MAIR: Multi-view Attention Inverse Rendering with 3D Spatially-Varying Lighting Estimation

JunYong Choi[1,2], SeokYeong Lee[1,2], Haesol Park[1], Seung-Won Jung[2], Ig-Jae Kim[1,3,4], Junghyun Cho[1,3,4]

[1]KIST, [2]Korea Univ., [3]KIST School(UST), [4]YU-KIST

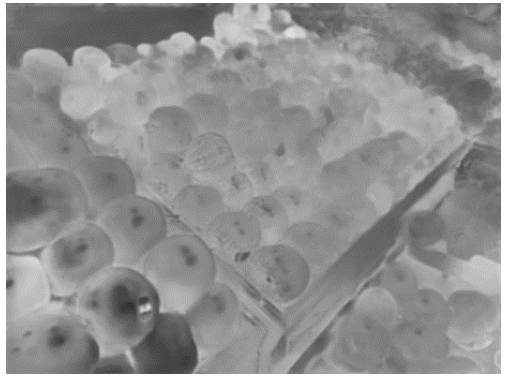
WED-AM-016

Learning-based method for inverse rendering with multi-view images

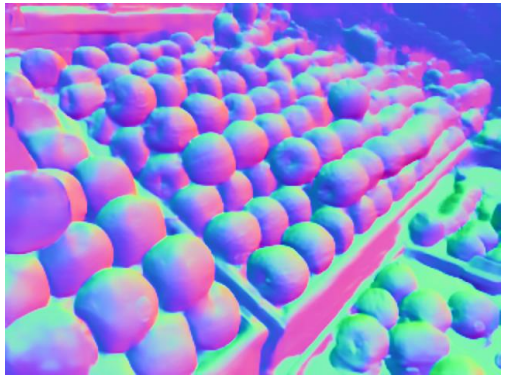
Target View

Multi-view Images

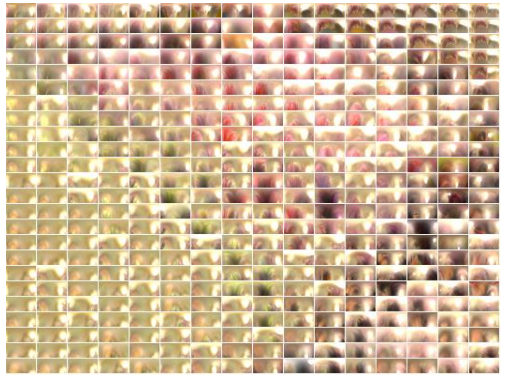
Albedo



Roughness



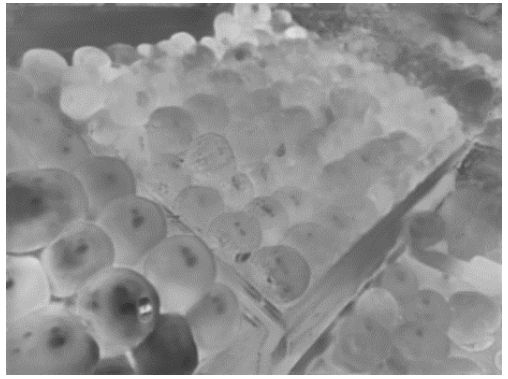
Normal



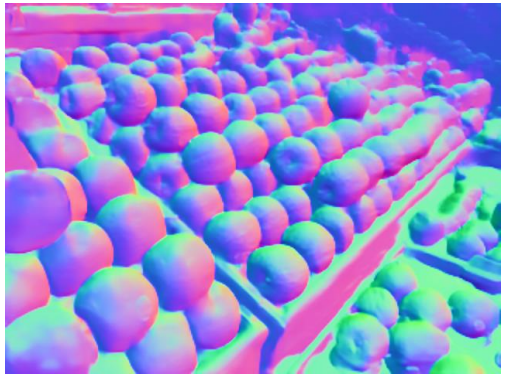
3D spatially-varying Lighting

Virtual object insertion

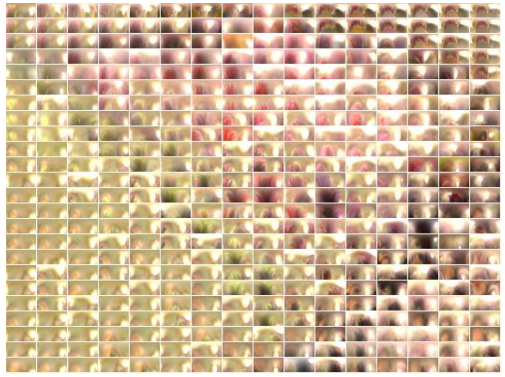
Albedo



Roughness



Normal

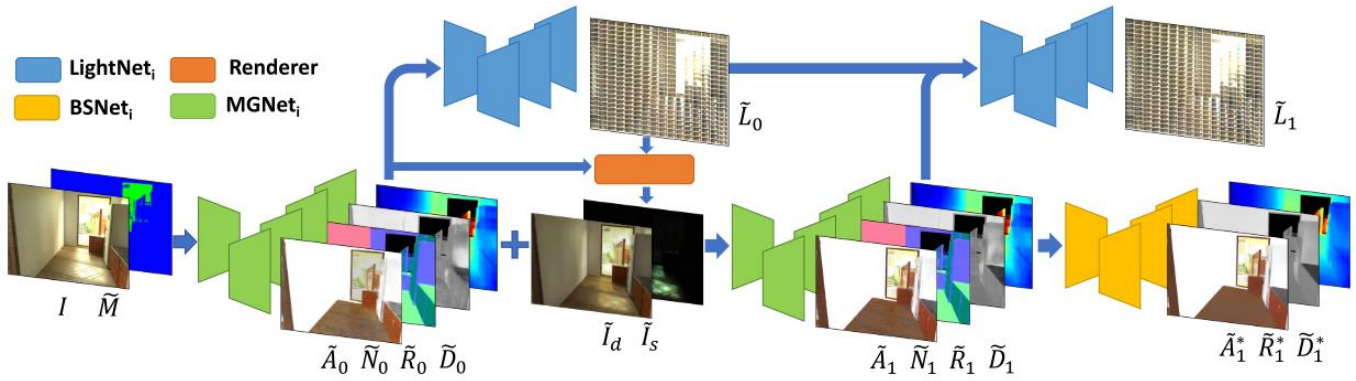


3D spatially-varying Lighting

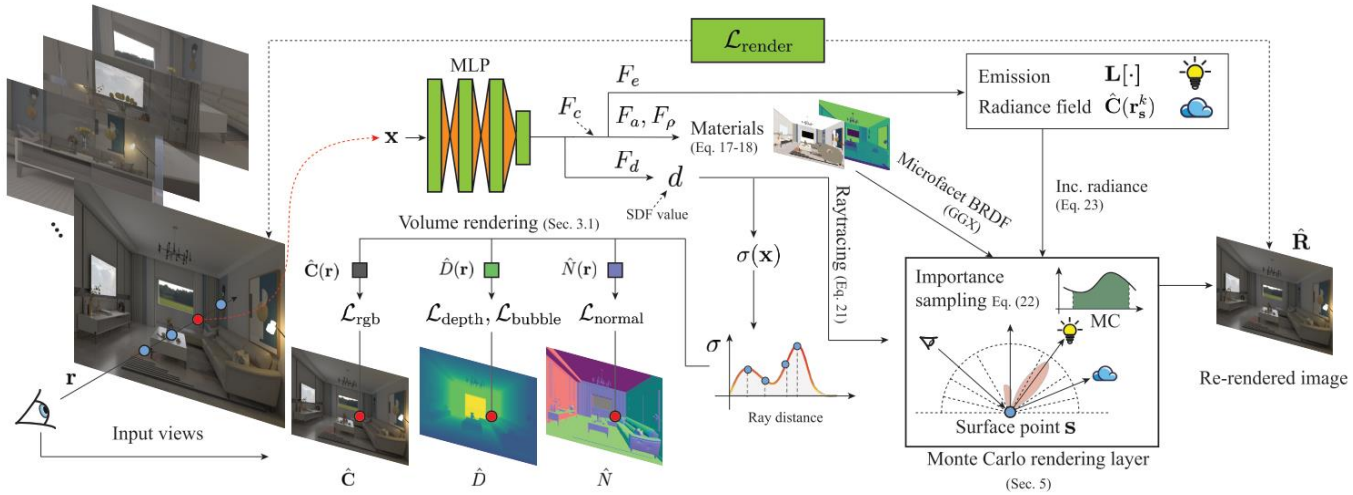
Chrome sphere Insertion with off-the-shelf view synthesis

Previous works for inverse rendering

- Learning-based method with single-view images



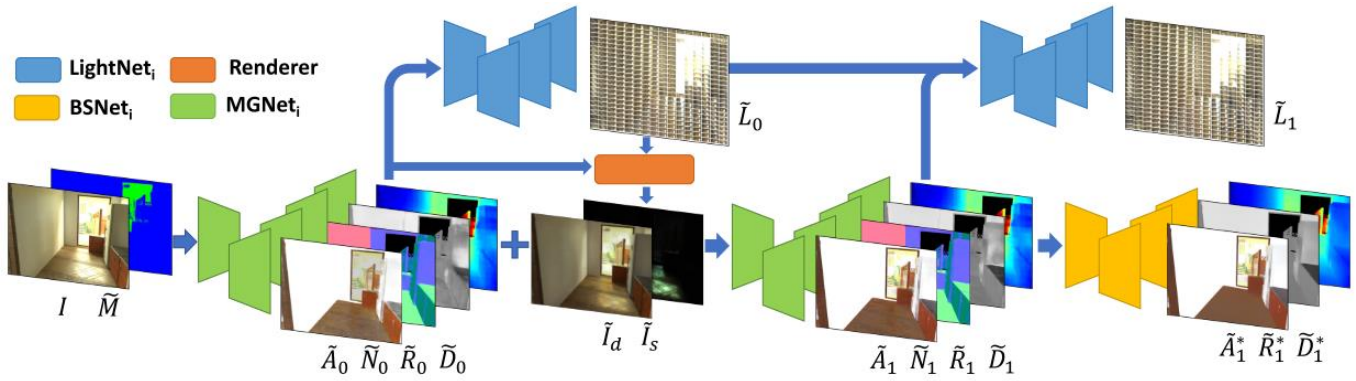
- Optimization-based method with multi-view images



Li et al. 2020, Inverse Rendering for Complex Indoor Scenes
Zhu et al. 2023, I^2-SDF

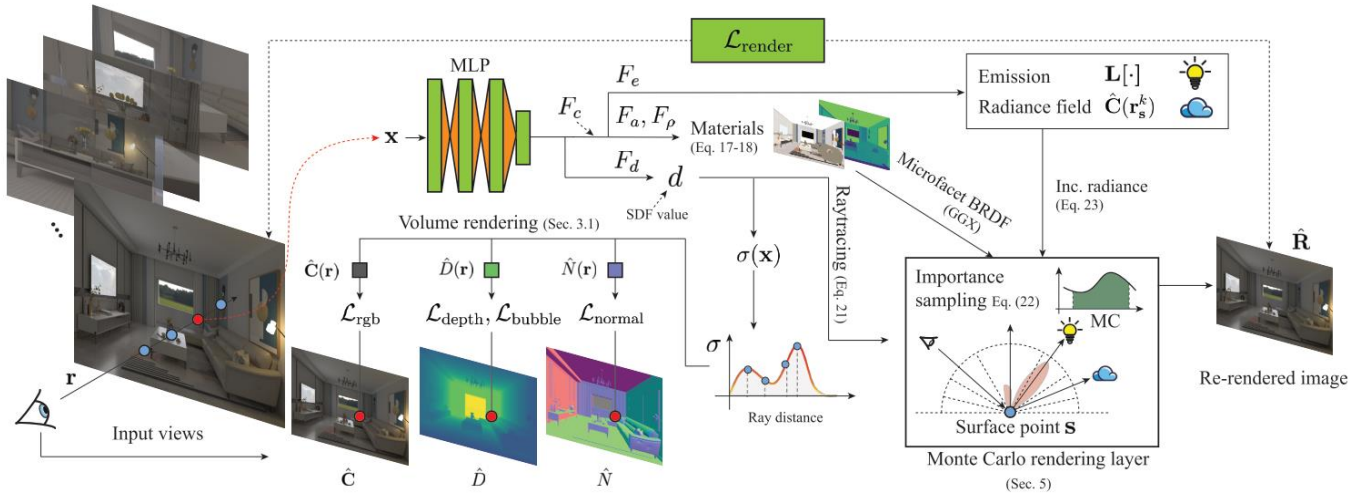
Previous works for inverse rendering

- Learning-based method with single-view images



Difficulty making accurate and robust predictions

- Optimization-based method with multi-view images

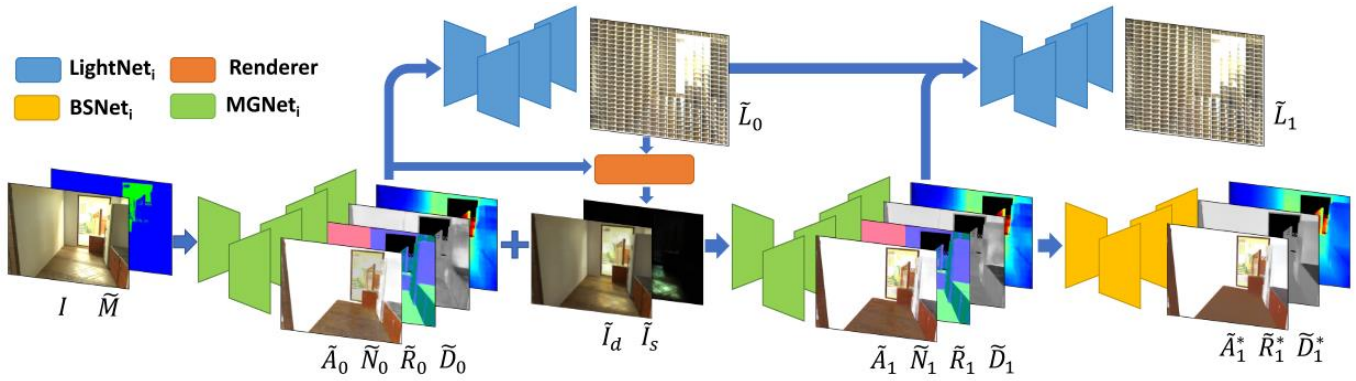


Requires long test-time optimization per scene.

Li et al. 2020, Inverse Rendering for Complex Indoor Scenes
Zhu et al. 2023, I^2-SDF

Previous works for inverse rendering

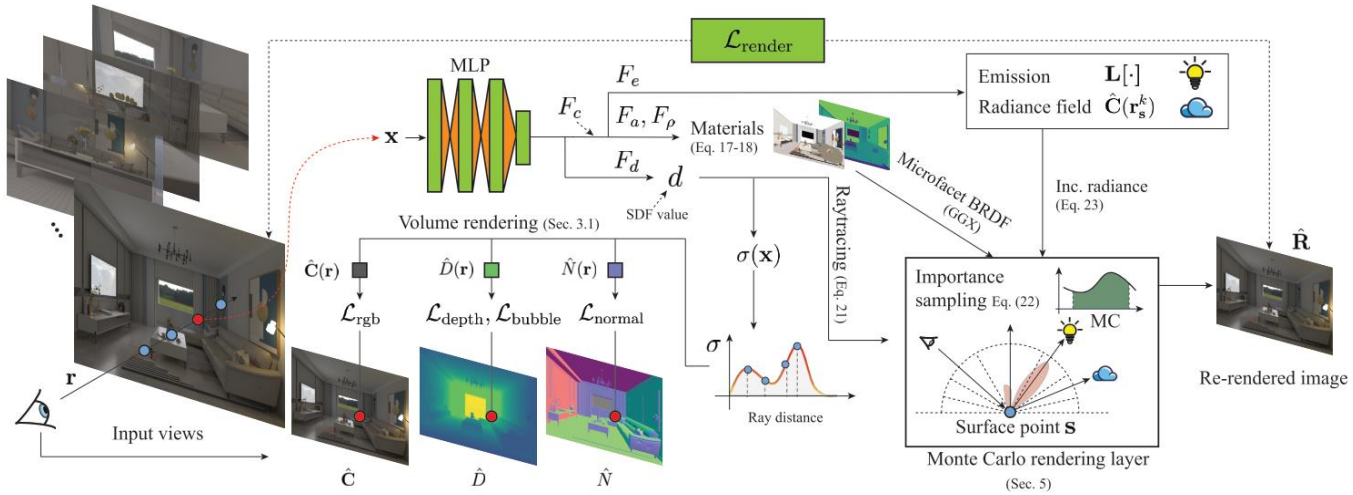
- Learning-based method with single-view images



Difficulty making accurate and robust predictions

***Learning-based* method with *multi-view* images**

- Optimization-based method with multi-view images

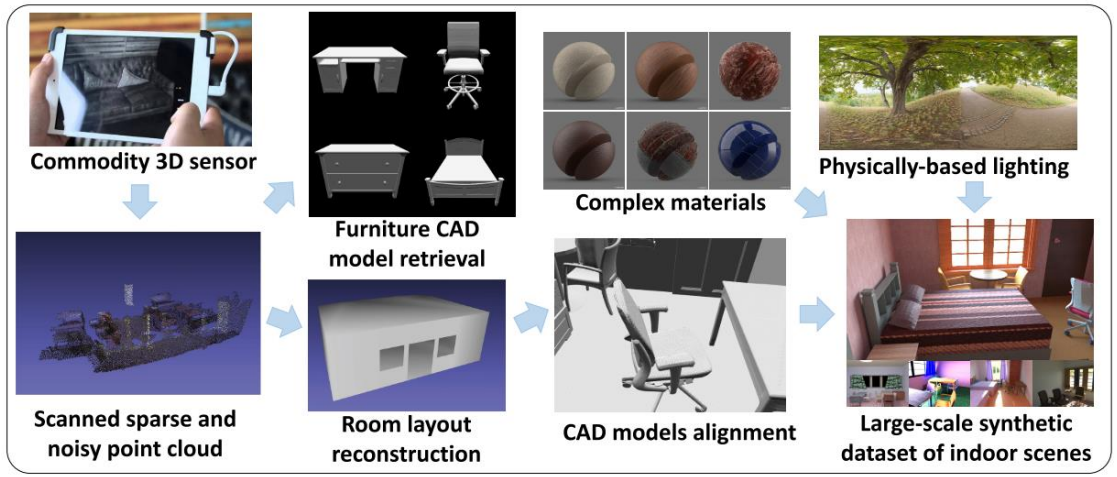


Requires long test-time optimization per scene.

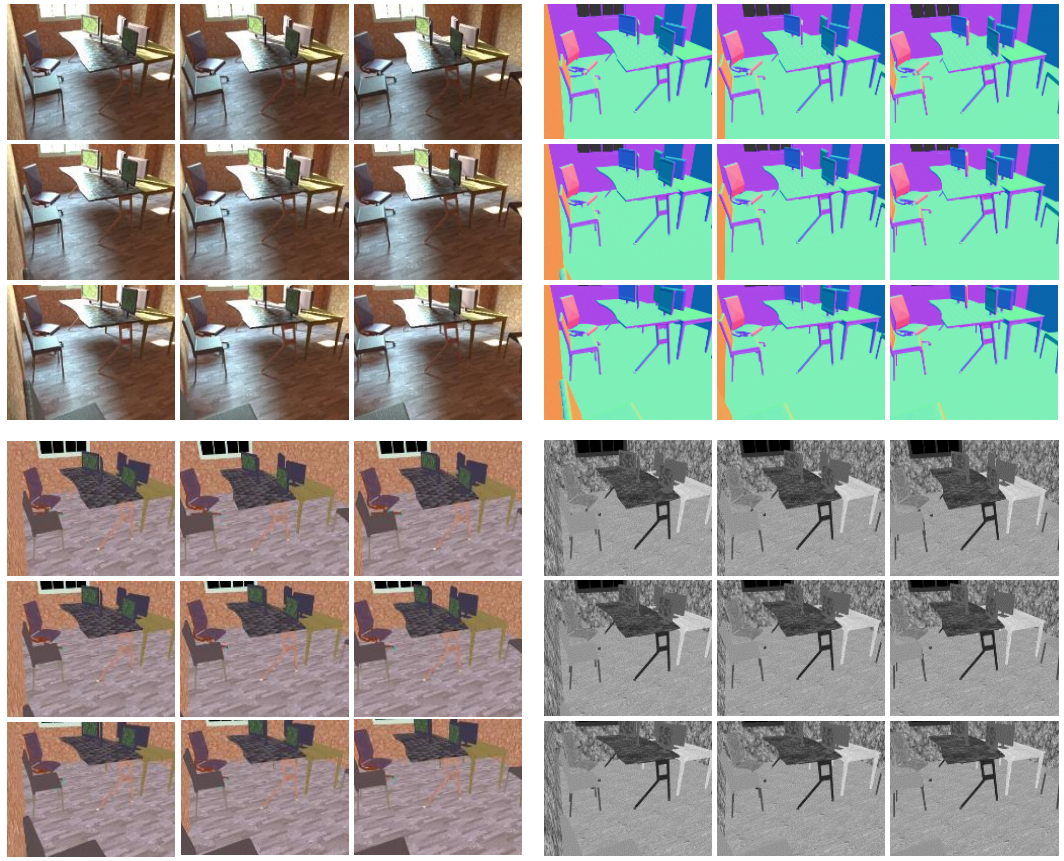
Li et al. 2020, Inverse Rendering for Complex Indoor Scenes
Zhu et al. 2023, I^2-SDF

Multi-view HDR synthetic dataset creation

- OpenRooms FF, an extension of OpenRooms to a multi-view setup.

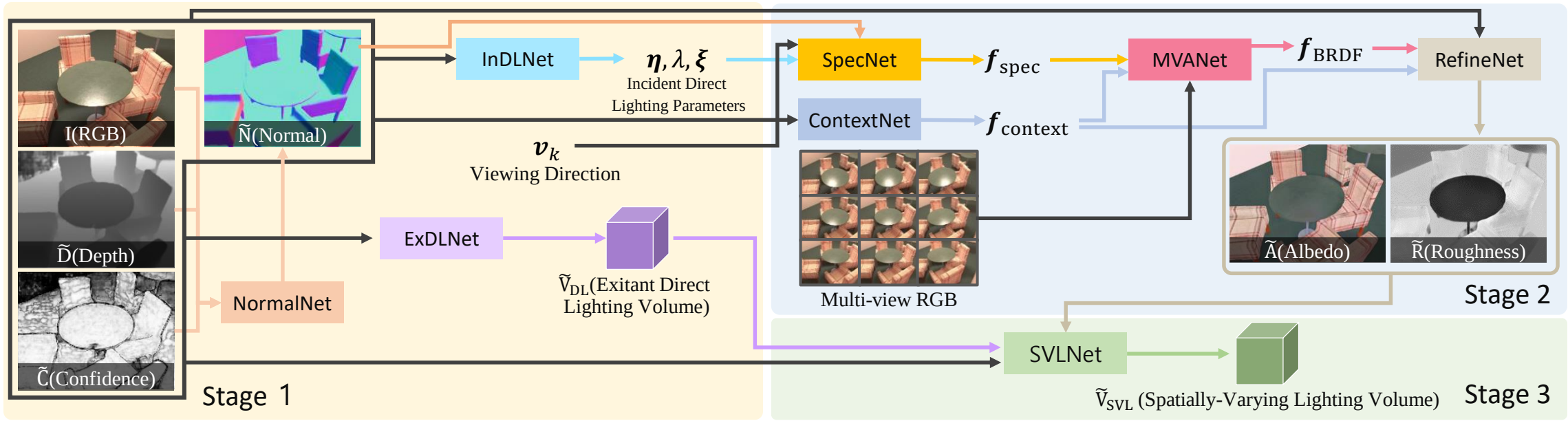


OpenRooms

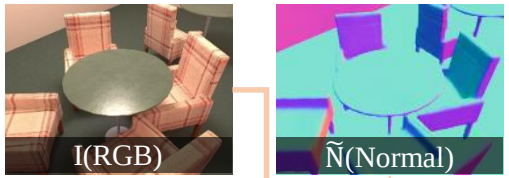


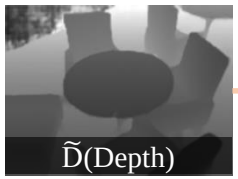
OpenRooms FF

Method


I(RGB)
D̃(Depth)
C̃(Confidence)
Ñ(Normal)
InDLNet
η, λ, ξ
Incident Direct Lighting Parameters
v_k
Viewing Direction
ExDLNet
Ṽ_DL(Exitant Direct Lighting Volume)
NormalNet
Stage 1
SpecNet
f_spec
ContextNet
f_context
MVANet
f_BRDF
RefineNet
Multi-view RGB
Ã(Albedo)
R̃(Roughness)
Stage 2
SVLNet
Ṽ_SVL (Spatially-Varying Lighting Volume)
Stage 3

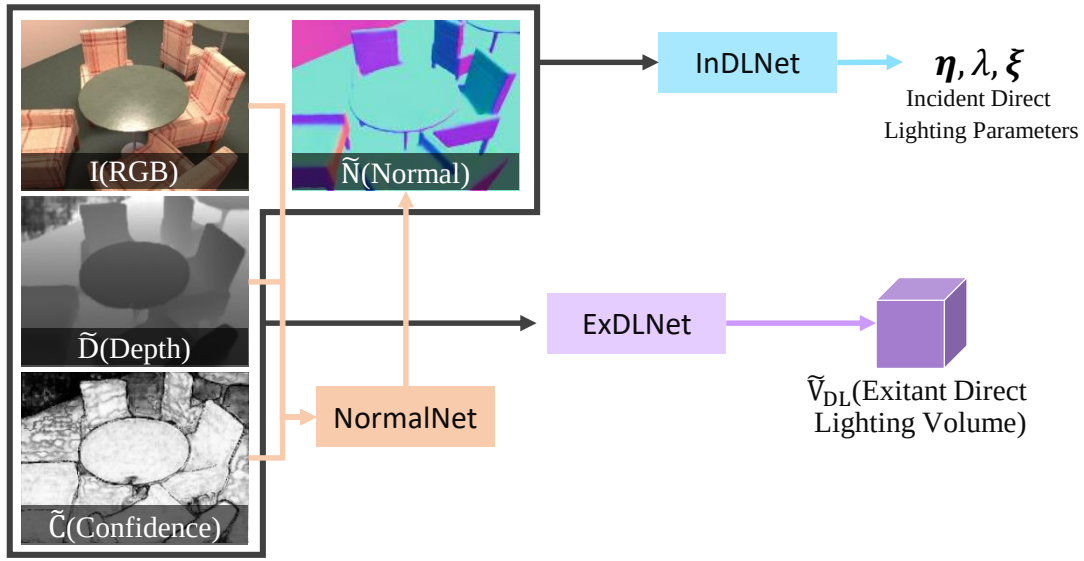
Method – target view analysis stage



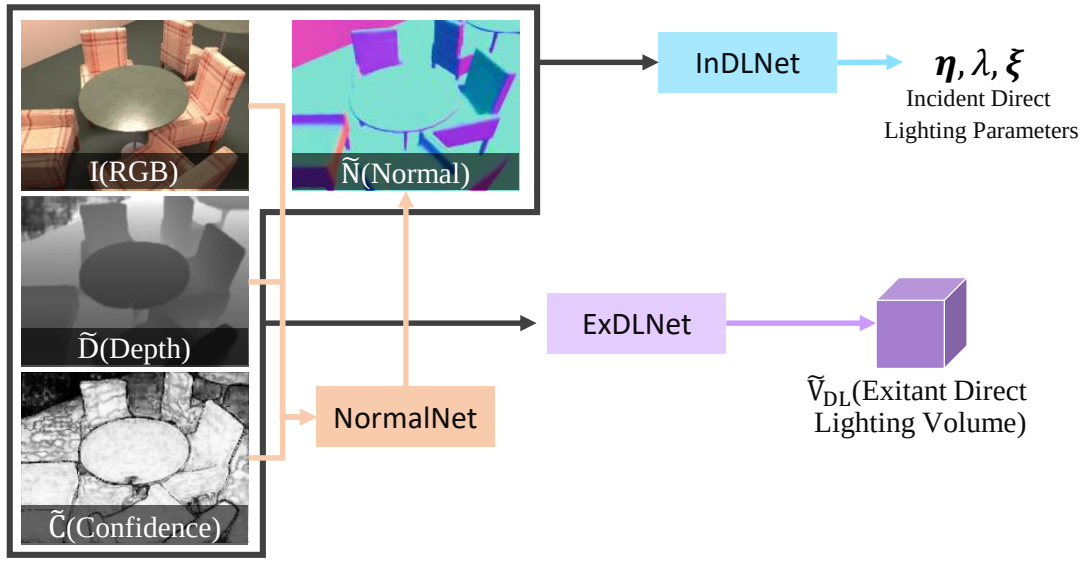


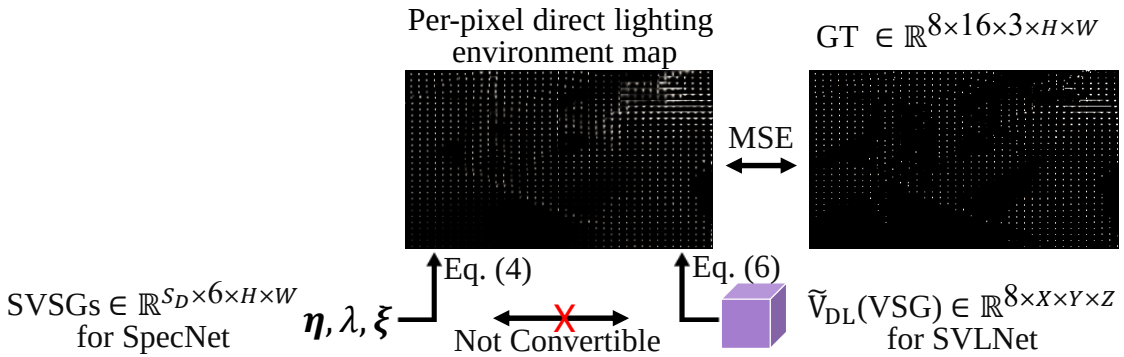
NormalNet

Method – target view analysis stage



Method – target view analysis stage

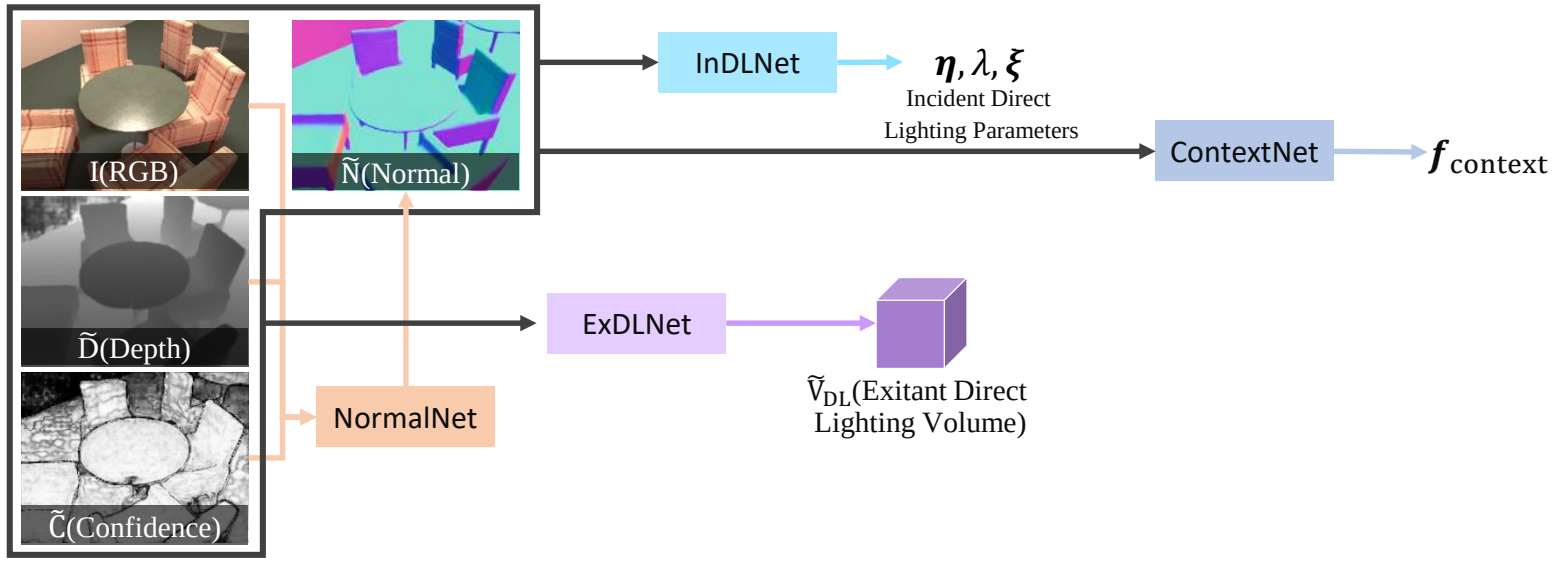




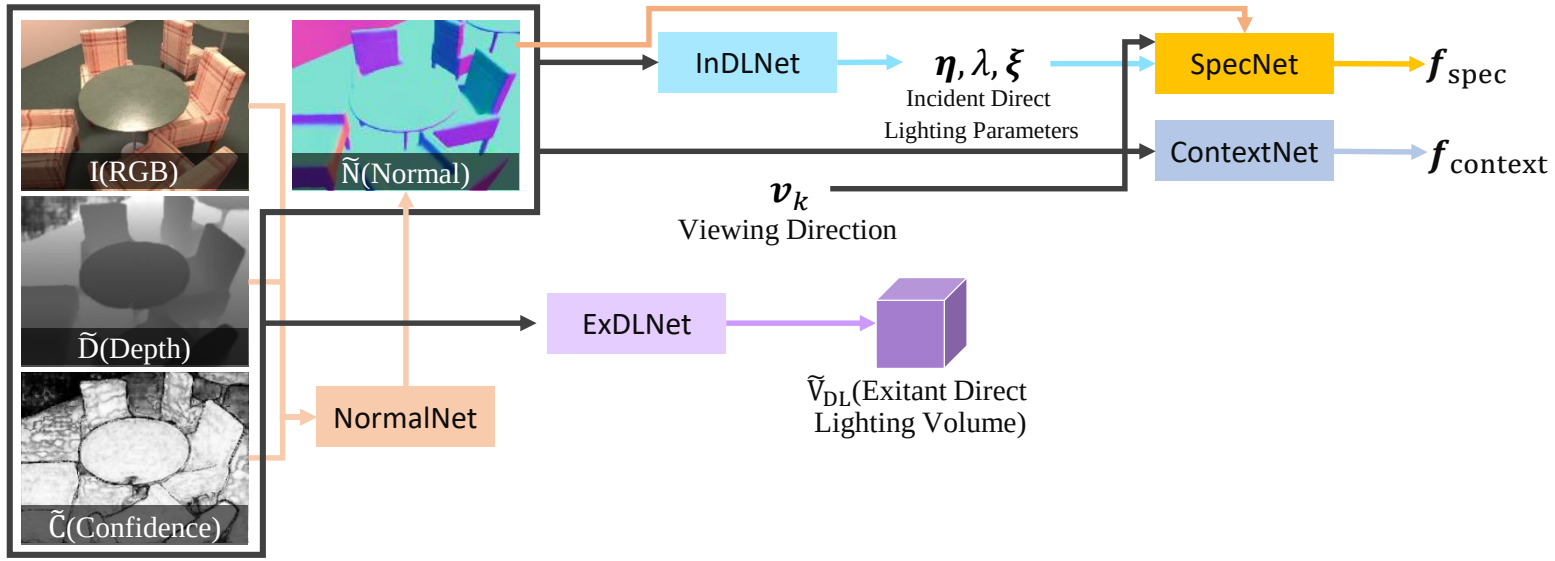
$$\mathcal{R}_i(\boldsymbol{l}) = \sum_{s=1}^{S_D} \mathcal{G}(\boldsymbol{l}; \boldsymbol{\eta}_s, \lambda_s, \boldsymbol{\xi}_s), \tag{4}$$

$$\mathcal{R}_e(\boldsymbol{l}) = \sum_{n=1}^{N_R} \prod_{m=1}^{n-1} (1-\alpha_m)\alpha_n \mathcal{G}(-\boldsymbol{l}; \boldsymbol{\eta}_n, \lambda_n, \boldsymbol{\xi}_n), \tag{6}$$

Method – material estimation stage



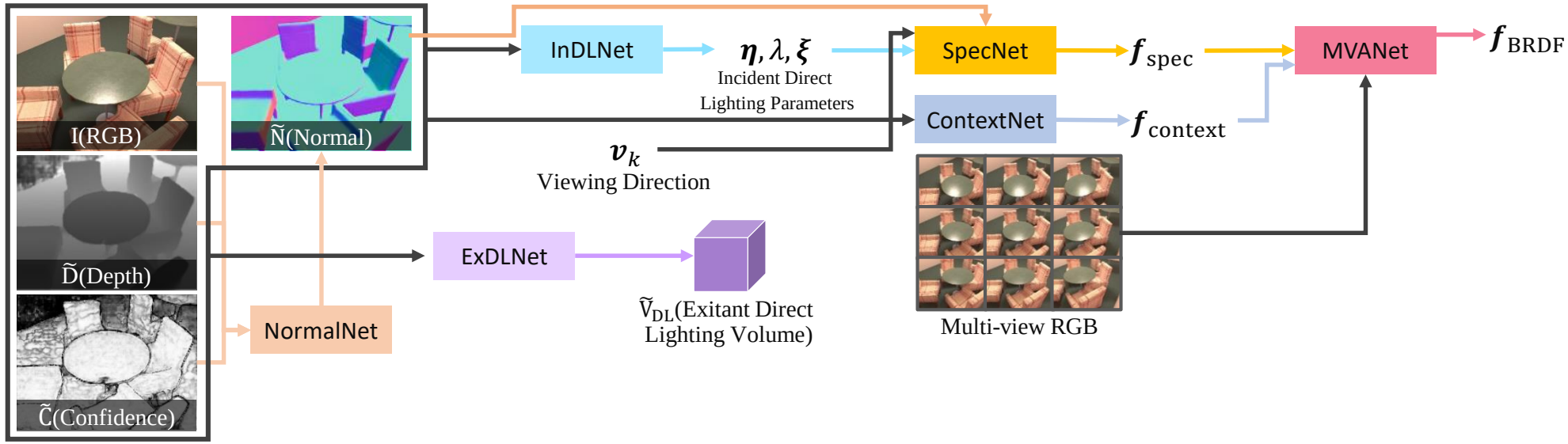
Method – material estimation stage

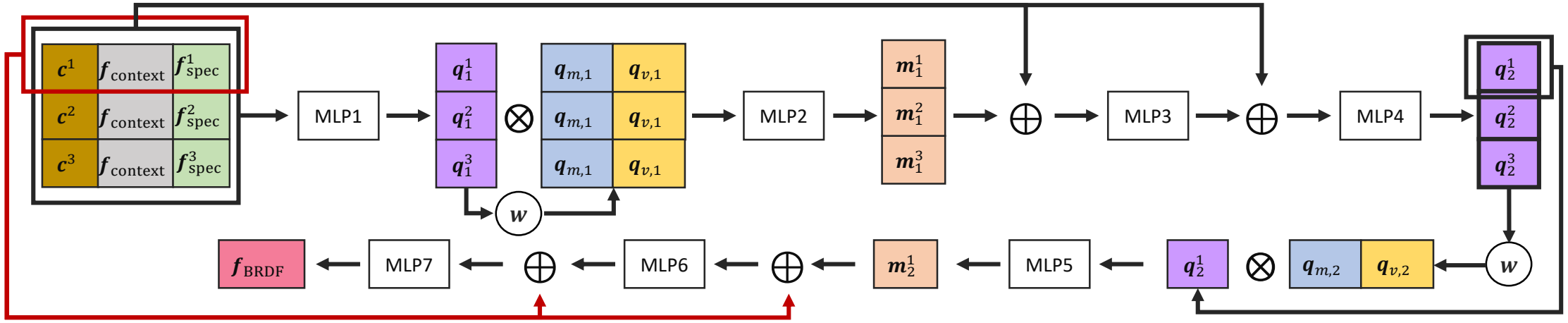


$$\boldsymbol{f}^{k}_{\text{spec}} = \sum_{s=1}^{S_D} m_s \text{SpecNet}(\mathcal{F}(\boldsymbol{v}_k, \boldsymbol{h}_{s,k}), (\tilde{\boldsymbol{n}} \cdot \boldsymbol{h}_{s,k})^2, \tilde{\boldsymbol{n}} \cdot \boldsymbol{\xi}_s, \tilde{\boldsymbol{n}} \cdot \boldsymbol{v}_k, \boldsymbol{\eta}_s, \lambda_s), \quad (11)$$

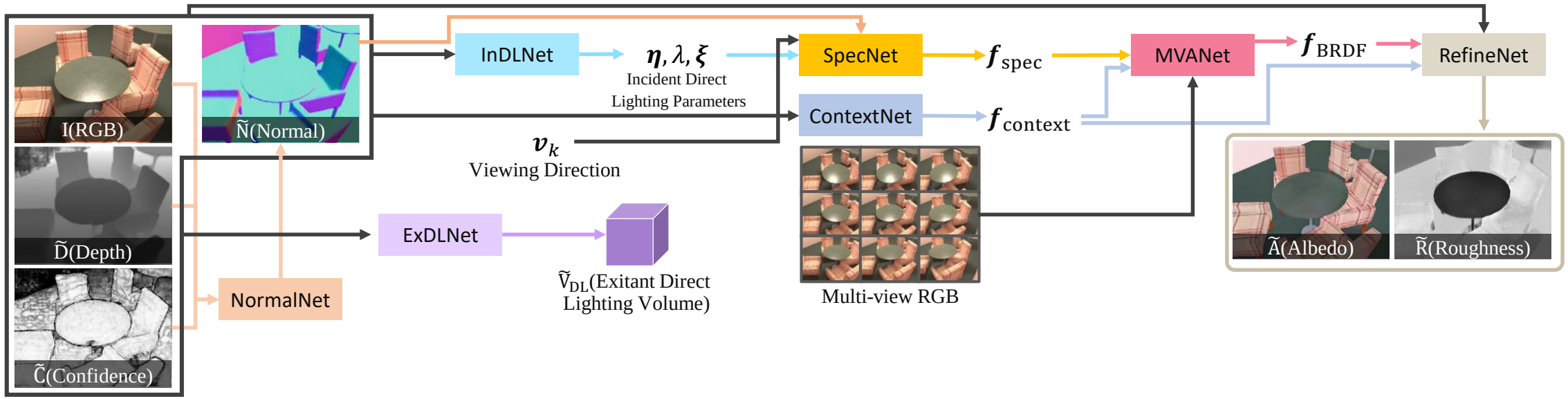
$$m_s = \begin{cases} 1 & \text{if } \|\boldsymbol{\eta}_s\|_1 \tilde{\boldsymbol{n}} \cdot \boldsymbol{\xi}_s > 0, \\ 0 & \text{else,} \end{cases} \quad (12)$$

Method – material estimation stage

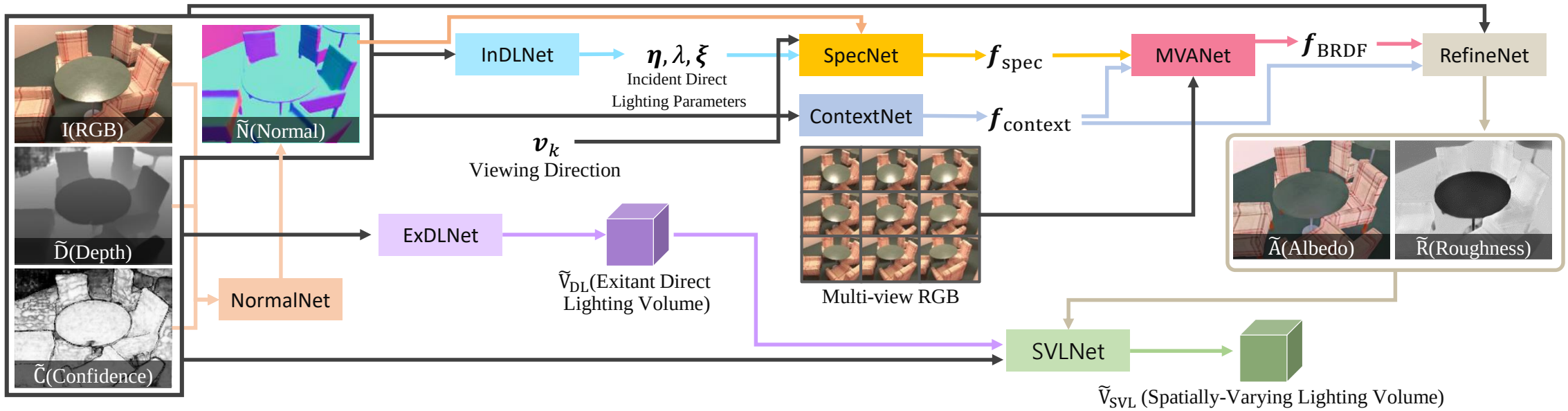




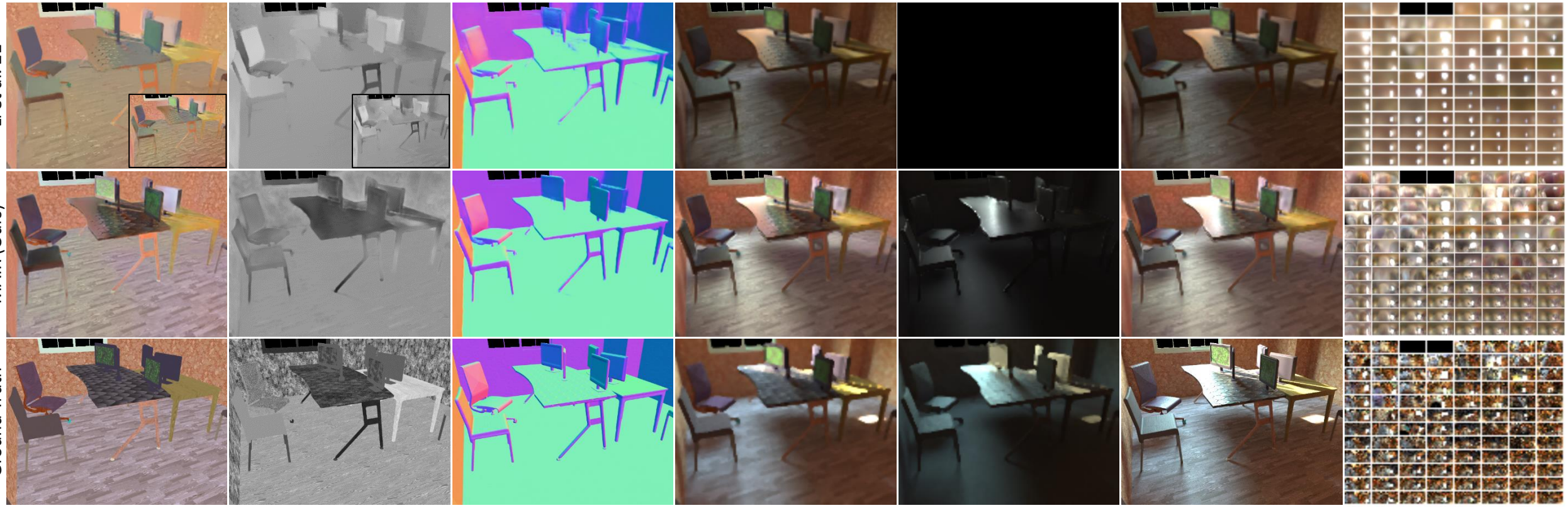
Method – material estimation stage



Method – lighting estimation stage

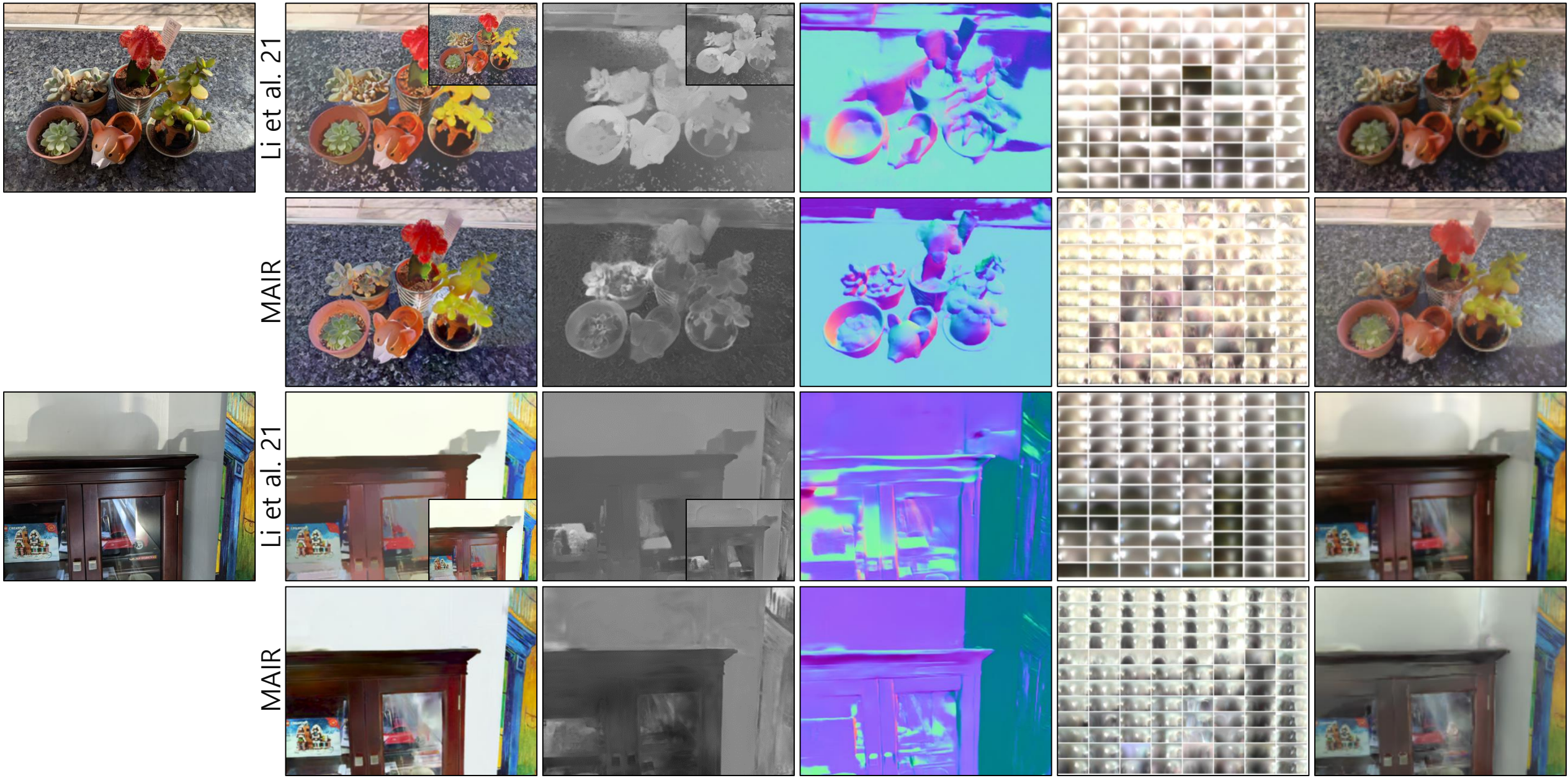


Evaluation – synthetic indoor scene



MSE ($\times 10^{-2}$)	Li *et al.* [19]	MAIR (Ours)	
Albedo ↓	0.569	**0.368**	(−0.201)
Normal ↓	2.71	**1.36**	(−1.35)
Roughness ↓	3.66	**2.70**	(−0.96)
Lighting ↓	13.74	**12.04**	(−1.70)
Re-rendering ↓	**0.554**	0.633	(+0.079)

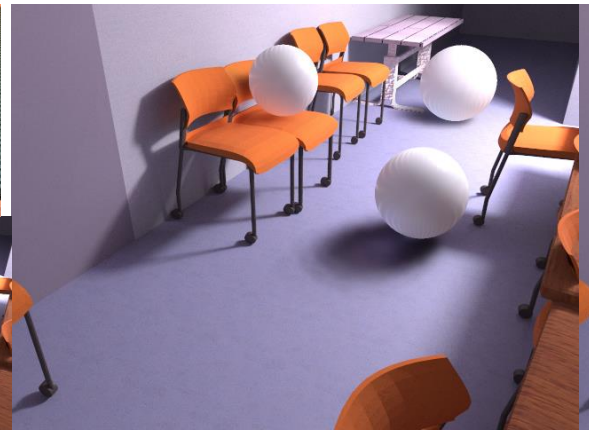
Evaluation – unseen real-world scene

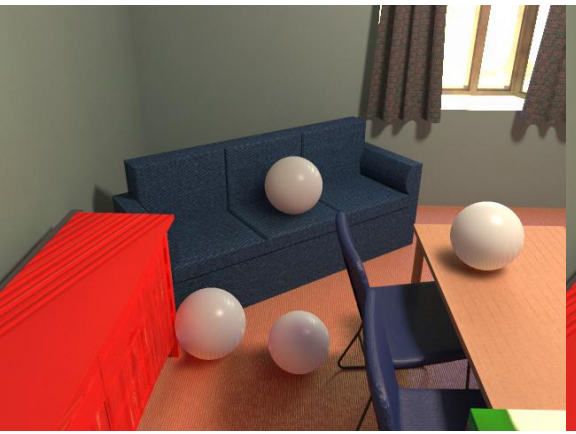
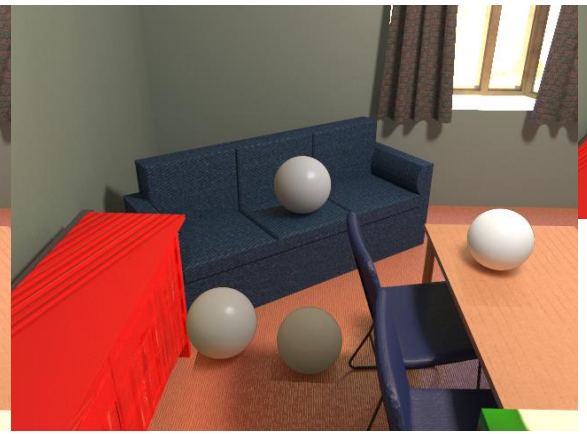


Evaluation – object insertion in synthetic indoor scene









MAIR(Ours) Li et al. 21 Input Image MAIR(Ours) Li et al. 21

Evaluation – object insertion in unseen real-world scene

Input Image | Lighthouse | Li et al. 21 | MAIR(Ours)

Additional object insertion results in unseen real-world scene

MAIR: Multi-view Attention Inverse Rendering with 3D Spatially-Varying Lighting Estimation

JunYong Choi[1,2], SeokYeong Lee[1,2], Haesol Park[1], Seung-Won Jung[2], Ig-Jae Kim[1,3,4], Junghyun Cho[1,3,4]

[1]KIST, [2]Korea Univ., [3]KIST School(UST), [4]YU-KIST

WED-AM-016