a noise from Gaussian distribution, representing the unknown lighting
  intrinsics = IR_model(input_image) # Get an one-step intrinsic properties observation.
  x = input_image
  z_t = alpha_t * x + sigma_t * eps # Diffuse observation.
  epshat_t = generative_renderer.epshat(z_t, intrinsics, t) # Score function evaluation.
  L_rec = epshat_t - eps # generative reconstuction loss
  g = grad(dot(stopgradient[L_rec], intrinsics), params)
  params, opt_state = optimizer.update(g, opt_state) # Update params with optimizer.
return params
```

Figure 2. Pseudo code for the SDS-based reconstruction loss via the generative renderer that defines a differentiable mapping from parameters to intrinsic properties.The gradient g is computed without backpropagating through the generative renderer's U-Net. We used the `stopgradient` operator to express the loss, but the gradient of the parameter can also be easily computed as: `g = matmul(`L_{rec}`, grad(intrinsics, params))`.

[5] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *CVPR*, pages 9492–9502, 2024. 9, 16

[6] Peter Kocsis, Vincent Sitzmann, and Matthias Nießner. Intrinsic image diffusion for single-view material estimation. In *CVPR*, 2024. 1, 3, 4, 5, 6, 10, 11, 12, 13, 14

[7] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *CVPR*, pages 2475–2484, 2020. 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 15, 16

[8] Ben Poole, Ajay Jain, Jonathan T Barron, and Ben Mildenhall. Dreamfusion: Text-to-3d using 2d diffusion. *arXiv preprint arXiv:2209.14988*, 2022. 1

[9] Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything: Unleashing the power of large-scale unlabeled data. In *CVPR*, 2024. 9, 16

[10] Lihe Yang, Bingyi Kang, Zilong Huang, Zhen Zhao, Xiaogang Xu, Jiashi Feng, and Hengshuang Zhao. Depth anything v2. *arXiv:2406.09414*, 2024. 9, 16

[11] Chongjie Ye, Lingteng Qiu, Xiaodong Gu, Qi Zuo, Yushuang Wu, Zilong Dong, Liefeng Bo, Yuliang Xiu, and Xiaoguang Han. Stablenormal: Reducing diffusion variance for stable and sharp normal. *ACM Transactions on Graphics (TOG)*, 2024. 8, 15

[12] 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 2024 Conference Papers*, pages 1–11, 2024. 3, 4, 5, 6, 7, 8, 10, 11, 12, 13, 14, 15

[13] 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 *SIGGRAPH Asia 2022 Conference Papers*, pages 1–8, 2022. 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16

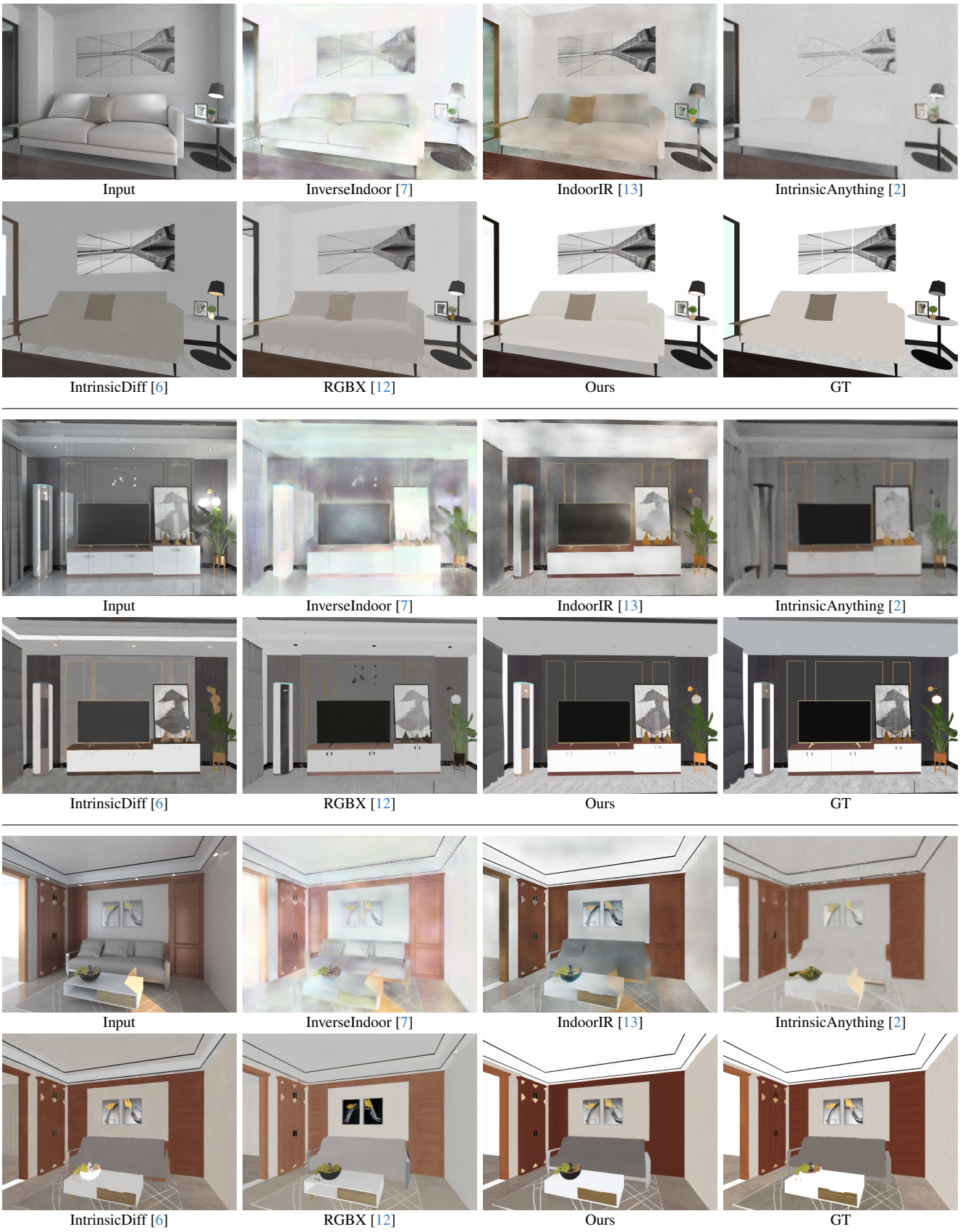


Figure 3. **More qualitative comparison of albedo estimation on the synthetic InteriorVerse dataset [13].**

Figure 4. **More qualitative comparison of albedo estimation on the synthetic InteriorVerse dataset [13].**

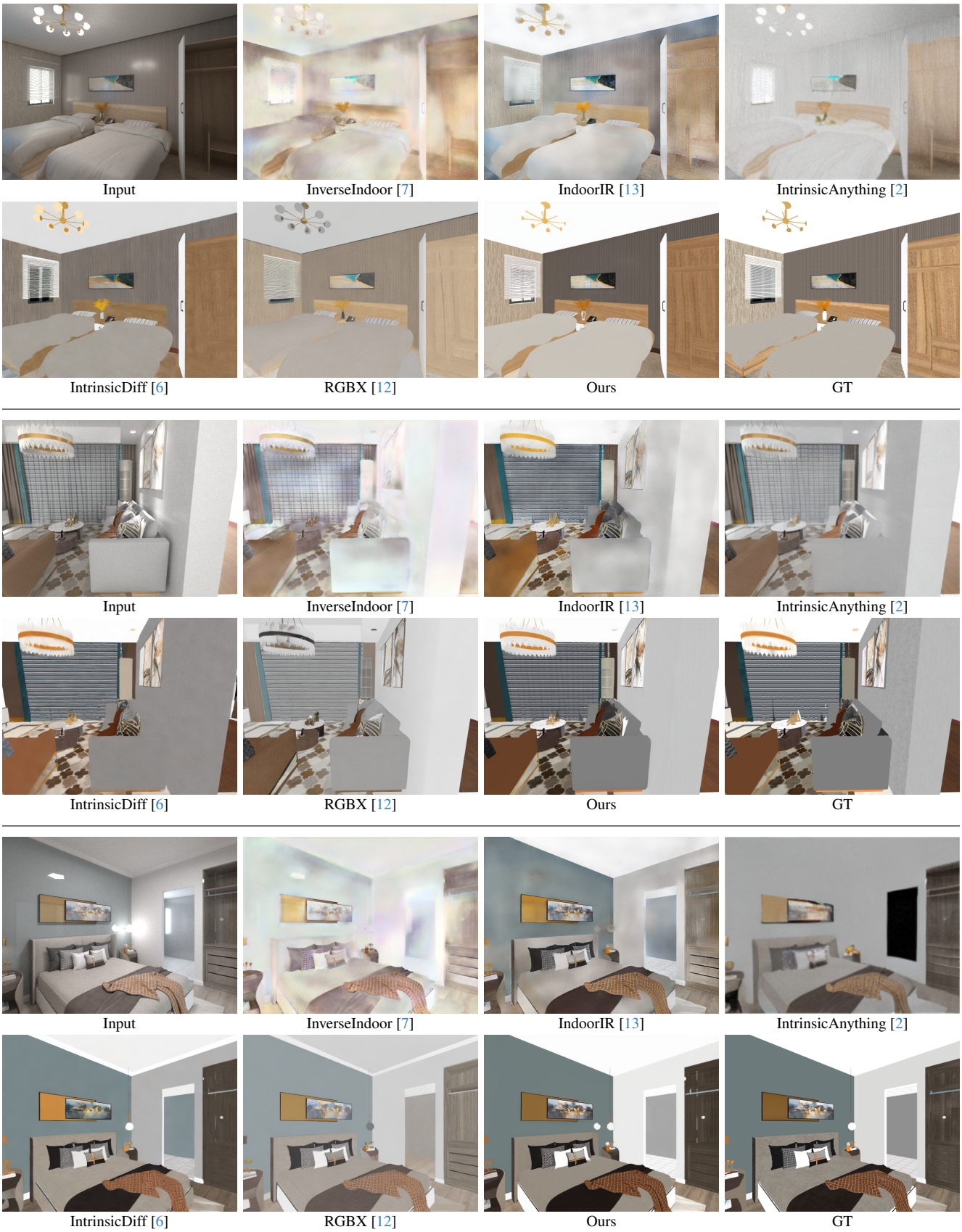


Figure 5. **More qualitative comparison of albedo estimation on the synthetic InteriorVerse dataset [13].**

Figure 6. **More qualitative comparison of albedo estimation on the synthetic InteriorVerse dataset [13].**

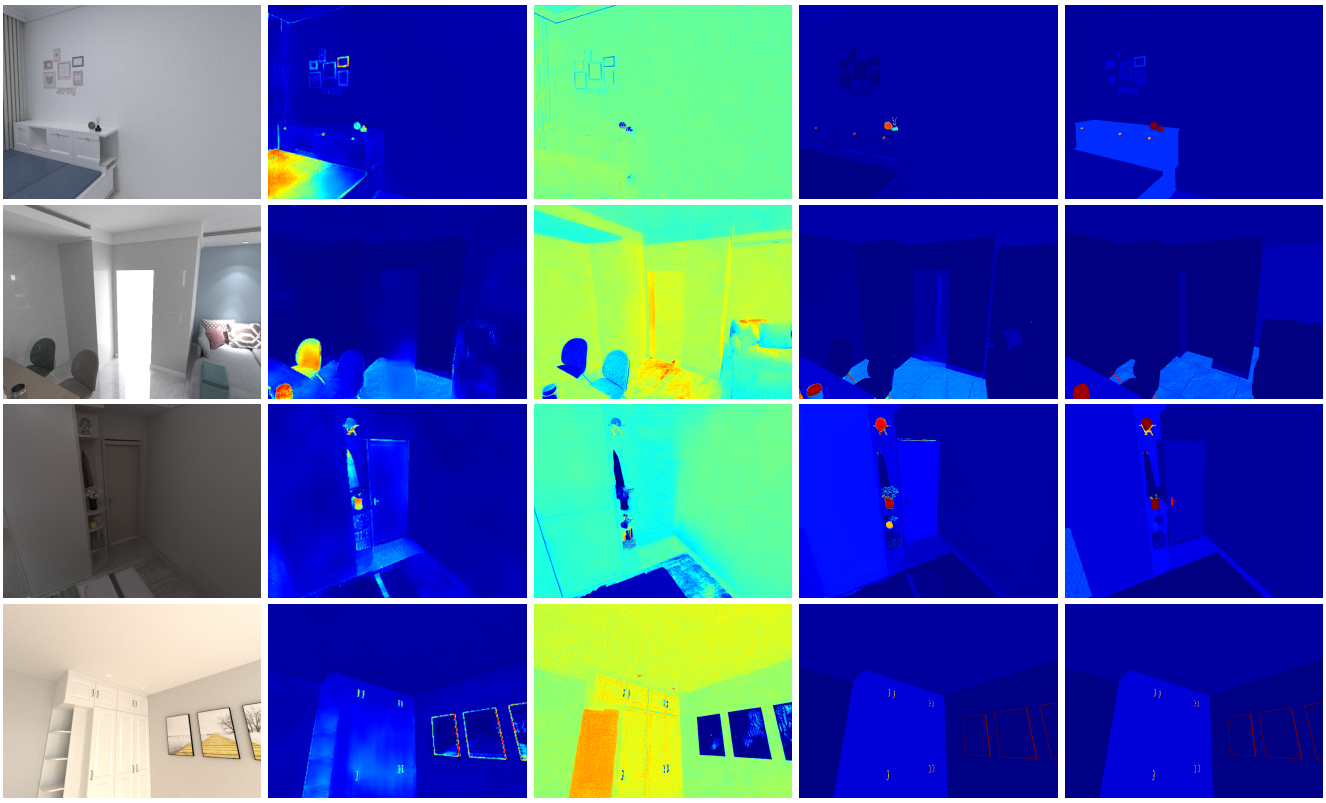


Figure 7. **More qualitative comparison of metallic estimation on the synthetic InteriorVerse dataset [13].**

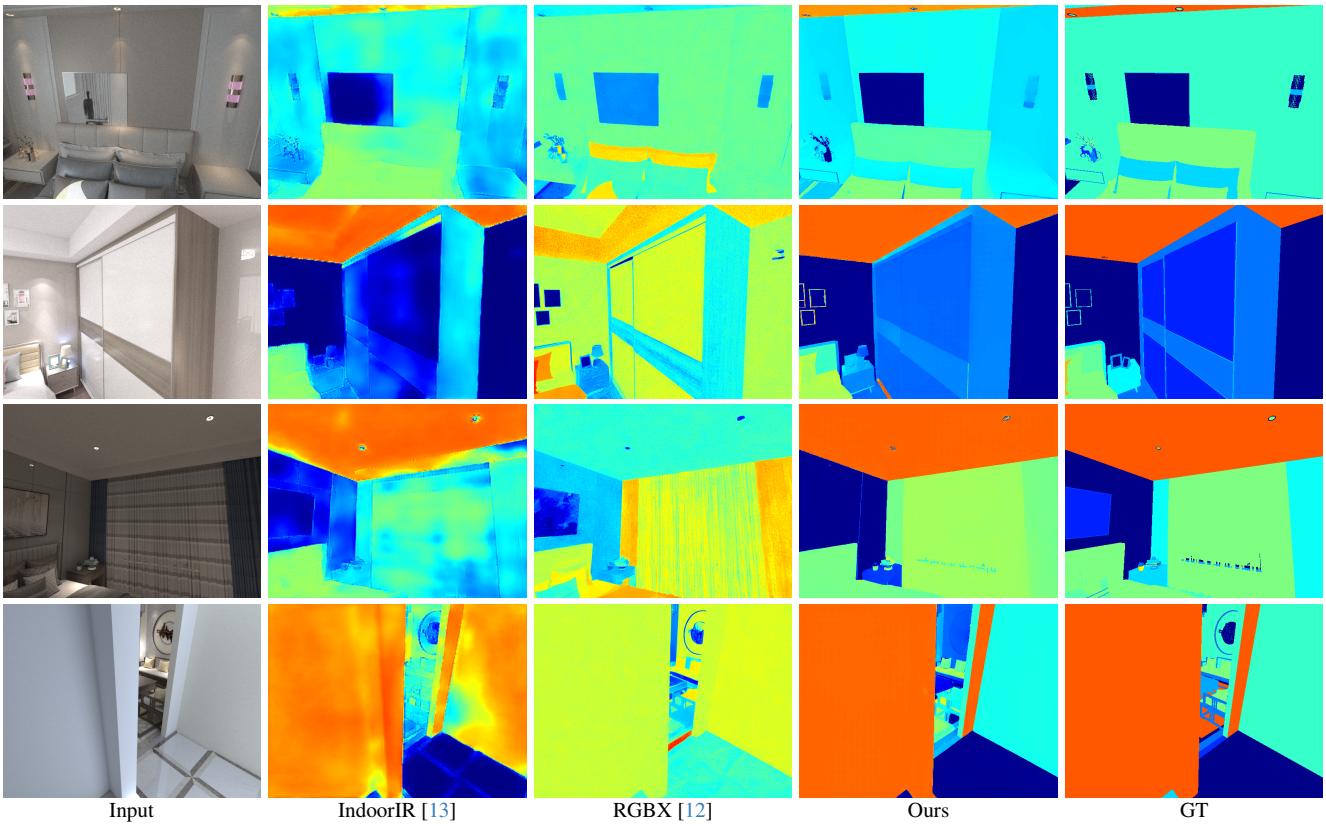


Figure 8. **More qualitative comparison of roughness estimation on the synthetic InteriorVerse dataset [13].**

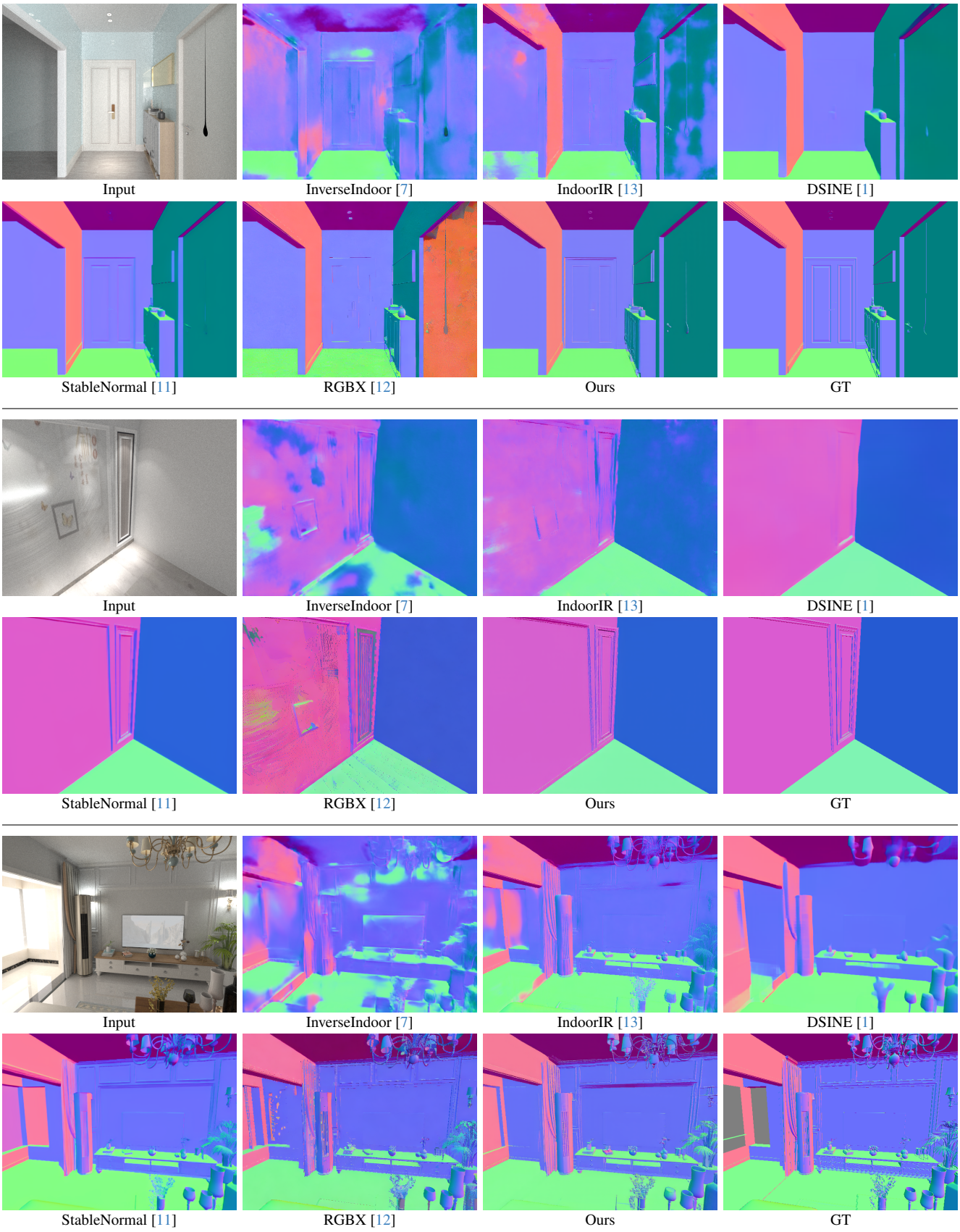


Figure 9. **More qualitative comparison of normal estimation on the synthetic InteriorVerse dataset [13].**

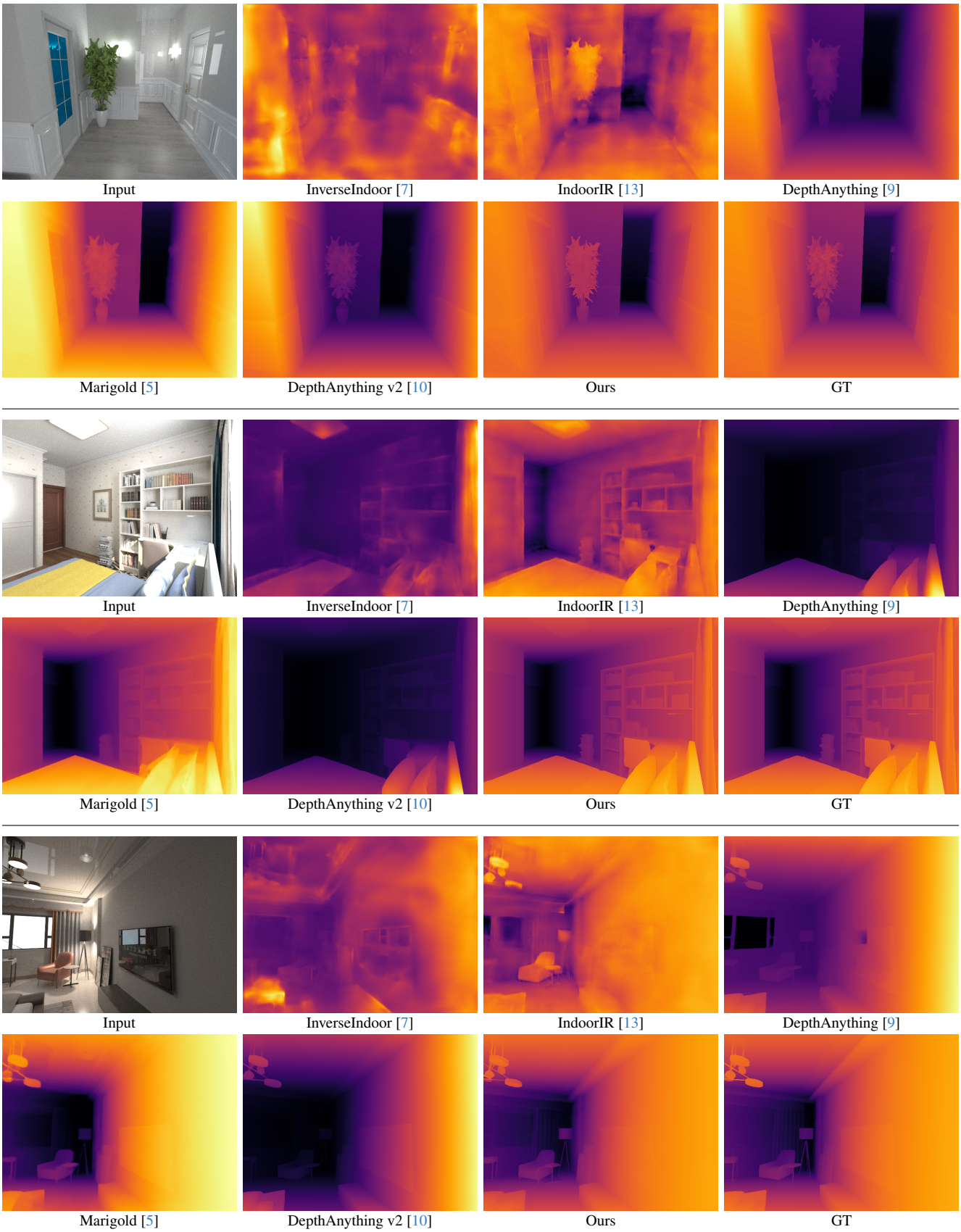


Figure 10. **More qualitative comparison of depth estimation on the synthetic InteriorVerse dataset [13].**

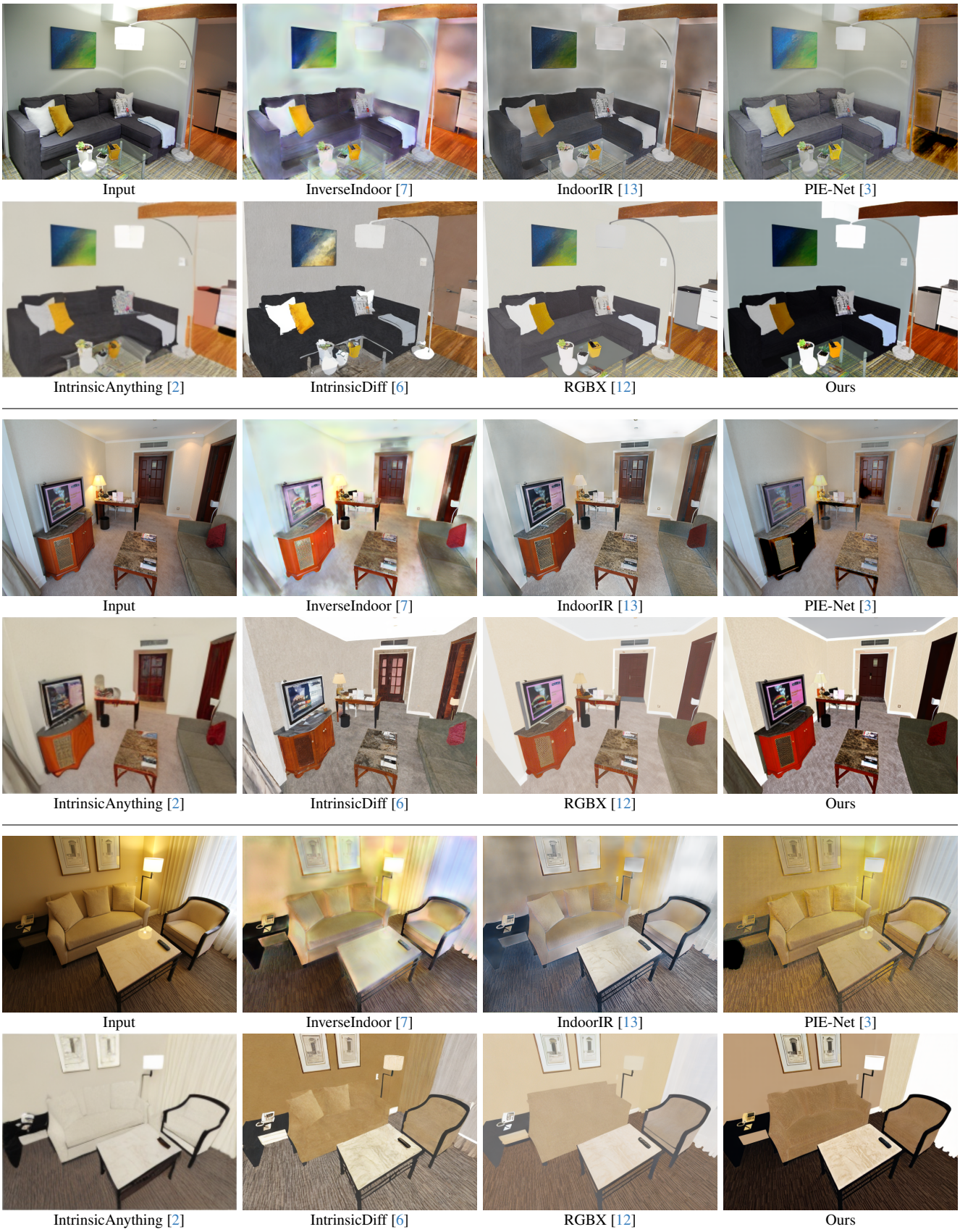


Figure 11. **More qualitative comparison of albedo estimation on real-world images.**

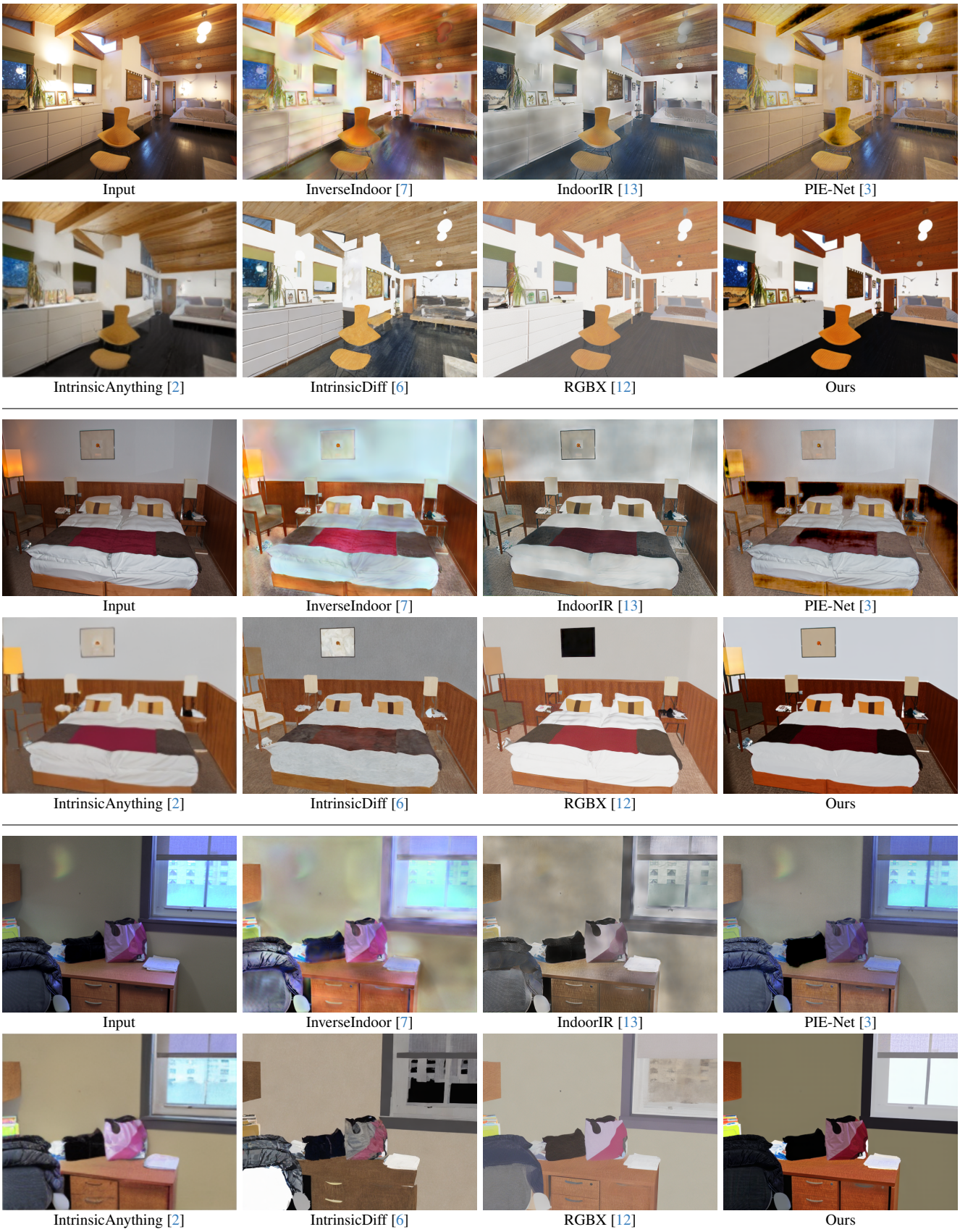


Figure 12. **More qualitative comparison of albedo estimation on real-world images.**

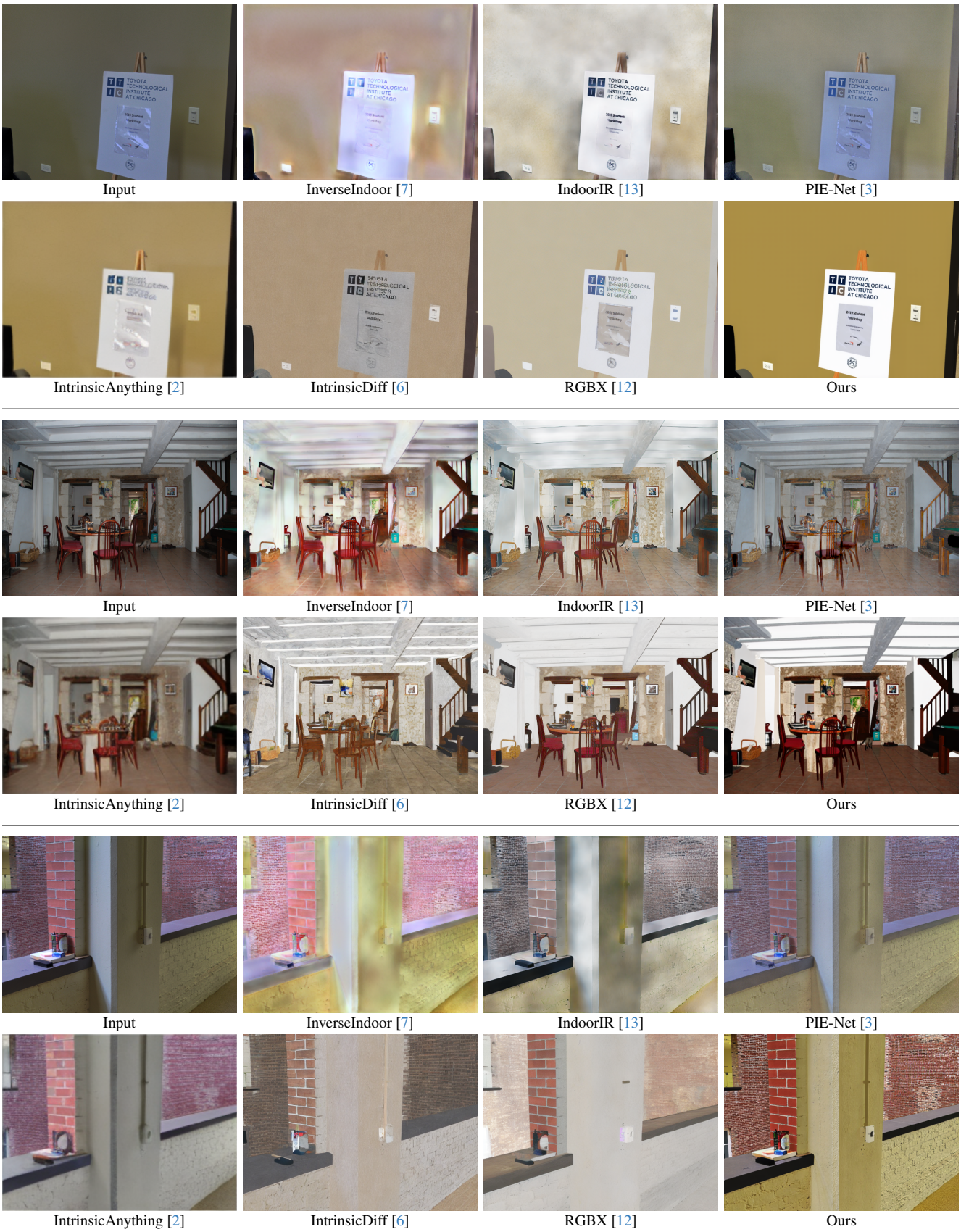


Figure 13. **More qualitative comparison of albedo estimation on real-world images.**

Figure 14. **More qualitative comparison of albedo estimation on real-world images.**

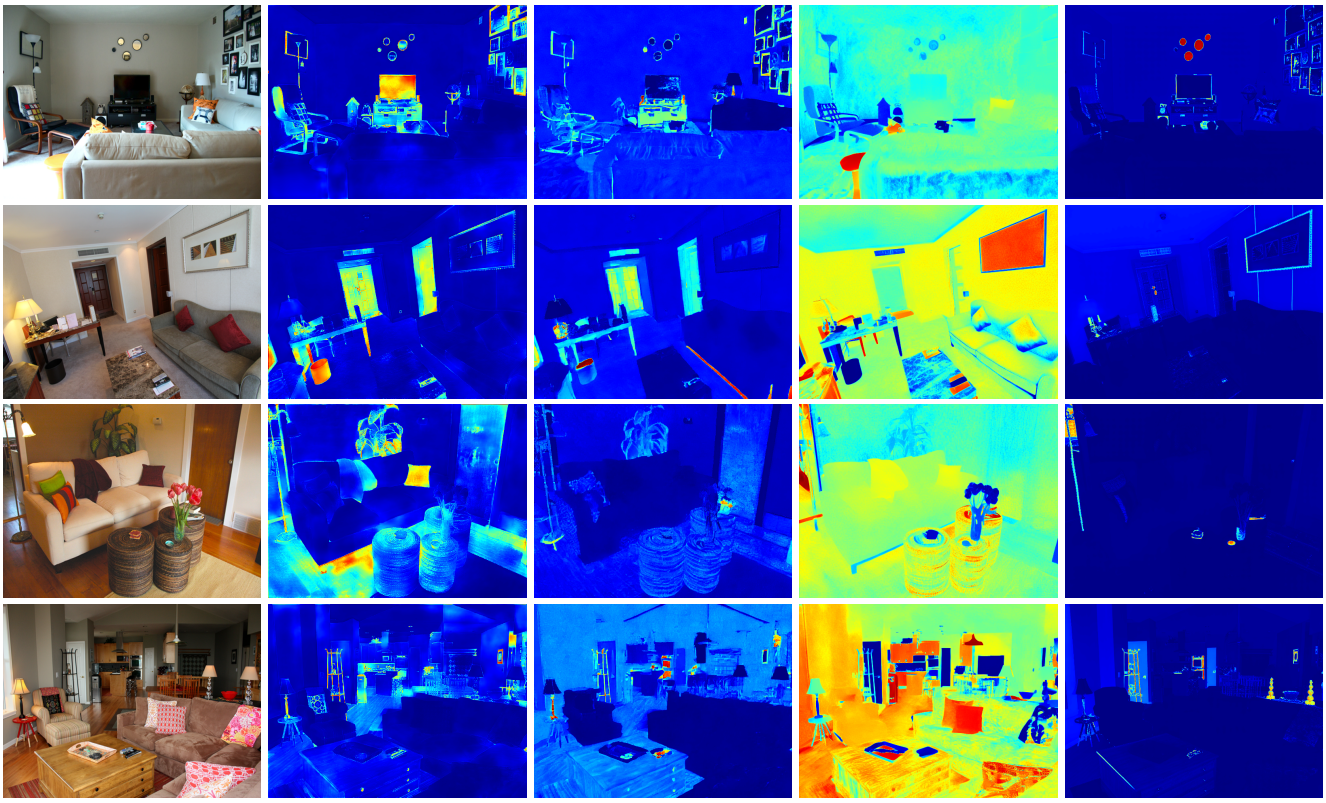


Figure 15. **More qualitative comparison of metallic estimation on real-world images.**

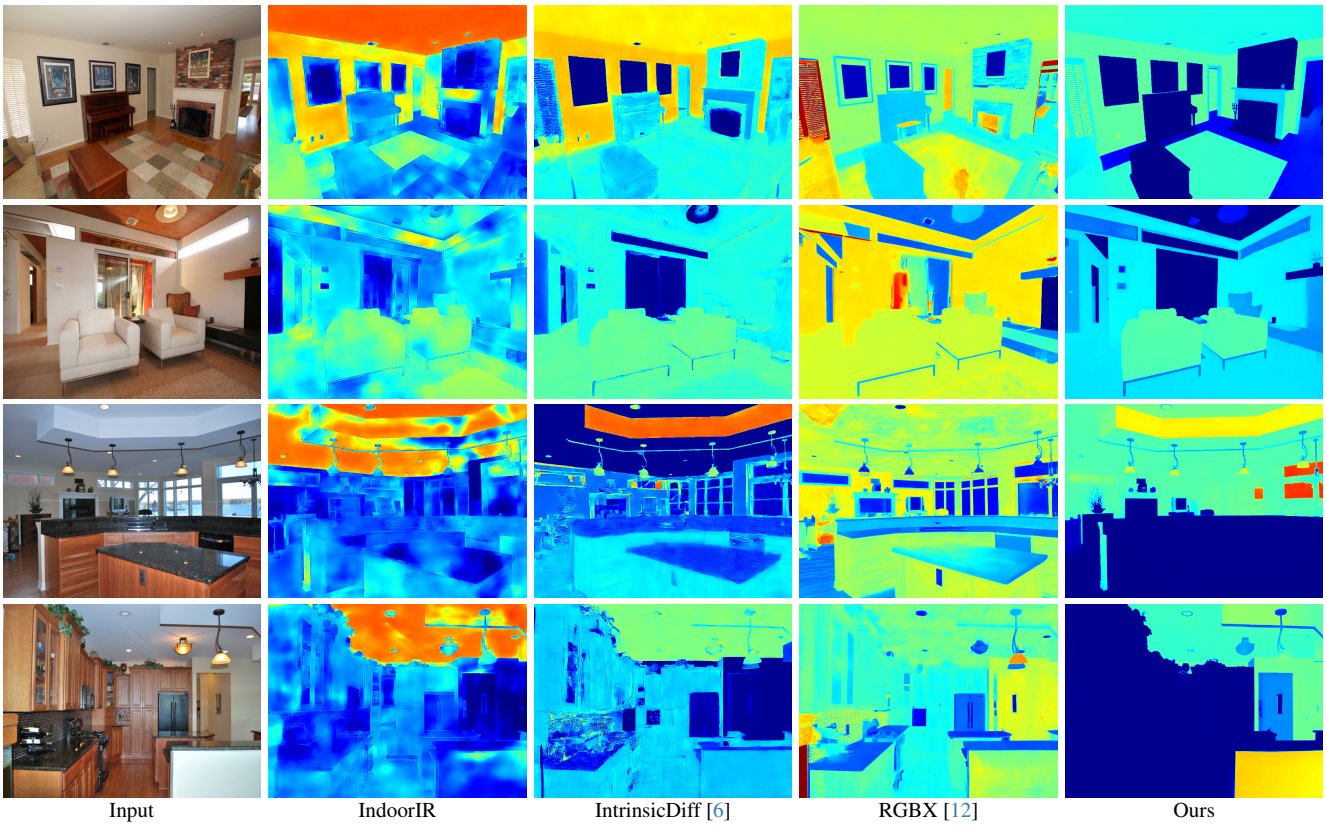


Figure 16. **More qualitative comparison of roughness estimation on real-world images.**

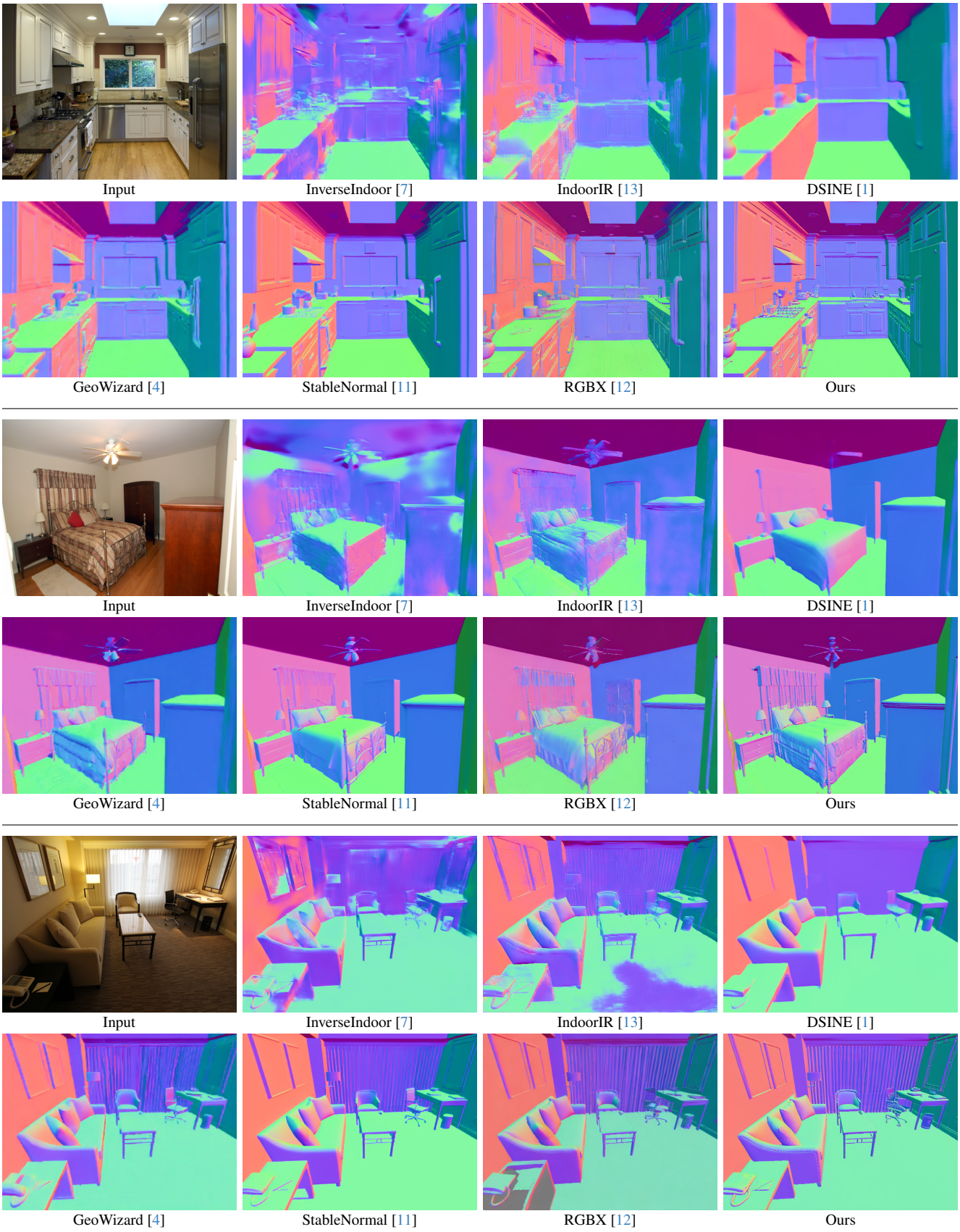


Figure 17. **More qualitative comparison of normal estimation on real-world images.**

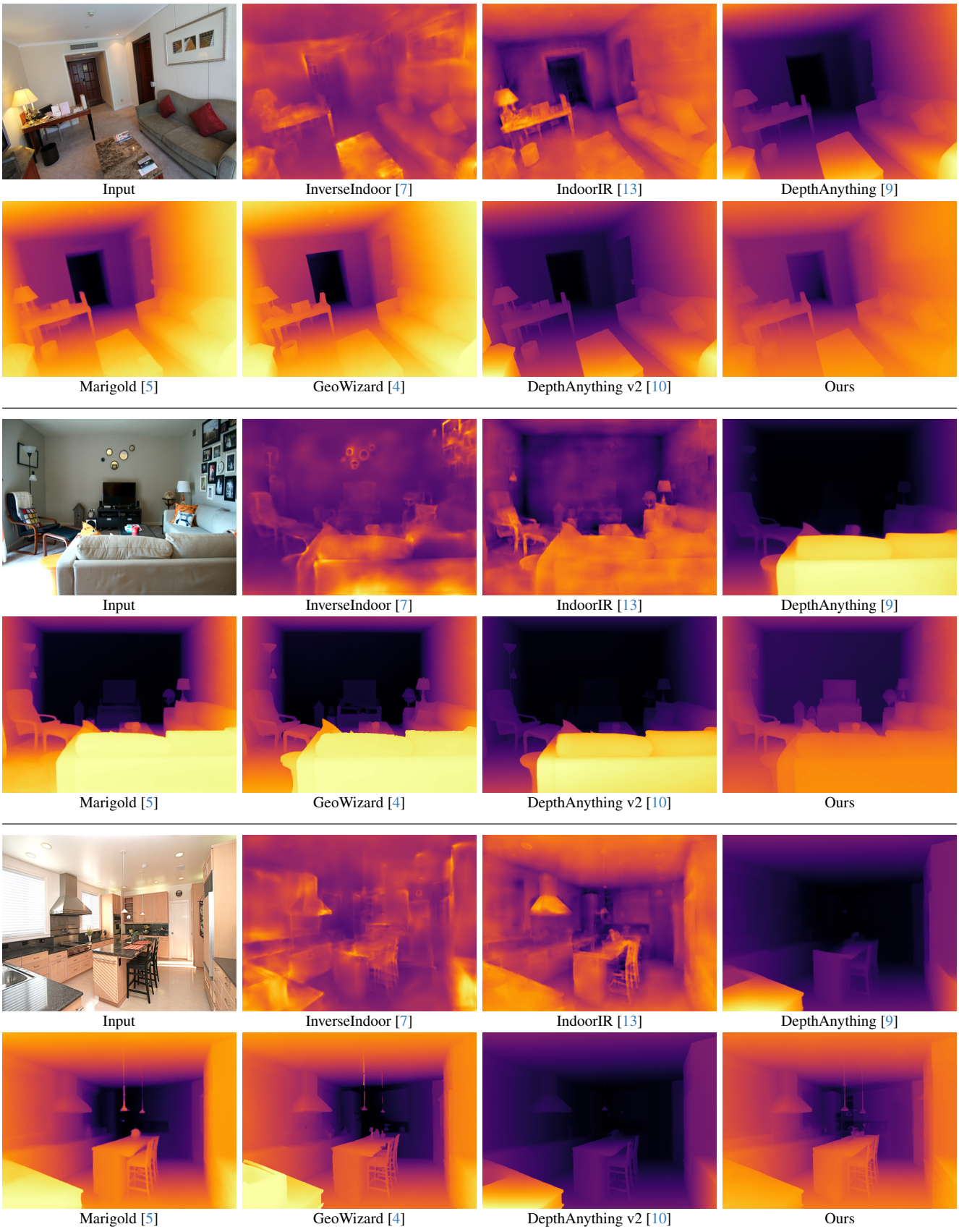
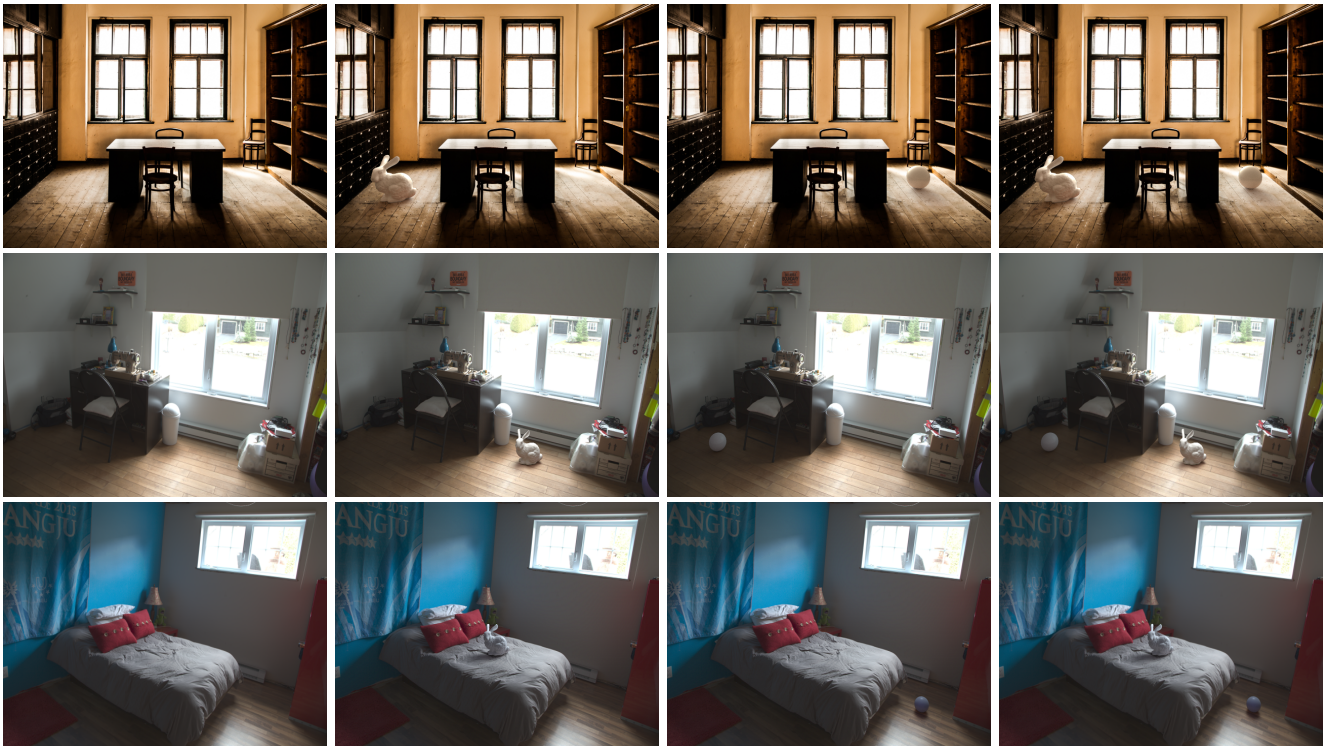


Figure 18. **More qualitative comparison of depth estimation on real-world images.**



Original image | Result 1 | Result 2 | Result 3

Figure 19. **More results of virtual object insertion.**

Original image | Material editing | Lighting editing: result 1 | Lighting editing: result 2

Figure 20. **More results of material and lighting editing.**