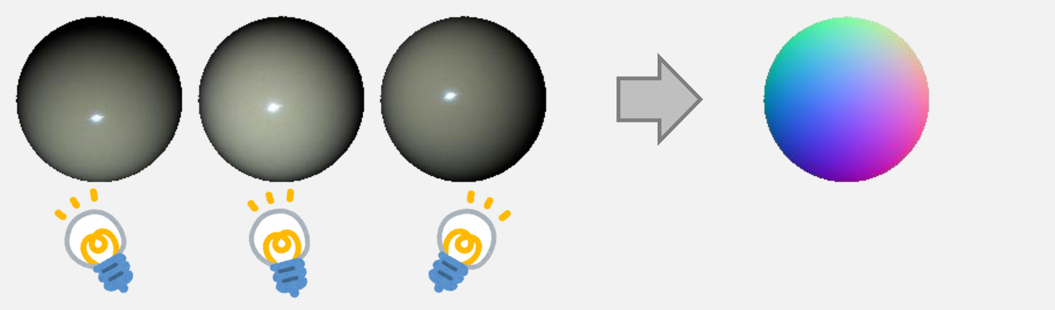
Neural Inverse Rendering for General Reflectance Photometric Stereo

Tatsunori Taniai (RIKEN AIP) **Takanori Maehara** (RIKEN AIP)

ICML | 2018
Thirty-fifth International Conference on Machine Learning

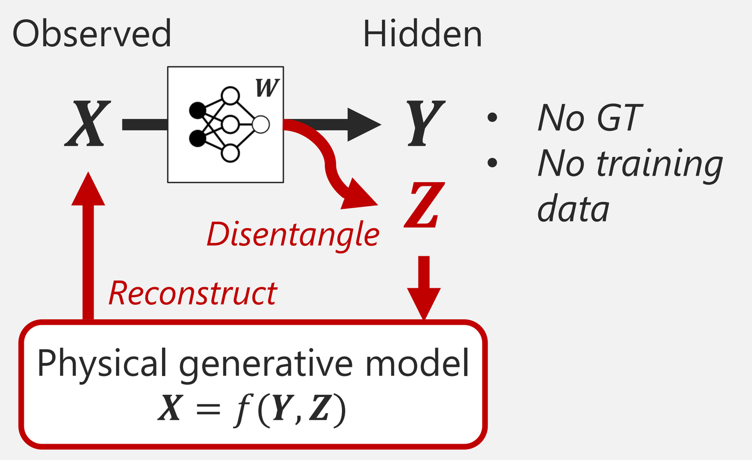
Introduction



Images observed under varying illuminations

Surface normals (3D orientations)

ML perspective



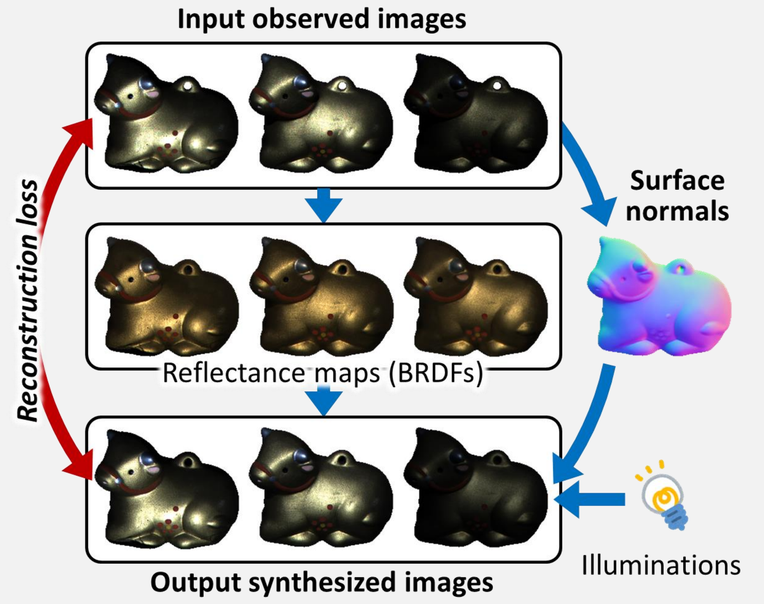
- *No GT*
- *No training data*

Challenges

Complex unknown non-linearity: Real objects have various reflectance properties (BRDFs) that are complex and unknown.

Lack of training data: Deeply learning complex relations of surface normals and BRDFs is promising, but accurately measuring ground truth of surface normals and BRDFs is difficult.

Permutation invariance: Permuting input images should not change the resulting surface normals.



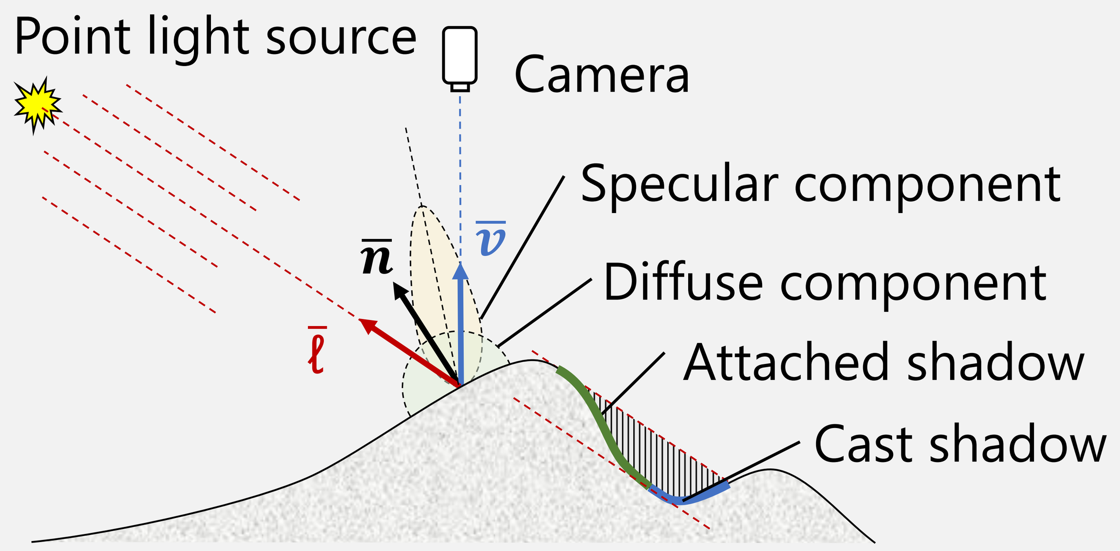
Our physics-embedded auto-encoder

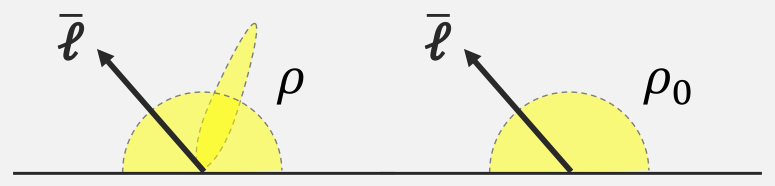
PS as inverse imaging process

Reflectance (rendering) equation

$$I = s\,\rho(\bar{\boldsymbol{n}}, \bar{\boldsymbol{\ell}}, \bar{\boldsymbol{v}}) \max\left(0, \boldsymbol{\ell}^T \bar{\boldsymbol{n}}\right)$$

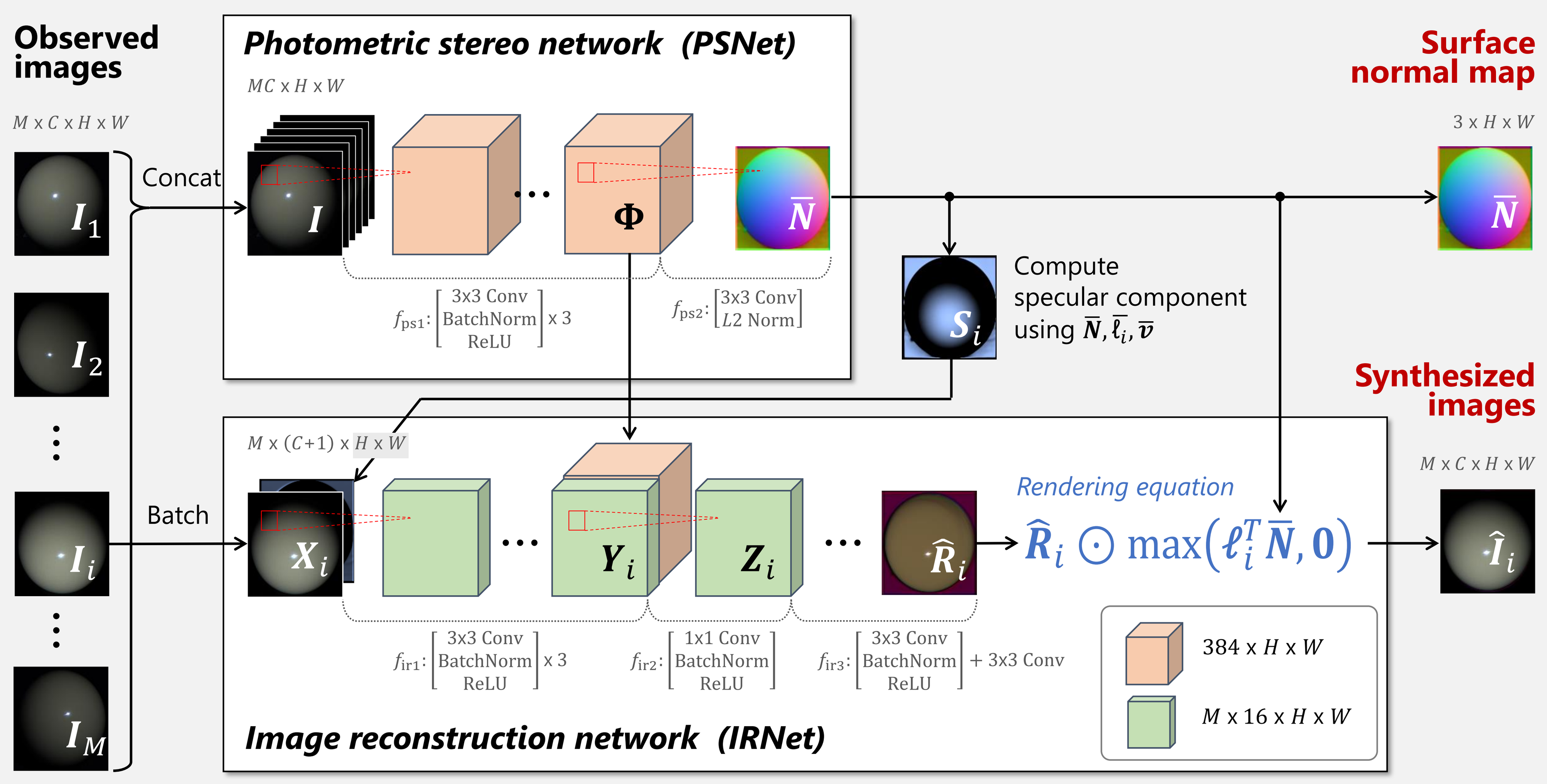
- I: Pixel intensity (known)
- s: {0,1} cast shadow (known)
- $\bar{\boldsymbol{v}}$: view direction (known)
- $\boldsymbol{\ell}$: lighting (known)
- ρ: BRDF (unknown)
- $\bar{\boldsymbol{n}}$: surface normal (unknown)





Physics-based unsupervised learning

Two-stream physics-embedded network



- **Global observation blending** ($\boldsymbol{\Phi}$) provides global information to enrich feature maps in IRNet.
- **Specularity input** ($\boldsymbol{S}_i$) gives a hint to IRNet to promote recovery of complex specular reflections.

Loss function

Image reconstruction loss   *Least squares (LS) prior*

$$L = \frac{1}{M}\sum_{i=1}^{M}\left\|\hat{\boldsymbol{I}}_i - \boldsymbol{I}_i\right\|_1 + \lambda_t\left\|\bar{\boldsymbol{N}} - \bar{\boldsymbol{N}}'\right\|_2^2$$

Minimize intensity differences btw synthesized $\hat{\boldsymbol{I}}_i$ and observed $\boldsymbol{I}_i$ images.

Constrain the output normals $\bar{\boldsymbol{N}}$ to be close to prior normals $\bar{\boldsymbol{N}}'$ obtained by the LS method.

Test-time learning with early-stage weak supervision

Initialize network parameters randomly.
Compute LS solution $\bar{\boldsymbol{N}}'$.
Repeat Adam's iterations

- Run PSNet to produce a normal map $\bar{\boldsymbol{N}}$.
- Run IRNet to reconstruct all input images as $\{\hat{\boldsymbol{I}}_i\}$.
- Compute the loss and update network parameters.
- Terminate the prior ($\lambda_t \leftarrow 0$) if iterations > 50 (because the prior has low accuracy)

Until convergence (1000 iterations)

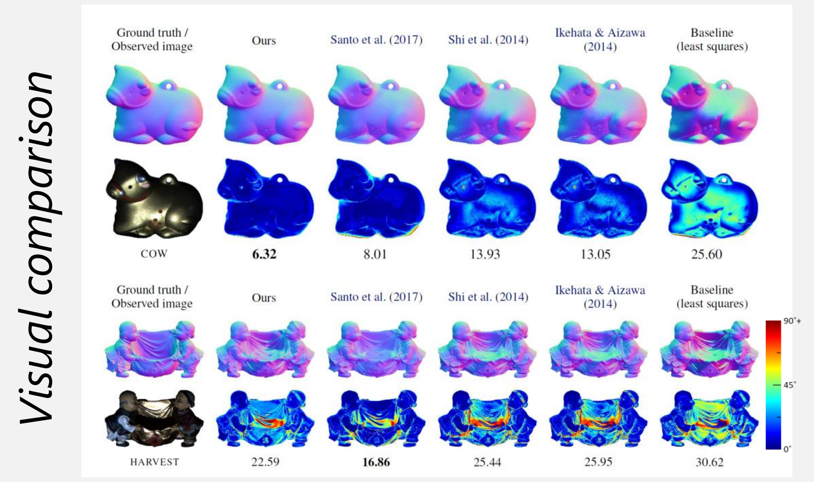
No pre-training
Directly optimize randomly-initialized network parameters for a given test scene images.
(*deep image prior [Dmitry+18]*)

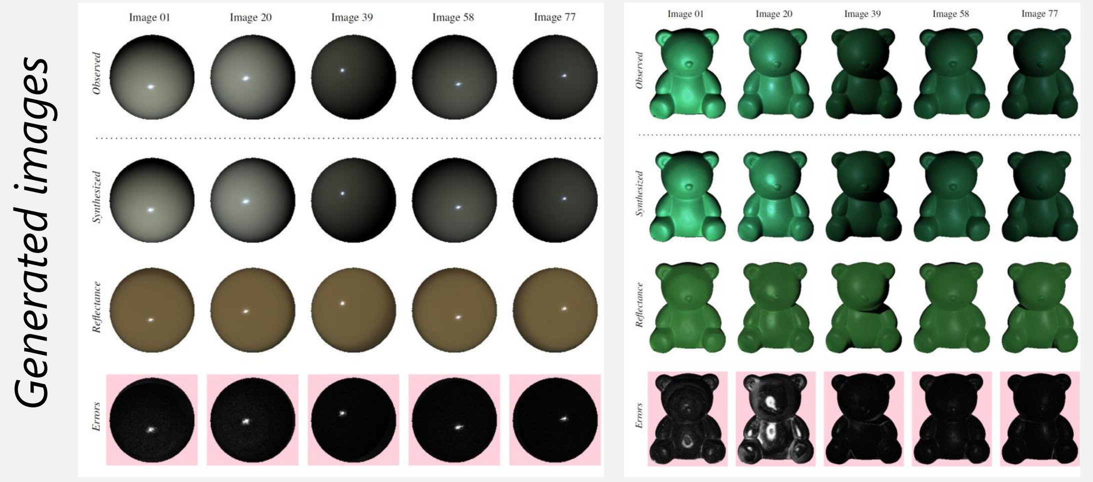
Experiments

Real-world scene benchmark (mean angular errors in degrees) [Shi+18]

	BALL	CAT	POT1	BEAR	POT2	BUDDHA	GOBLET	READING	COW	HARVEST	AVG.
Proposed	**1.47**	**5.44**	**6.09**	**5.79**	**7.76**	**10.36**	11.47	**11.03**	**6.32**	22.59	**8.83**
Santo et al. (2017)	2.02	6.54	7.05	6.31	7.86	12.68	11.28	15.51	8.01	**16.86**	9.41
Shi et al. (2014)	1.74	6.12	6.51	6.12	8.78	10.60	10.09	13.63	13.93	25.44	10.30
Ikehata & Aizawa (2014)	3.34	6.74	6.64	7.11	8.77	10.47	**9.71**	14.19	13.05	25.95	10.60
Goldman et al. (2010)	3.21	8.22	8.53	6.62	7.90	14.85	14.22	19.07	9.55	27.84	12.00
Alldrin et al. (2008)	2.71	6.53	7.23	5.96	11.03	12.54	13.93	14.17	21.48	30.50	12.61
Higo et al. (2010)	3.55	8.40	10.85	11.48	16.37	13.05	14.89	16.82	14.95	21.79	13.22
Wu et al. (2010)	2.06	6.73	7.18	6.50	13.12	10.91	15.70	15.39	25.89	30.01	13.35
Ikehata et al. (2012)	2.54	7.21	7.74	7.32	14.09	11.11	16.25	16.17	25.70	29.26	13.74
Shi et al. (2012)	13.58	12.34	10.37	19.44	9.84	18.37	17.80	17.17	7.62	19.30	14.58
Baseline (least squares)	4.10	8.41	8.89	8.39	14.65	14.92	18.50	19.80	25.60	30.62	15.39

- Each scene is provided 96 images with known lightings.
- Santo et al. (2017) use a supervised DNN method pre-trained on synthetic data.
- Others are classical physics-based unsupervised methods.





Analysis of network architecture and early-stage weak supervision

	S	G	WS	BALL		CAT		POT1		BEAR		POT2		BUDDHA		GOBLET		READING		COW		HARVEST		AVG.	
Proposed	✓	✓	✓	**1.47**	**1.50**	5.44	**5.38**	**6.09**	**6.15**	5.79	5.84	7.76	7.71	10.36	10.22	11.47	**11.35**	11.03	10.98	6.32	6.26	22.59	**22.63**	**8.83**	**8.80**
No supecular input		✓	✓	1.64	1.63	7.09	7.06	7.78	7.77	5.53	**5.55**	8.47	8.34	11.23	11.22	14.53	14.59	**10.71**	**10.75**	19.04	18.83	26.75	26.71	11.28	11.25
No global observation	✓		✓	1.50	1.50	13.18	15.12	8.47	8.50	5.76	5.74	**7.50**	**7.51**	12.76	12.68	12.50	12.54	16.81	20.20	**5.40**	**5.44**	25.12	25.34	10.90	11.46
No supervision	✓	✓		1.61	1.58	**5.30**	5.97	6.25	10.91	**5.53**	8.10	8.18	8.70	**10.08**	**10.16**	11.67	14.29	11.20	20.27	6.03	6.72	**22.48**	32.12	**8.83**	11.88
All-stage supervision	✓	✓	★	1.65	1.63	5.50	5.55	6.20	6.16	5.56	5.55	8.12	8.12	10.18	10.22	**11.34**	11.54	12.98	13.37	9.56	9.80	24.05	23.90	9.51	9.58

+10° 0° -10°