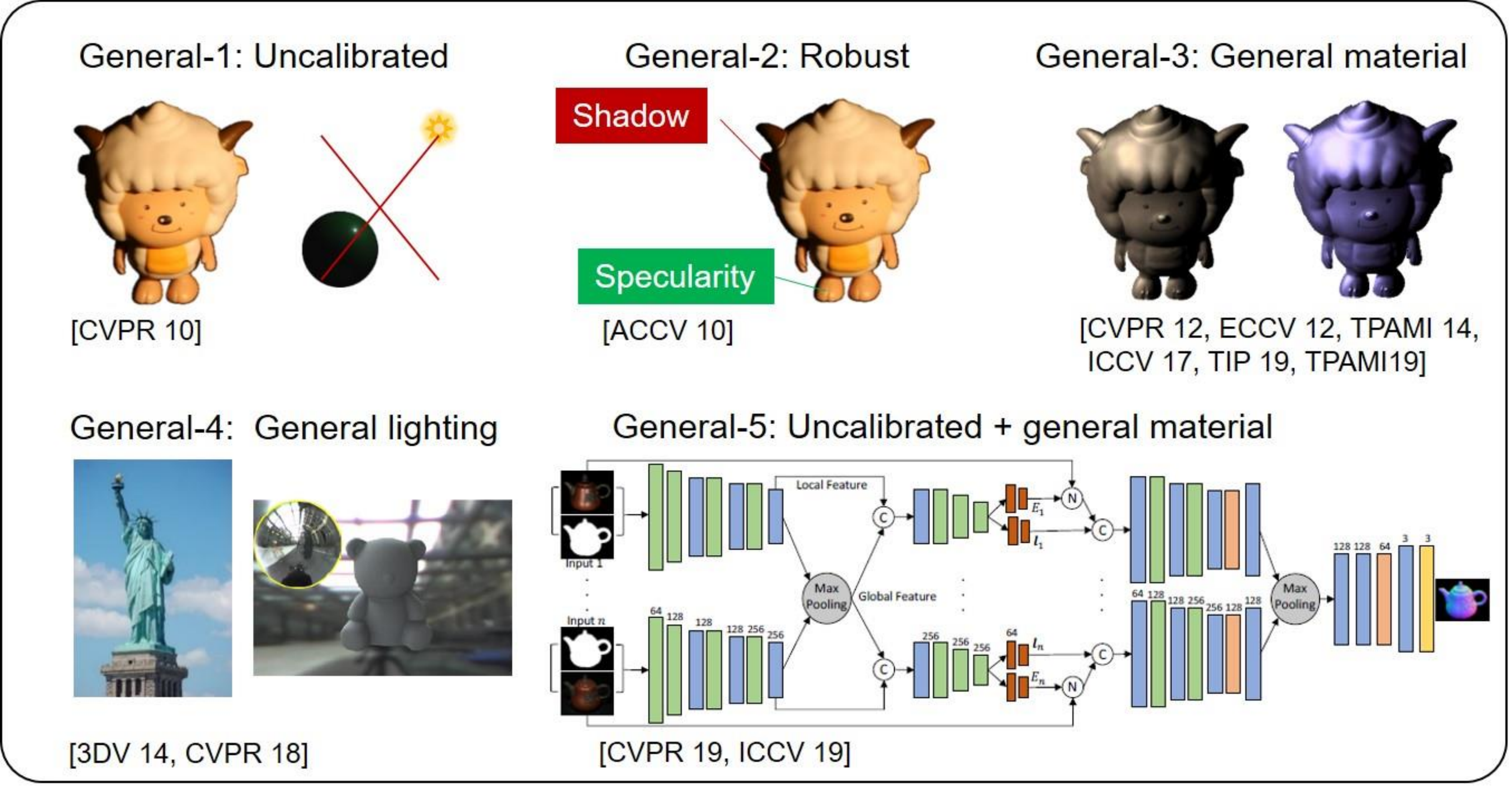
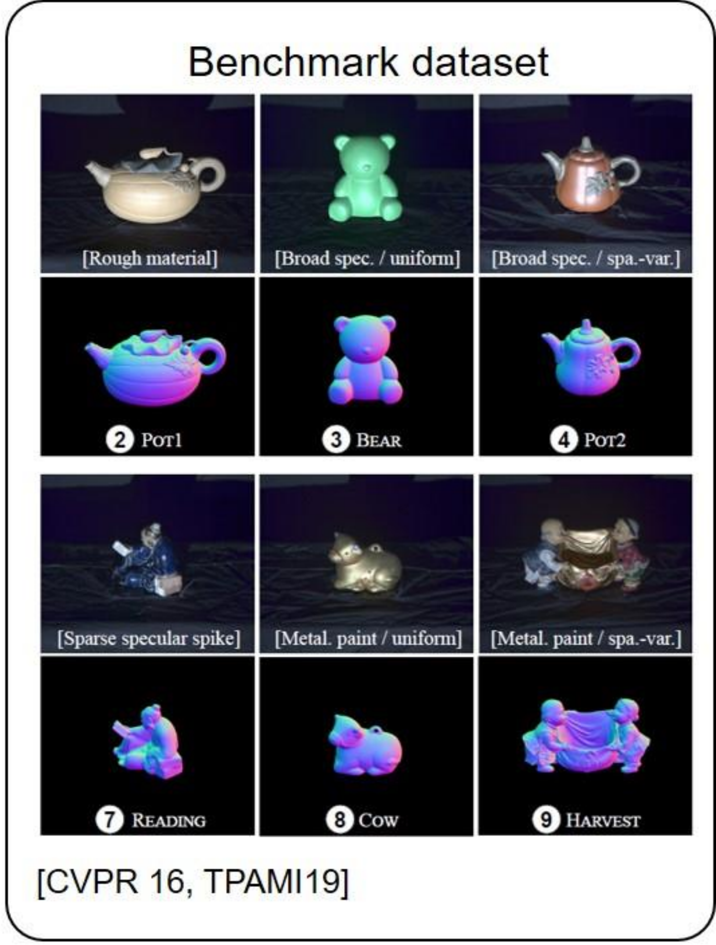




Data-driven Photometric 3D Modeling

Boxin Shi (Peking University)

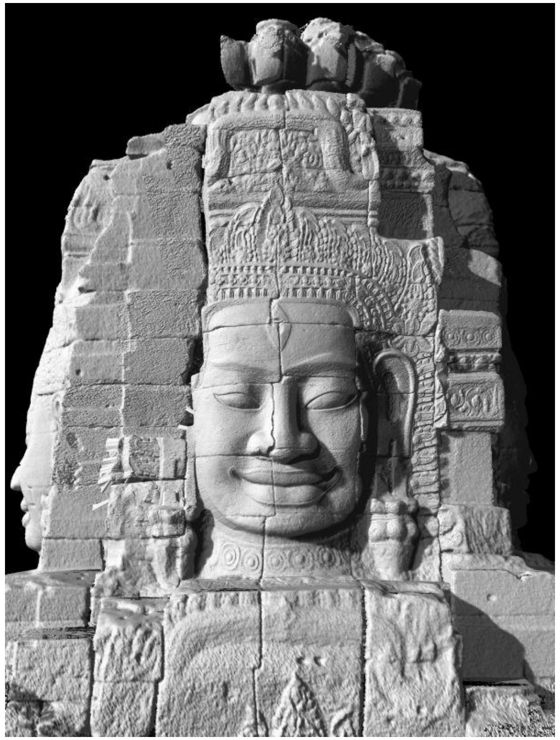
http://www.shiboxin.com / shiboxin@pku.edu.cn

Photometric Stereo Basics

SA '19 Courses, November 17-20, 2019, Brisbane, QLD, Australia
ACM 978-1-4503-6941-1/19/11.
10.1145/3355047.3359422

3D imaging

3D modeling methods



Laser range scanning
Bayon Digital Archive Project
Ikeuchi lab., UTokyo

3D modeling methods

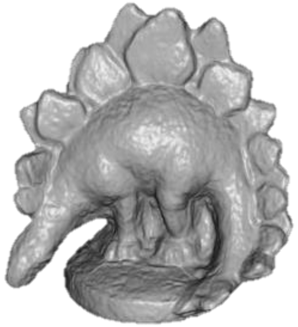
Multiview stereo

Reconstruction
[Furukawa 10]

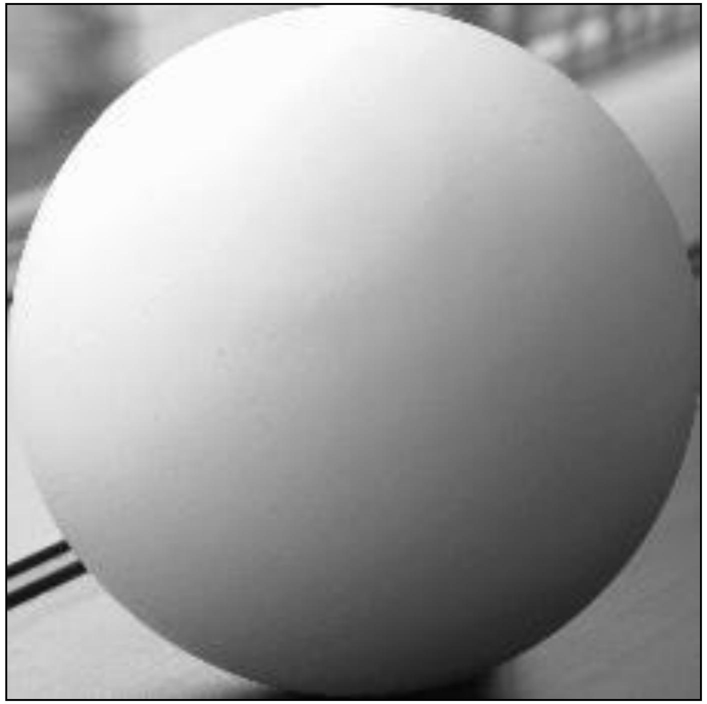
Ground truth

Geometric vs. photometric approaches

	Geometric approach	Photometric approach
Gross shape	○	×
Detailed shape	×	○



Shape from image intensity



149	127	168	210	222	232	239	233	200	152	145	144	134	88
147	113	184	252	255	254	248	239	232	220	188	150	178	115
113	145	248	254	251	245	235	226	215	203	188	173	190	104
130	239	255	250	245	236	224	212	197	181	170	150	144	86
188	255	248	243	236	225	212	197	177	163	150	136	124	70
213	250	241	234	226	214	197	179	162	148	135	122	114	57
222	240	233	224	211	195	178	164	149	134	121	110	104	57
216	231	223	213	197	181	165	149	134	120	107	98	94	52
206	219	209	195	180	165	150	135	121	107	95	86	82	53
201	199	187	175	162	149	134	121	108	95	84	75	73	62
182	163	164	155	143	131	118	105	92	81	72	59	85	71
111	160	150	129	121	110	98	86	75	65	52	67	120	71
186	211	187	129	94	81	71	60	50	41	68	116	123	67
192	199	195	180	126	85	79	72	61	76	130	137	136	96
213	214	212	209	204	200	199	200	200	204	208	209	209	206

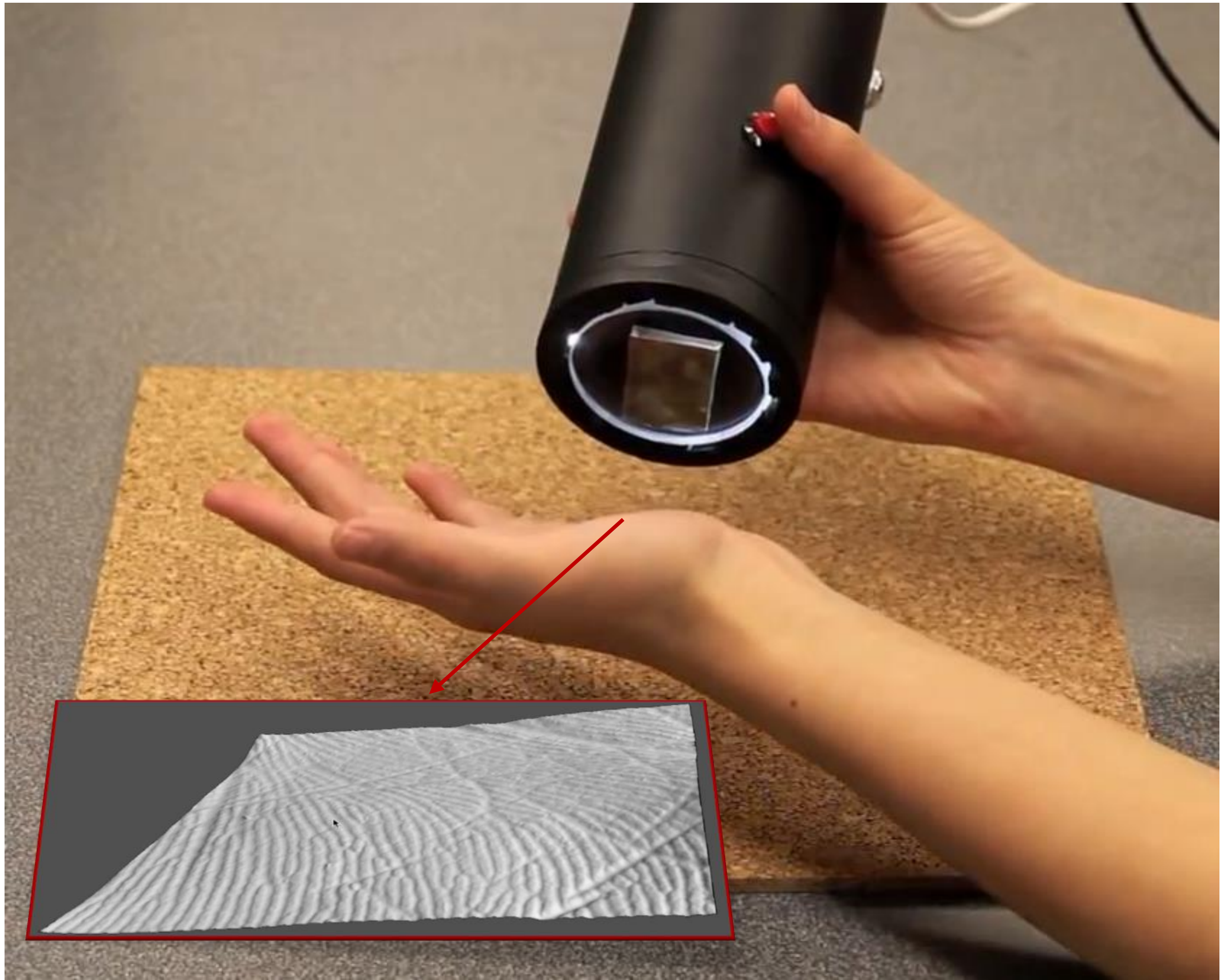
How can machine understand the shape from image intensities ?

Photometric 3D modeling

3D Scanning the President of the United States
P. Debevec et al., USC, 2014

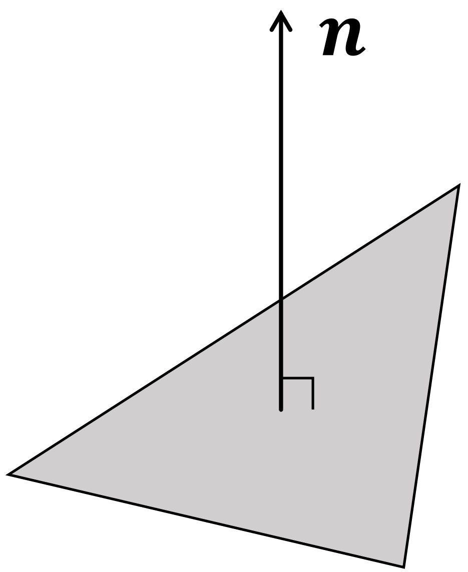
Photometric 3D modeling

GelSight Microstructure 3D Scanner
E. Adelson et al., MIT, 2011



Preparation 1: Surface normal

A surface normal $\boldsymbol{n}$ to a surface is a vector that is **perpendicular** to the tangent plane to that surface.



$$\boldsymbol{n} \in \mathcal{S}^2 \subset \mathbb{R}^3, \|\boldsymbol{n}\|_2 = 1$$

$$\boldsymbol{n} = \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}$$

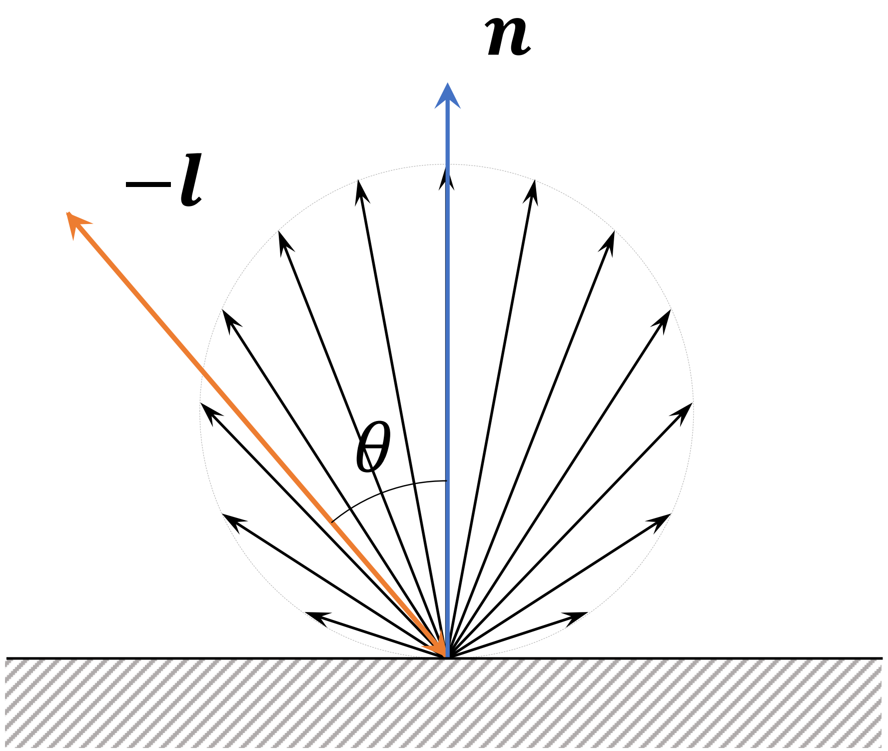
Preparation 2: Lambertian reflectance

- Amount of reflected light proportional to $\boldsymbol{l}^T\boldsymbol{n}$ $(= \cos\theta)$

- Apparent brightness does not depend on the viewing angle.

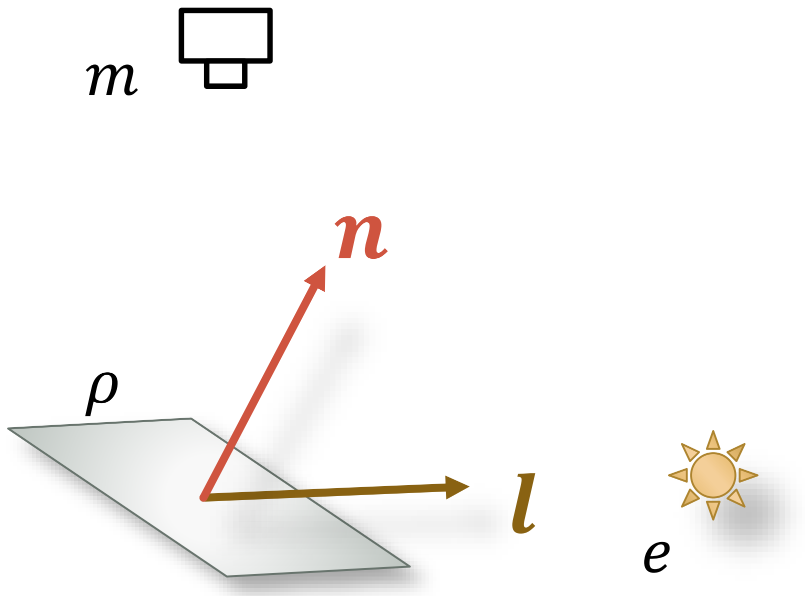
$$\boldsymbol{l} \in \mathcal{S}^2 \subset \mathbb{R}^3, \|\boldsymbol{l}\|_2 = 1$$

$$\boldsymbol{l} = \begin{bmatrix} l_x \\ l_y \\ l_z \end{bmatrix}$$



Lambertian image formation model

$$m \propto e\rho \boldsymbol{l}^T \boldsymbol{n} = e\rho \begin{bmatrix} l_x & l_y & l_z \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}$$



$m \in \mathbb{R}_+$: Measured intensity for a pixel
$e \in \mathbb{R}_+$: Light source intensity (or radiant intensity)
$\rho \in \mathbb{R}_+$: Lambertian diffuse reflectance (or albedo)
$\boldsymbol{l}$: 3-D unit light source vector
$\boldsymbol{n}$: 3-D unit surface normal vector

Simplified Lambertian image formation model

$$m \propto e\rho \boldsymbol{l}^T \boldsymbol{n} = e\rho \begin{bmatrix} l_x & l_y & l_z \end{bmatrix} \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix}$$

$$m = \rho \boldsymbol{l}^T \boldsymbol{n}$$

Photometric stereo

[Woodham 80]

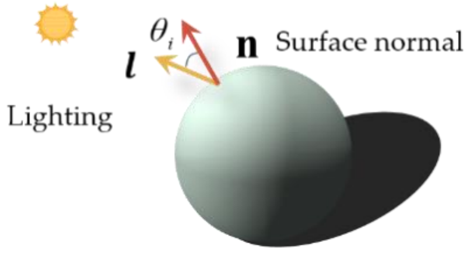
Assuming $\rho = 1$

j-th image under
j-th lightings l_j,
In total f images

For a pixel with
normal direction $\mathbf{n}$

$$\begin{cases} I_1 = \boldsymbol{n} \cdot \boldsymbol{l}_1 \\ I_2 = \boldsymbol{n} \cdot \boldsymbol{l}_2 \\ \cdots \\ I_f = \boldsymbol{n} \cdot \boldsymbol{l}_f \end{cases}$$

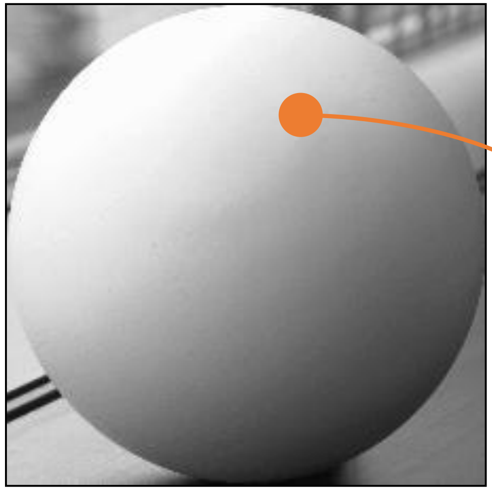
$$I = \cos\theta_i = \mathbf{n} \cdot \mathbf{l}$$



$$[I_1, I_2, \cdots, I_f] = [n_x, n_y, n_z] \begin{bmatrix} l_{1x} & l_{2x} & & l_{fx} \\ l_{1y} & l_{2y} & \cdots & l_{fy} \\ l_{1z} & l_{2z} & & l_{fz} \end{bmatrix}$$

Photometric stereo

Matrix form $\boldsymbol{I} = \boldsymbol{N}\boldsymbol{L}$



$$\underbrace{\boldsymbol{I}}_{p \times f} = \underbrace{\boldsymbol{N}}_{p \times 3}\ \underbrace{\boldsymbol{L}}_{3 \times f}$$

p: Number of pixels

f: Number of images

Least squares solution : $\boldsymbol{N} = \boldsymbol{I}\boldsymbol{L}^{+}$

Photometric stereo: An example

Captured

$$\begin{bmatrix} I \end{bmatrix} = \begin{bmatrix} N \end{bmatrix} \begin{bmatrix} L \end{bmatrix}$$

Calibrated

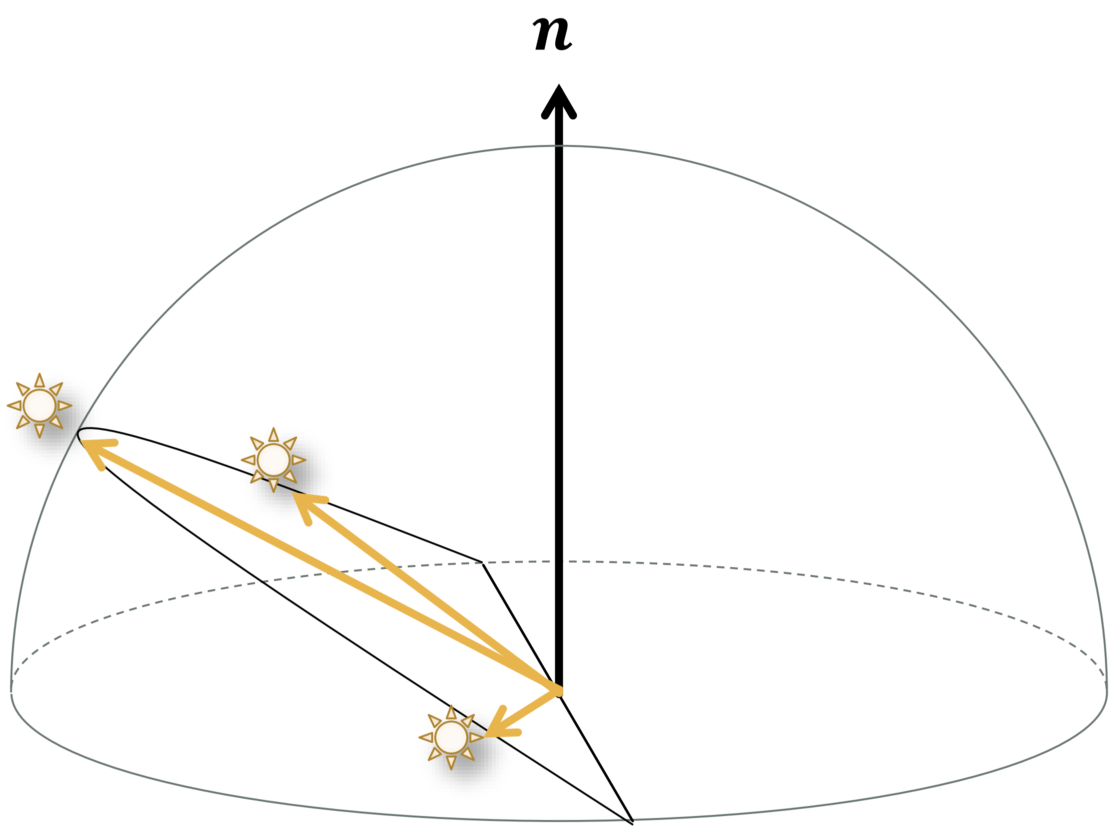
To estimate

$$\boxed{N = IL^{+}}$$

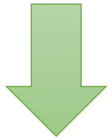
Normal map

Degenerate case

- Light sources locate on a plane (co-planar)



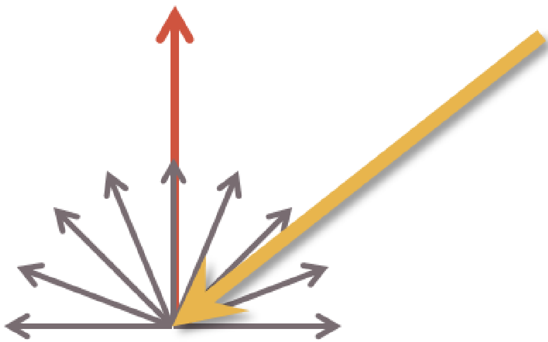
$$\begin{bmatrix} \\ \boldsymbol{I} \\ \\ \end{bmatrix} = \begin{bmatrix} \\ \boldsymbol{N} \\ \\ \end{bmatrix} \underline{\begin{bmatrix} \boldsymbol{L} \end{bmatrix}}$$



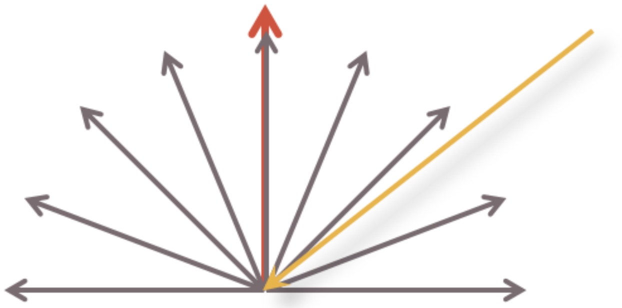
Not invertible

Diffuse albedo

- We have ignored diffuse albedo so far
  - $\boldsymbol{I} = \boldsymbol{NL}$
- Normalizing the surface normal $\boldsymbol{n}$ to 1, we obtain diffuse albedo (magnitude of $\boldsymbol{n}$)
  - $\rho = |\boldsymbol{n}|$
- Diffuse albedo is a relative value



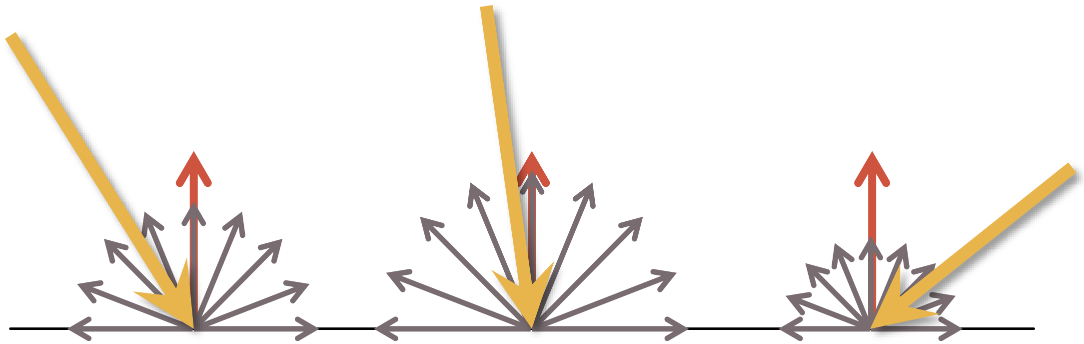
Low reflectance
High light-source intensity



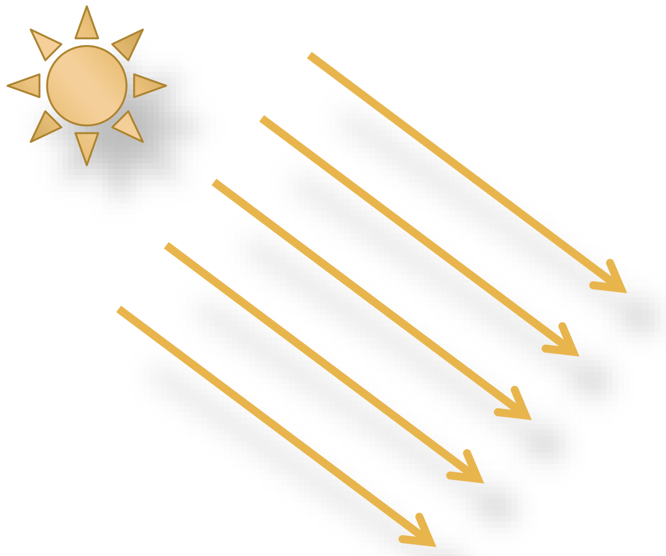
High reflectance
Low light-source intensity

So far, limited to…

- Lambertian reflectance



- Known, distant lighting



Generalization of photometric stereo

- ~~Lambertian reflectance~~

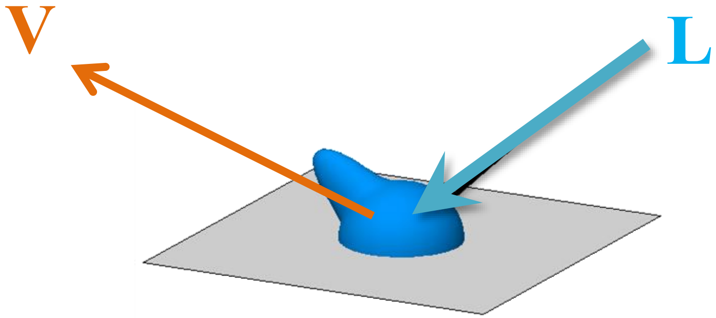
Outliers beyond Lambertian

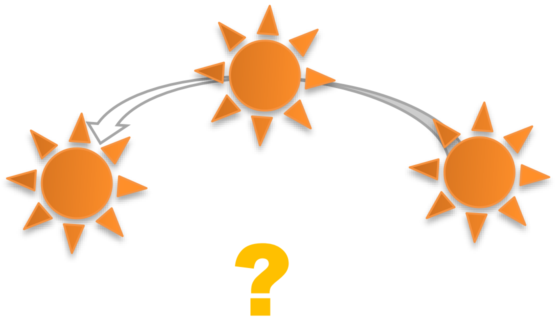
General BRDF

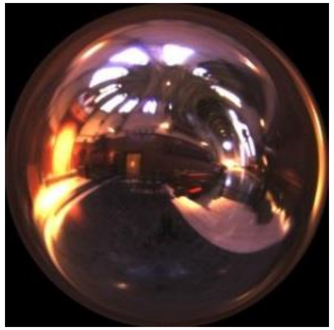
- ~~Known, distant lighting~~

Unknown distant lighting

Unknown general lighting





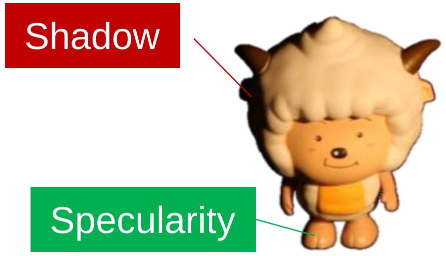


Generalization of photometric stereo

General-1: Uncalibrated

[CVPR 10]

General-2: Robust



[ACCV 10]

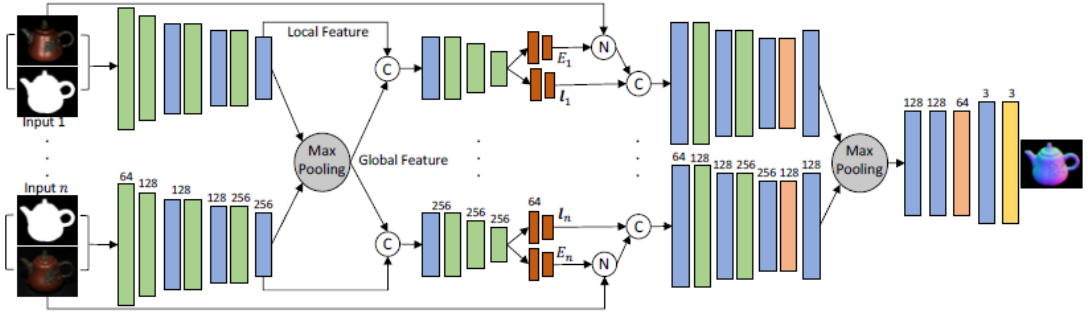
General-3: General material

[CVPR 12, ECCV 12, TPAMI 14, ICCV 17, TIP 19, TPAMI19]

General-4: General lighting

[3DV 14, CVPR 18]

General-5: Uncalibrated + general material



[CVPR 19, ICCV 19]

Benchmark dataset

[CVPR 16, TPAMI19]

Photometric Stereo Taxonomy

Photometric stereo taxonomy

- Reflectance model

L Lambert's model

R Robust methods

(Lambert's model + outliers)

A Analytic model

G General properties of BRDF

Photometric stereo taxonomy

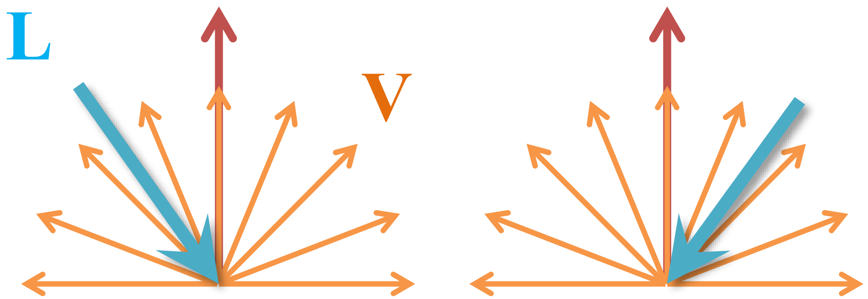
- Reflectance model

L Lambert's model

BASELINE

Simplest

Most widely used



R Robust methods

(Lambert's model + outliers)

A Analytic model

G General properties of BRDF

Photometric stereo taxonomy

- Reflectance model
  - **L** Lambert's model
  - **R** Robust methods (Lambert's model + outliers)
  - **A** Analytic model
  - **G** General properties of BRDF

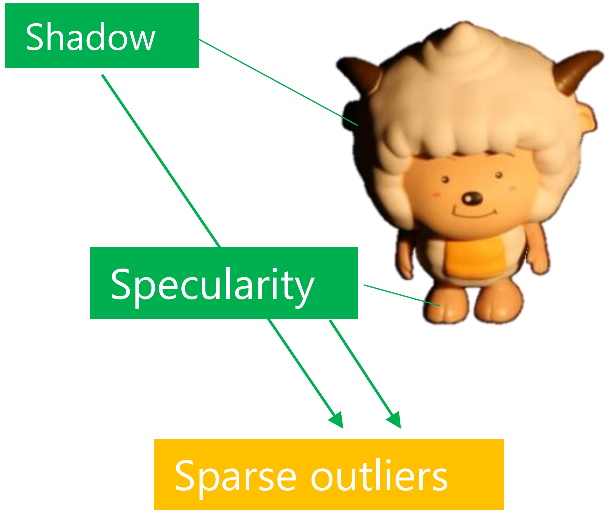
Outlier rejection:

Early four-lights method
[Solomon 96]
[Barsky 03]

RANSAC
[Mukaigawa 07]
Median approach
[Miyazaki 10]

Rank minimization
[Wu 10] WG10
[Ikehata 12] IW12

$$\mathbf{I} = \mathbf{NL} + \mathbf{E}$$



Non-Lambertian outliers

Cast shadow

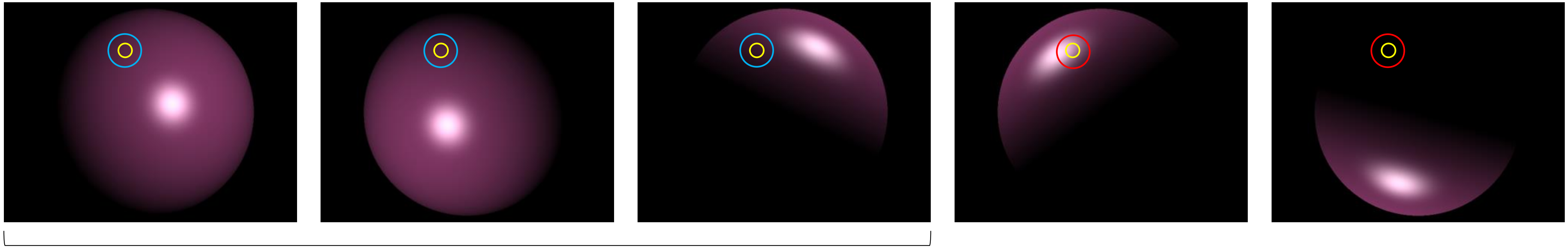
Attached shadow

Specularity

Non-Lambertian photometric stereo

[Coleman 82, Barsky 03]

- Robust approach
  - More than 3 images for removing non-Lambertian effects as outliers



$$N = IL^{-1}$$

Robust PCA approach

[Wu 10]

- Traditional solution method
  - Least-squares solution
  - $\boldsymbol{D} = \boldsymbol{NL} \rightarrow \widehat{\boldsymbol{N}} = \boldsymbol{DL}^{+}$
    - Observation matrix $\boldsymbol{D} \in \mathbb{R}^{p \times f}$ (p pixels, f images)
    - Normal matrix $\boldsymbol{N} \in \mathbb{R}^{p \times 3}$
    - Light matrix $\boldsymbol{L} \in \mathbb{R}^{3 \times f}$

➡ $\boldsymbol{D}$ has a low-rank structure

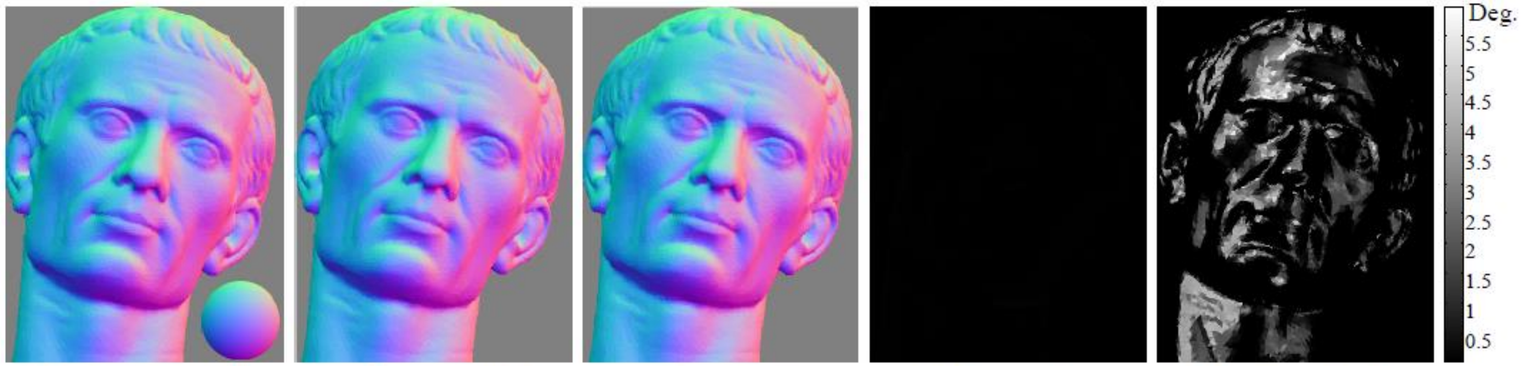
Robust PCA approach

- Low-rank matrix structure
  - $\boldsymbol{D} = \boldsymbol{N}\boldsymbol{L}$
  - Rank of $\boldsymbol{D}$ should be at most 3, irrespective to p and f
- Modeling corruptions as sparse errors
  - Shadows, specularities breaks the low-rank structure
  - Model these corruptions as $\boldsymbol{E}$
  - $\boldsymbol{D} = \boldsymbol{N}\boldsymbol{L} + \boldsymbol{E} \ \ (= \boldsymbol{A} + \boldsymbol{E})$

Robust PCA approach

- Formulation
  - $\min_{\boldsymbol{A},\boldsymbol{E}} \operatorname{rank}(\boldsymbol{A}) + \gamma \|\boldsymbol{E}\|_0$ s.t. $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$
- Solution via convex programming
  - $\min_{\boldsymbol{A},\boldsymbol{E}} \|\boldsymbol{A}\|_* + \gamma \|\boldsymbol{E}\|_1$ s.t. $\boldsymbol{D} = \boldsymbol{A} + \boldsymbol{E}$

Robust PCA approach



(a) Ground truth (b) Our method (c) Least Squares (d) Error map (our method) (e) Error map (LS)

Object	Mean error (in degrees)		Max. error (in degrees)		Avg. % of corrupted pixels	
	LS	Our method	LS	Our method	Shadow	Specularity
Sphere	0.99	$\mathbf{5.1 \times 10^{-3}}$	8.1	**0.20**	18.4	16.1
Caesar	0.96	$\mathbf{1.4 \times 10^{-2}}$	8.0	**0.22**	20.7	13.6
Elephant	0.96	$\mathbf{8.7 \times 10^{-3}}$	8.0	**0.29**	18.1	16.5

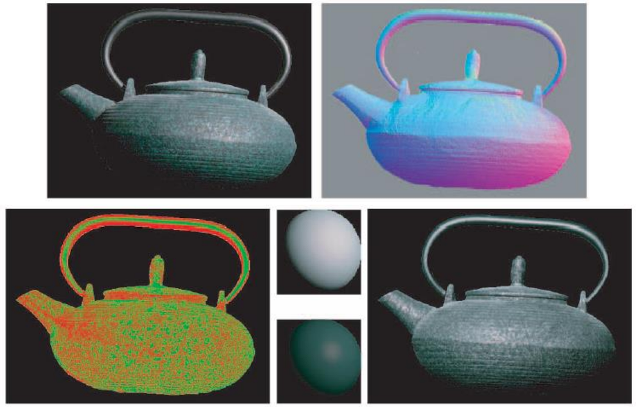
Photometric stereo taxonomy

- Reflectance model
  - L Lambert's model
  - R Robust methods (Lambert's model + outliers)
  - A Analytic model
  - G General properties of BRDF

Torrance-Sparrow model
[Georghiades 03]

Ward model
[Chung 08]

Mixture of Ward lobes
[Goldman 10] GC10



Photometric stereo taxonomy

- Reflectance model

L Lambert's model

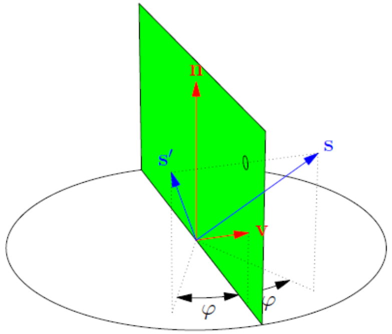
R Robust methods

(Lambert's model + outliers)

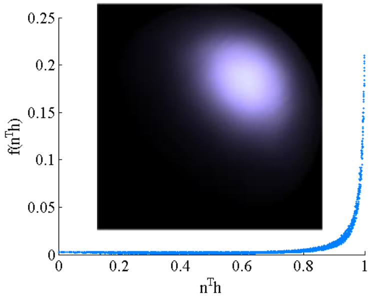
A Analytic model

G General properties of BRDF

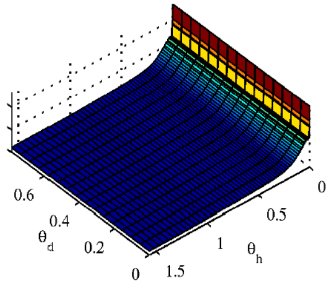
Isotropy
[Alldrin 07]
[Tan 11]
[Chandraker 13]



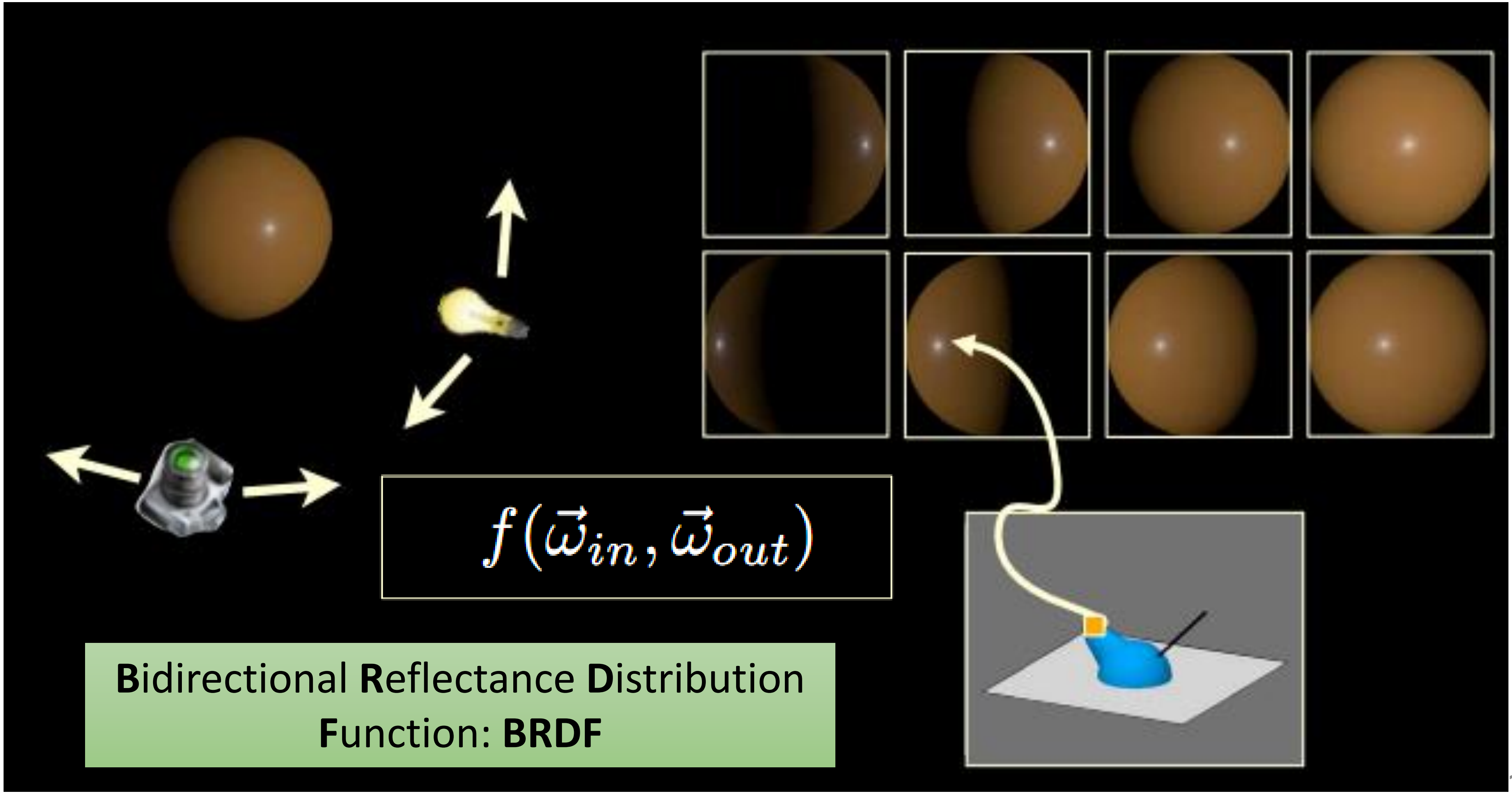
HM10 Monotonicity
[Higo 10]
[Shi 12] ST12



AZ08 Bi-variate model
[Alldrin 08]
[Shi 14] ST14
IA14 [Ikehata 14]



General material reflectance modeling



Bi-polynomial model for photometric stereo

[Shi 12, 14]

Disregarded by thresholding

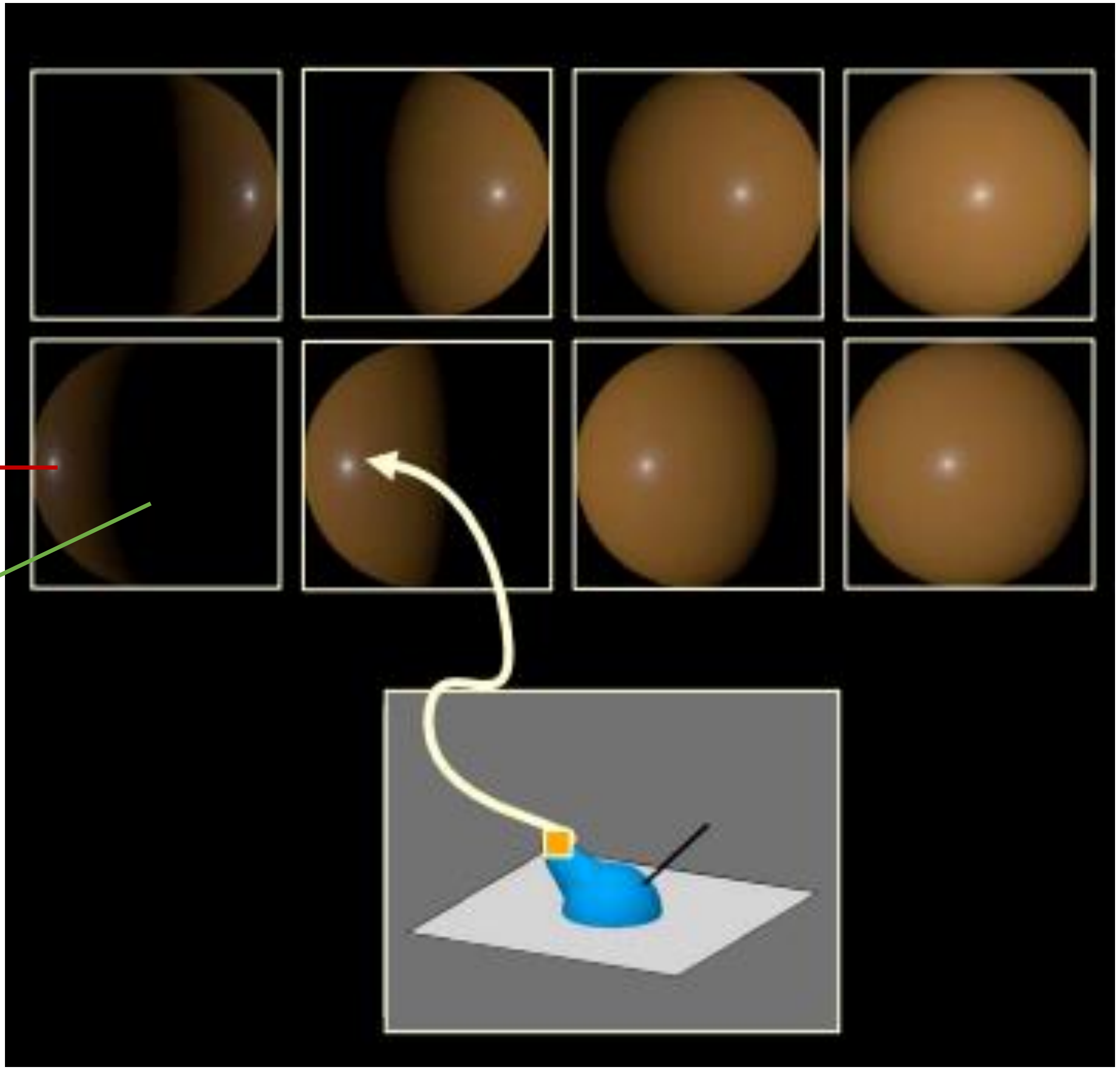
High-frequency reflectance

$$\frac{\rho_s}{4\pi\alpha^2}\sqrt{\frac{1}{(\mathbf{n}^T\mathbf{l})(\mathbf{n}^T\mathbf{v})}}\, e^{\frac{1}{\alpha^2}\left(1-\frac{1}{\mathbf{n}^T\mathbf{h}}\right)}$$

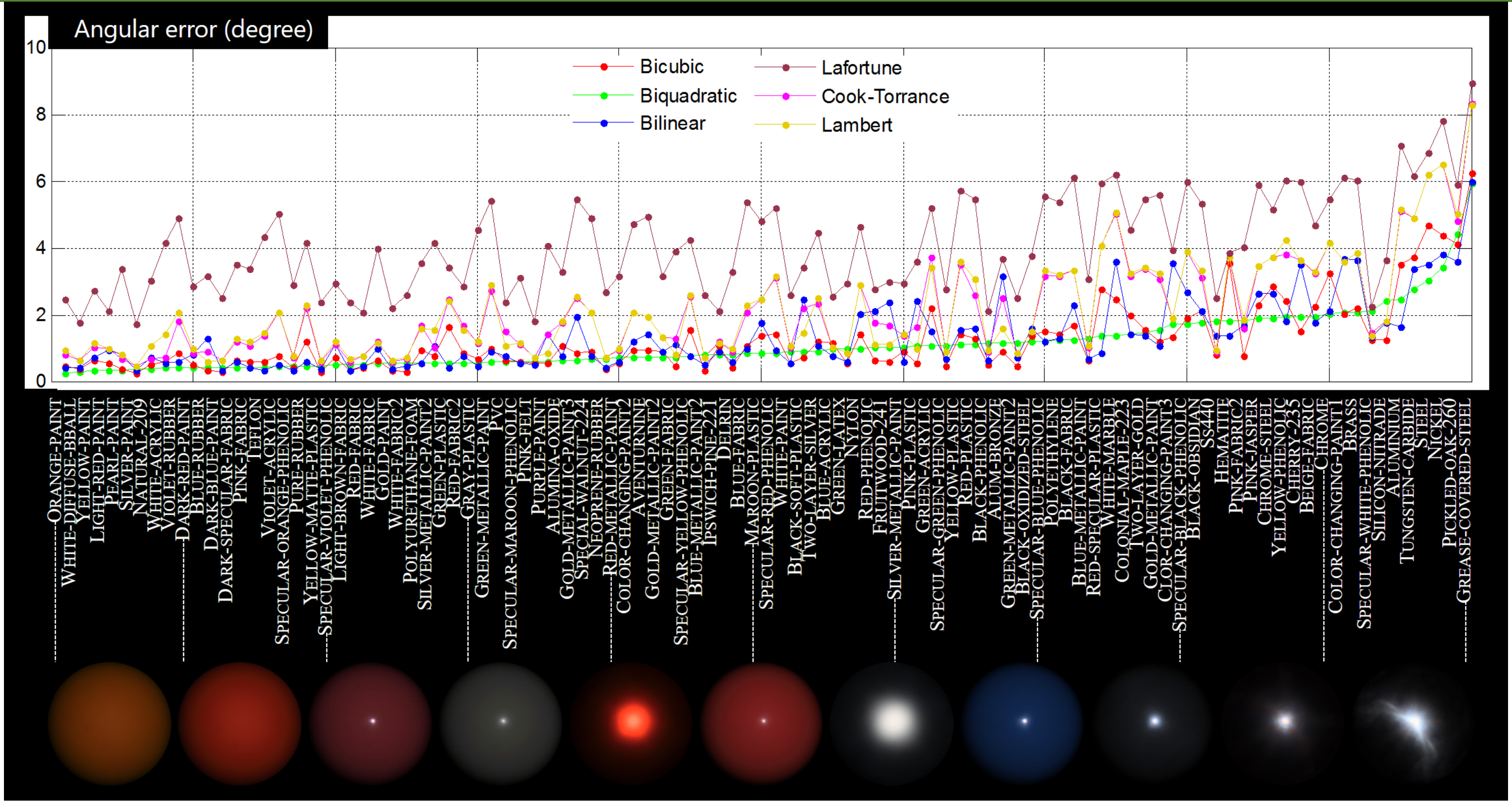
***h** is the bisector of lighting direction **l** and viewing direction **v**

Low-frequency reflectance

- Conventional approach: Lambertian
  - ρ is a constant
  - Simplest but inaccurate
- Proposed approach: Bi-polynomial
  - $(A_2(\mathbf{n}^T\mathbf{h})^2 + A_1(\mathbf{n}^T\mathbf{h}) + A_0)(B_2(\mathbf{l}^T\mathbf{h})^2 + B_1(\mathbf{l}^T\mathbf{h}) + B_0)$
  - Simple equation with general modeling ability

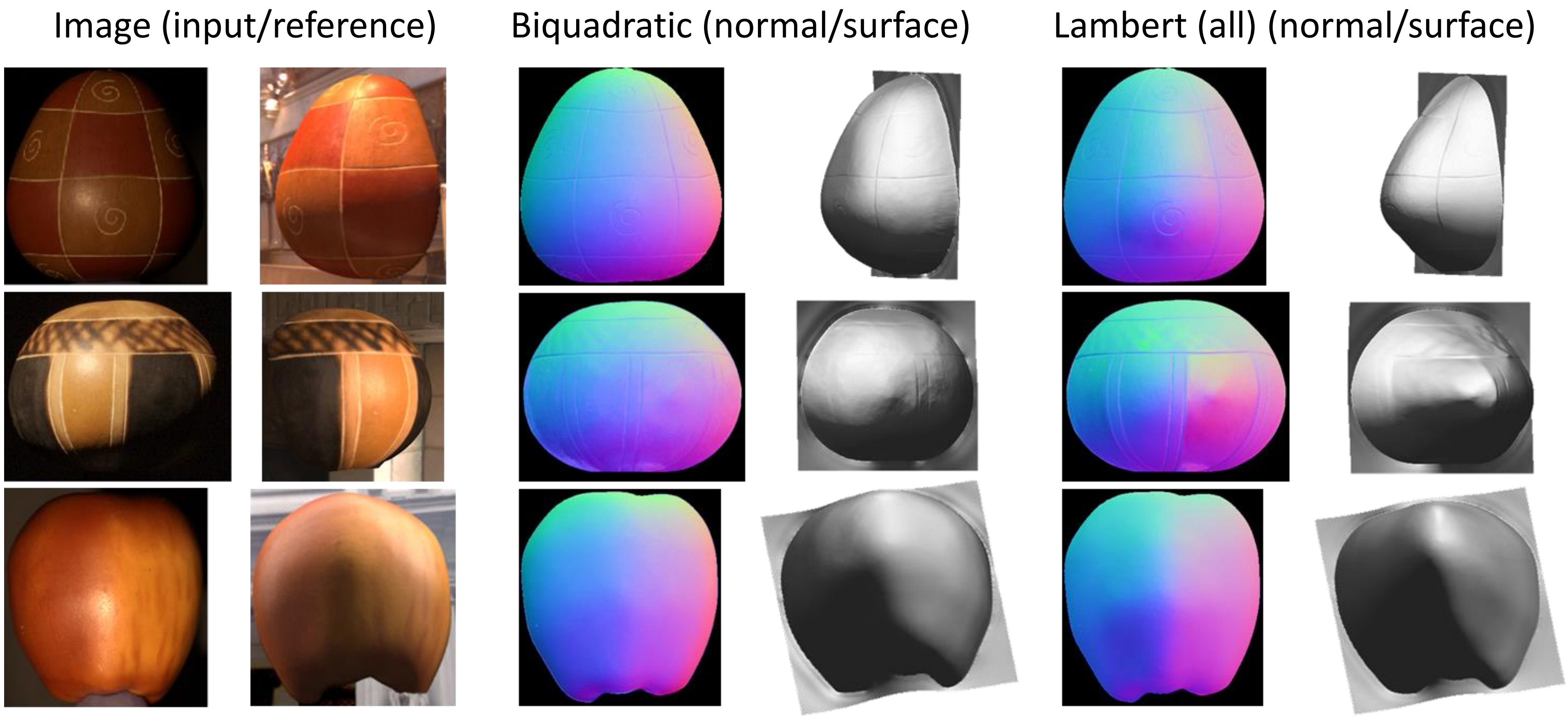


Accuracy on MERL BRDF dataset [Matusik 03]



AVG. 100	Bicubic	Biquadratic	Bilinear	Lafortune	C.-Torrance	Lambert
Ang. Err.	1.25	1.12	1.37	4.07	2.13	2.14

Photometric stereo results on real objects



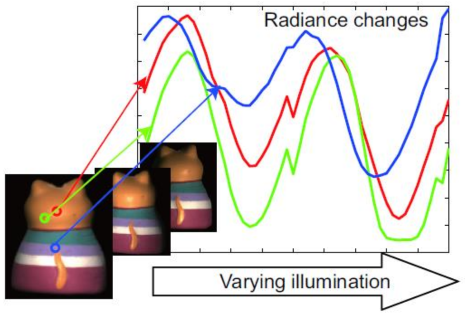
* Data courtesy of N. Alldrin

Photometric stereo taxonomy

- Reflectance model
  - L Lambert's model
  - R Robust methods (Lambert's model + outliers)
  - A Analytic model
  - G General properties of BRDF

Cont.

Manifold embedding
[Sato 07]
[Okabe 09]
[Lu 13] LM13



Example-based
[Hertzmann 05]
[Johnson 11]

Non-Lambertian methods

Solve $\mathbf{N}$ from $\mathbf{I} = \max\{\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \circ (\mathbf{N}^\top \mathbf{L}), 0\}$ by using different assumptions and constraints on $\boldsymbol{\rho}(\mathbf{n},\mathbf{l})$ Notations: $\mathbf{h} = (\mathbf{l}+\mathbf{v})/\lVert\mathbf{l}+\mathbf{v}\rVert$, $\theta_h = \langle \mathbf{n}, \mathbf{h} \rangle = \arccos(\mathbf{n}^\top \mathbf{h})$, $\theta_d = \langle \mathbf{l}, \mathbf{h} \rangle = \arccos(\mathbf{l}^\top \mathbf{h})$	
BASELINE	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \approx \mathbf{D}$, where each row of $\mathbf{D}$ is a constant representing the albedo of a Lambertian surface
WG10	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \approx \mathbf{D} + \mathbf{E}$, where $\mathbf{E}$ is sparse and $\mathrm{rank}(\mathbf{I})$ is minimized
IW12	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \approx \mathbf{D} + \mathbf{E}$, where $\mathbf{E}$ is sparse and $\mathrm{rank}(\mathbf{I}) = 3$
GC10	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \approx \sum_i \mathbf{w}_i \circ \boldsymbol{\rho}_i(d_i, s_i, \alpha_i)$, where $\boldsymbol{\rho}_i(d_i, s_i, \alpha_i) = \frac{d_i}{\pi} + \frac{s_i}{4\pi\alpha_i^2\sqrt{(\mathbf{n}^\top\mathbf{l})(\mathbf{n}^\top\mathbf{v})}} \exp\left(\frac{(1-1/\mathbf{n}^\top\mathbf{h})}{\alpha_i^2}\right)$
AZ08	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l})$ is isotropic and depends only on (θ_h, θ_d)
ST12	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l})$ is isotropic, depends only on θ_h, and is monotonic about $\mathbf{n}^\top\mathbf{h}$
HM10	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l})$ is isotropic, monotonic about $\mathbf{n}^\top\mathbf{l}$, and $\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) = 0$ for $\mathbf{n}^\top\mathbf{l} \leq 0$
ST14	The low-frequency part of $\boldsymbol{\rho}(\mathbf{n},\mathbf{l})$ is a bi-polynomial $A(\cos(\theta_h))B(\cos(\theta_d))$, where A and B are polynomials
IA14	$\boldsymbol{\rho}(\mathbf{n},\mathbf{l}) \approx \sum_i \boldsymbol{\rho}_i(\mathbf{n}^\top\alpha_i)$, where $\alpha_i = (p_i\mathbf{l} + q_i\mathbf{v})/\lVert p_i\mathbf{l} + q_i\mathbf{v}\rVert$, $p_i,\ q_i$ are nonnegative unknown values

Photometric stereo taxonomy

- Lighting calibration
  - **C** Calibrated
  - **U** Uncalibrated

Photometric stereo taxonomy

- Lighting calibration
  - C Calibrated
  - U Uncalibrated

Using a mirror sphere

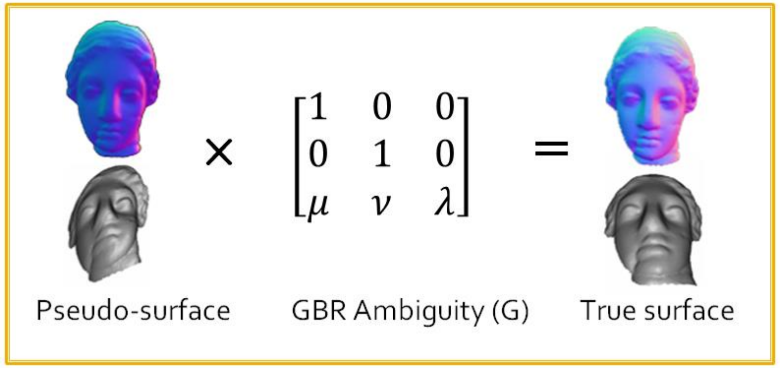
Accurate but tedious
Most methods are calibrated

Photometric stereo taxonomy

- Lighting calibration
  - C Calibrated
  - U Uncalibrated

Unknown lighting condition

Factorization based
Resolving RGB
AM07 [Alldrin 07]
[Shi 10] SM10
WT13 [Wu 13]
[Papadhimitri 14] PF14



$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}$$

SH lighting model
[Basri 07]
[Shi 14]

Manifold embedding based methods

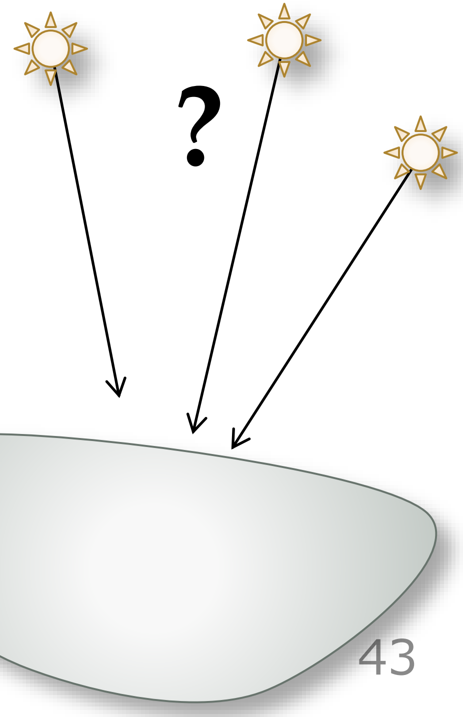
Uncalibrated photometric stereo

- Photometric stereo with **unknown** directional lighting
- Given three or more images $\boldsymbol{I}$, estimate $\boldsymbol{N}$ and $\boldsymbol{L}$

$$\underset{p \times f}{\boldsymbol{I}} = \underbrace{\underset{p \times 3}{\boldsymbol{N}}}_{\text{Unknown}} \; \underbrace{\underset{3 \times f}{\boldsymbol{L}}}_{\text{Unknown}}$$

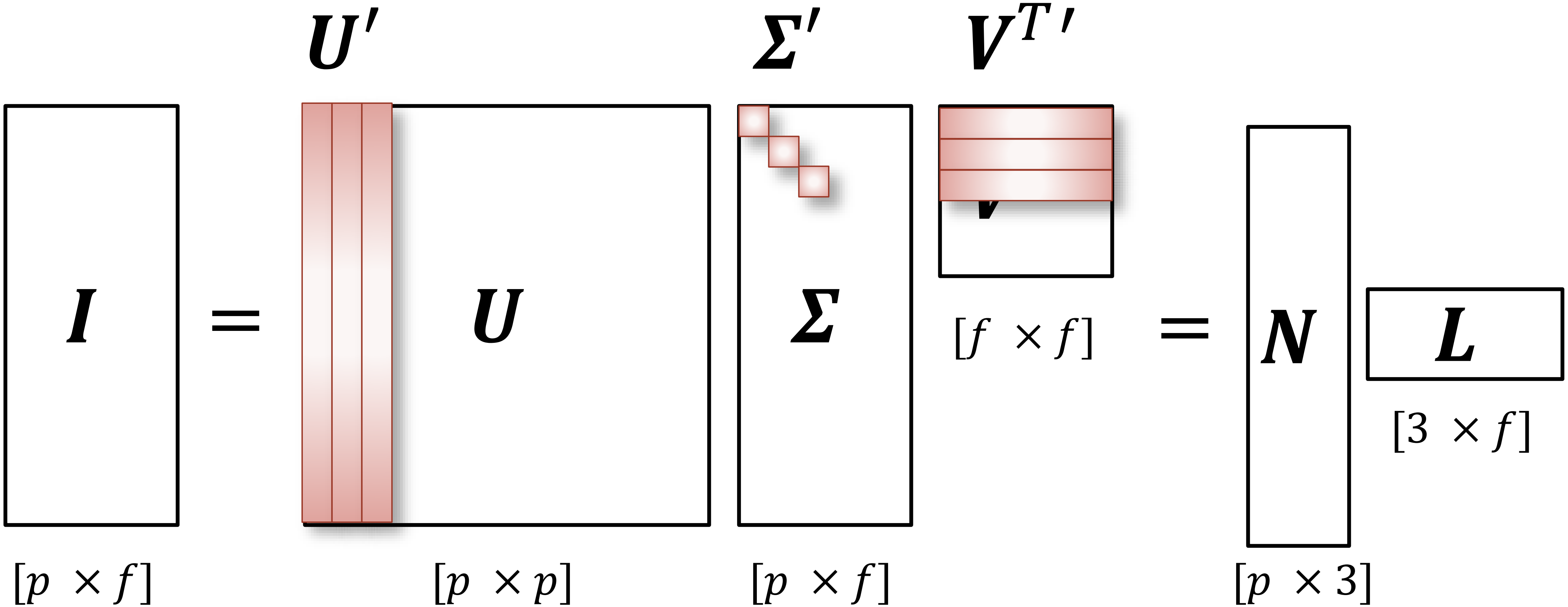
p: Number of pixels

f: Number of images



SVD approach

[Hayakawa 94]



Rank-3 approximation: $I' = U'\Sigma'V^{T\prime}$

SVD approach

- Rank-3 approximation: $\boldsymbol{I}' = \boldsymbol{U}'\boldsymbol{\Sigma}'\boldsymbol{V}^{T'}$

Surface normal	$\widehat{\boldsymbol{N}} = \boldsymbol{U}'(\boldsymbol{\Sigma}')^{\frac{1}{2}}$
Light source	$\widehat{\boldsymbol{L}} = (\boldsymbol{\Sigma}')^{\frac{1}{2}}\boldsymbol{V}'$

Is this solution unique?

No, there are ambiguities.

Ambiguities in SVD-based solution

- For any $\boldsymbol{A} \in GL(3)$, $\boldsymbol{N}^* = \widehat{\boldsymbol{N}}\boldsymbol{A}$ is also a solution, because $\boldsymbol{I}' = \widehat{\boldsymbol{N}}\widehat{\boldsymbol{L}} = \left(\widehat{\boldsymbol{N}}\boldsymbol{A}\right)\left(\boldsymbol{A}^{-1}\widehat{\boldsymbol{L}}\right) = \boldsymbol{N}^*\boldsymbol{L}^*$

$$\times \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} =$$

Pseudo-normal Ambiguity Correct solution

Generalized Bas-Relief ambiguity

[Belhumeur 97, Yuille 99]

- In general, the pseudo-normal field $\boldsymbol{N}^*$ does not have a corresponding surface
  - Subset of the solutions satisfies **integrability constraint** ($Z_{xy} = Z_{yx}$)

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}$$

Linear ambiguity → GBR ambiguity

Resolving the GBR ambiguity

[Shi 10]

- Pixels with the same albedo but different surface normals should satisfy

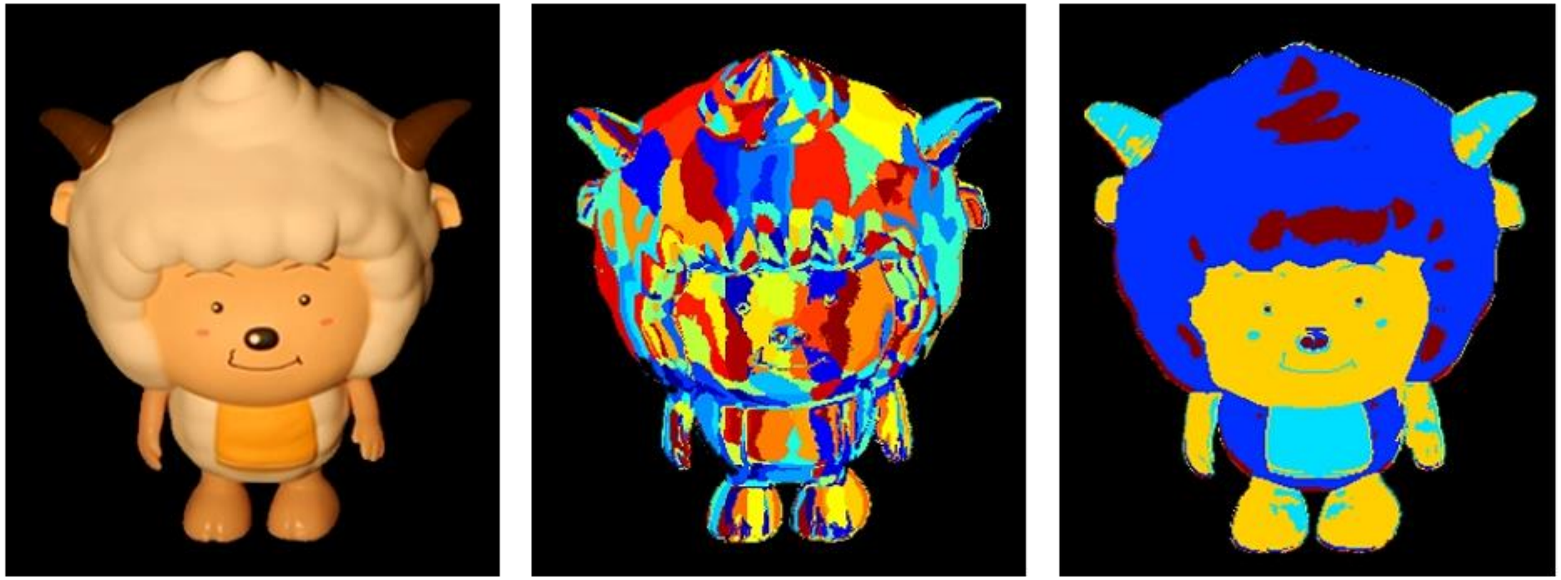
$$\mathbf{s}_i \mathbf{C} \mathbf{s}_i^T = a^2 \quad \text{where} \quad \mathbf{C} = \mathbf{G}\mathbf{G}^\mathrm{T} = \begin{pmatrix} 1 & 0 & \mu \\ 0 & 1 & \nu \\ \mu & \nu & \mu^2 + \nu^2 + \lambda^2 \end{pmatrix}$$

- 4 unknowns: can be resolved if at least 4 pixels are selected

$(\mathbf{s}_i = a\mathbf{n}_i)$

Resolving the GBR ambiguity

- K-means clustering of pixels
- Surface normal grouping using intensity profiles
- Albedo grouping using chromaticity

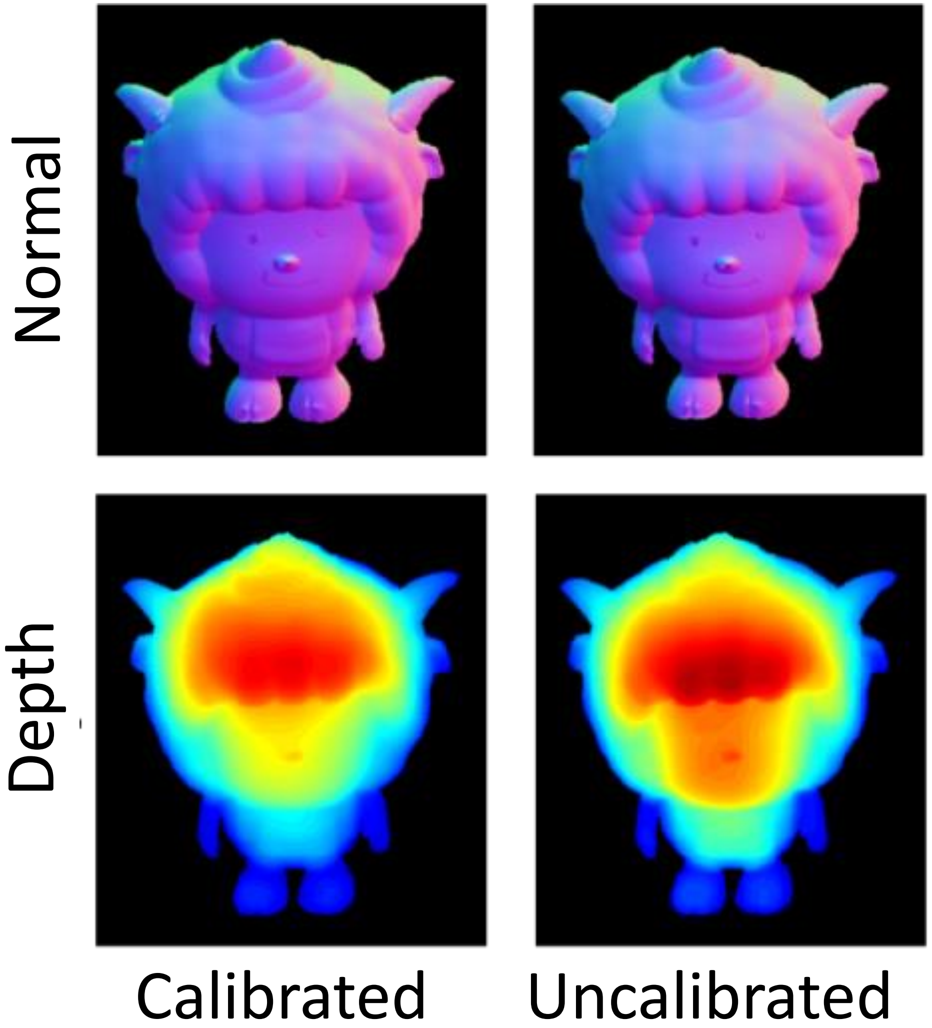


Input Normal grouping Albedo grouping

Results: resolving the ambiguity

Sheep scene

12 images



Angular error	(degree)
mean	7.30
std. dev.	3.02

Uncalibrated methods

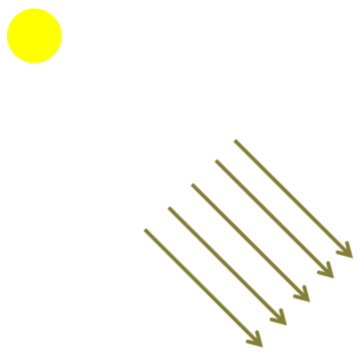
Solve $\mathbf{N}$ from $\mathbf{I} = \max\{\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \circ (\mathbf{N}^\top \mathbf{L}), 0\}$ when $\mathbf{L}$ is unknown For Lambertian objects, $\mathbf{I} = \max\{\mathbf{D} \circ (\mathbf{N}^\top \mathbf{L}), 0\} = \mathbf{S}^\top \mathbf{L} = \tilde{\mathbf{S}}^\top \mathbf{A}^\top \mathbf{A}^{-1} \tilde{\mathbf{L}} = \hat{\mathbf{S}}^\top \mathbf{G}^\top \mathbf{G}^{-1} \hat{\mathbf{L}}$	
AM07	$\mathbf{D}$ has only a few different albedos, *i.e.*, the rows of $\mathbf{S}$ have only a few different lengths
SM10	Several surface points have equal albedo, *i.e.*, several rows of $\mathbf{S}$ having equal length are identified
PF14	Several points with locally maximum intensity on a Lambertian surface, *i.e.*, points with $\mathbf{n} = \mathbf{l}$ are identified
WT13	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \mathbf{D} + \boldsymbol{\rho}_s(\theta_h, \theta_d)$, *i.e.*, the specular reflection depends only on $\{\theta_h, \theta_d\}$

Photometric stereo taxonomy

- Lighting model
  - D Directional lighting
  - P Point lighting
  - G General (environment) lighting

Photometric stereo taxonomy

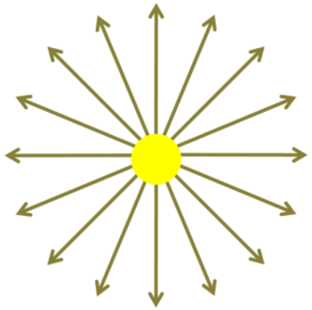
- Lighting model
  - D Directional lighting
  - P Point lighting
  - G General (environment) lighting



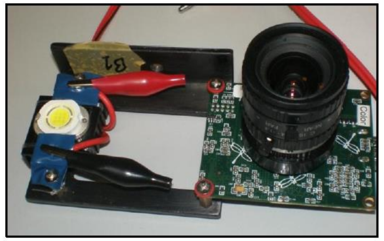
Simplest lighting model
Most works

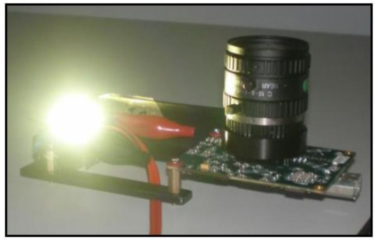
Photometric stereo taxonomy

- Lighting model
  - D Directional lighting
  - P Point lighting
  - G General (environment) lighting



More realistic
Spatially-varying direction
Intensity fall-off
[Iwahori 90]
[Clark 92]
[Higo 09]



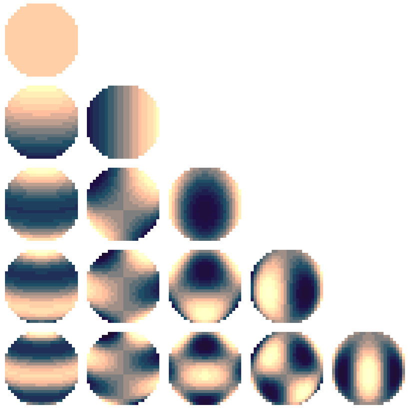


Photometric stereo taxonomy

- Lighting model
  - D Directional lighting
  - P Point lighting
  - G General (environment) lighting

Most general
Sum of directional lighting
[Yu 13]

Spherical harmonics model
[Basri 07]
[Shen 09]
[Shi 14]

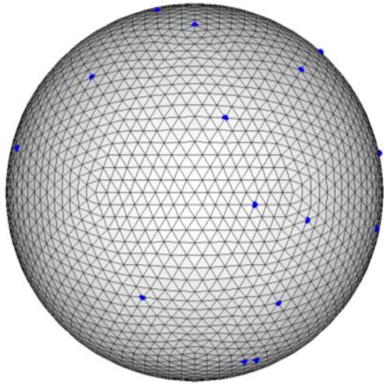


Photometric stereo taxonomy

- Number of images
  - **S** Small (10-20)
  - **M** Medium (50-100)
  - **L** Large (500+)

Photometric stereo taxonomy

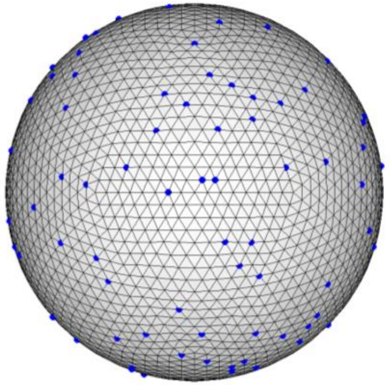
- Number of images
  - **S** Small (10-20)
  - **M** Medium (50-100)
  - **L** Large (500+)



Four/five lights for robustness
General lighting (1st/2nd order SH)
Fitting analytic BRDF
Near point lighting methods
Most uncalibrated methods

Photometric stereo taxonomy

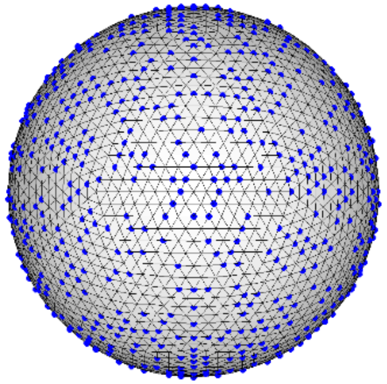
- Number of images
  - S Small (10-20)
  - M Medium (50-100)
  - L Large (500+)



Outlier rejection
Handing general isotropic BRDFs
Multi-view for Lambertian surface

Photometric stereo taxonomy

- Number of images
  - S Small (10-20)
  - M Medium (50-100)
  - L Large (500+)



Outdoor scenario
Manifold embedding
Handle anisotropic BRDF
Multi-view for non-Lambertian surfaces

Photometric stereo taxonomy

- Additional features
  - **PC** Perspective Camera
  - **NL** Non-Linear camera
  - **CL** Color Lighting
  - **DP** Depth Prior
  - **MV** Multi-View setup
  - **OM** Object Motion

Label the category of each work

- E.g., conventional photometric stereo [Woodham 80]
  - Lambertian, calibrated, directional lighting, a small number of (3) images

L C D S

References

* *The letters in brackets are category labels defined in Section 2.5.*

[1] R. J. Woodham. Photometric method for determining surface orientation from multiple images. *Optical Engineering* 19(1):139–144, 1980, [LCDS]. 1, 2, 5, 7

[2] D. B. Goldman, B. Curless, A. Hertzmann, and S. M. Seitz. Shape and spatially-varying BRDFs from photometric stereo. *IEEE TPAMI* 32(6):1060–1071, 2010, [ACDS]. 1, 2, 5, 7

[3] N. G. Alldrin, S. P. Mallick, and D. J. Kriegman. Resolving the generalized bas-relief ambiguity by entropy minimization. In *Proc. CVPR*, 2007, [LUDS]. 1, 4, 5, 8

[4] R. Basri, D. Jacobs, and I. Kemelmacher. Photometric stereo with general, unknown lighting. *IJCV* 72(3):239–257, 2007, [LUGS]. 1, 4, 5

[LUDS]. 4, 8

[19] Z. Wu and P. Tan. Calibrating photometric stereo by holistic reflectance symmetry analysis. In *Proc. CVPR*, 2013, [BUDS]. 3, 4, 5, 8

[20] F. Lu, Y. Matsushita, I. Sato, T. Okabe, and Y. Sato. Uncalibrated photometric stereo for unknown isotropic reflectances. In *Proc. CVPR*, 2013, [BUDM]. 1, 3, 4, 5, 8

[21] S. Herbort and C. Wöhler. An introduction to image-based 3D surface reconstruction and a survey of photometric stereo methods. *3D Research* 2(3):1–17, 2011. 1

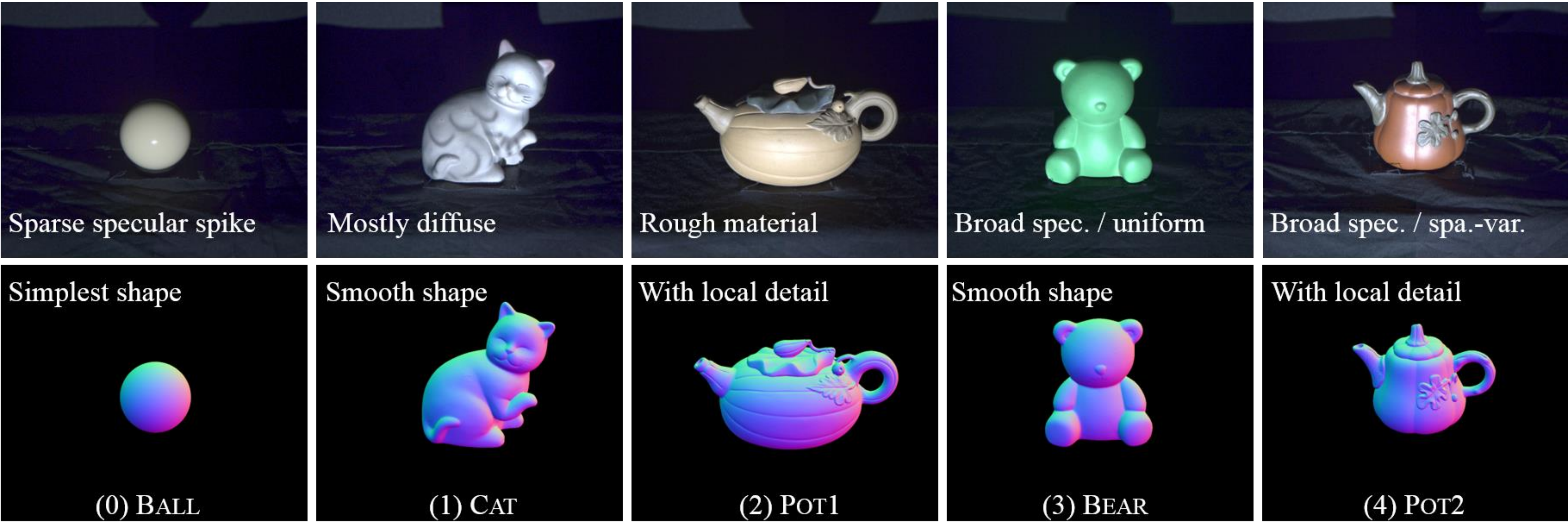
[22] J. Ackermann and M. Goesele. A survey of photometric stereo techniques. *Foundations and Trends in Computer Graphics and Vision* 9(3-4):149–254, 2015. 1

[23] E. N. Coleman and R. Jain. Obtaining 3-dimensional shape of textured and specular surfaces using four-source photometry. *CGIP* 18(4):309–328, 1982, [RCDS]. 2, 5

Benchmark Datasets and Evaluation

“DiLiGenT” photometric stereo datasets

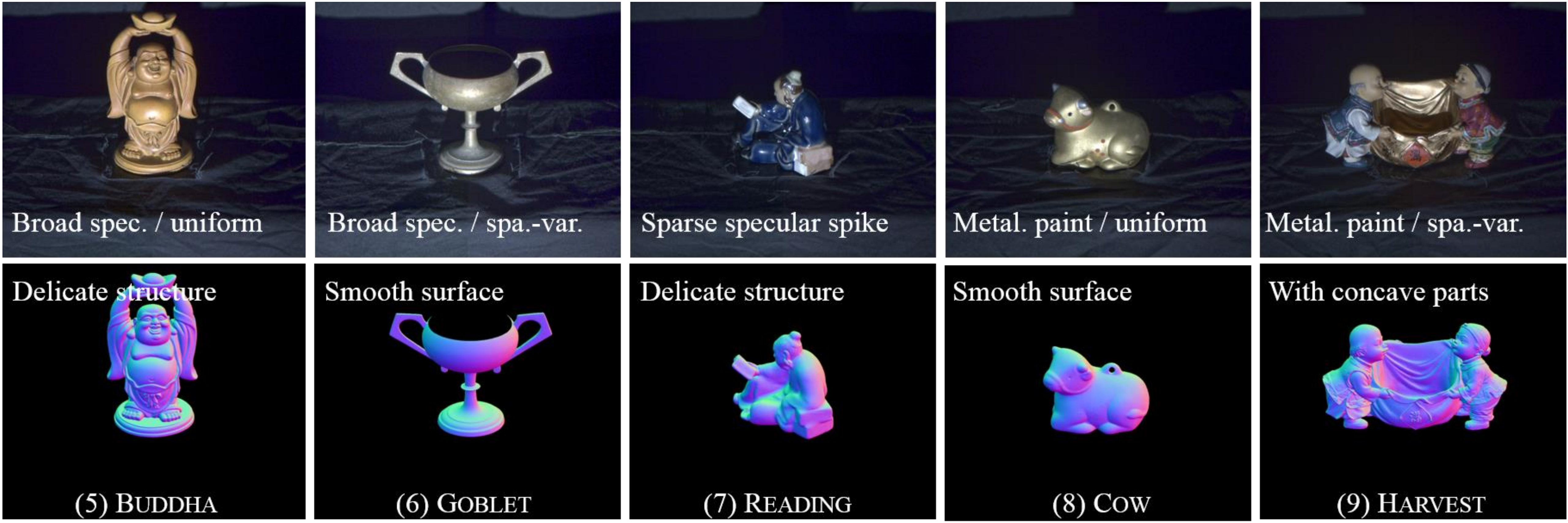
[Shi 16, 19] https://sites.google.com/site/photometricstereodata



Directional Lighting, General reflectance, with ground “Truth” shape

"DiLiGenT" photometric stereo datasets

[Shi 16, 19] https://sites.google.com/site/photometricstereodata



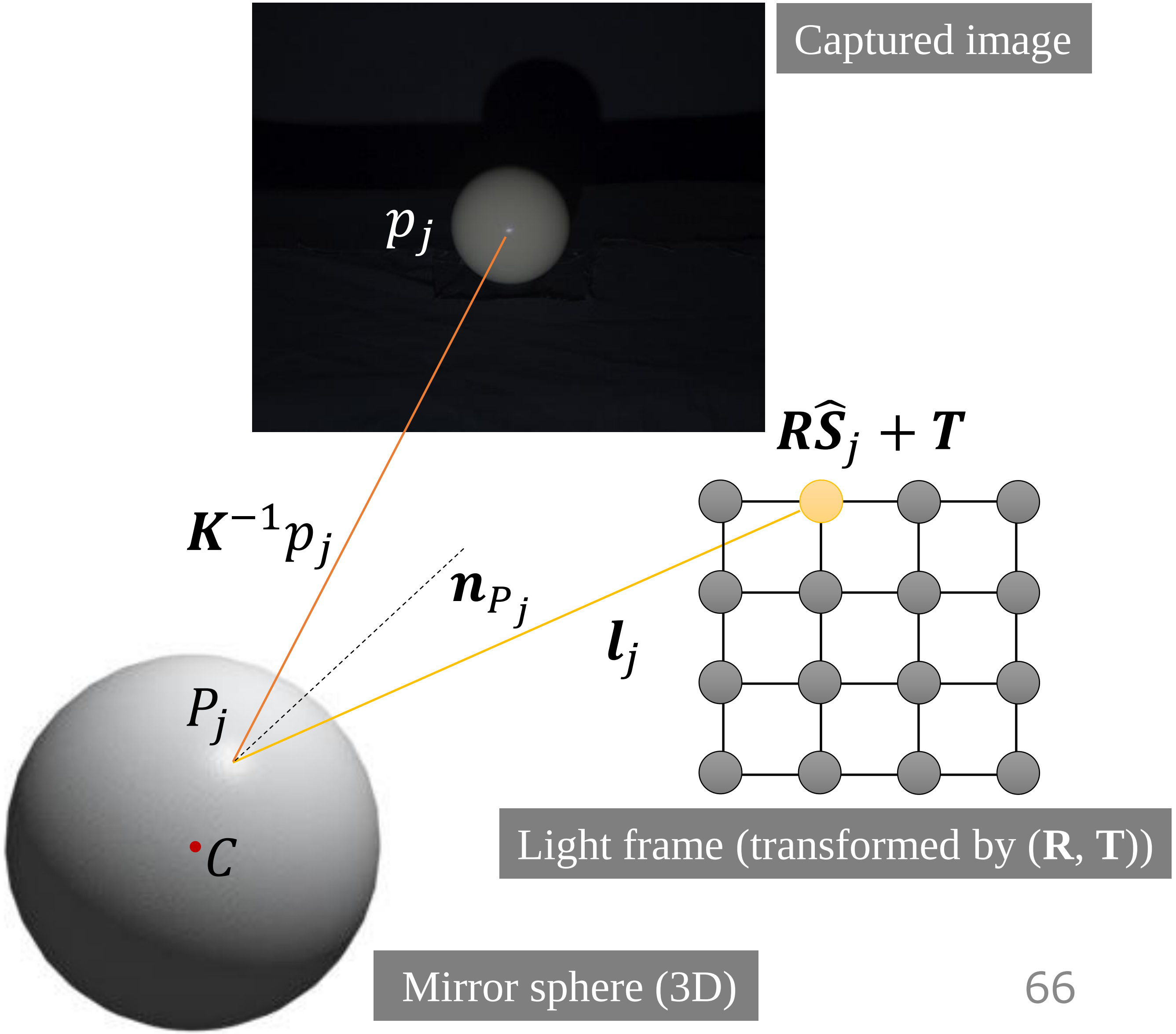
Directional Lighting, General reflectance, with ground "Truth" shape

Data capture

- Point Grey Grasshopper + 50*mm* lens
- Resolution: 2448 x 2048
- Object size: 20*cm*
- Object to camera distance: 1.5*m*
- 96 white LED in an 8 x 12 grid

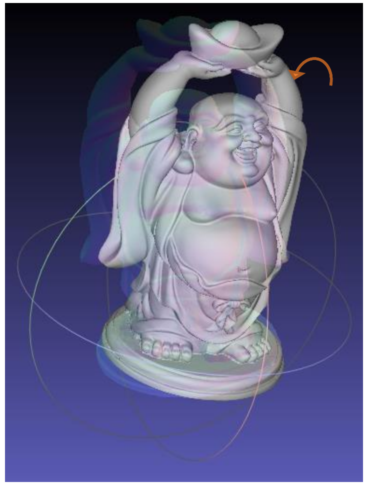
Lighting calibration

- Intensity
  - Macbeth white balance board
- Direction
  - From 3D positions of LED bulbs for higher accuracy



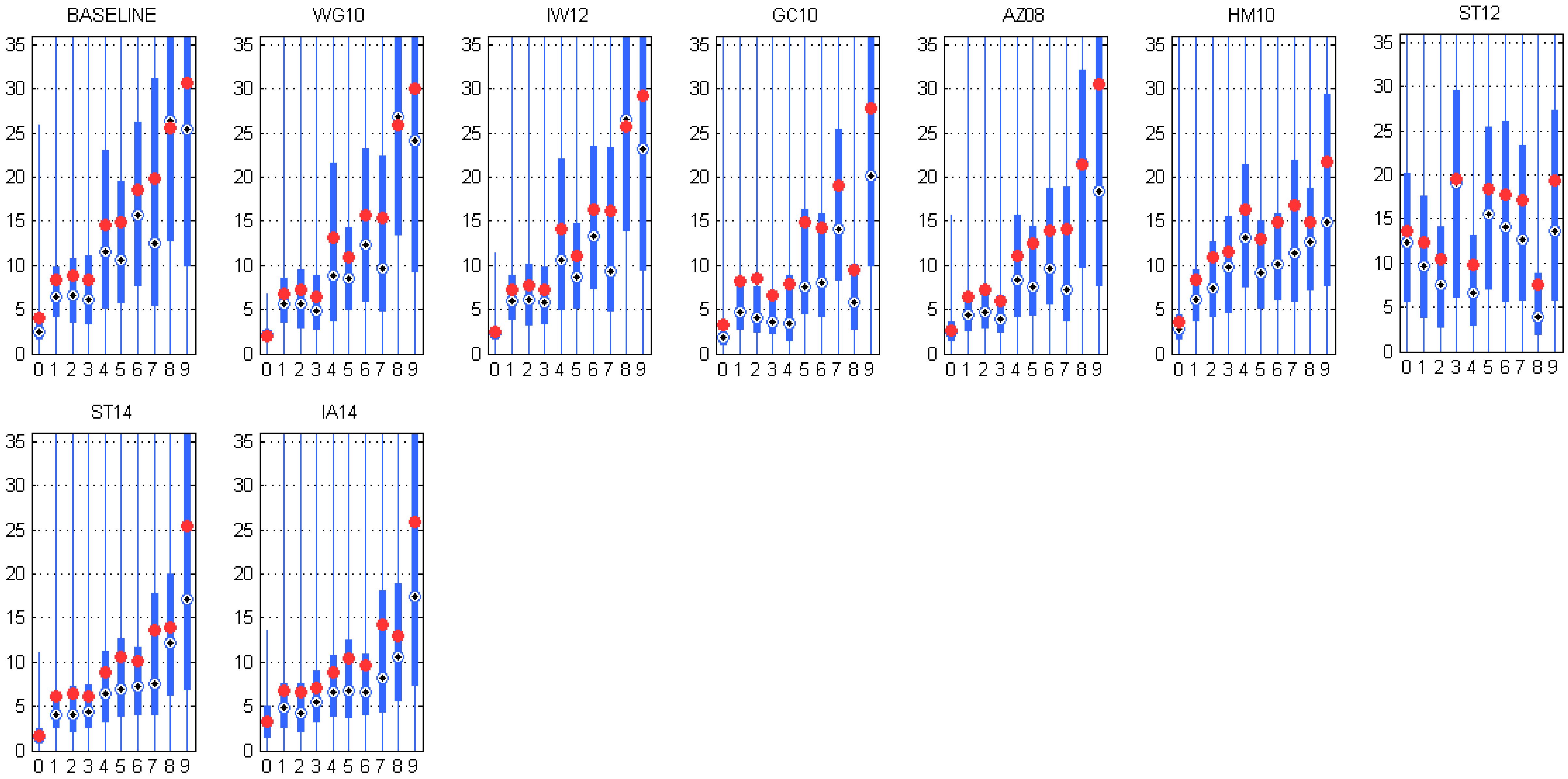
“Ground truth” shapes

- 3D shape
  - Scanner: Rexcan CS+ (res. 0.01*mm*)
  - Registration: EzScan 7
  - Hole filling: Autodesk Meshmixer 2.8
- Shape-image registration
  - Mutual information method [Corsini 09]
  - Meshlab + manual adjustment
- Evaluation criteria
  - Statistics of angular error (degree)
    - Mean, median, min, max, 1st quartile, 3rd quartile



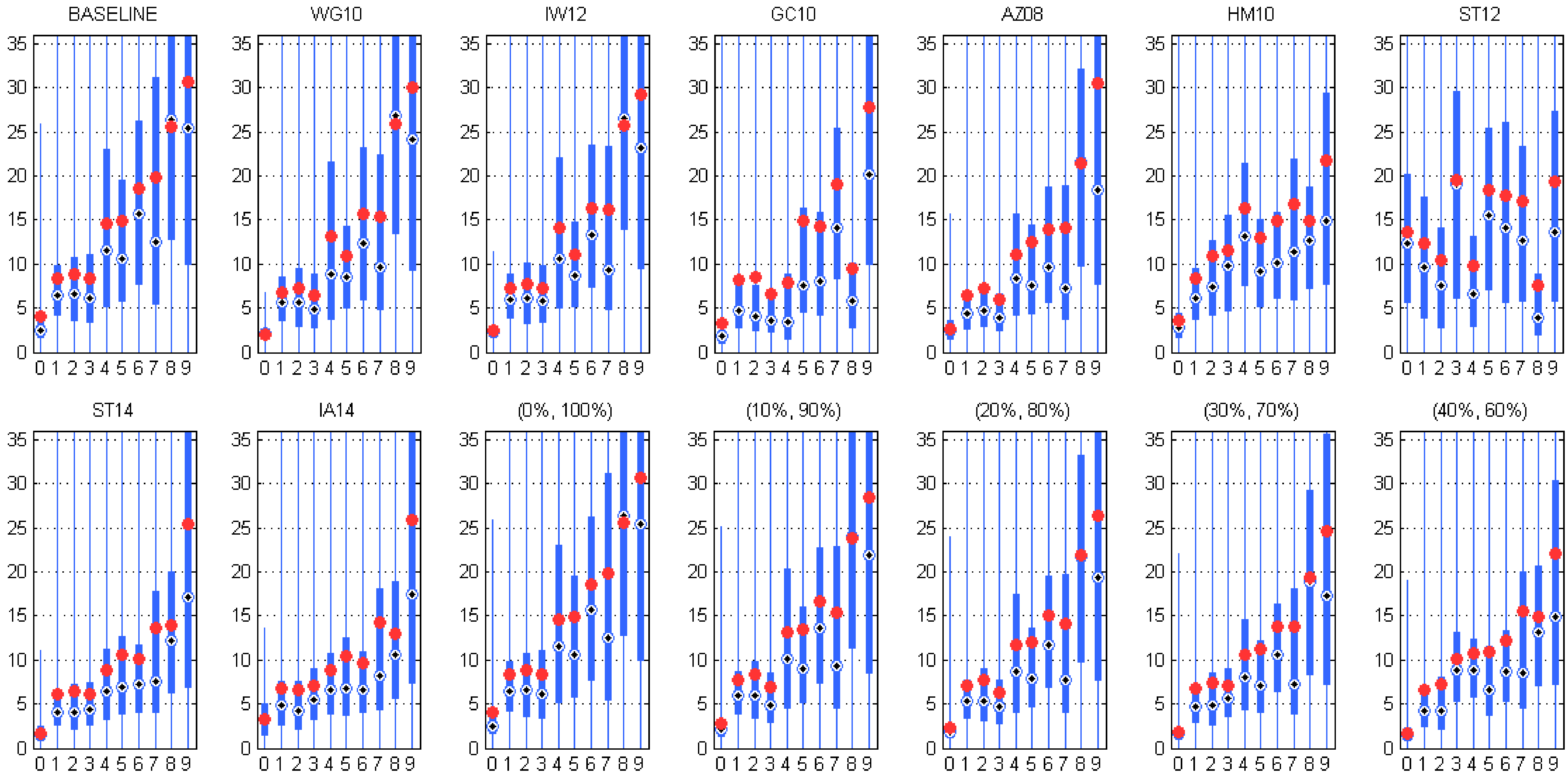
Evaluation for non-Lambertian methods

Solve $\mathbf{N}$ from $\mathbf{I} = \max\{\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \circ (\mathbf{N}^\top \mathbf{L}), 0\}$ by using different assumptions and constraints on $\boldsymbol{\rho}(\mathbf{n}, \mathbf{l})$ Notations: $\mathbf{h} = (\mathbf{l} + \mathbf{v}) / \|\mathbf{l} + \mathbf{v}\|$, $\theta_h = \langle \mathbf{n}, \mathbf{h} \rangle = \arccos(\mathbf{n}^\top \mathbf{h})$, $\theta_d = \langle \mathbf{l}, \mathbf{h} \rangle = \arccos(\mathbf{l}^\top \mathbf{h})$	
BASELINE	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \mathbf{D}$, where each row of $\mathbf{D}$ is a constant representing the albedo of a Lambertian surface
WG10	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \mathbf{D} + \mathbf{E}$, where $\mathbf{E}$ is sparse and $\text{rank}(\mathbf{I})$ is minimized
IW12	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \mathbf{D} + \mathbf{E}$, where $\mathbf{E}$ is sparse and $\text{rank}(\mathbf{I}) = 3$
GC10	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \sum_i \mathbf{w}_i \circ \boldsymbol{\rho}_i(d_i, s_i, \alpha_i)$, where $\boldsymbol{\rho}_i(d_i, s_i, \alpha_i) = \frac{d_i}{\pi} + \frac{s_i}{4\pi\alpha_i^2 \sqrt{(\mathbf{n}^\top \mathbf{l})(\mathbf{n}^\top \mathbf{v})}} \exp\left(\frac{(1 - 1/\mathbf{n}^\top \mathbf{h})}{\alpha_i^2}\right)$
AZ08	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l})$ is isotropic and depends only on (θ_h, θ_d)
ST12	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l})$ is isotropic, depends only on θ_h, and is monotonic about $\mathbf{n}^\top \mathbf{h}$
HM10	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l})$ is isotropic, monotonic about $\mathbf{n}^\top \mathbf{l}$, and $\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) = 0$ for $\mathbf{n}^\top \mathbf{l} \leq 0$
ST14	The low-frequency part of $\boldsymbol{\rho}(\mathbf{n}, \mathbf{l})$ is a bi-polynomial $A(\cos(\theta_h))B(\cos(\theta_d))$, where A and B are polynomials
IA14	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \sum_i \boldsymbol{\rho}_i(\mathbf{n}^\top \alpha_i)$, where $\alpha_i = (p_i \mathbf{l} + q_i \mathbf{v}) / \|p_i \mathbf{l} + q_i \mathbf{v}\|$, $p_i,\ q_i$ are nonnegative unknown values


BASELINE
WG10
IW12
GC10
AZ08
HM10
ST12
ST14
IA14
0 5 10 15 20 25 30 35
0 1 2 3 4 5 6 7 8 9

Evaluation for non-Lambertian methods

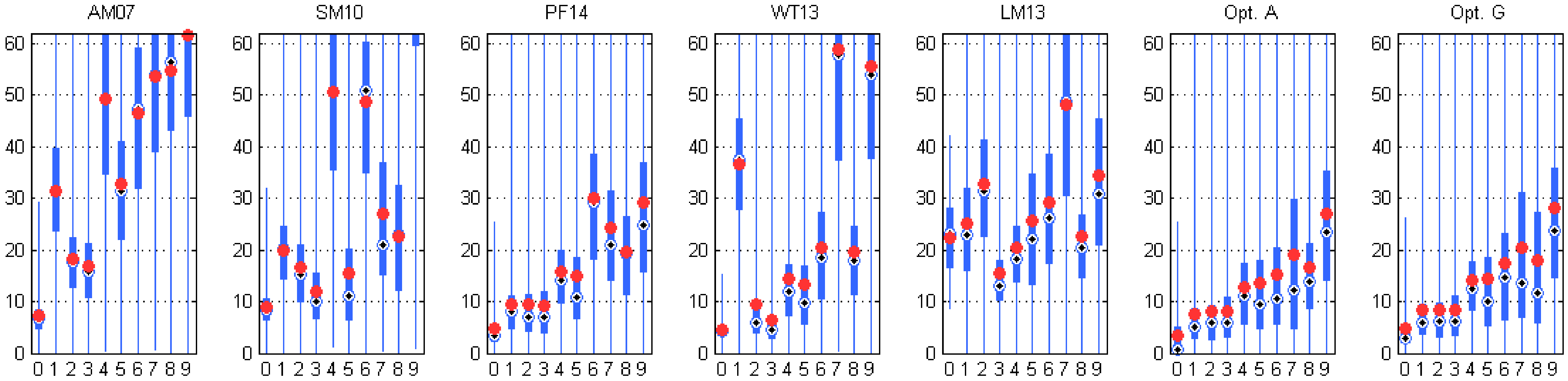
- Sort each intensity profile in ascending order
- Only use the data ranked between (T_{low}, T_{high})


BASELINE
WG10
IW12
GC10
AZ08
HM10
ST12
ST14
IA14
(0%, 100%)
(10%, 90%)
(20%, 80%)
(30%, 70%)
(40%, 60%)
0 5 10 15 20 25 30 35
0 1 2 3 4 5 6 7 8 9

Evaluation for uncalibrated methods

Solve $\mathbf{N}$ from $\mathbf{I} = \max\{\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \circ (\mathbf{N}^\top \mathbf{L}), 0\}$ when $\mathbf{L}$ is unknown For Lambertian objects, $\mathbf{I} = \max\{\mathbf{D} \circ (\mathbf{N}^\top \mathbf{L}), 0\} = \mathbf{S}^\top \mathbf{L} = \tilde{\mathbf{S}}^\top \mathbf{A}^\top \mathbf{A}^{-1} \tilde{\mathbf{L}} = \hat{\mathbf{S}}^\top \mathbf{G}^\top \mathbf{G}^{-1} \hat{\mathbf{L}}$	
AM07	$\mathbf{D}$ has only a few different albedos, *i.e.*, the rows of $\mathbf{S}$ have only a few different lengths
SM10	Several surface points have equal albedo, *i.e.*, several rows of $\mathbf{S}$ having equal length are identified
PF14	Several points with locally maximum intensity on a Lambertian surface, *i.e.*, points with $\mathbf{n} = \mathbf{l}$ are identified
WT13	$\boldsymbol{\rho}(\mathbf{n}, \mathbf{l}) \approx \mathbf{D} + \boldsymbol{\rho}_s(\theta_h, \theta_d)$, *i.e.*, the specular reflection depends only on $\{\theta_h, \theta_d\}$

Opt. A	Fitting an optimal linear transform after factorization (pseudo-normal with 3×3 ambiguity)
Opt. G	Fitting an optimal GBR transform after applying integrability constraint (pseudo-normal up to GBR)

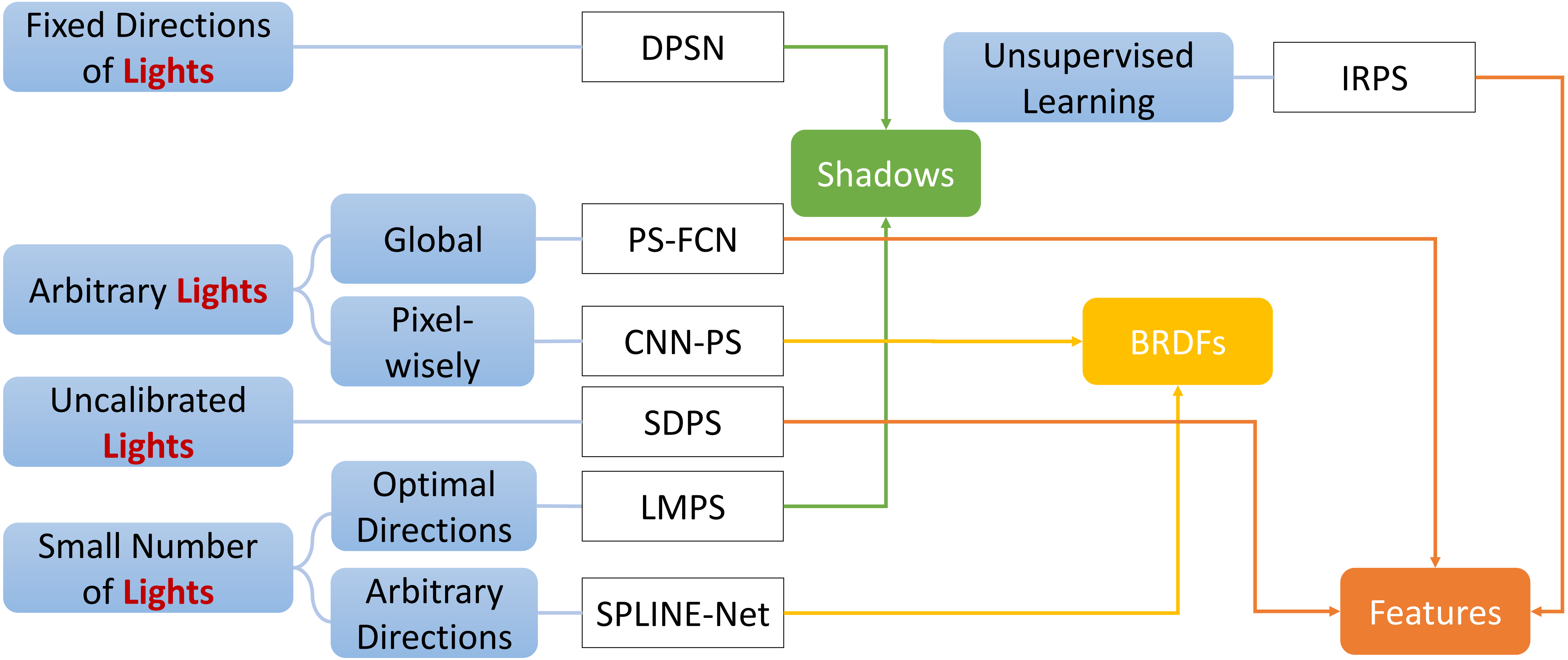

AM07
SM10
PF14
WT13
LM13
Opt. A
Opt. G
0
10
20
30
40
50
60
0 1 2 3 4 5 6 7 8 9

Photometric Stereo Meets Deep Learning

Photometric stereo + Deep learning

- [ICCV 17 Workshop]
  - Deep Photometric Stereo Network (DPSN)
- [ICML 18]
  - Neural Inverse Rendering for General Reflectance Photometric Stereo (IRPS)
- [ECCV 18]
  - PS-FCN: A Flexible Learning Framework for Photometric Stereo
- [ECCV 18]
  - CNN-PS: CNN-based Photometric Stereo for General Non-Convex Surfaces
- [CVPR 19]
  - Self-calibrating Deep Photometric Stereo Networks (SDPS)
- [CVPR 19]
  - Learning to Minify Photometric Stereo (LMPS)
- [ICCV 19]
  - SPLINE-Net: Sparse Photometric Stereo through Lighting Interpolation and Normal Estimation Networks

Photometric stereo + Deep learning



[ICCV 17 Workshop]
Deep Photometric Stereo Network

Deep Photometric Stereo Network

Hiroaki Santo[*,1], Masaki Samejima[†,1], Yusuke Sugano[‡,1], Boxin Shi[§,2], and Yasuyuki Matsushita[¶,1]

[1]Graduate School of Information Science and Technology, Osaka University
[2]Artificial Intelligence Research Center, National Institute of AIST

Abstract

This paper presents a photometric stereo method based on deep learning. One of the major difficulties in photometric stereo is designing can appropriate reflectance model that is both capable of representing real-world reflectances and computationally tractable in terms of deriving surface normal. Unlike previous photometric stereo methods that rely on a simplified parametric image formation model, such as the Lambert's model, the proposed method

it is known difficult to directly work with general non-parametric BRDFs in the context of photometric stereo. To ease the problem, there have been studies to use parametric representations to approximate BRDFs. However, so far, known parametric models have been only accurate for a limited class of materials, and the solution methods suffer from unstable optimization, which prohibits obtaining accurate estimates. Thus, it is needed to develop a photometric stereo method that is both computationally tractable and capable of handling diverse BRDFs.

Research background

- f : reflectance model
- $\boldsymbol{m}$: measurement vector
- $\boldsymbol{L}$: light source direction
- $\boldsymbol{n}$: normal vector

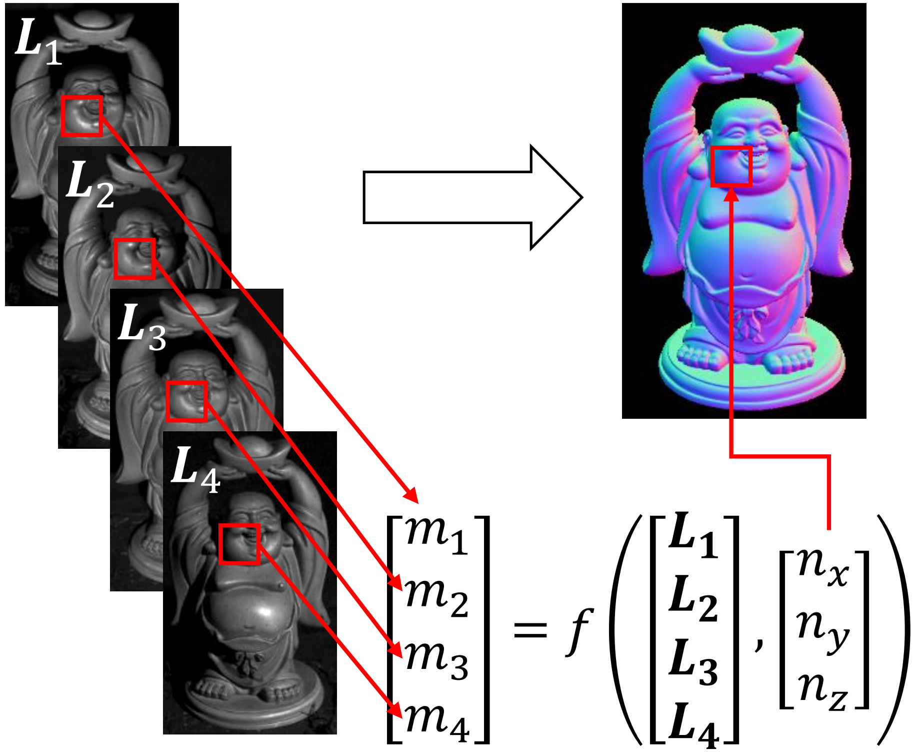
Photometric Stereo

Measurements

Normal map

Image formation

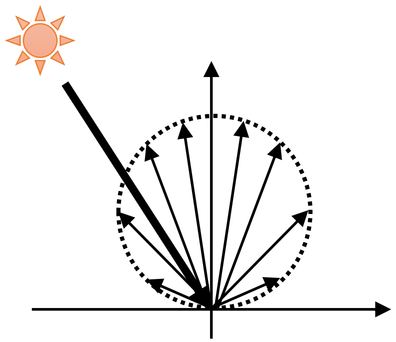
$$\boldsymbol{m} = f(\boldsymbol{L}, \boldsymbol{n})$$



$$\begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ m_4 \end{bmatrix} = f\left(\begin{bmatrix} \boldsymbol{L_1} \\ \boldsymbol{L_2} \\ \boldsymbol{L_3} \\ \boldsymbol{L_4} \end{bmatrix}, \begin{bmatrix} n_x \\ n_y \\ n_z \end{bmatrix} \right)$$

Motivations

Parametric reflectance model



Lambertian model
(Ideal diffuse reflection)

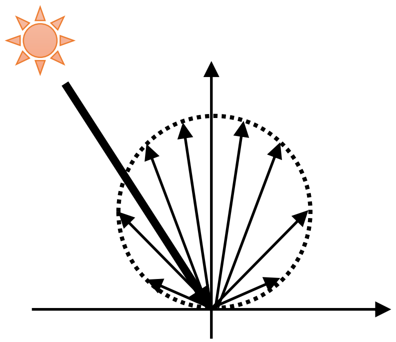
only accurate for a limited class of materials

Metal

rough surface

Motivations

Parametric reflectance model



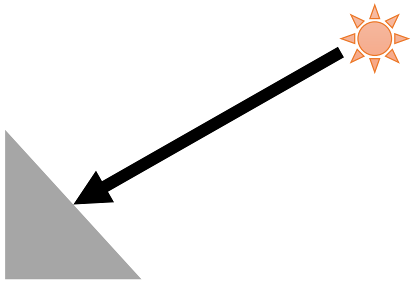
Lambertian model
(Ideal diffuse reflection)

only accurate for a limited class of materials

Metal

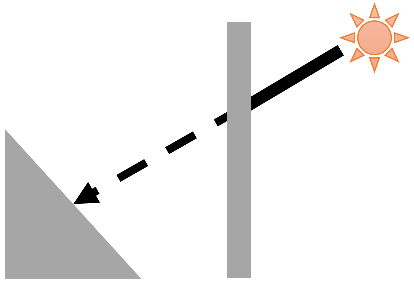
rough surface

Local illumination model



Model direct illumination only

Global illumination effects cannot be modeled



Cast shadow

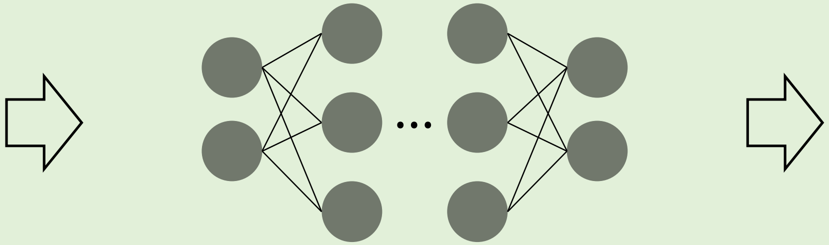
Motivations

Para

- Model the mapping from measurements to surface normal directly using Deep Neural Network (DNN)
- DNN can express more flexible reflection phenomenon compared to existing models designed based on physical phenomenon

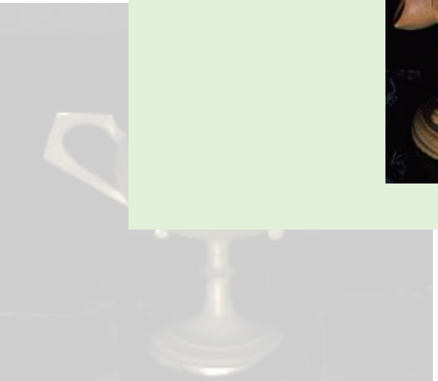
Measurements

Deep Neural Network



Normal map

only acc



Metal

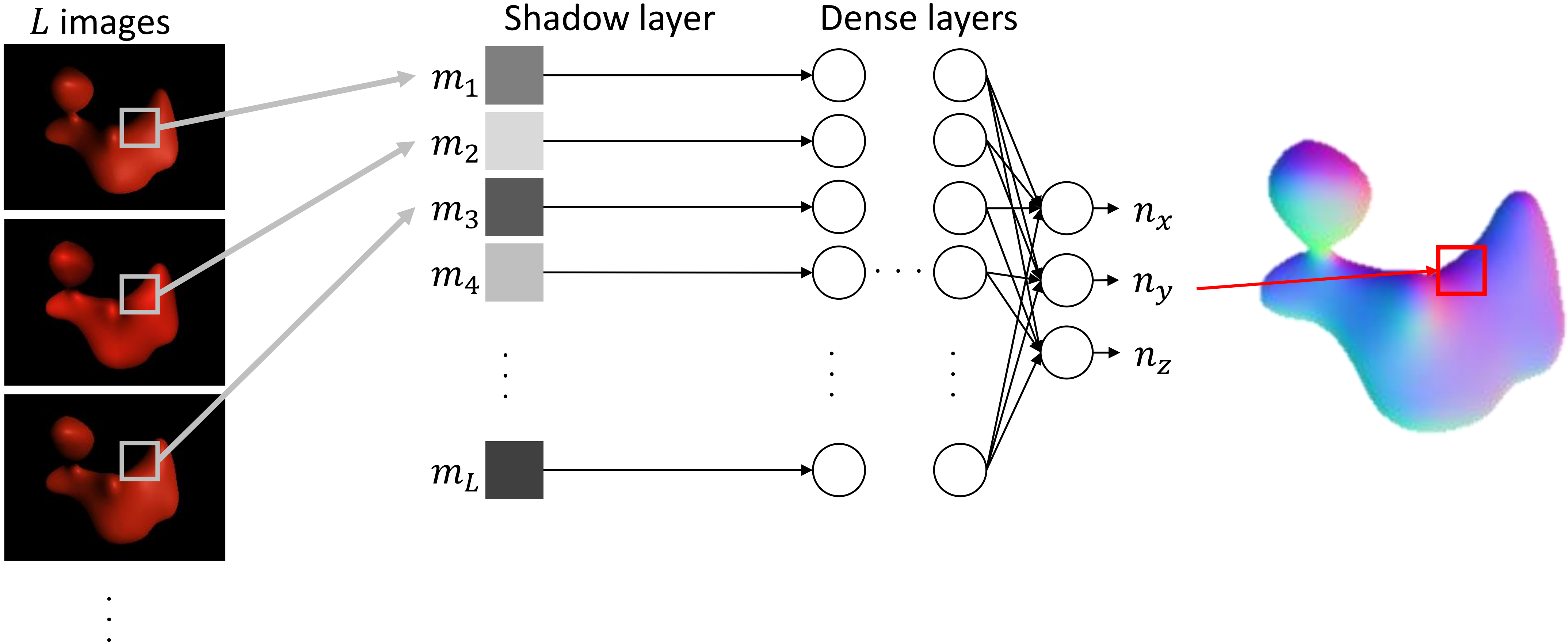
rough surface

Global illumination effects cannot be modeled

Proposed method

<u>Reflectance model with Deep Neural Network</u>

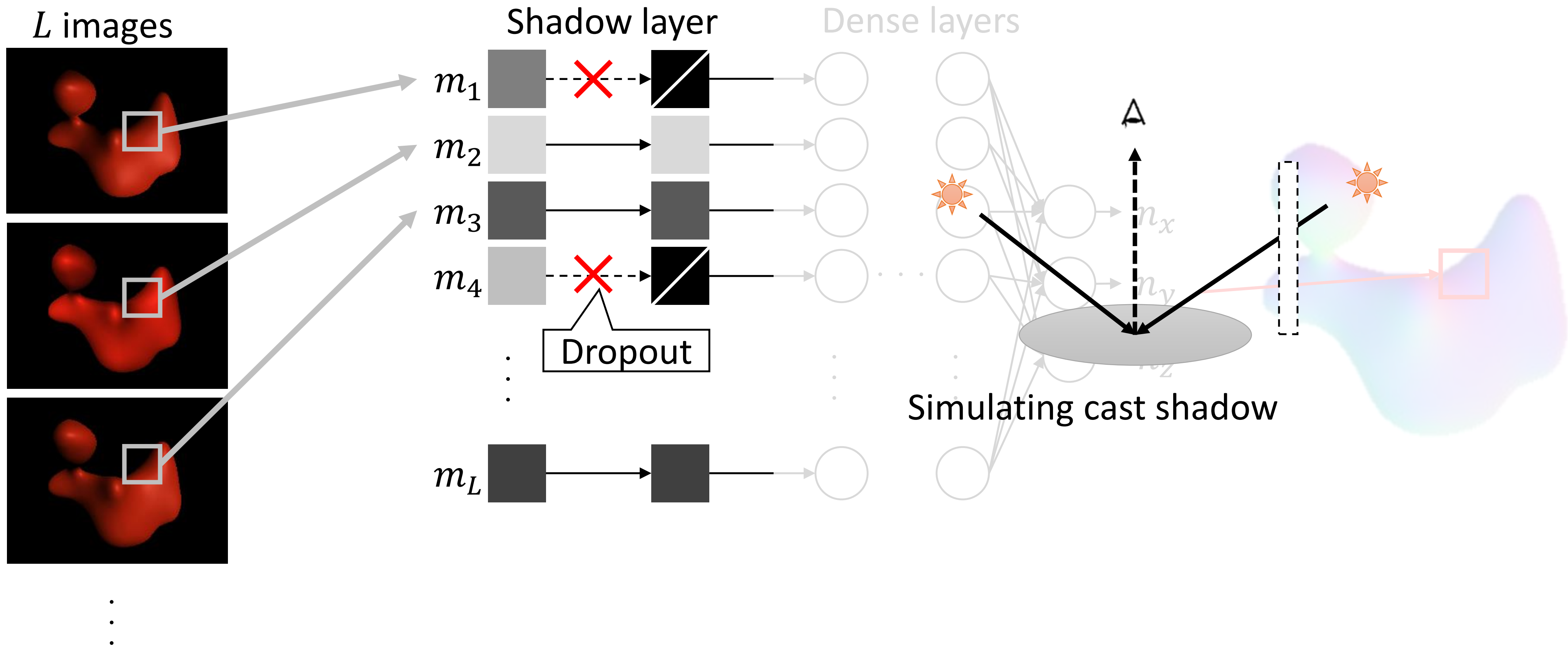
- mappings from measurement ($\boldsymbol{m} = [m_1, m_2, \ldots, m_L]^{\mathrm{T}}$) to surface normal ($\boldsymbol{n} = [n_x, n_y, n_z]^{\mathrm{T}}$)



Proposed method

Reflectance model with Deep Neural Network

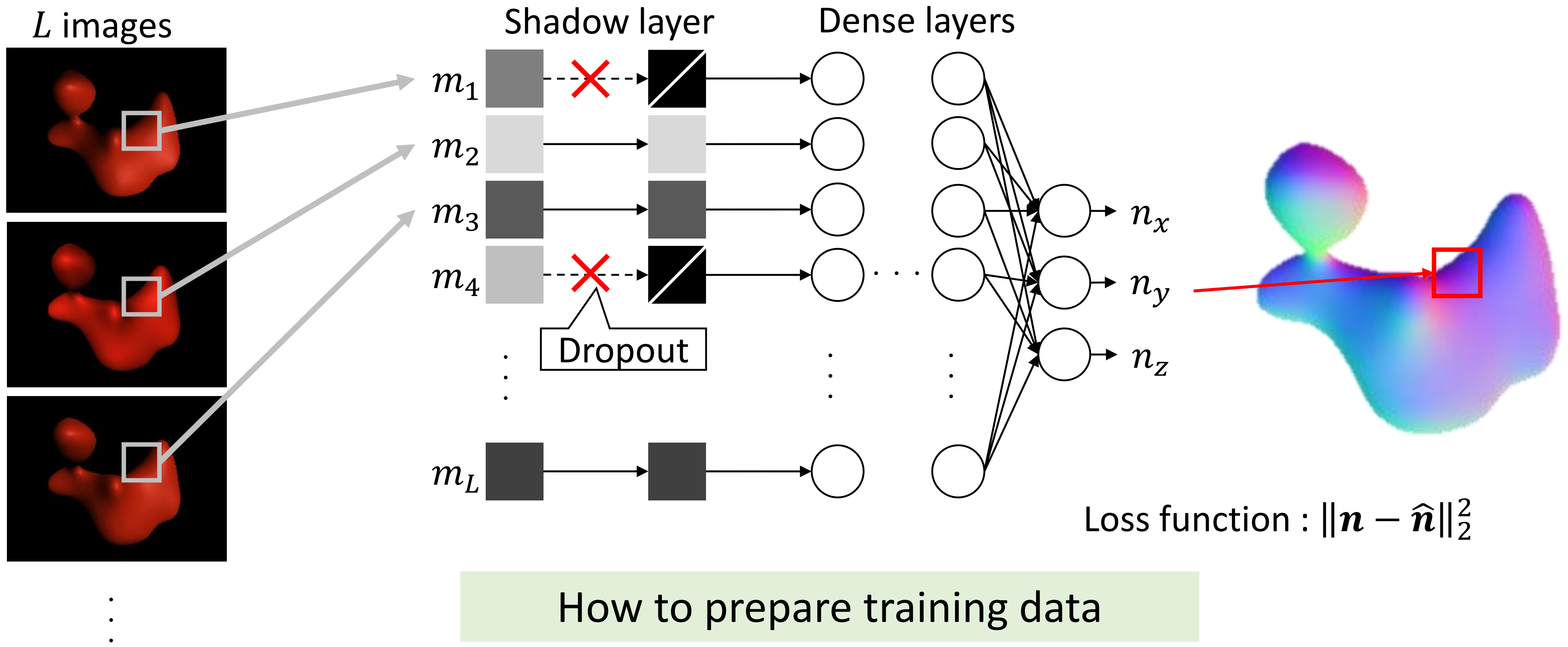
- mappings from measurement ($\boldsymbol{m} = [m_1, m_2, \dots, m_L]^{\mathrm{T}}$) to surface normal ($\boldsymbol{n} = [n_x, n_y, n_z]^{\mathrm{T}}$)



Proposed method

Reflectance model with Deep Neural Network

- mappings from measurement ($\boldsymbol{m} = [m_1, m_2, \dots, m_L]^{\mathrm{T}}$) to surface normal ($\boldsymbol{n} = [n_x, n_y, n_z]^{\mathrm{T}}$)



How to prepare training data

Training data

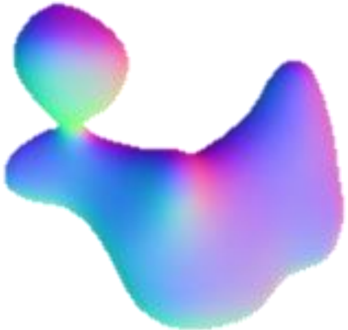
Rendering synthetic images

- Rendering with database (MERL BRDF database), which stores reflectance functions of 100 different real-world materials [Matusik 03]

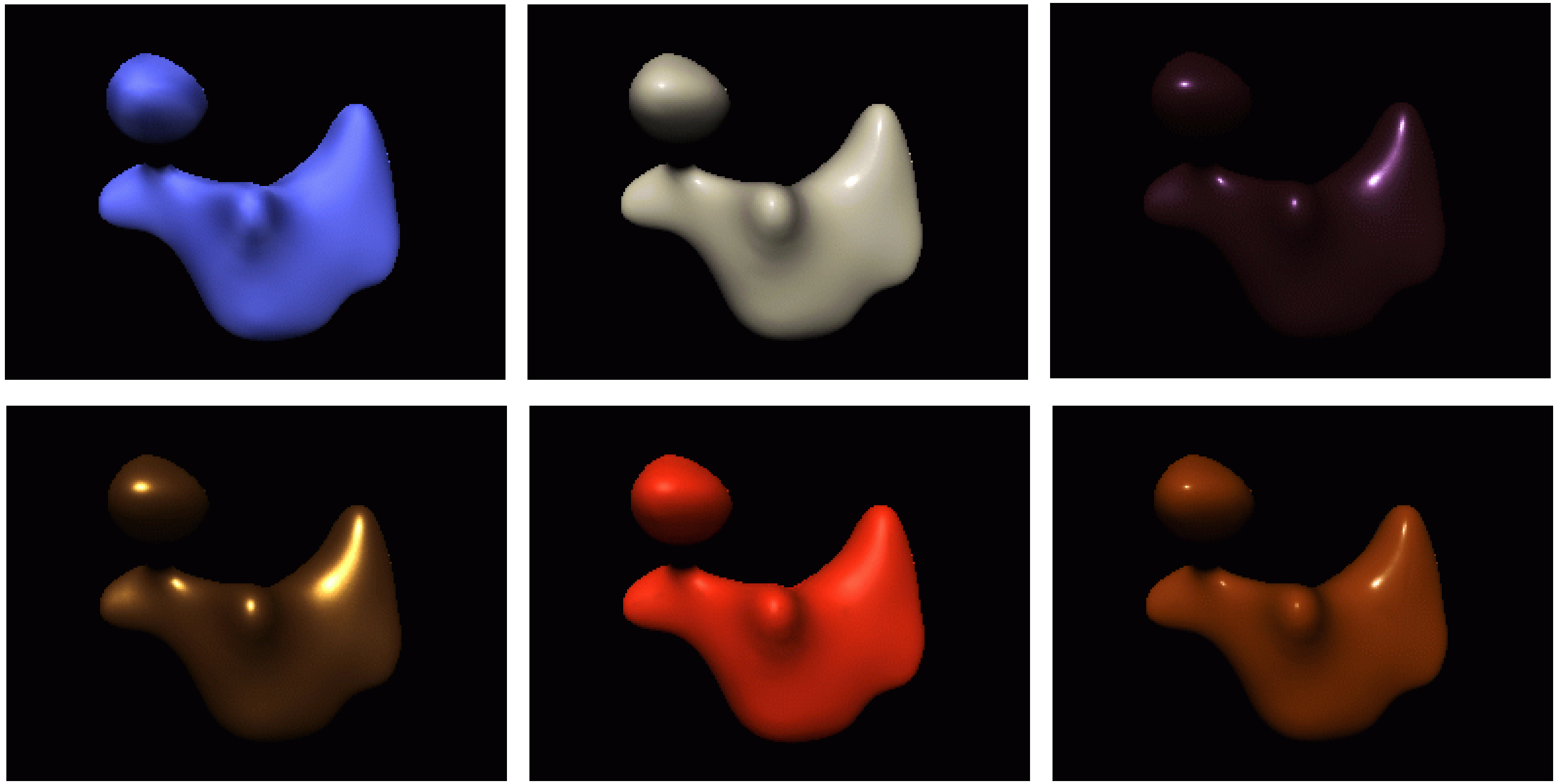
Training data

Rendering synthetic images

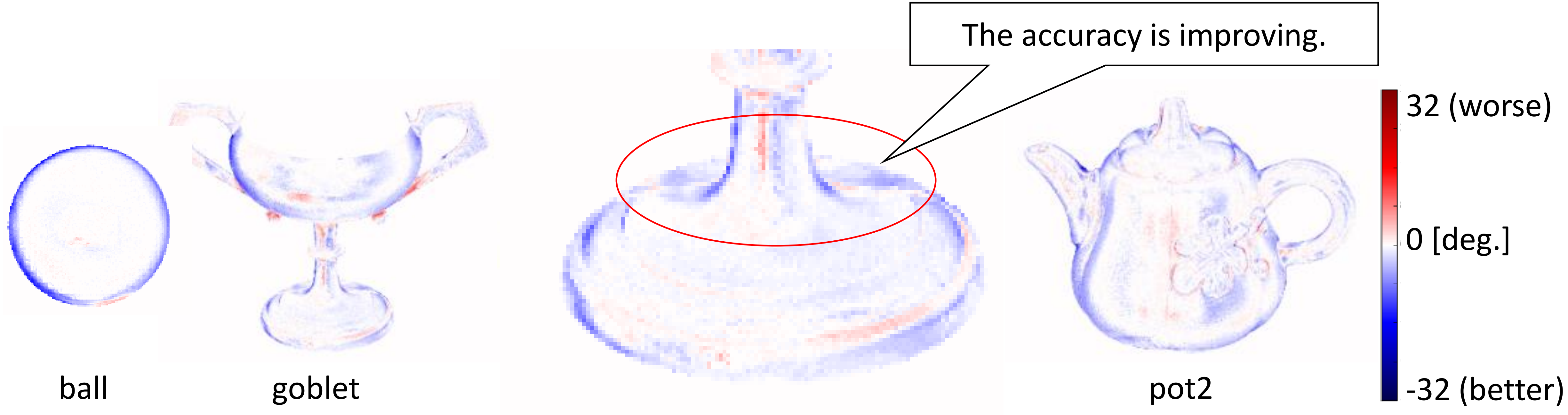
- Rendering with database (MERL BRDF database), which stores reflectance functions of 100 different real-world materials [Matusik'03].



Given normal map



Effectiveness of the shadow layer



The difference map of error map between "Proposed" and "Proposed W/ SL"

Blue pixels ：The estimation accuracy is improved by shadow layer

Red pixels ：The estimation accuracy is NOT improved by shadow layer

Benchmark results using “DiLiGenT”

	ball	cat	pot1	bear	buddha	cow	goblet	harvest	pot2	reading	AVG.
Proposed	3.44	7.21	7.90	7.20	13.30	8.49	12.35	**16.81**	8.80	17.47	10.30
Proposed W/ SL	2.02	6.54	7.05	6.31	12.68	8.01	11.28	16.86	**7.86**	15.51	**9.41**
ST14 (Shi+, PAMI, 2014)	**1.74**	**6.12**	**6.51**	6.12	10.60	13.93	10.09	25.44	8.78	**13.63**	10.30
IA14 (Ikehata+, CVPR, 2014)	3.34	6.74	6.64	7.11	**10.47**	13.05	**9.71**	25.95	8.77	14.19	10.60
WG10 (Wu+, ACCV, 2010)	2.06	6.73	7.18	6.50	10.91	25.89	15.70	30.01	13.12	15.39	13.35
AZ08 (Alldrin+, CVPR, 2008)	2.71	6.53	7.23	**5.96**	12.54	21.48	13.93	30.50	11.03	14.17	12.61
HM10 (Higo+, CVPR, 2010)	3.55	8.40	10.85	11.48	13.05	14.95	14.89	21.79	16.37	16.82	13.22
IW12 (Ikehata+, CVPR, 2012)	2.54	7.21	7.74	7.32	11.11	25.70	16.25	29.26	14.09	16.17	13.74
ST12 (Shi+, ECCV, 2012)	13.58	12.34	10.37	19.44	18.37	**7.62**	17.80	19.30	9.84	17.17	14.58
GC10 (Goldman+, PAMI, 2010)	3.21	8.22	8.53	6.62	14.85	9.55	14.22	27.84	7.90	19.07	12.00
BASELINE (L2)	4.10	8.41	8.89	8.39	14.92	25.60	18.50	30.62	14.65	19.80	15.39

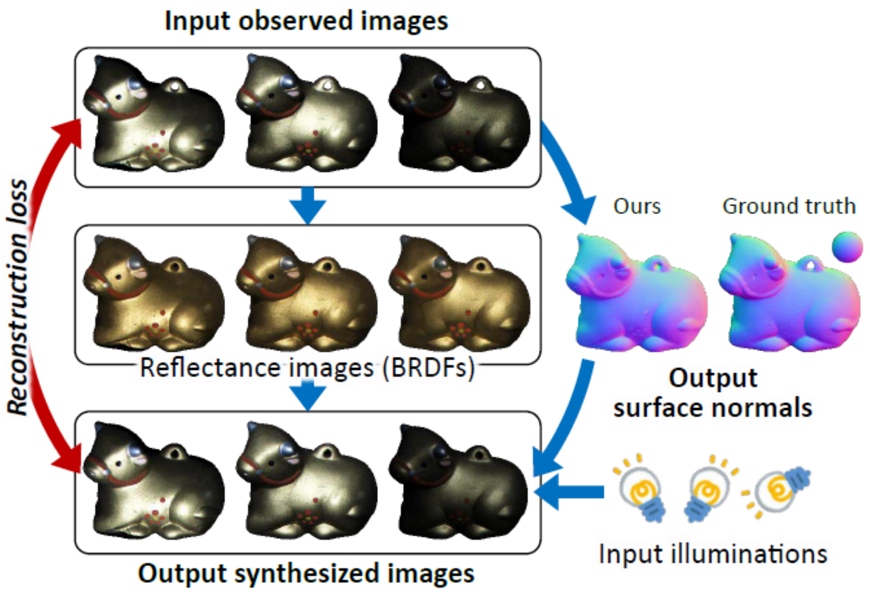
[ICML 18]
Neural Inverse Rendering for General Reflectance Photometric Stereo

Neural Inverse Rendering for General Reflectance Photometric Stereo

Tatsunori Taniai [1] Takanori Maehara [1]

Abstract

We present a novel convolutional neural network architecture for photometric stereo (Woodham, 1980), a problem of recovering 3D object surface normals from multiple images observed under varying illuminations. Despite its long history in computer vision, the problem still shows fundamental challenges for surfaces with unknown general reflectance properties (BRDFs). Leveraging deep neural networks to learn complicated reflectance models is promising, but studies in this direction are very limited due to difficulties in acquiring accurate ground truth for training and also

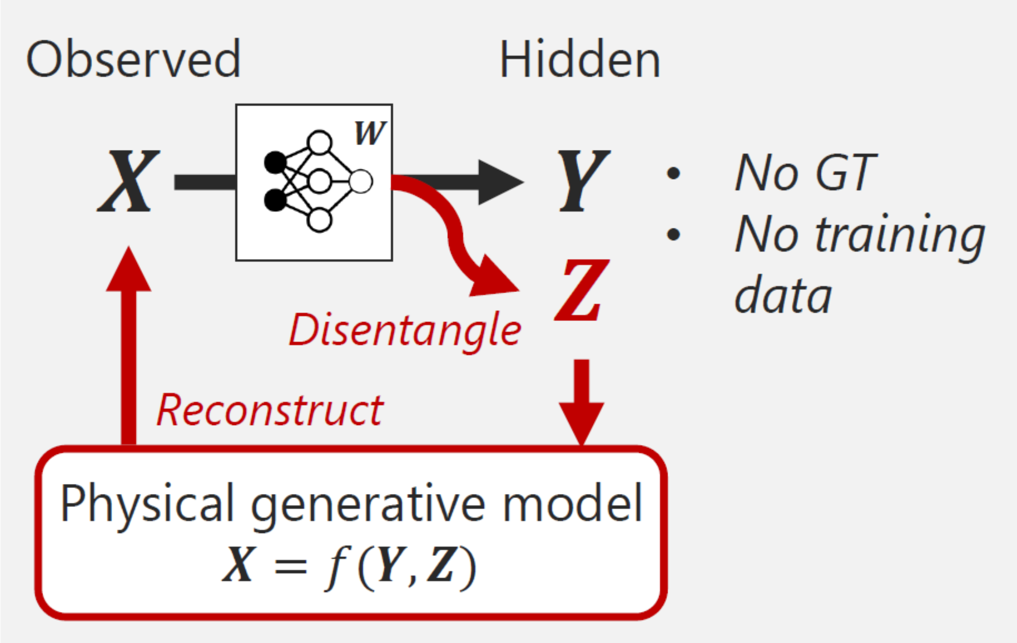


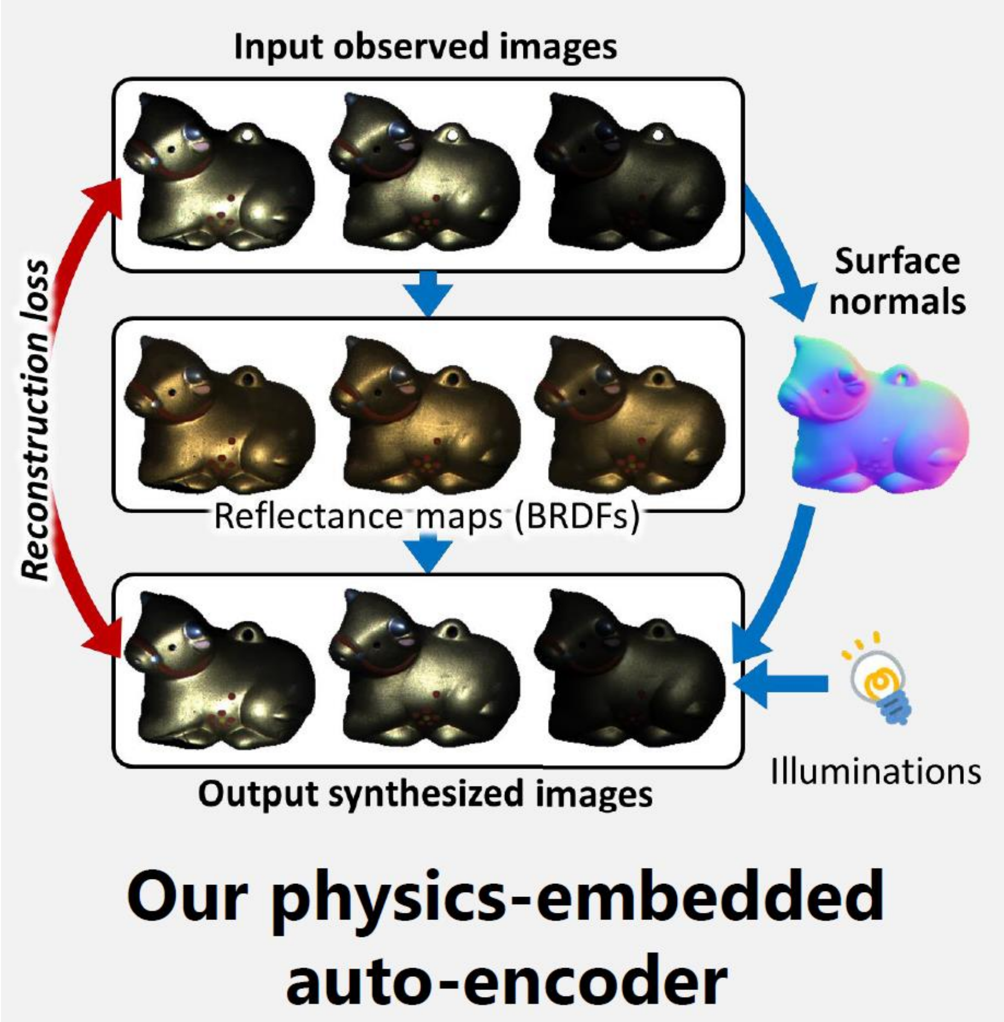
Challenges

- **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 normal and BRDFs is promising, but accurately measuring ground truth of surface normal and BRDFs is difficult
- **Permutation invariance**: Permuting input images should not change the resulting surface normals

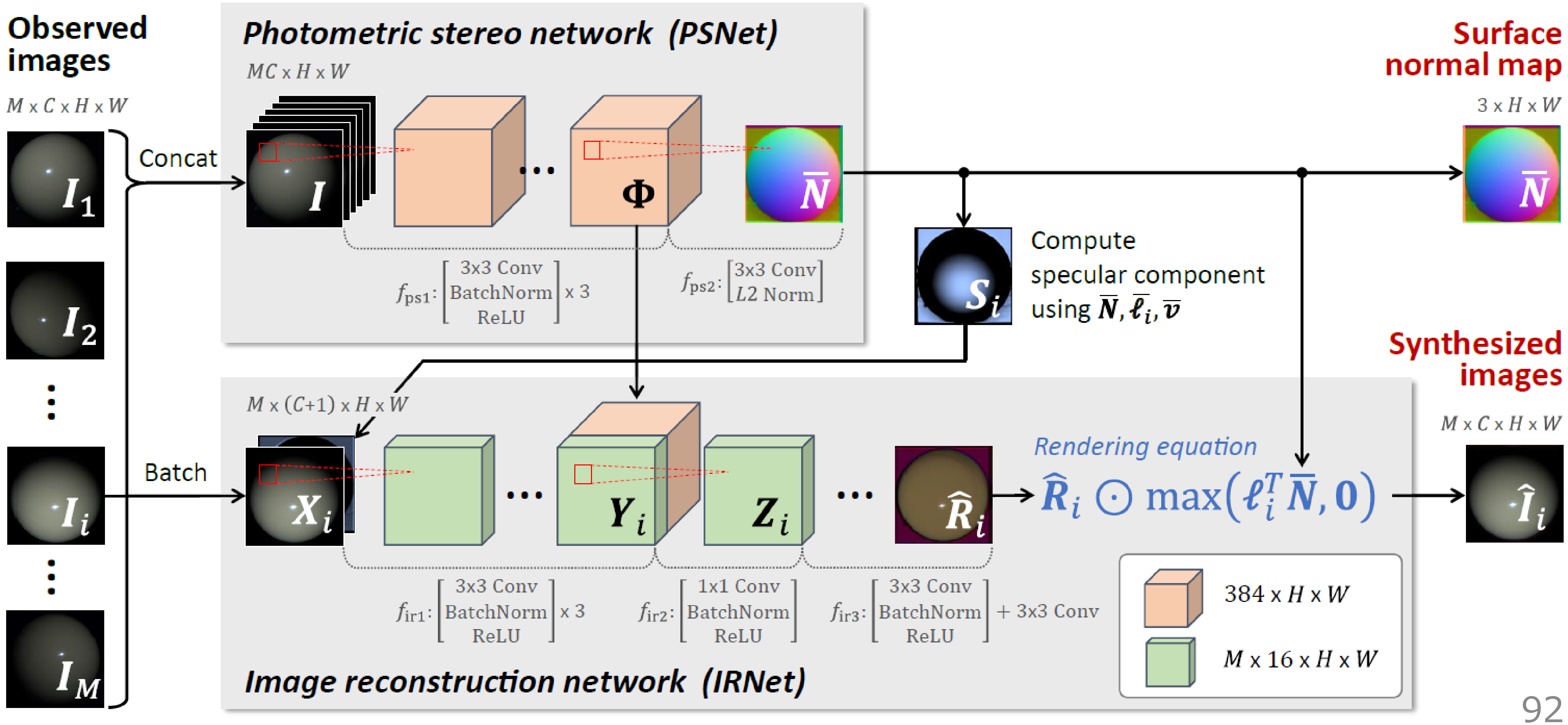
Key ideas

- Inverse rendering
- Reconstruction loss
- Unsupervised

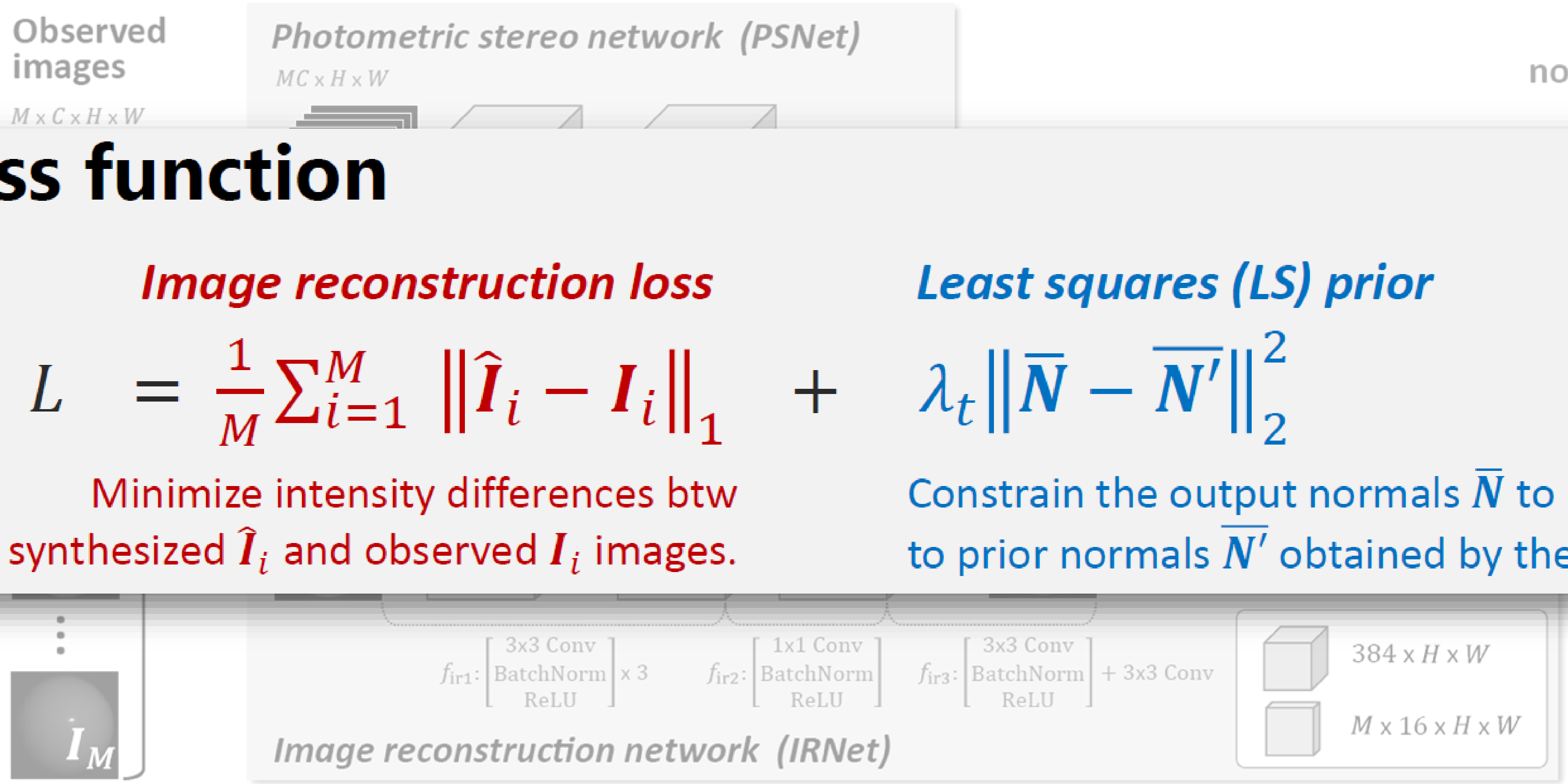




Network architecture



Network architecture



Loss function

Image reconstruction loss — ***Least squares (LS) prior***

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

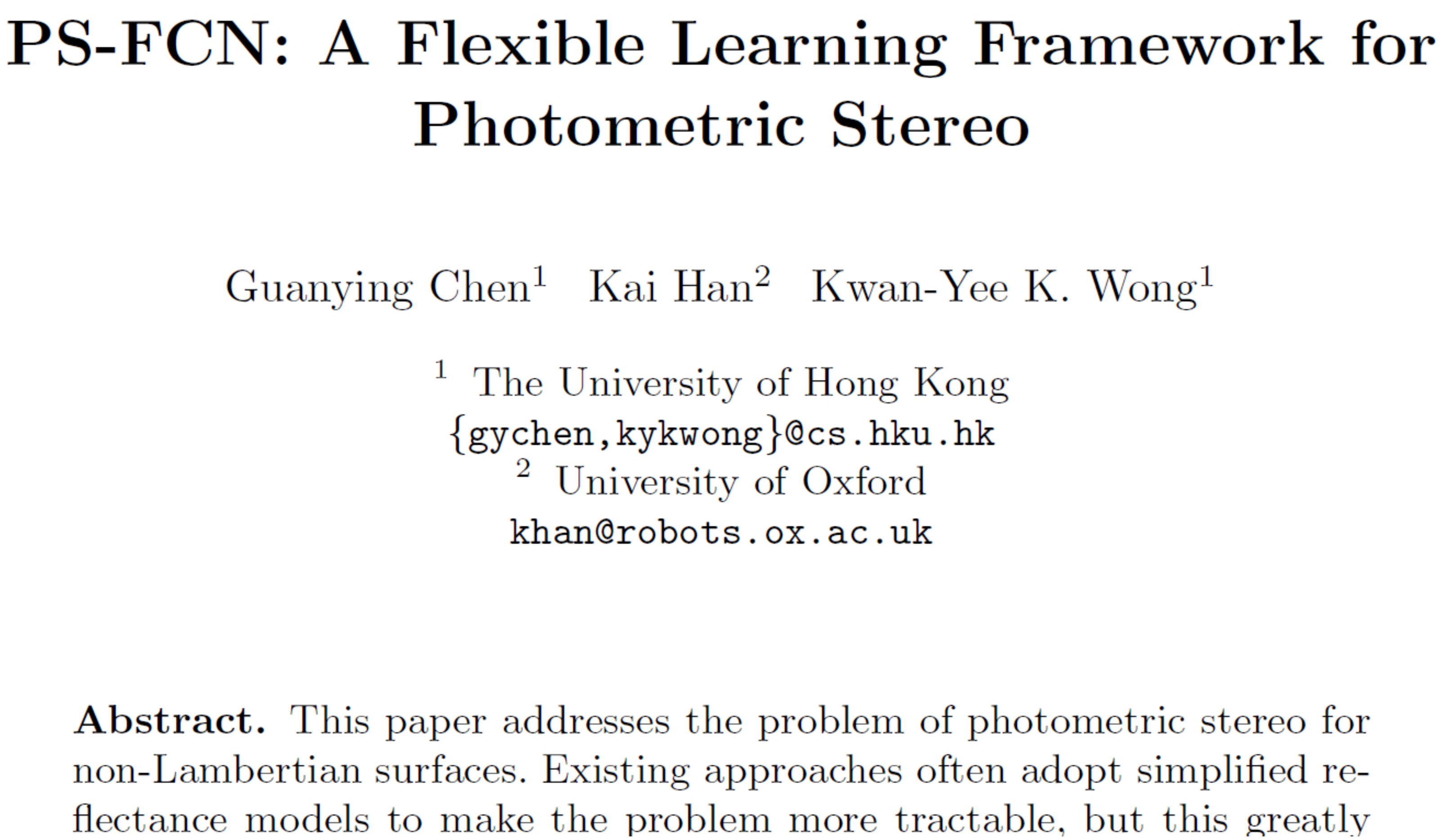
Minimize intensity differences btw synthesized $\hat{\boldsymbol{I}}_i$ and observed $\boldsymbol{I}_i$ images.

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

Benchmark results using "DiLiGenT"

	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

[ECCV 18]
PS-FCN: A Flexible Learning Framework for Photometric Stereo



PS-FCN: A Flexible Learning Framework for Photometric Stereo

Guanying Chen[1] Kai Han[2] Kwan-Yee K. Wong[1]

[1] The University of Hong Kong
{gychen,kykwong}@cs.hku.hk
[2] University of Oxford
khan@robots.ox.ac.uk

Abstract. This paper addresses the problem of photometric stereo for non-Lambertian surfaces. Existing approaches often adopt simplified reflectance models to make the problem more tractable, but this greatly

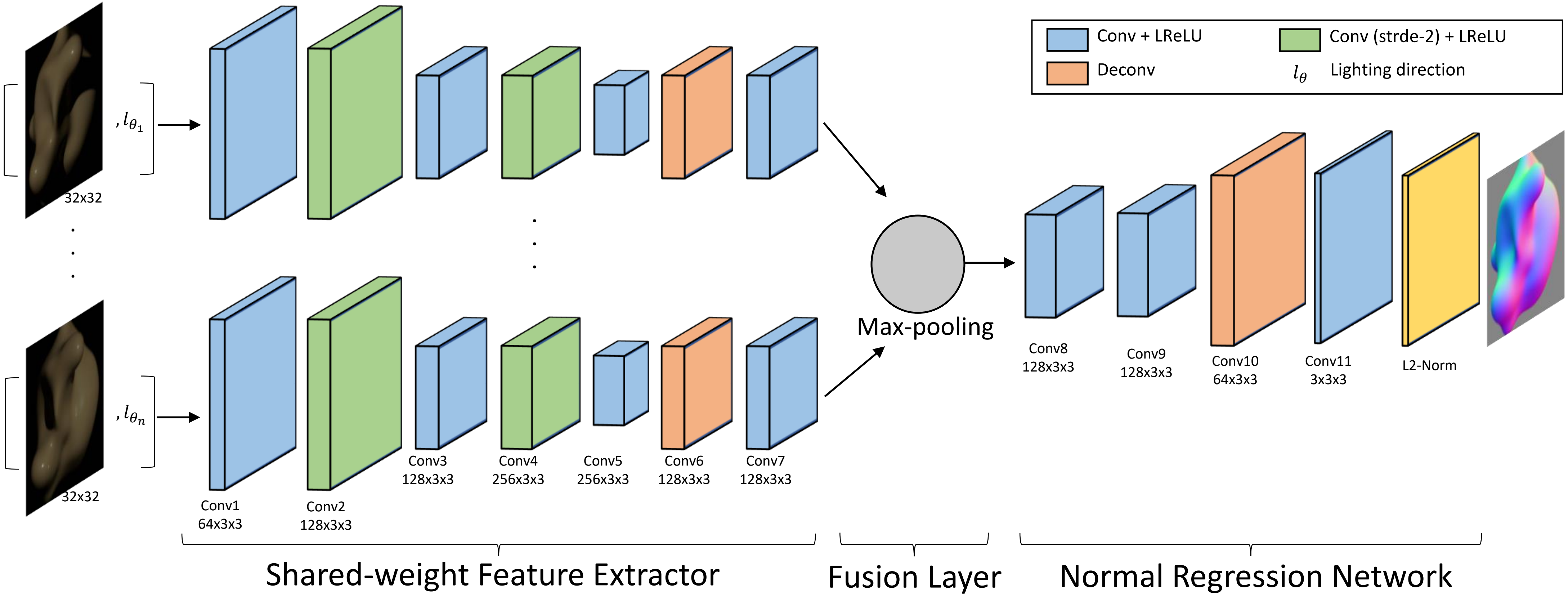
Overview of PS-FCN

Given an arbitrary number of images and their associated light directions as input, PS-FCN estimates a normal map of the object in a fast feed-forward pass.

Advantages:

- Does not depend on a pre-defined set of light directions
- Can handle input images in an order-agnostic manner

Network architecture



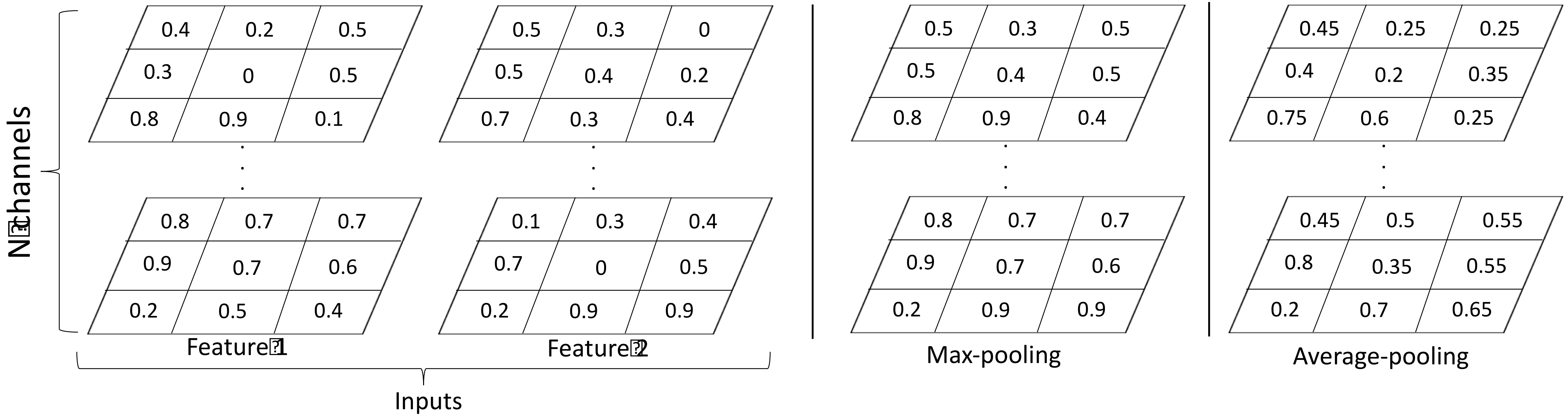
PS-FCN consists of three components:

- A Shared-weight Feature Extractor
- A Fusion Layer
- A Normal Regression Network

Loss function:

$$L_{normal} = \frac{1}{HW} \sum_{i,j} (1 - N_{ij} \cdot \tilde{N}_{ij})$$

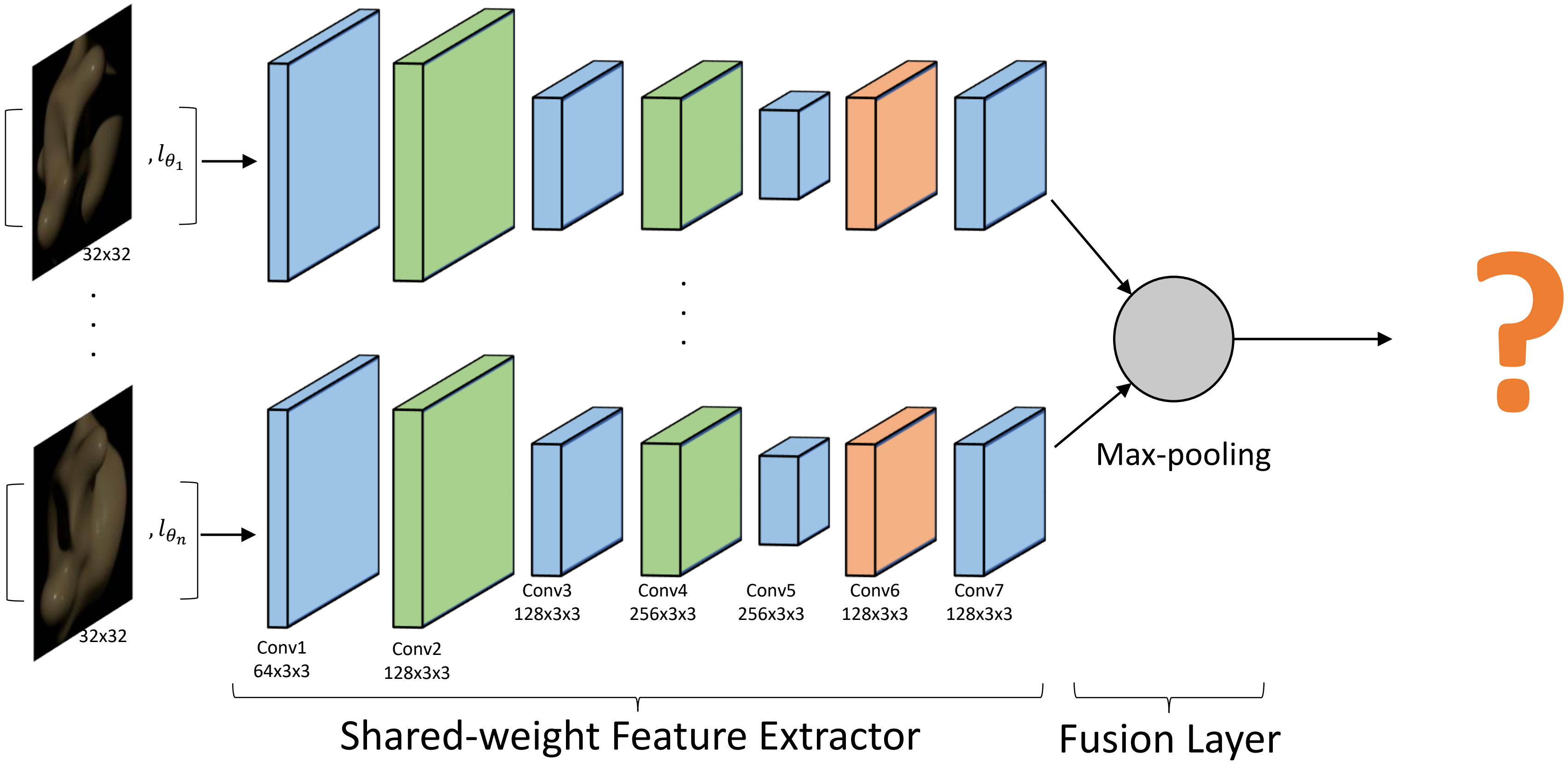
Max-pooing for multi-feature fusion



Max-pooling is well-suited for this task:

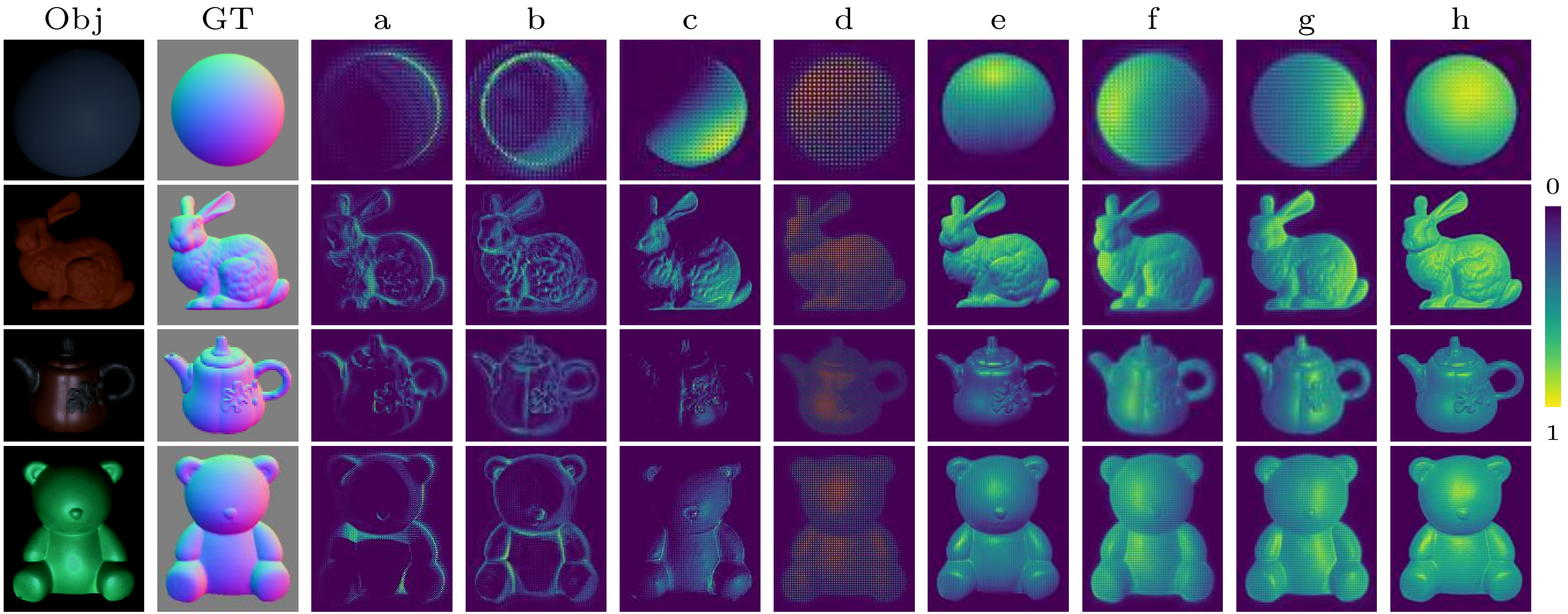
- Order-agnostic operation (compared with RNNs)
- Can fused an arbitrary number of features into a single feature
- Can extract the most salient information from all the features

Feature visualization



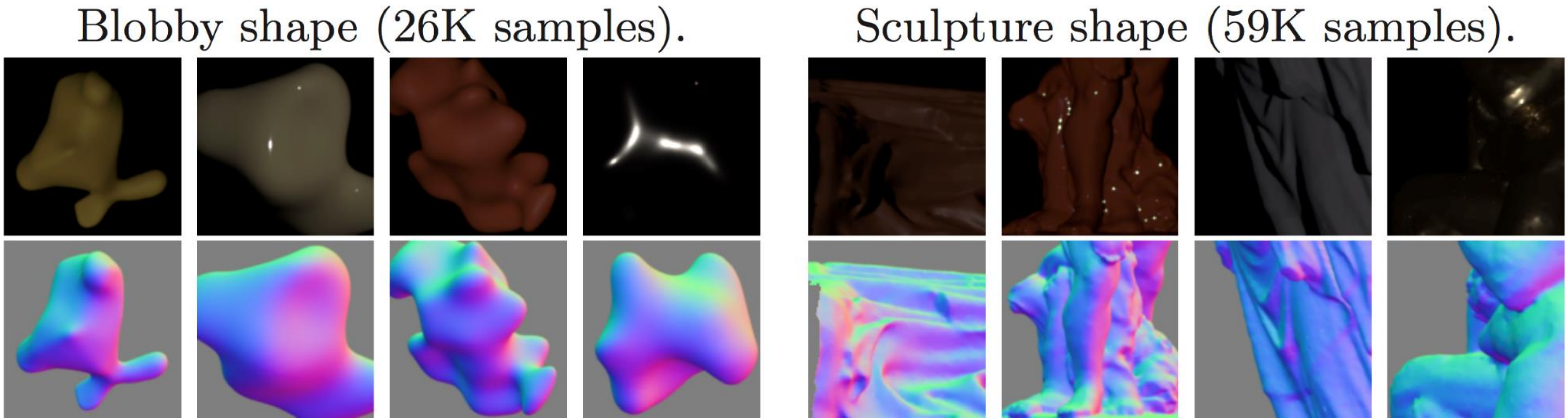
What is encoded in the fused feature?

Visualization for the fused features



- Different regions with similar normal directions are fired in different channels
- Each channel can be interpreted as the probability of the normal belonging to a certain direction

Two synthetic training datasets



- 100 BRDFs from MERL dataset [Matusik 03]
- Rendered with the physically based raytracer Mitsuba
- Trained only on the synthetic data, PS-FCN generalizes well on real data

Benchmark results using "DiLiGenT"

Method	ball	cat	pot1	bear	pot2	buddha	goblet	reading	cow	harvest	Avg.
L2	4.10	8.41	8.89	8.39	14.65	14.92	18.50	19.80	25.60	30.62	15.39
AZ08	2.71	6.53	7.23	**5.96**	11.03	12.54	13.93	14.17	21.48	30.50	12.61
WG10	2.06	6.73	7.18	6.50	13.12	10.91	15.70	15.39	25.89	30.01	13.35
IA14	3.34	6.74	6.64	7.11	8.77	10.47	9.71	14.19	13.05	25.95	10.60
ST14	**1.74**	**6.12**	**6.51**	6.12	8.78	10.60	10.09	13.63	13.93	25.44	10.30
DPSN	2.02	6.54	7.05	6.31	7.86	12.68	11.28	15.51	8.01	16.86	9.41
PS-FCN (16)	3.31	7.64	8.14	7.47	8.22	8.76	9.81	14.09	8.78	17.48	9.37
PS-FCN (96)	2.82	6.16	7.13	7.55	**7.25**	**7.91**	**8.60**	**13.33**	**7.33**	**15.85**	**8.39**

[ECCV 18]
CNN-PS: CNN-based Photometric Stereo for General Non-Convex Surfaces

CNN-PS: CNN-based Photometric Stereo for General Non-Convex Surfaces

Satoshi Ikehata

National Institute of Informatics, Tokyo, Japan
sikehata@nii.ac.jp

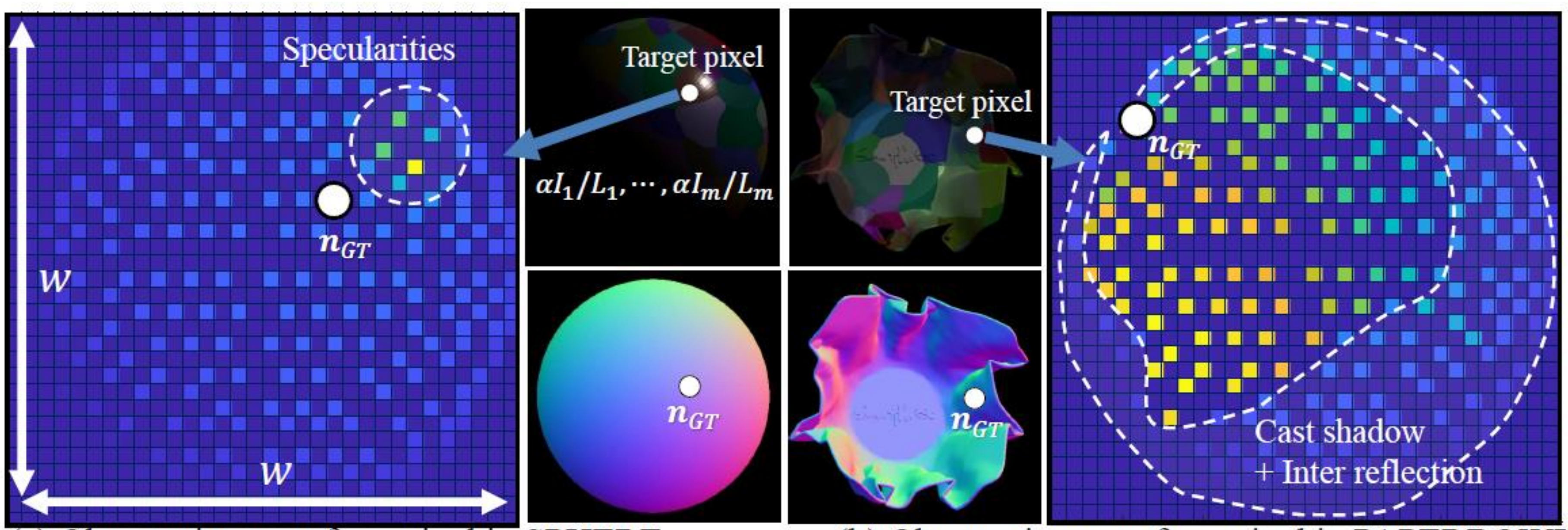
Abstract. Most conventional photometric stereo algorithms inversely solve a BRDF-based image formation model. However, the actual imaging process is often far more complex due to the global light transport on

Observation map (per-pixel)

- Find an easy-to-learn representation

Definition of an observation map (α is normalizing factor, L is light intensity)

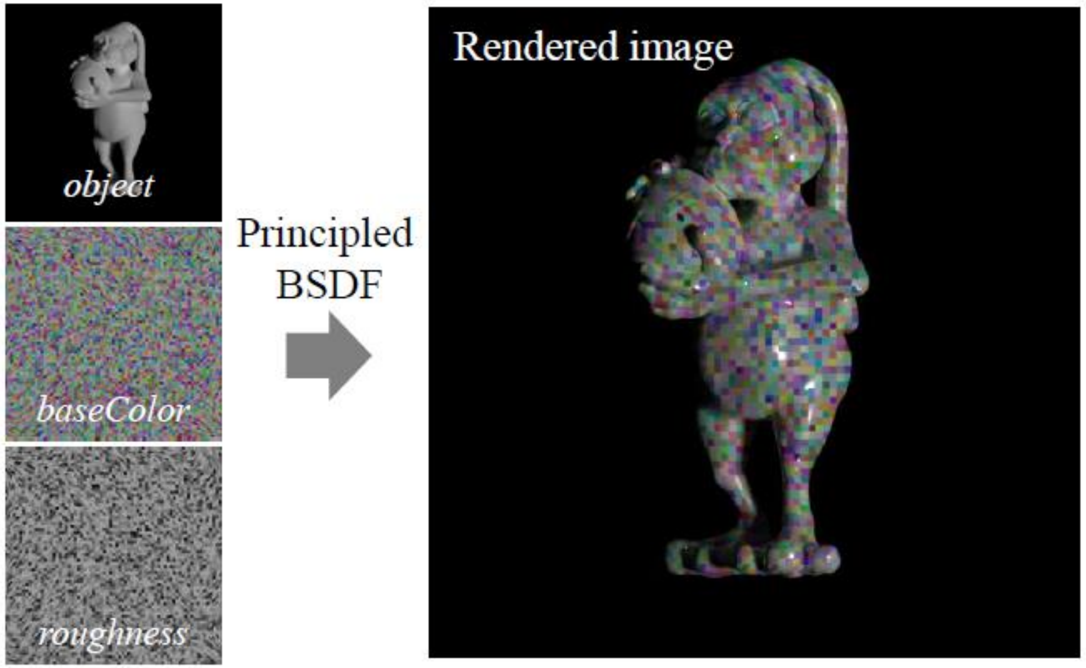
$$O_{\text{int}(w(l_x+1)/2),\text{int}(w(l_y+1)/2)} = \alpha I_j / L_j \;\; \forall\, j \in 1, \cdots, m,$$



(a) Observation map for a pixel in SPHERE (b) Observation map for a pixel in PAPERBOWL

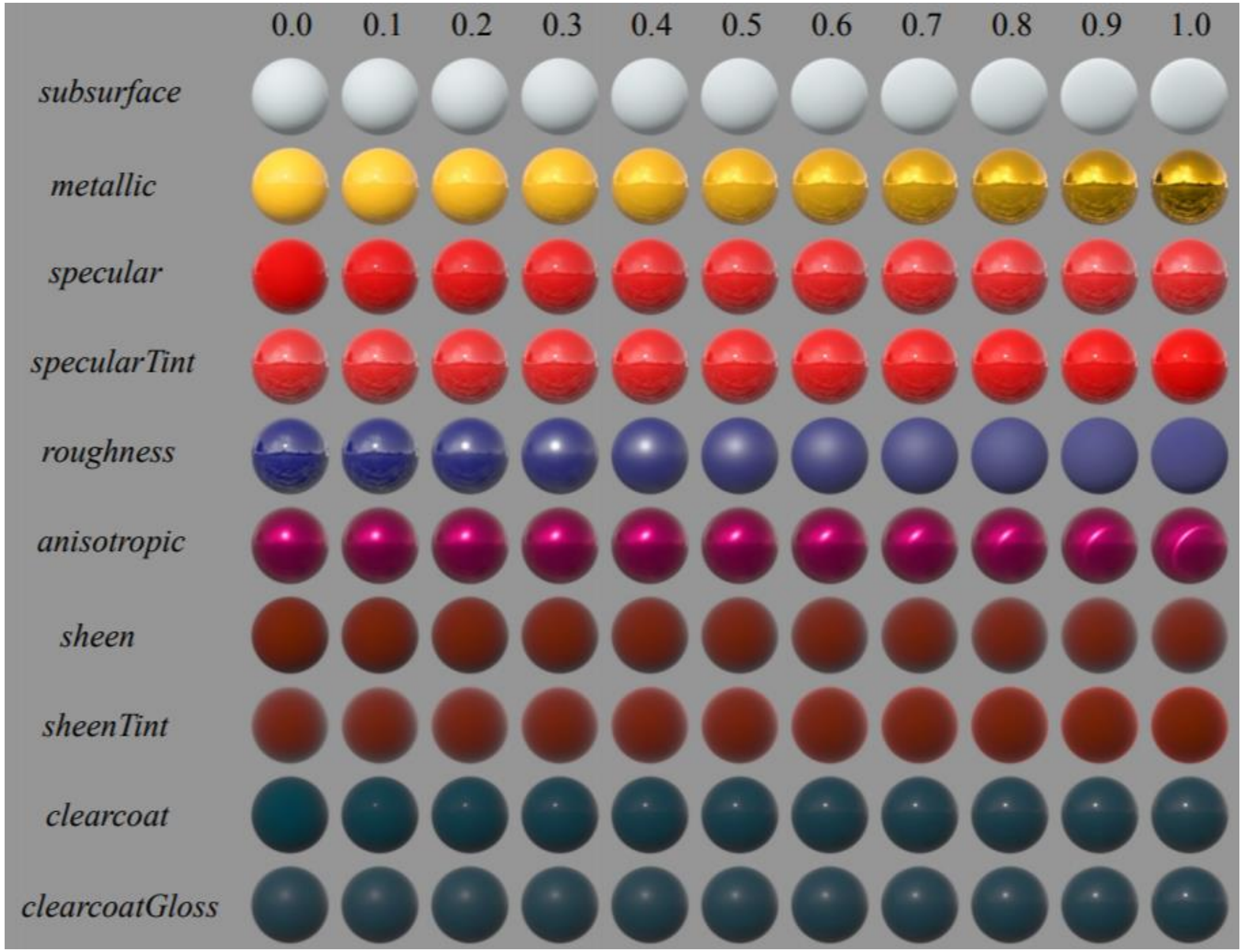
Training dataset

- Cycles renderer in Blender
- A a set of 3-D model, BSDF parameter maps (Disney's Principled BSDS model), and lighting configuration
- Generate observation map pixelwisely



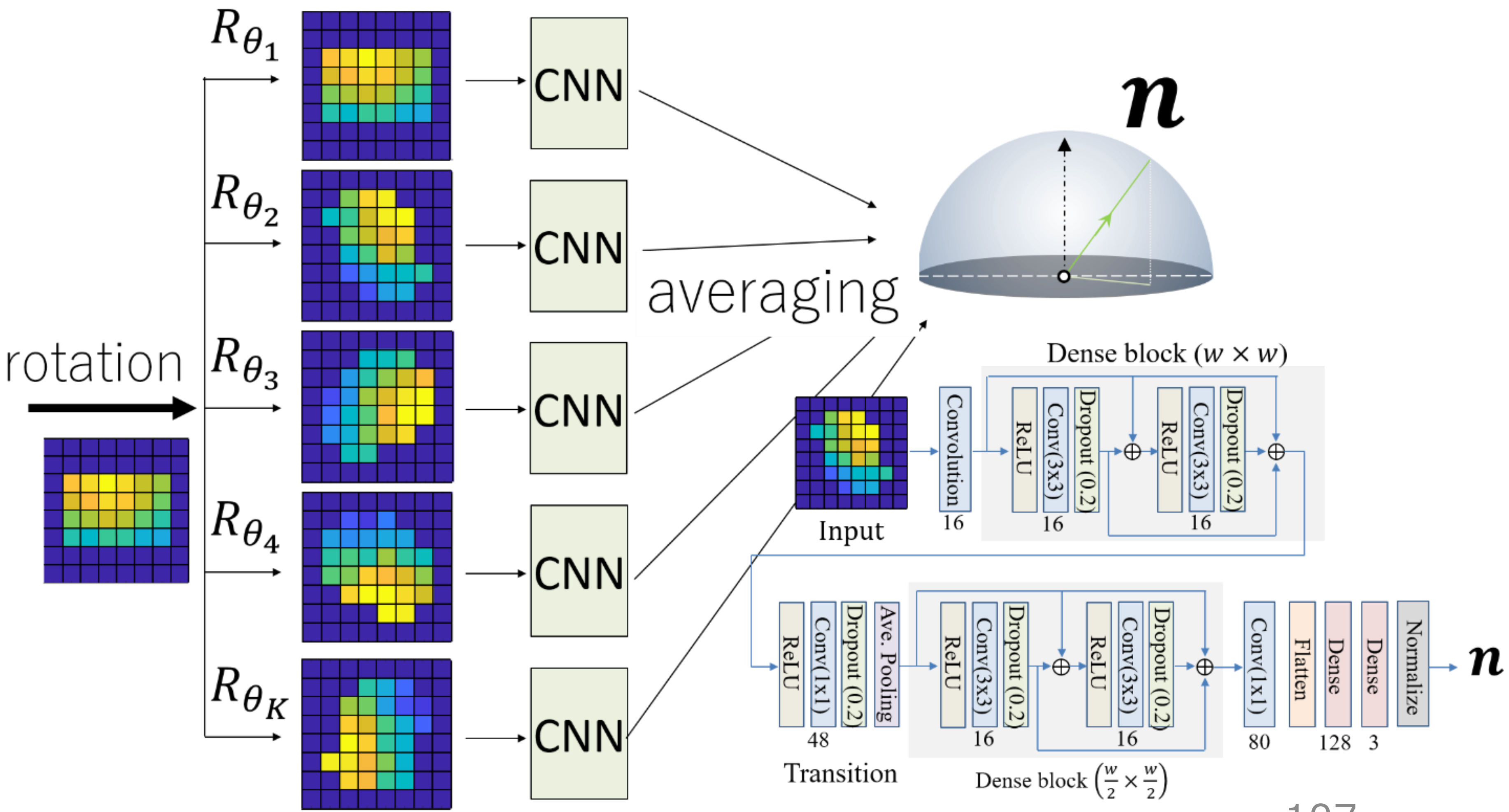
Disney's principled BSDS model

- Intuitive rather than physical parameters should be used
- As few parameters as possible
- Parameters should be zero to one over their plausible range
- Parameters should be allowed to be pushed beyond their plausible range where it makes sense
- All combinations of parameters should be as robust and plausible as possible



Normal prediction

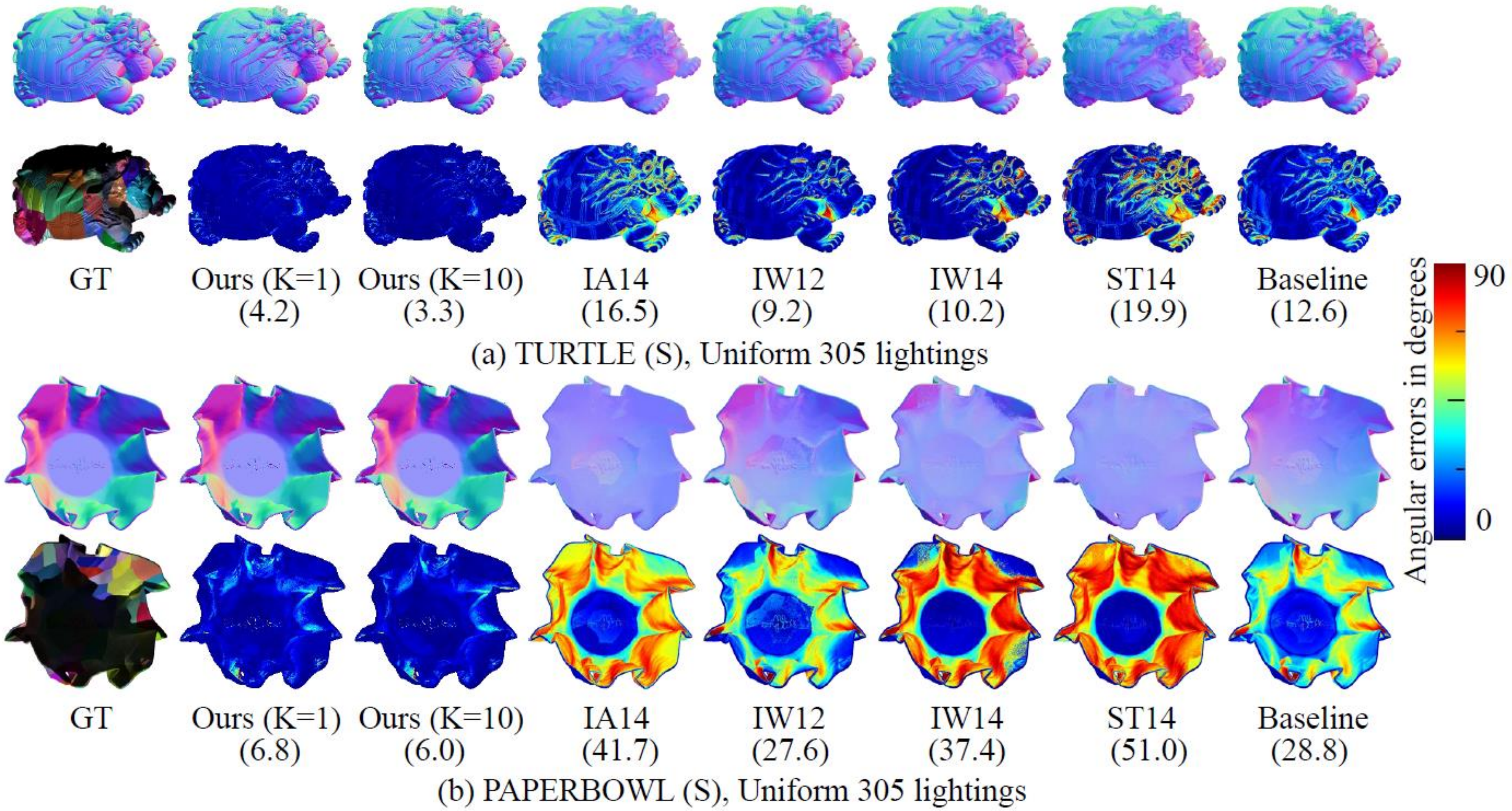
Observation map



Benchmark results using “DiLiGenT”

	BALL	BEAR	BUDDHA	CAT	COW	GOBLET	HARVEST	POT1	POT2	READING	AVE. ERR	RANK
OURS (K=10)	2.2	**4.1** *	**7.9**	**4.6**	8.0	7.3	**14.0**	5.4	**6.0**	12.6	**7.2**	1
OURS (K=1)	2.7	4.5 *	8.6	5.0	8.2	**7.1**	14.2	5.9	6.3	13.0	7.6	2
HS17 [20]	**1.3**	5.6	8.5	4.9	8.2	7.6	15.8	**5.2**	6.4	12.1	7.6	2
TM18 [21]	1.5	5.8	10.4	5.4	**6.3**	11.5	22.6	6.1	7.8	**11.0**	8.8	4
IW14 [7]	2.0	4.8	8.4	5.4	13.3	8.7	18.9	6.9	10.2	12.0	9.0	5
SS17 [20]	2.0	6.3	12.7	6.5	8.0	11.3	16.9	7.1	7.9	15.5	9.4	6
ST14 [18]	1.7	6.1	10.6	6.1	13.9	10.1	25.4	6.5	8.8	13.6	10.3	7
SH17 [25]	2.2	5.3	9.3	5.6	16.8	10.5	24.6	7.3	8.4	13.0	10.3	7
IA14 [17]	3.3	7.1	10.5	6.7	13.1	9.7	26.0	6.6	8.8	14.2	10.6	9
GC10 [14]	3.2	6.6	14.9	8.2	9.6	14.2	27.8	8.5	7.9	19.1	12.0	10
BASELINE [12]	4.1	8.4	14.9	8.4	25.6	18.5	30.6	8.9	14.7	19.8	15.4	-

Results: CyclePS test dataset



[CVPR 19]
Self-calibrating Deep Photometric Stereo Networks

Self-calibrating Deep Photometric Stereo Networks

Guanying Chen[1] Kai Han[2] Boxin Shi[3,4] Yasuyuki Matsushita[5] Kwan-Yee K. Wong[1]
[1]The University of Hong Kong [2]University of Oxford
[3]Peking University [4]Peng Cheng Laboratory [5]Osaka University

Abstract

This paper proposes an uncalibrated photometric stereo method for non-Lambertian scenes based on deep learning. Unlike previous approaches that heavily rely on assumptions of specific reflectances and light source distributions, our method is able to determine both shape and light directions of a scene with unknown arbitrary reflectances observed under unknown varying light directions. To achieve this goal, we propose a two-stage deep learning architecture, called SDPS-Net, *which can effectively take advantage of intermediate supervision, resulting in reduced learning difficulty compared to a single-stage model. Experiments on both synthetic and real datasets show that our proposed approach significantly outperforms previous uncalibrated photometric stereo methods.*

31, 15, 5]. Instead of explicitly modeling complex surface reflectances, they directly learn the mapping from reflectance observations to surface normal given light directions. Although they have obtained promising results in a calibrated setting, they cannot handle the more challenging problem of *uncalibrated* photometric stereo, where light directions are unknown. One simple strategy to handle uncalibrated photometric stereo with deep learning is to directly learn the mapping from images to surface normal without taking the light directions as input. However, as reported in [5], the performance of such a model lags far behind those which take both images and light directions as input.

In this paper, we propose a two-stage model named Self-calibrating Deep Photometric Stereo Networks (SDPS-Net) to tackle this problem. The first stage of SDPS-Net, denoted as *Lighting Calibration Network* (LCNet), takes an

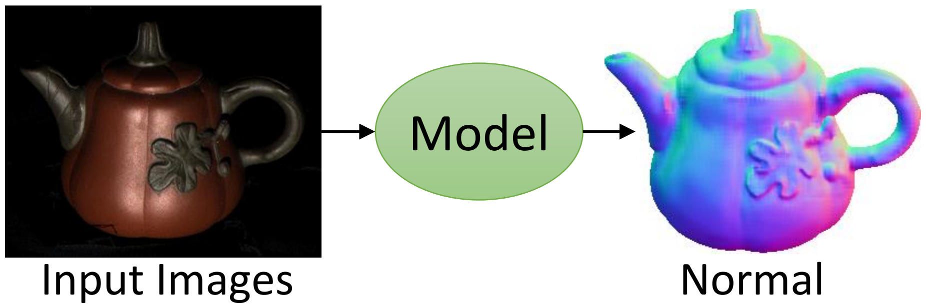
Motivation

- Recent learning based methods for PS often assume known light directions
  - DPSN
  - IRPS
  - CNN-PS
  - PS-FCN

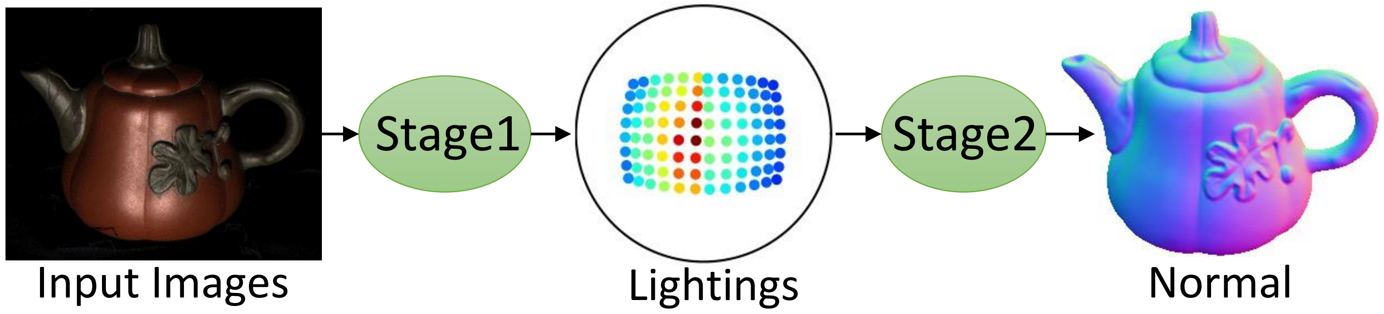
- The performance of the existing learning based method for UPS is far from satisfactory
  - PS-FCN + uncalibrated setting

Main idea of SDPS-Net

Single-stage method:



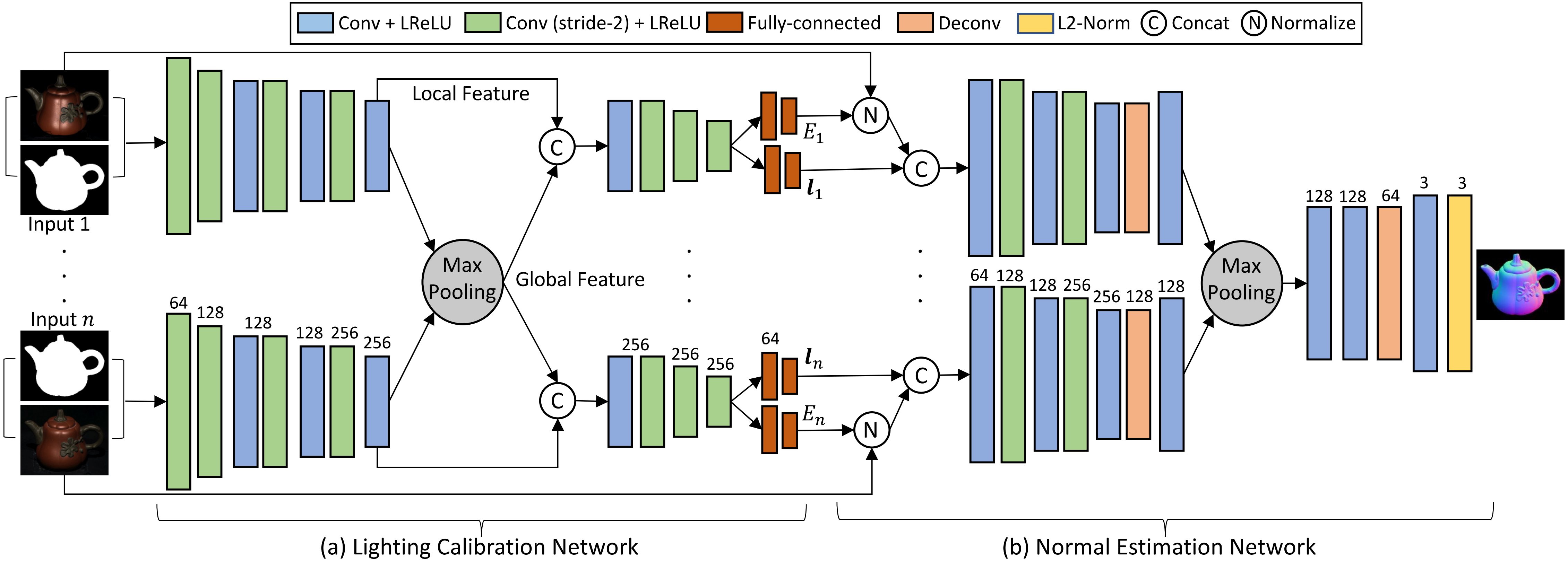
Two-stage method:



Advantages of the proposed two-stage method:

- Directional lightings are much easier to estimate than surface normals
- Take advantage of the intermediate supervision (more interpretable)
- The estimated lightings can be utilized by existing calibrated methods

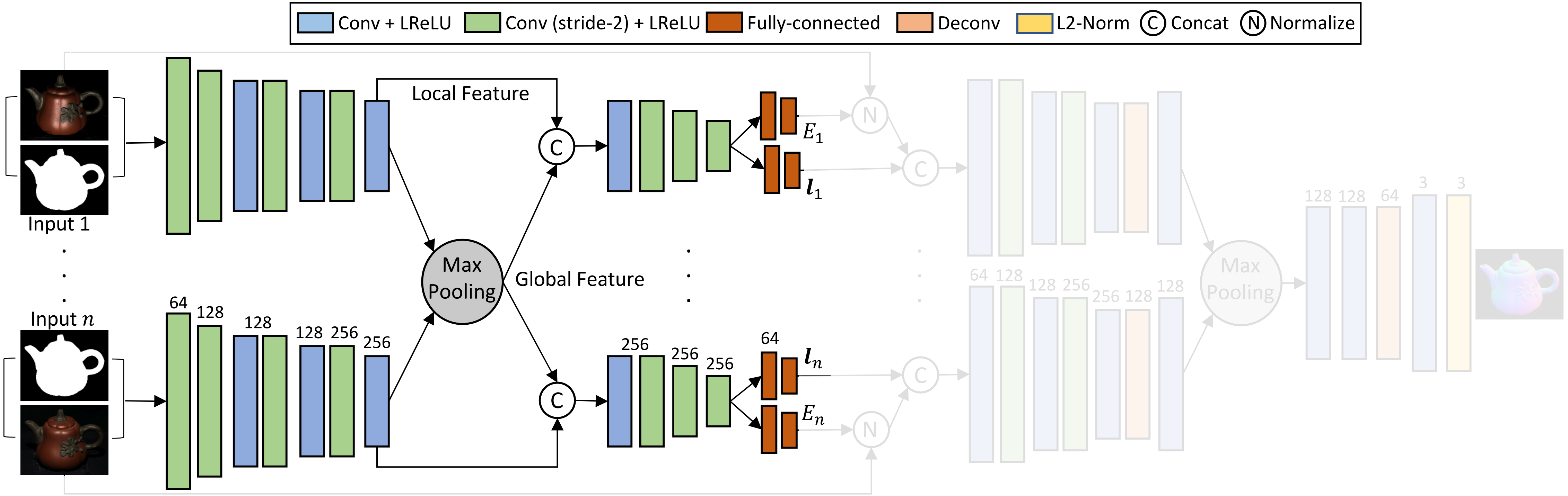
The proposed two-stage framework



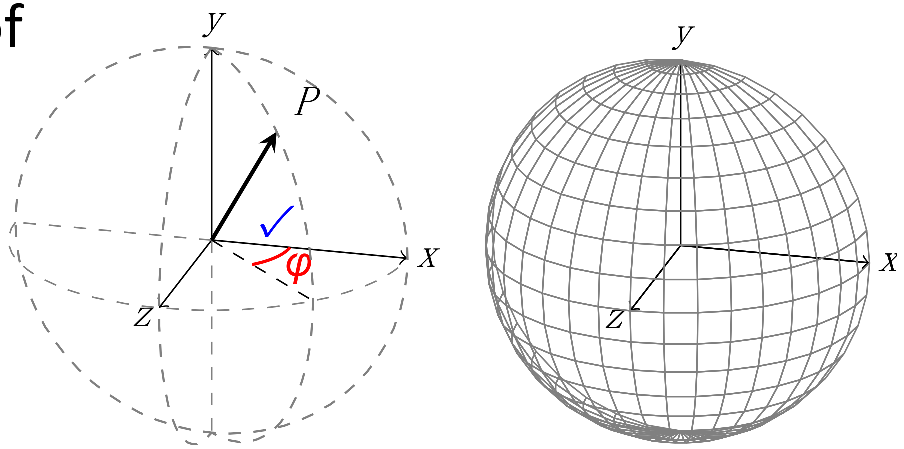
SDPS-Net consists of two stages:

- Stage 1: Lighting Calibration Network (LCNet) for lighting estimation
- Stage 2: Normal Estimation Network (NENet) for normal estimation

Stage 1: Lighting calibration network



Discretization of lighting space:

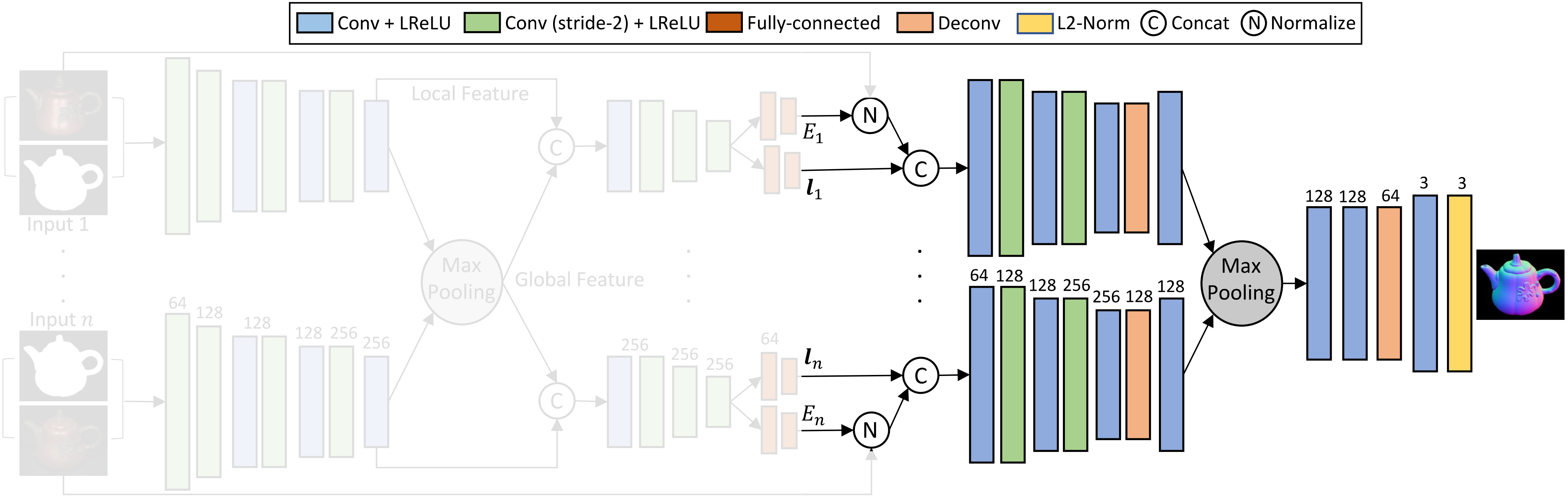


Loss function:

$$\mathcal{L}_{\text{Light}} = \lambda_{l_a}\mathcal{L}_{l_a} + \lambda_{l_e}\mathcal{L}_{l_e} + \lambda_e\mathcal{L}_e$$

- $\mathcal{L}_{la}$: azimuth classification loss
- $\mathcal{L}_{le}$: elevation classification loss
- $\mathcal{L}_{e}$: light intensity classification loss

Stage 2: Normal estimation network



Loss function:

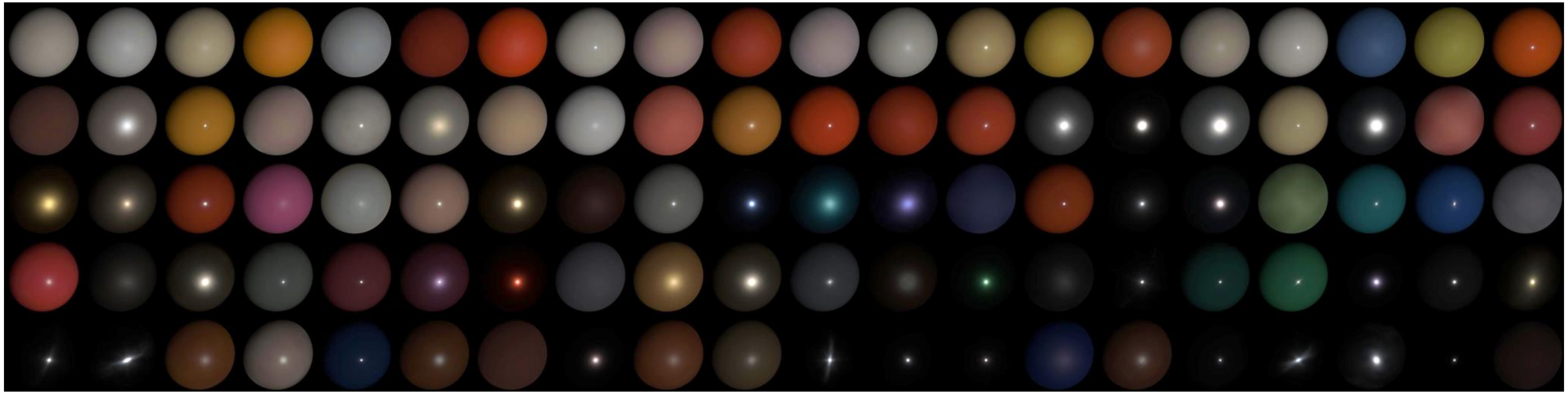
$$\mathcal{L}_{\text{Normal}} = \frac{1}{hw} \sum_{i}^{hw} \left(1 - \boldsymbol{n}_i^{\top} \tilde{\boldsymbol{n}}_i\right)$$

- Cosine similarity loss

- Our framework can handle an arbitrary number of images in an order agnostic manner.

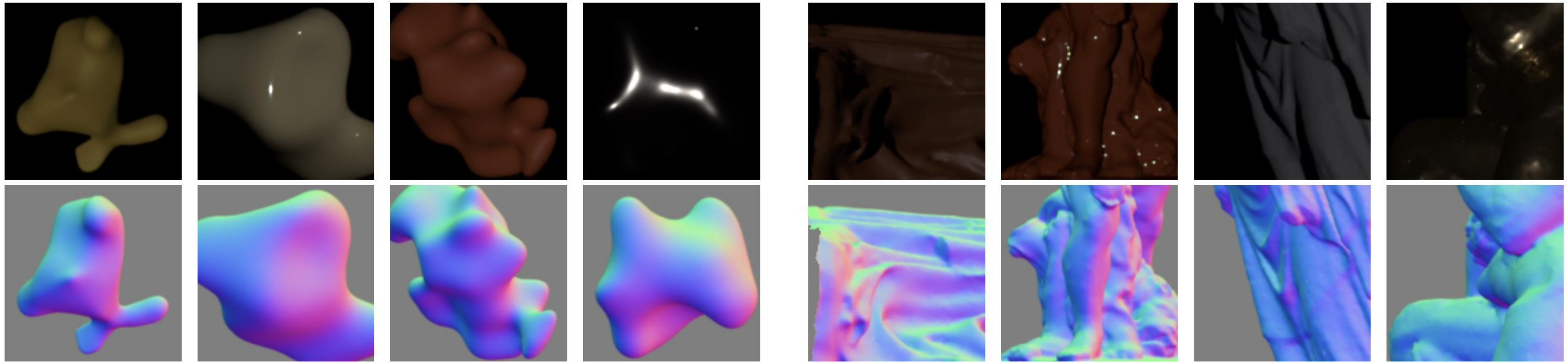
Synthetic training dataset [Chen 18]

- 100 measured BRDFs from MERL dataset



- Cast-shadow and inter-reflection are considered using Mitsuba.

Blobby shape (26K samples). Sculpture shape (59K samples).

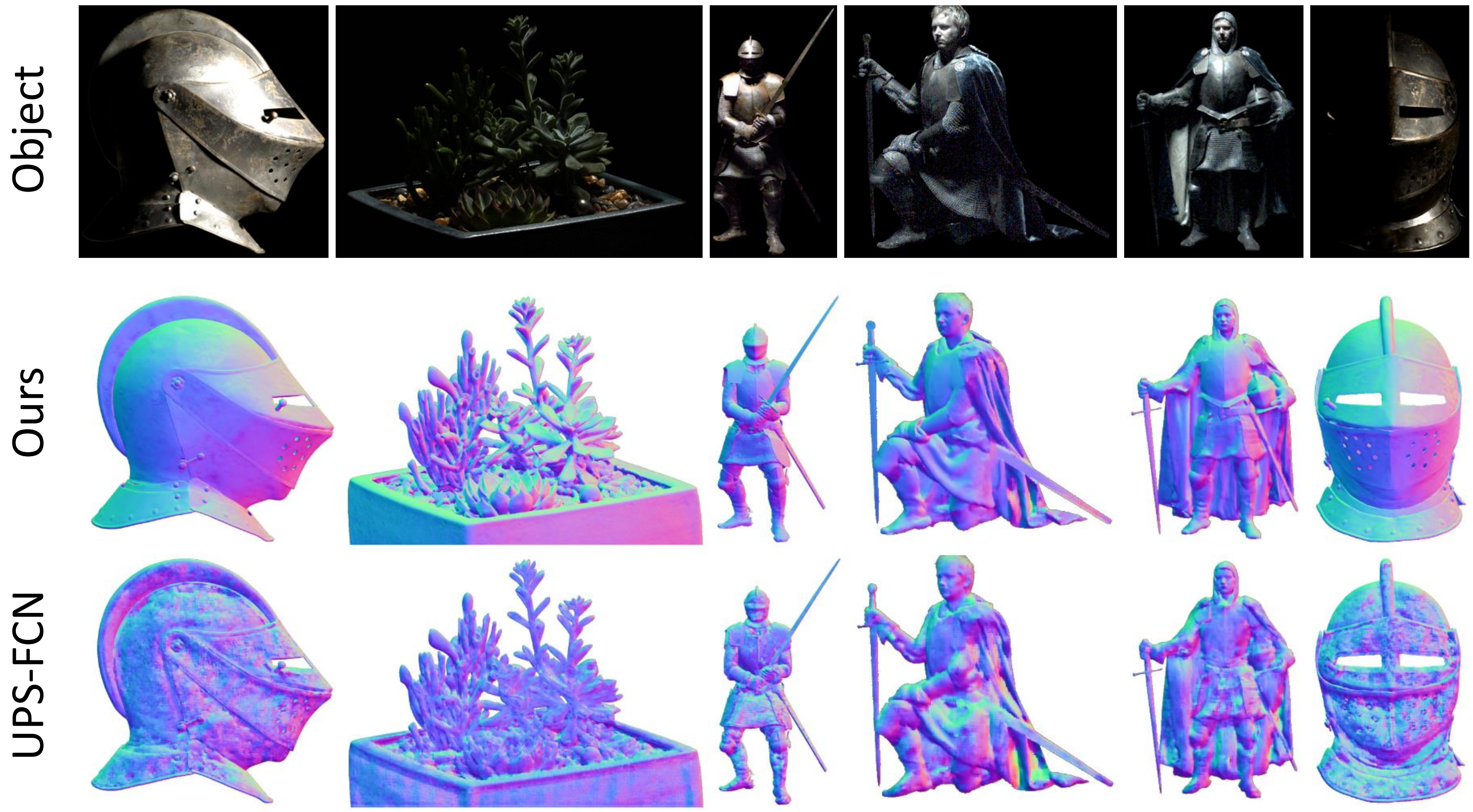


Benchmark results using “DiLiGenT”

Method	BALL	CAT	POT1	BEAR	POT2	BUDD.	GOBL.	READ.	COW	HARV.	Avg.
AM07	7.3	31.5	18.4	16.8	49.2	32.8	46.5	53.7	54.7	61.7	37.3
SM10	8.9	19.8	16.7	12.0	50.7	15.5	48.8	26.9	22.7	73.9	29.6
WT13	4.4	36.6	9.4	**6.4**	14.5	13.2	20.6	59.0	19.8	55.5	23.9
LM13	22.4	25.0	32.8	15.4	20.6	25.8	29.2	48.2	22.5	34.5	27.6
PF14	4.8	9.5	9.5	9.1	15.9	14.9	29.9	24.2	19.5	29.2	16.7
LC18	9.3	12.6	12.4	10.9	15.7	19.0	18.3	22.3	15.0	28.0	16.3
UPS-FCN	6.6	14.7	14.0	11.2	14.2	15.9	20.7	23.3	11.9	27.8	16.0
LCNet + L2	4.9	11.1	9.7	9.4	14.7	14.9	18.3	20.1	25.1	29.2	15.7
SDPS-Net	**2.8**	**8.1**	**8.1**	6.9	**7.5**	**9.00**	**11.9**	**14.9**	**8.5**	**17.4**	**9.5**

- Our method achieves state-of-the-art results (value the lower the better)
- The proposed LCNet can be integrated with the previous calibrated methods

Qualitative results on light stage data gallery



[CVPR 19]
Learning to Minify Photometric Stereo

Learning to Minify Photometric Stereo

Junxuan Li[1,2] Antonio Robles-Kelly[3] Shaodi You[1,2] Yasuyuki Matsushita[4]

[1]Australian National University, College of Eng. and Comp. Sci., Acton, ACT 2601, Australia
[2]Data61-CSIRO, Black Mountain Laboratories, Acton, ACT 2601, Australia
[3]Deakin University, Faculty of Sci., Eng. and Built Env., Waurn Ponds, VIC 3216, Australia
[4]Osaka University, Graduate School of Information Science and Technology, Osaka 565-0871, Japan

Abstract

Photometric stereo estimates the surface normal given a set of images acquired under different illumination conditions. To deal with diverse factors involved in the image formation process, recent photometric stereo methods demand a large number of images as input. We propose a method that can dramatically decrease the demands on the number of images by learning the most informative ones under different illumination conditions. To this end, we use a deep learning framework to automatically learn the critical illumination conditions required at input. Furthermore, we present an occlusion layer that can synthesize cast shad-

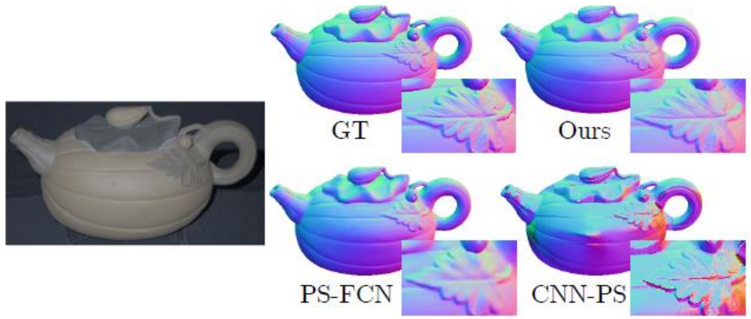
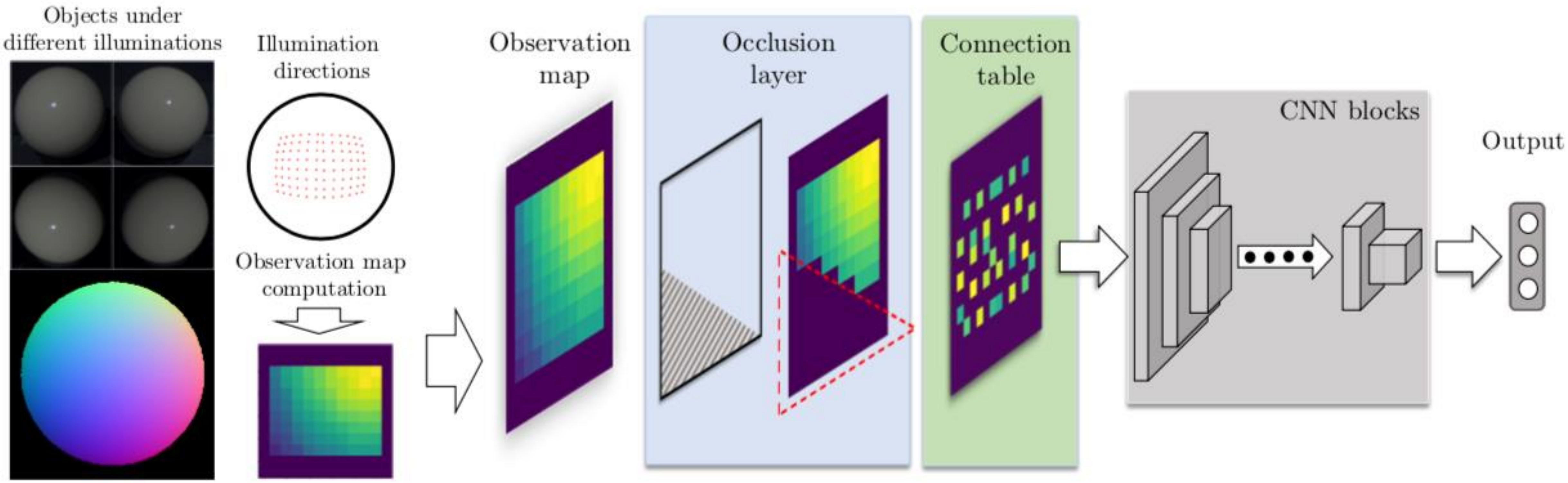


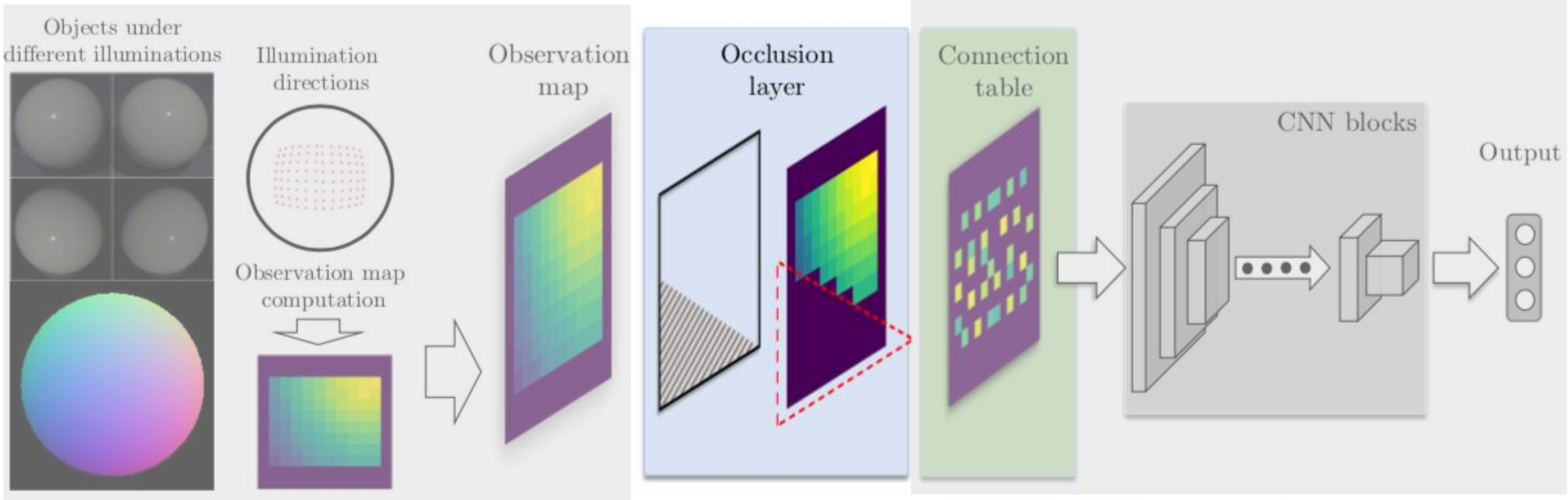
Figure 1. Performance with only 8 inputs for our method, PS-FCN [3] and CNN-PS [8] on the "pot1" from DiLiGenT [17]. Note we outperform the alternatives.

ages so as to minify the photometric stereo input. We ap-

Main idea

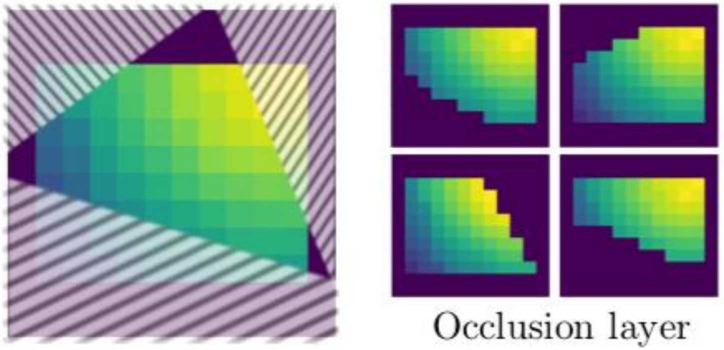

Objects under different illuminations
Illumination directions
Observation map computation
Observation map
Occlusion layer
Connection table
CNN blocks
Output

Main idea

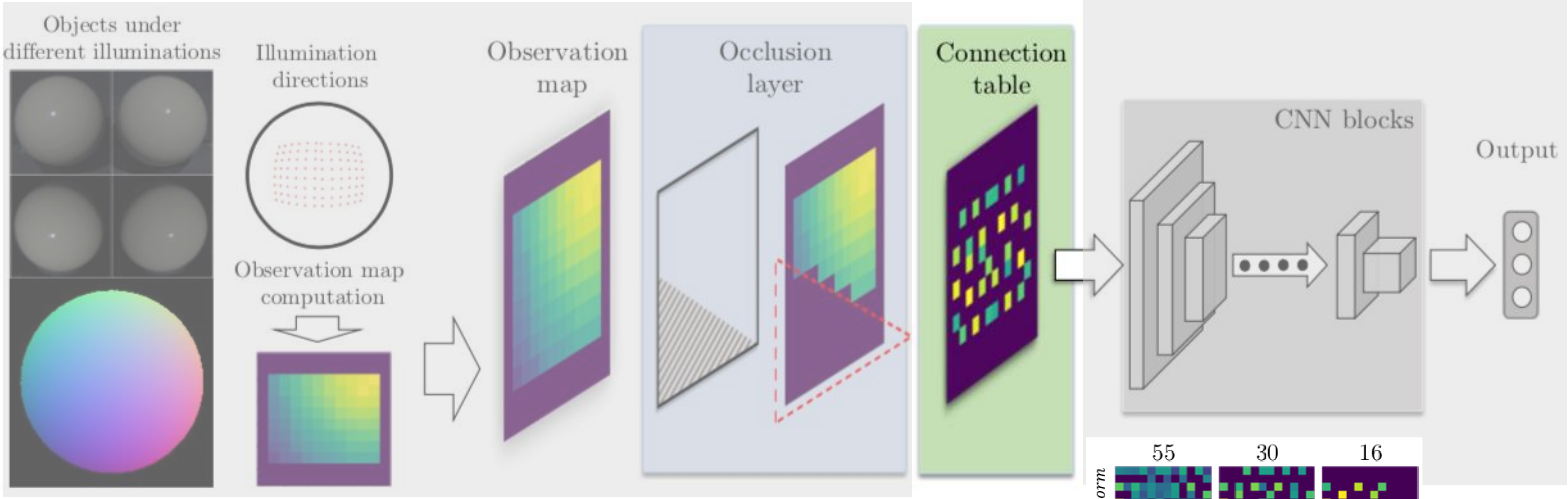


Occlusion layer

- Cast-shadows are consistent patterns with a relatively sharp and straight boundary
- Randomly select two sides of the map, and randomly picks a point on each side



Main idea



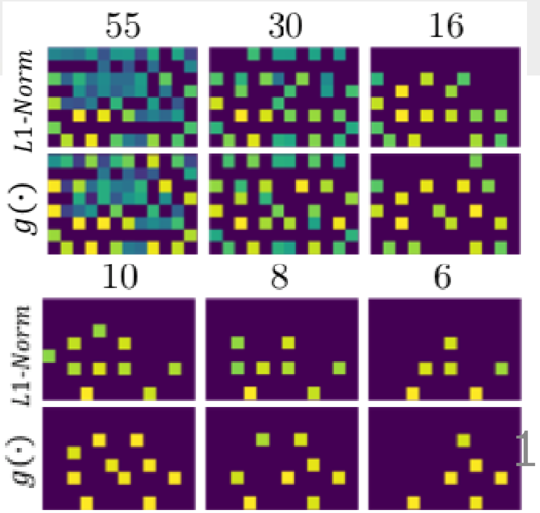
Sparse connection table

- Select the most relevant illuminant directions at input
- Fixed after training

Loss functions

$$\mathcal{L} = \|\mathbf{n} - \mathbf{n}'\|_2^2 + \lambda g(\boldsymbol{C})$$

$$g(\boldsymbol{C}) = \sum_{i,j} \left(2\boldsymbol{C}_{i,j} - \frac{\boldsymbol{C}_{i,j}^2}{2\alpha} \right)$$



Effectiveness of occlusion layer

- Compared with random zeroing in DPSN

Benchmark results using “DiLiGenT”

***10 selected lights**

Light-Config	Proposed	PS-FCN	CNN-PS	IW12	LS
Random (10 trials)	**10.02**	10.51	14.34	16.37	17.31
Selected by Proposed method		11.35	13.02	15.83	17.12
Optimal [Drbohlav 05]		8.73	13.35	15.50	16.57

[ICCV 19]
SPLINE-Net: Sparse Photometric Stereo through Lighting Interpolation and Normal Estimation Networks

SPLINE-Net: Sparse Photometric Stereo through Lighting Interpolation and Normal Estimation Networks

Qian Zheng[1♯*] Yiming Jia[2♯†] Boxin Shi[3,4*] Xudong Jiang[1] Ling-Yu Duan[3,4] Alex C. Kot[1]
[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]Department of Precision Instrument, Tsinghua University, Beijing, China
[3]National Engineering Laboratory for Video Technology, Department of CS, Peking University, Beijing, China
[4]Peng Cheng Laboratory, Shenzhen, China
{zhengqian, exdjiang, eackot}@ntu.edu.sg, jiaym15@outlook.com, {shiboxin, lingyu}@pku.edu.cn

Abstract

This paper solves the Sparse Photometric stereo through Lighting Interpolation and Normal Estimation using a generative Network (SPLINE-Net). SPLINE-Net contains a lighting interpolation network to generate dense lighting observations given a sparse set of lights as inputs followed by a normal estimation network to estimate surface normals. Both networks are jointly constrained by the proposed symmetric and asymmetric loss functions to enforce isotropic constrain and perform outlier rejection of global illumination effects. SPLINE-Net is verified to outperform existing methods for photometric stereo of general BRDFs by using only ten images of different lights instead of using nearly one hundred images.

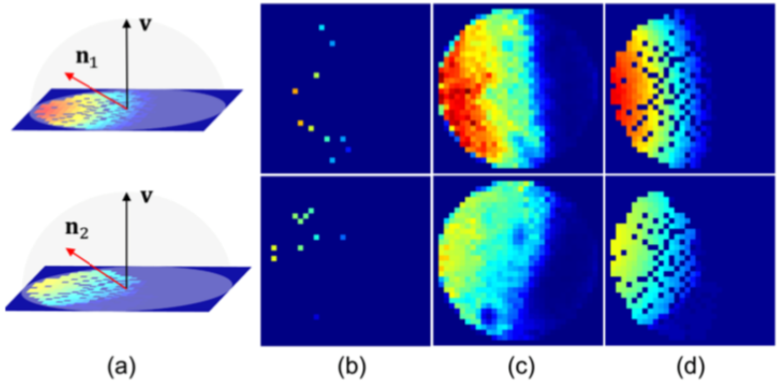


Figure 1. An illustration of observation maps corresponding to two surface normals (a brief introduction of observation maps can be found in Section 3.2 and [18]). (a) Two surface normals and their observation maps with dense lights, (b) sparse observation maps with 10 order-agnostic lights, (c) dense observation maps generated by our SPLINE-Net given sparse observation maps in (b) as

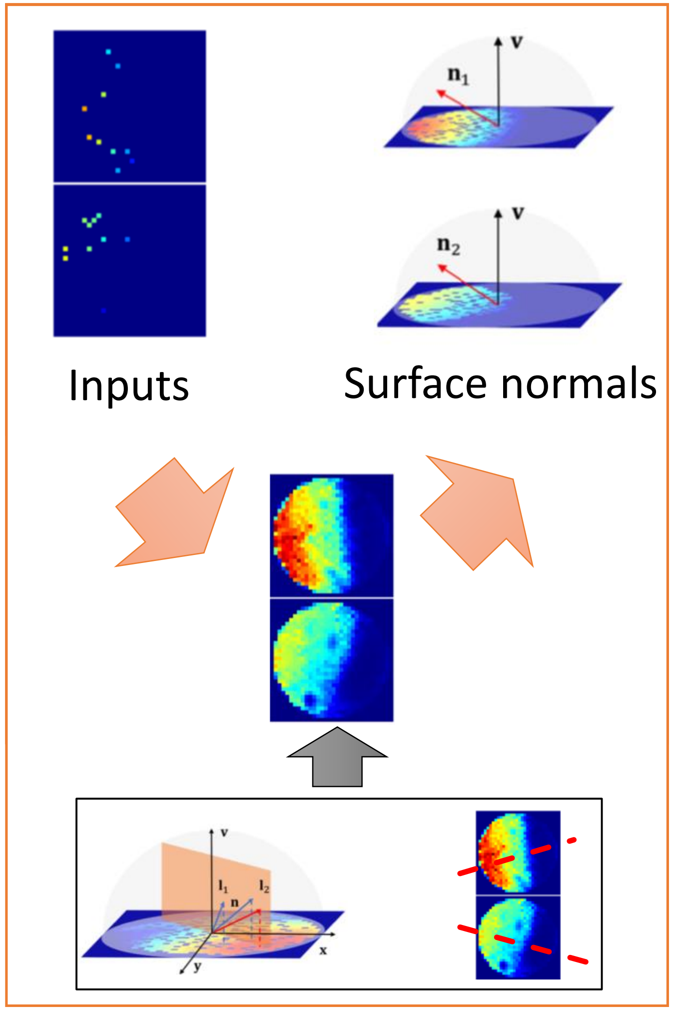
Key idea

- Sparse photometric stereo
  - Fixed number of inputs with **arbitrary** lightings
- Basic idea
  - Spatial continuity: dense interpolation
  - Isotropy of BRDFs: physics constraint

Random positions of valid pixels in observation maps

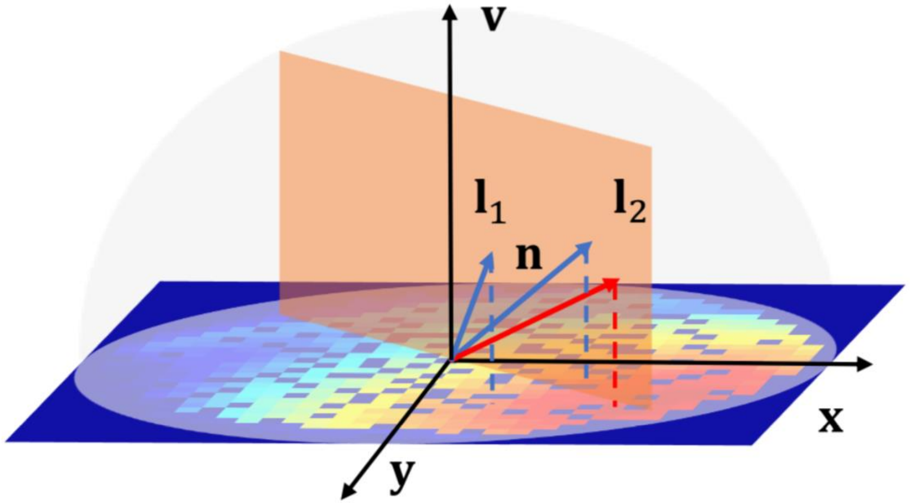
Lighting interpolation **guides** normal estimation

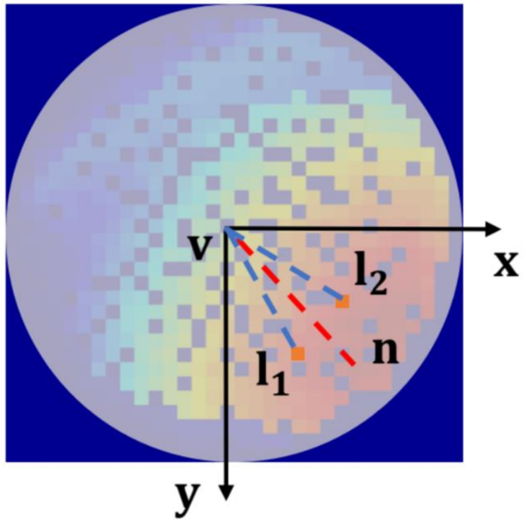
Symmetric pattern in observation maps

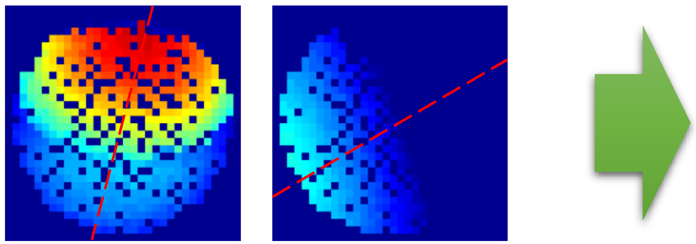


Isotropic BRDFs in observation maps

- $\rho(\mathbf{n}^{\mathrm{T}}\mathbf{l}, \mathbf{n}^{\mathrm{T}}\mathbf{v}, \mathbf{v}^{\mathrm{T}}\mathbf{l})$







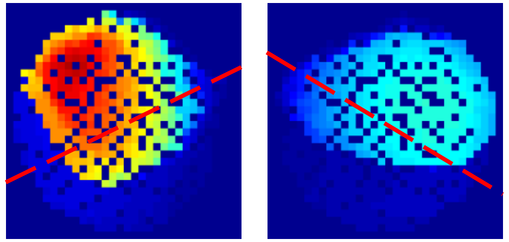
Loss functions of symmetric

$$\mathcal{L}_s = \mathcal{L}_s(\mathbf{D}, \mathbf{n}) = |\mathbf{D} - r(\mathbf{D}, \mathbf{n})|_1$$

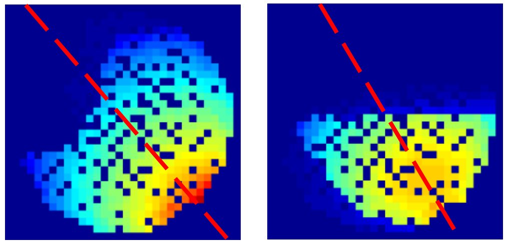
$r(\cdot)$ is a mirror function

Global illumination effects in observation maps

- Inter-reflections



- Cast shadows

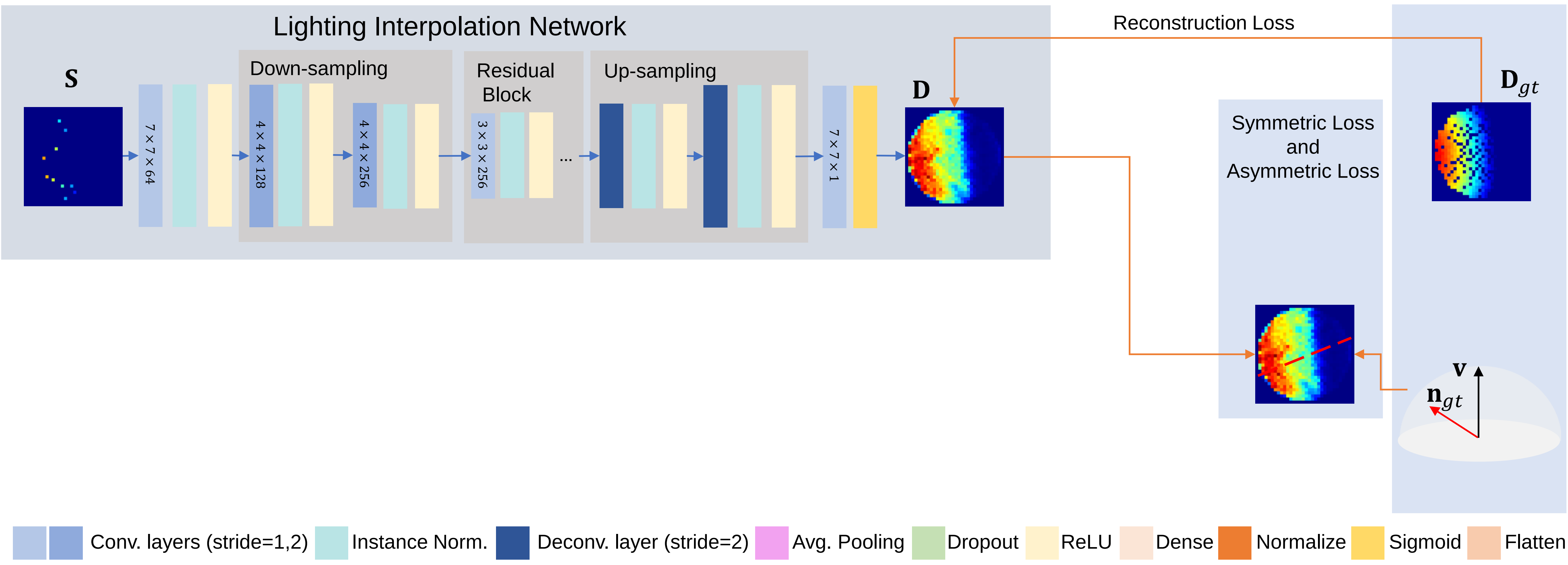


Loss functions of asymmetric

$$\mathcal{L}_a = \mathcal{L}_a(\mathbf{D}, \mathbf{n}) = \left| |\mathbf{D} - r(\mathbf{D}), \mathbf{n}|_1 - \eta \right|_1 + \lambda_c \left| |p(\mathbf{D}) - r(p(\mathbf{D}), \mathbf{n})|_1 - \eta \right|_1$$

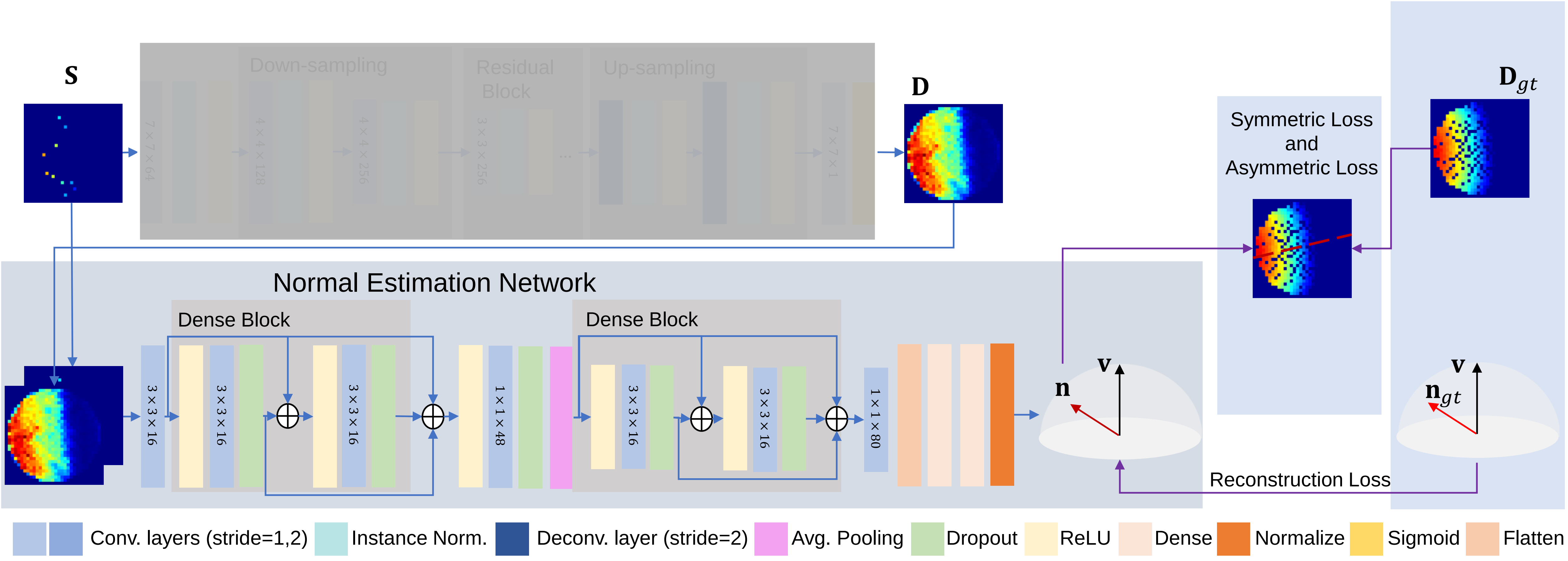
$p(\cdot)$ is a max pooling operation

Framework

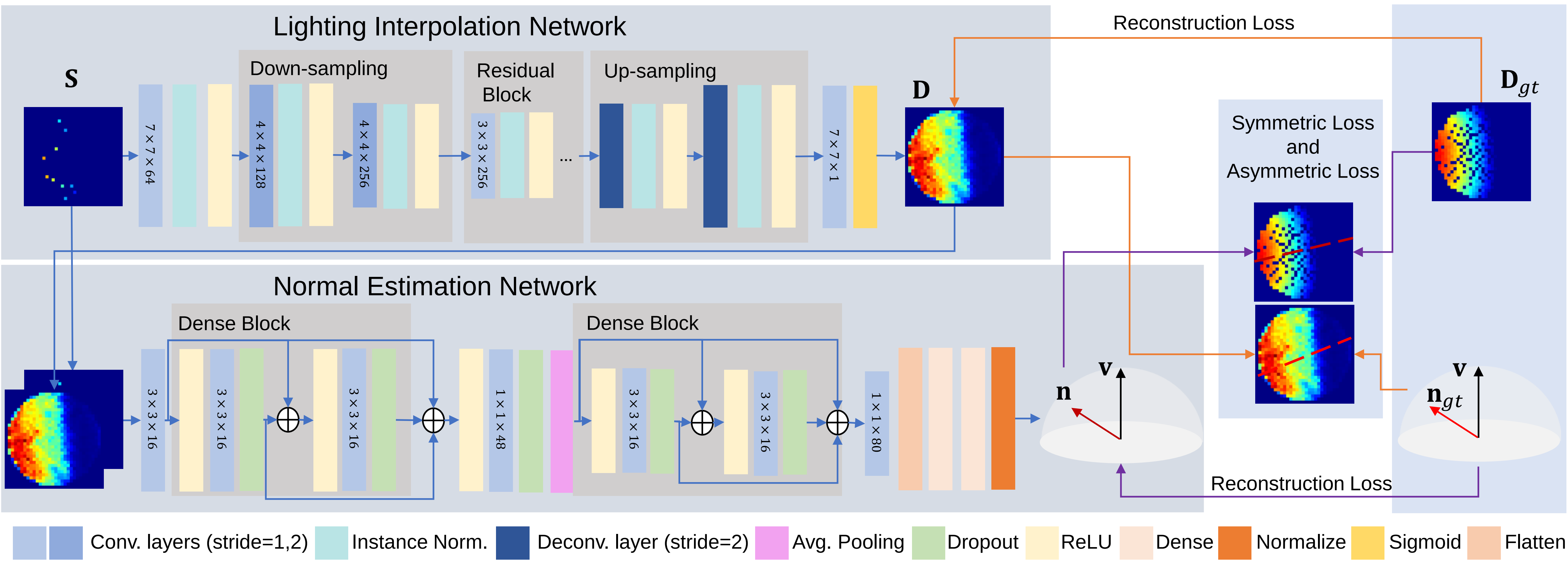


Lighting Interpolation Network
Reconstruction Loss
S
Down-sampling
Residual Block
Up-sampling
D
D_{gt}
7 × 7 × 64
4 × 4 × 128
4 × 4 × 256
3 × 3 × 256
...
7 × 7 × 1
Symmetric Loss and Asymmetric Loss
v
n_{gt}
Conv. layers (stride=1,2)
Instance Norm.
Deconv. layer (stride=2)
Avg. Pooling
Dropout
ReLU
Dense
Normalize
Sigmoid
Flatten

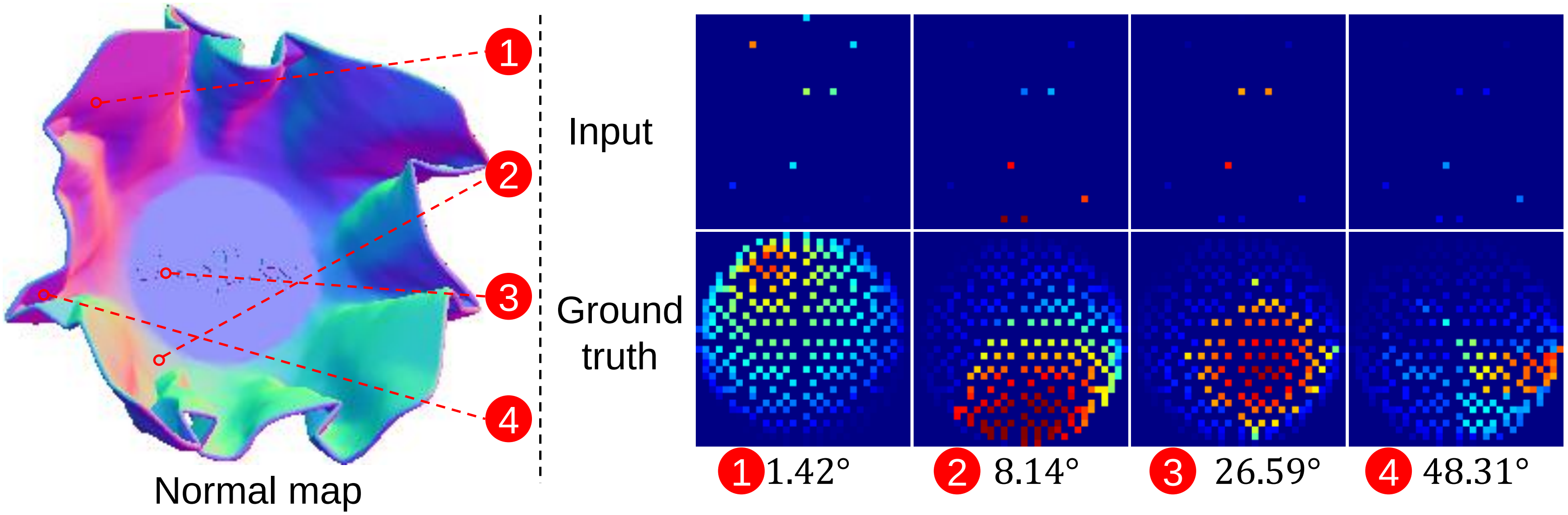
Framework


S
Down-sampling
Residual Block
Up-sampling
7 × 7 × 64
4 × 4 × 128
4 × 4 × 256
3 × 3 × 256
7 × 7 × 1
D
D_{gt}
Symmetric Loss and Asymmetric Loss
Normal Estimation Network
Dense Block
Dense Block
3 × 3 × 16
3 × 3 × 16
3 × 3 × 16
1 × 1 × 48
3 × 3 × 16
3 × 3 × 16
1 × 1 × 80
n
v
n_{gt}
v
Reconstruction Loss
Conv. layers (stride=1,2)
Instance Norm.
Deconv. layer (stride=2)
Avg. Pooling
Dropout
ReLU
Dense
Normalize
Sigmoid
Flatten

Framework

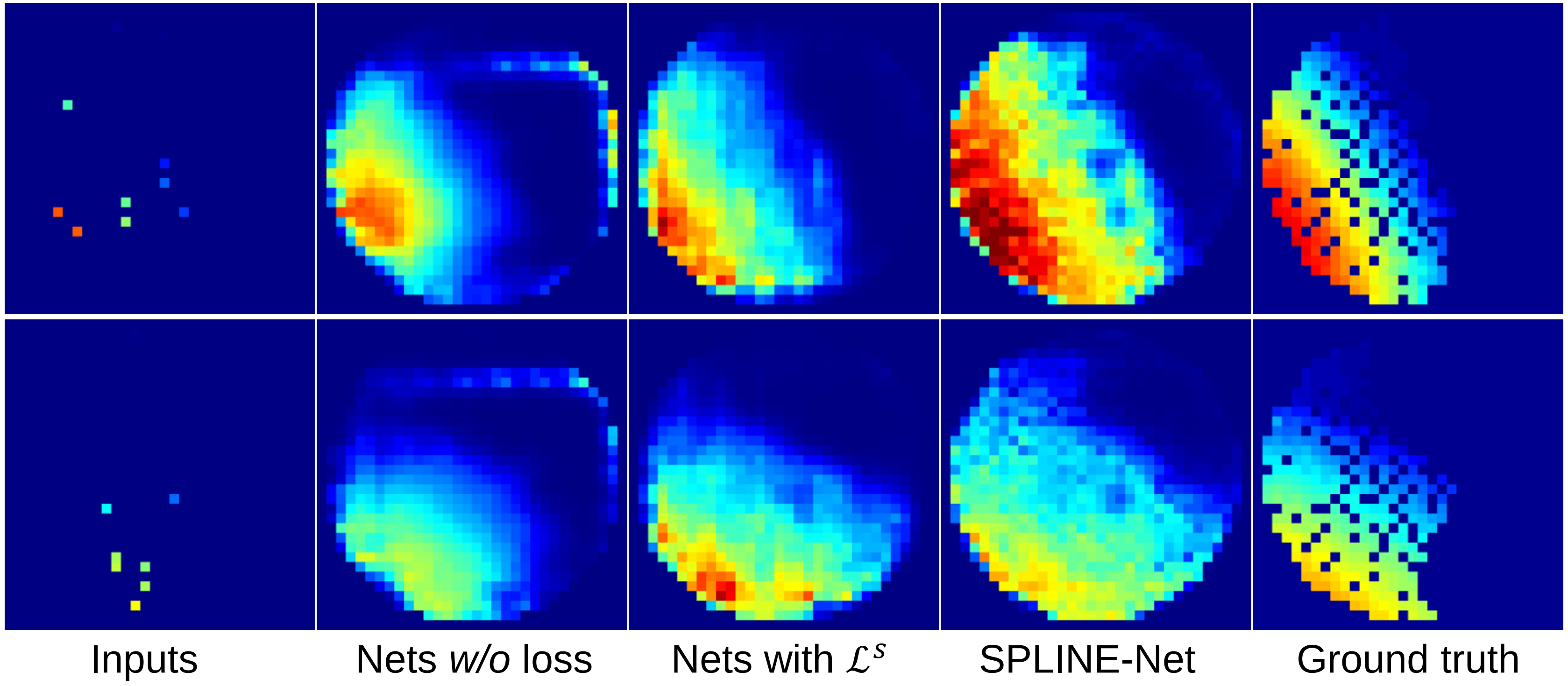

Lighting Interpolation Network
S
Down-sampling
Residual Block
Up-sampling
D
7×7×64
4×4×128
4×4×256
3×3×256
...
7×7×1
Reconstruction Loss
D_{gt}
Symmetric Loss and Asymmetric Loss
Normal Estimation Network
Dense Block
Dense Block
3×3×16
3×3×16
3×3×16
1×1×48
3×3×16
3×3×16
1×1×80
n
v
n_{gt}
v
Reconstruction Loss
Conv. layers (stride=1,2)
Instance Norm.
Deconv. layer (stride=2)
Avg. Pooling
Dropout
ReLU
Dense
Normalize
Sigmoid
Flatten

Noise in sparse observation maps (inputs)



- More brighter pixels, less shadows
- More ‘valid’ pixels, more accurate results

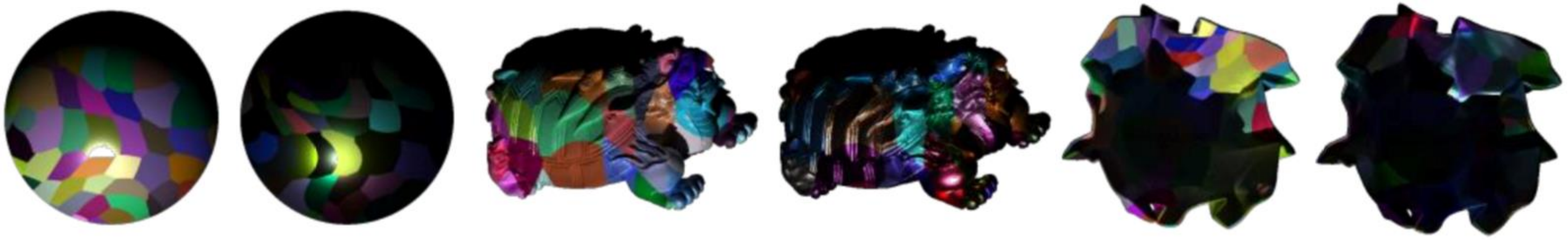
Generated dense observation maps



- Symmetric loss and asymmetric loss help generate more accurate dense observation maps

Benchmark results using Cycle-PS dataset

***10 selected lights, 100 random trials**



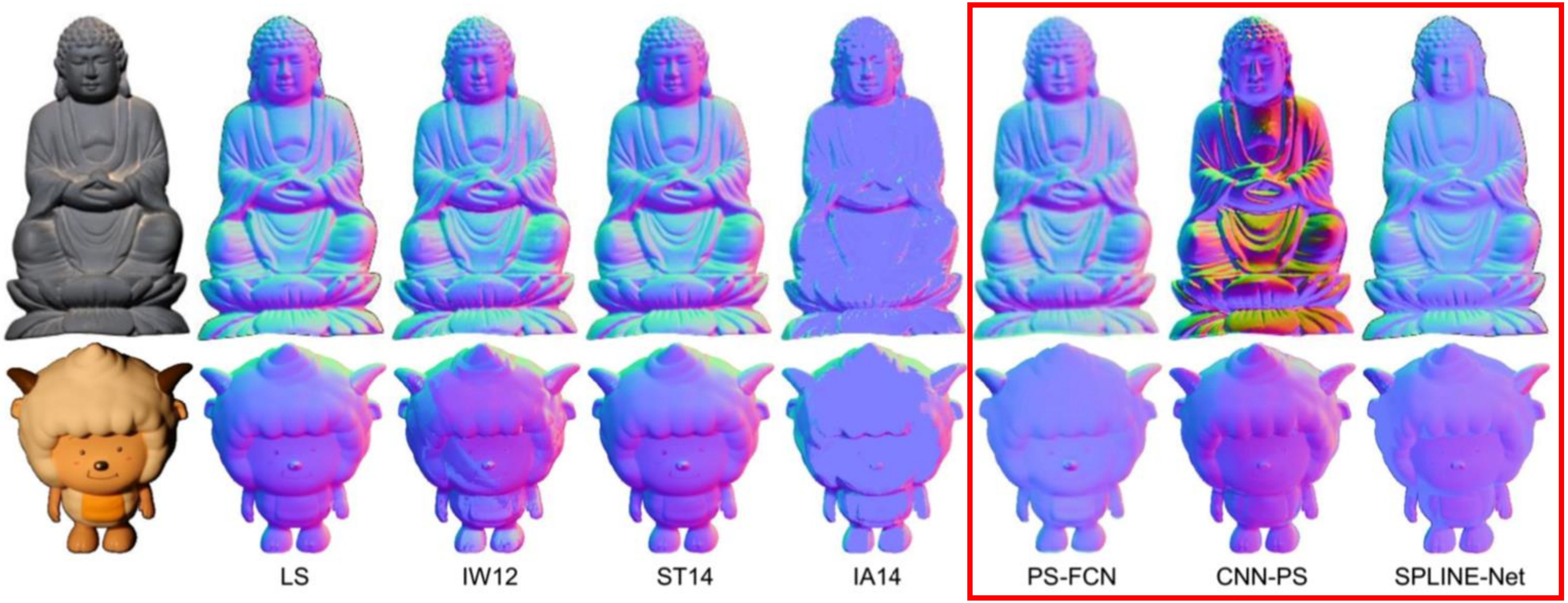
	PAPERBOWL		SPHERE		TURTLE		Avg.	PAPERBOWL		SPHERE		TURTLE		Avg.
	M	S	M	S	M	S		M	S	M	S	M	S	
LS	41.47	35.09	18.85	10.76	27.74	19.89	25.63	43.09	37.36	20.19	12.79	28.51	21.76	27.28
IW12	46.68	33.86	16.77	**2.23**	31.83	12.65	24.00	48.01	37.10	21.93	**3.19**	34.91	16.32	26.91
ST14	42.94	35.13	22.58	4.18	34.30	17.01	26.02	44.44	37.35	25.41	4.89	36.01	19.06	27.86
IA14	48.25	43.51	18.62	11.71	30.59	23.55	29.37	49.01	45.37	21.52	13.63	32.82	26.27	31.44
CNN-PS	37.14	23.40	17.44	6.99	22.86	10.74	19.76	38.45	26.90	18.25	9.04	23.91	14.36	21.82
SPLINE-Net	**29.87**	**18.65**	**6.59**	3.82	**15.07**	**7.85**	**13.64**	**33.99**	**23.15**	**9.21**	6.69	**17.35**	**12.01**	**17.07**

Benchmark results using "DiLiGenT"

***10 selected lights, 100 random trials**

Methods	Ball	Bear	Buddha	Cat	Cow	Goblet	Harvest	Pot1	Pot2	Reading	Avg.
LS	4.41	9.05	15.62	9.03	26.42	19.59	31.31	9.46	15.37	20.16	16.04
IW12	**3.33**	7.62	13.36	8.13	25.01	18.01	29.37	8.73	14.60	16.63	14.48
ST14	5.24	9.39	15.79	9.34	26.08	19.71	30.85	9.76	15.57	20.08	16.18
IA14	12.94	16.40	20.63	15.53	18.08	18.73	32.50	**6.28**	14.31	24.99	19.04
CNN-PS	17.86	13.08	19.25	15.67	19.28	21.56	21.52	16.95	18.52	21.30	18.50
Nets *w/o* loss	6.06	7.01	10.69	8.38	10.39	11.37	**19.02**	9.42	12.34	16.18	11.09
Nets with $\mathcal{L}^s$	5.04	**5.89**	10.11	7.79	9.38	10.84	19.03	8.91	**11.47**	**15.87**	10.43
SPLINE-Net	4.96	5.99	**10.07**	**7.52**	**8.80**	**10.43**	19.05	8.77	11.79	16.13	**10.35**

Open problems for data-driven methods



- When input light becomes sparse, data-driven methods does not outperform baseline (L2) for diffuse datasets

Open problems for dataset

- "DiLiGenT" only provides the "ground truth" of scanned shape
  - How to measure the true surface normal precisely
- For more delicate structures, a scanned shape to too "blurred" to evaluate photometric stereo
  - Integrating scanned shapes and photometric stereo for very high quality 3D modeling

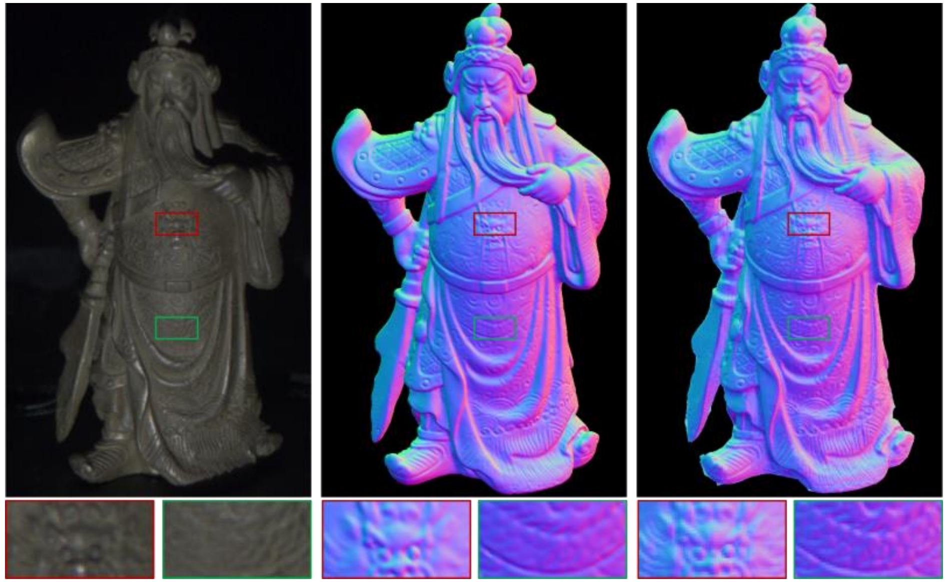
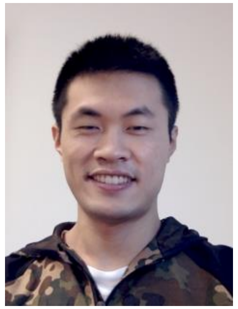


Image Scanned Photometric stereo

Acknowledgement

- Many slides are adopted from my collaborators:

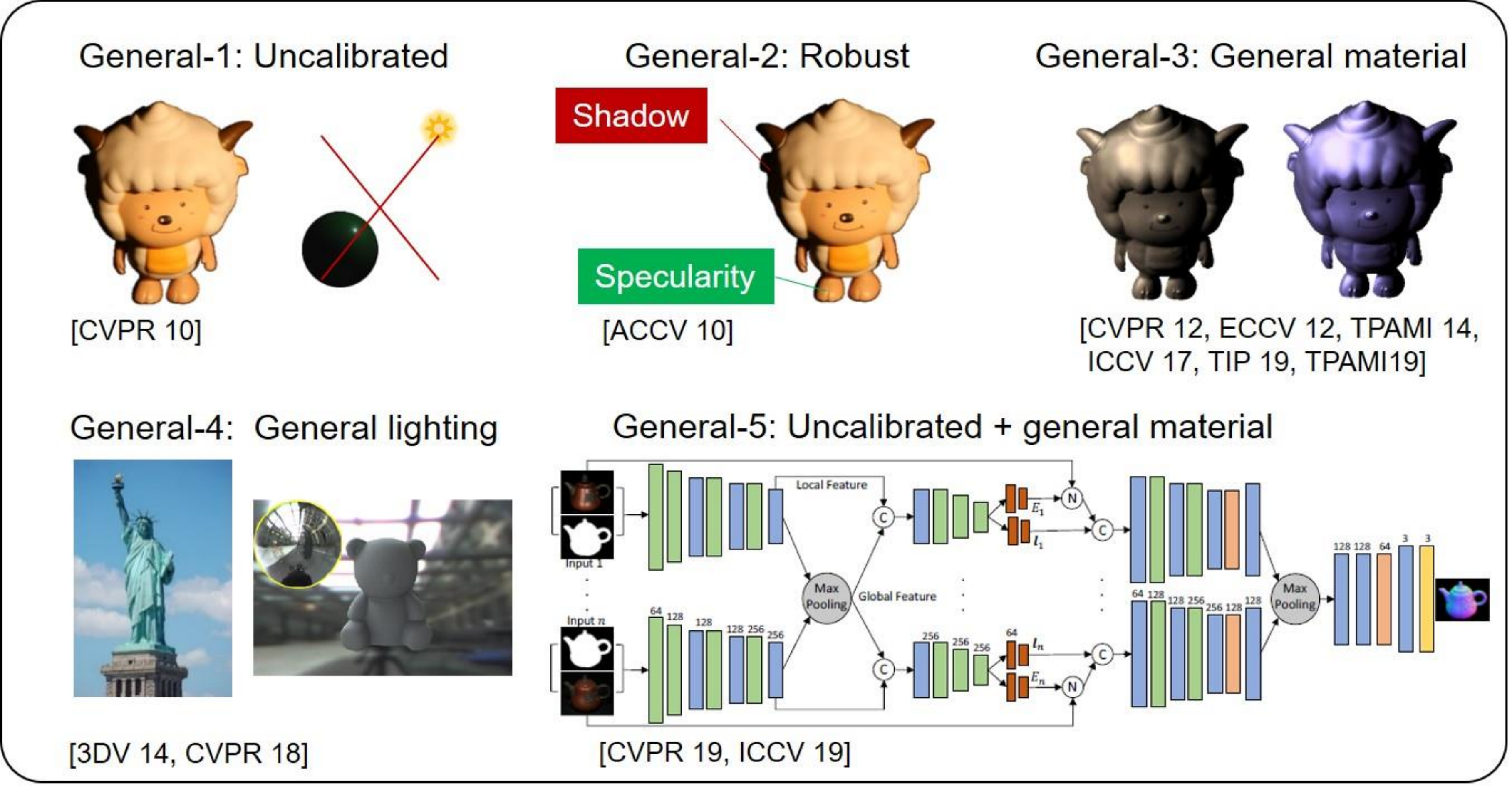
Yasuyuki Matsushita
Osaka University

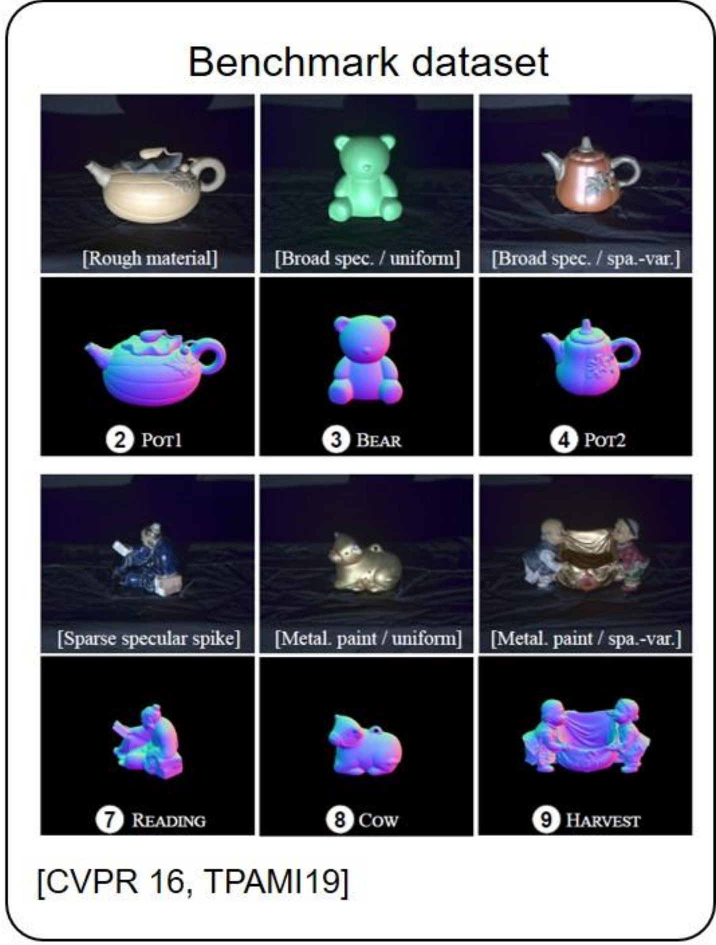


Qian Zheng
Nanyang Technological University

Hiroaki Santo
Osaka University

Guanying Chen
University of Hong Kong





Thank You!

Q&A

Boxin Shi (Peking University)

http://www.shiboxin.com / shiboxin@pku.edu.cn