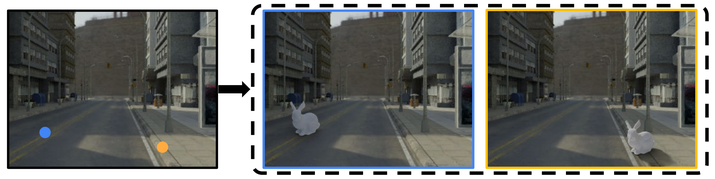
Estimating Spatially-Varying Lighting in Urban Scenes with Disentangled Representation

Jiajun Tang[1], Yongjie Zhu[2], Haoyu Wang[1], Jun Hoong Chan[1], Si Li[2], Boxin Shi[1, 3]

ECCV TEL AVIV 2022

PROBLEM&CONTRIBUTION

Goal: Estimating spatially-varying lighting in urban scenes.

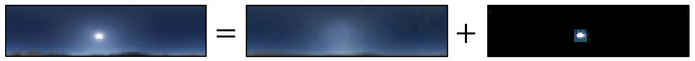


Contributions: A flexible and editable local lighting estimation by

- using disentangled global and local lighting representation
- designing SOLD-Net network architecture for spatially-varying outdoor lighting estimation with disentangled representation
- enhancing synthetic data and capturing unclipped HDR real data

FORMULATION

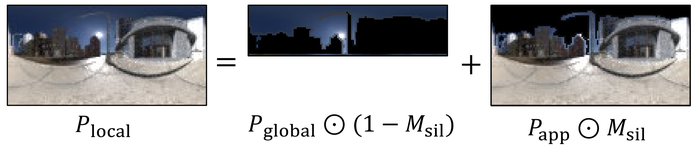
Global lighting disentanglement:



$P_{\text{global}} = P_{\text{sky}} + z_{\text{vis}}(P_{\text{sun}} \odot M_{\text{sun}})$

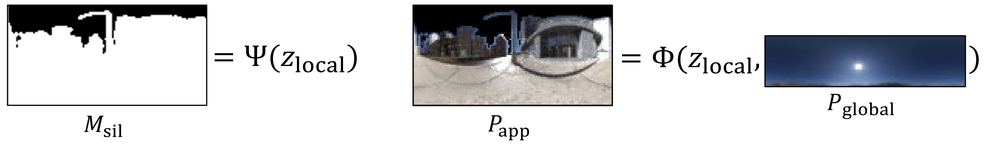
$z_{\text{vis}} \in \mathbb{R}$: sun visibility, M_{sun}: determined by $z_{\text{pos}} \in \mathbb{R}^2$

Spatially-varying lighting disentanglement:



$P_{\text{local}} = P_{\text{global}} \odot (1 - M_{\text{sil}}) + P_{\text{app}} \odot M_{\text{sil}}$

Local property disentanglement:



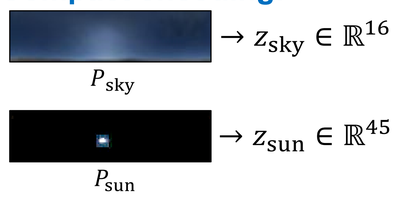
$M_{\text{sil}} = \Psi(z_{\text{local}})$

$P_{\text{app}} = \Phi(z_{\text{local}}, P_{\text{global}})$

$z_{\text{local}} \in \mathbb{R}^{64}$: local content information

Φ : lighting (rendering) process

Compact encoding:

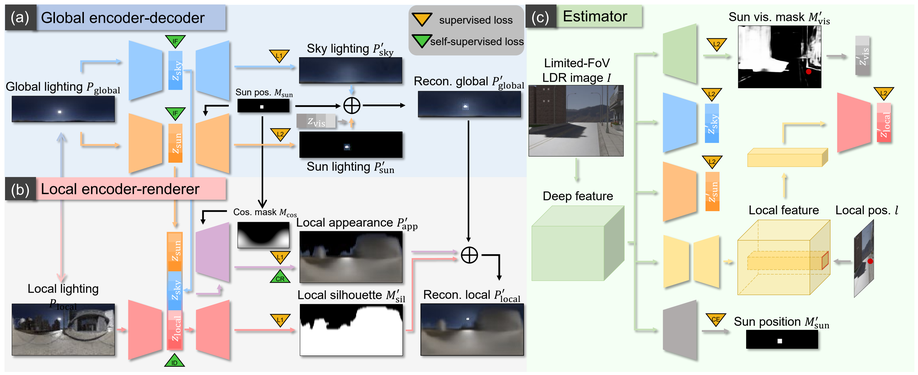


$P_{\text{sky}} \rightarrow z_{\text{sky}} \in \mathbb{R}^{16}$

$P_{\text{sun}} \rightarrow z_{\text{sun}} \in \mathbb{R}^{45}$

METHOD

Network Architecture: SOLD-Net consists of a global lighting encoder-decoder, a local content encoder-renderer, and a spatially-varying lighting estimator.



Self-supervised Losses: Info loss (IF) [2], local identity loss (ID), and cross rendering loss (CR) are used to ensure the disentanglement of different lighting components.

$$\mathcal{L}_{\text{IF}} = \max \mathbb{E}_{z_{\text{sky}}}\left[\log E^i_{z_{\text{sky}}}\left(z'_{\text{sky}}|P'_{\text{sky}}\right)\right] + \max \mathbb{E}_{z_{\text{sun}},z_{\text{pos}}}\left[\log E^i_{z_{\text{sun}}}\left(z'_{\text{sun}}|P'_{\text{sun}}\right)\right]$$

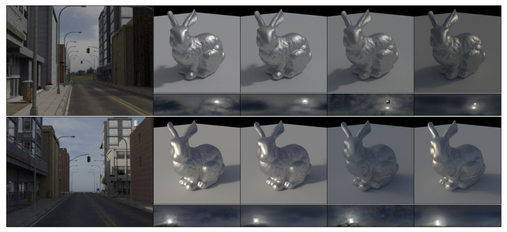
$$\mathcal{L}_{\text{ID}} = \| z^{\prime 1}_{\text{local}} - z^{\prime 2}_{\text{local}} \|_1$$

$$\mathcal{L}_{\text{CR}} = \| P^1_{\text{local}} \odot M_{\text{sil}} - P^{\prime 1}_{\text{local}}\left(z^{\prime 2}_{\text{local}}\right) \odot M_{\text{sil}} \|_1 + \| P^2_{\text{local}} \odot M_{\text{sil}} - P^{\prime 2}_{\text{local}}(z^{\prime 1}_{\text{local}}) \odot M_{\text{sil}} \|_1$$

$E^i_{z_{\text{sky}}}$ and $E^i_{z_{\text{sun}}}$ are encoders, $\{P^1_{\text{local}}, P^2_{\text{local}}\}$ are paired local lighting (same local content, different global lighting), $P'_{\text{local}}(z^i_{\text{local}})$ is rendered appearance.

EXPERIMENT RESULTS

Global lighting estimation:



Methods	Panorama		Relighting	
	MAE↓	RMSE↓	MAE↓	RMSE↓
SkyNet[3]	0.431	8.357	0.226	0.253
SOLID-Net[1]	**0.384**	**6.360**	0.153	0.174
Ours	0.439	7.607	**0.098**	**0.119**

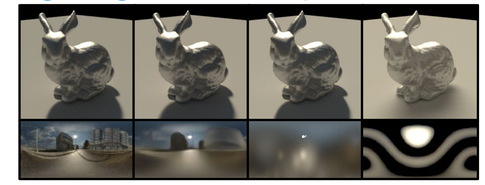
Quantitative evaluation on synthetic dataset

Methods	ξ_{ang} ↓	ξ_{az} ↓	ξ_{el} ↓
SkyNet[3]	27.38°/32.83°	30.11°/32.81°	6.93°/14.20°
SOLID-Net[1]	23.90°/30.82°	26.36°/33.92°	5.00°/11.95°
Ours	**19.97°/21.15°**	**20.31°/20.76°**	**3.61°/8.44°**

Sun position estimation on synthetic/real dataset

ξ_{ang}: angular error ξ_{az}: azimuth error ξ_{el}: elevation error

Lighting model comparison:



Spatially-varying lighting estimation:

Methods	Panorama		Relighting	
	MAE↓	RMSE↓	MAE↓	RMSE↓
SOLID-Net	0.308/0.390	3.384/4.206	0.186/0.310	0.337/0.601
Ours w/o M_{vis}	0.140/0.274	2.814/6.496	0.081/0.299	0.159/0.552
Ours_{SH}	0.190/0.244	**1.943/2.367**	0.081/0.269	**0.139**/0.469
Ours_{SG}	0.170/**0.179**	2.785/2.748	0.093/0.318	0.179/0.560
Ours	**0.128**/0.240	2.394/4.872	**0.075/0.259**	0.145/**0.437**

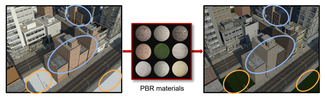
Quantitative evaluation on synthetic/real dataset

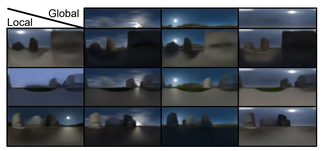
Ablation study:

Methods	Reconstruction		Cross Rendering	
	MAE↓	RMSE↓	MAE↓	RMSE↓
Ours w/o M_{cos}	0.034	0.083	0.036	0.085
Ours w/o $\mathcal{L}_{\text{CR}}$	0.029	0.083	0.048	0.101
Ours w/o $\mathcal{L}_{\text{ID}}$	0.032	0.087	0.035	0.089
Ours	**0.028**	**0.075**	**0.031**	**0.079**

Local appearance evaluation on synthetic dataset

Dataset:





[1] Zhu *et al.*, Spatially-varying outdoor lighting estimation from intrinsics, CVPR 2021.
[2] Yu *et al.*, Hierarchical disentangled representation learning for outdoor illumination estimation and editing, ICCV 2021.
[3] Hold-Geoffroy *et al.*, Deep sky modeling for single image outdoor lighting estimation, CVPR 2019.