LIME: Live Intrinsic Material Estimation

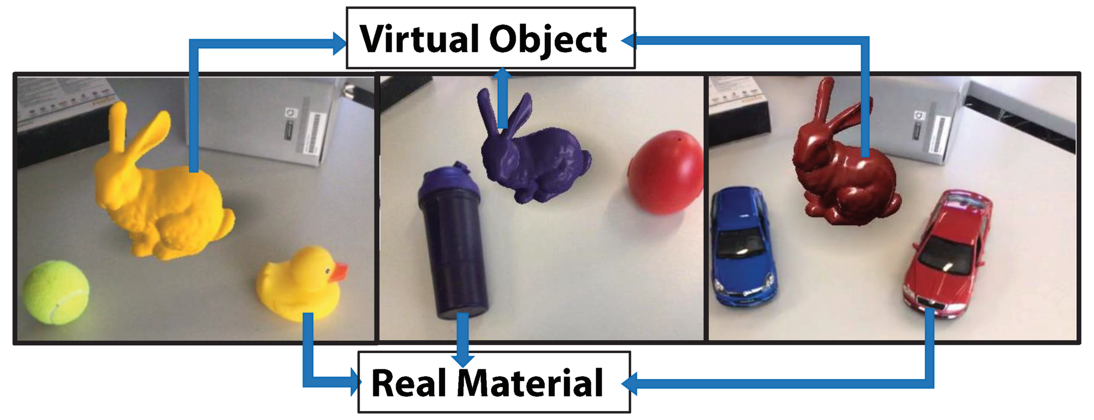
Abhimitra Meka[1,2] Maxim Maximov[1,2] Michael Zollhöfer[1,2,3] Avishek Chatterjee[1,2] Hans-Peter Seidel[1,2] Christian Richardt[4] Christian Theobalt[1,2]

[1]Max-Planck-Institute for Informatics [2]Saarland Informatics Campus [3]Stanford University [4]University of Bath

2018 CVPR SALT LAKE CITY • JUNE 18-22

1. Goal & Motivation

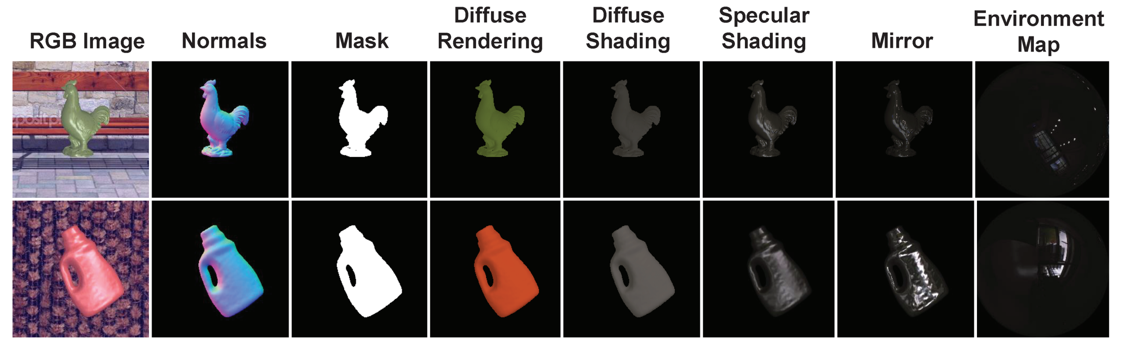
- **Estimating material (BRDF/reflectance) of general-shape objects from a single RGB image in real time**
- AR applications impose new contraints on the problem of material estimation
  - *Real-time*
  - *Uncontrolled lighting*
  - *In-the-wild setting*
  - *Limited sensor input*
  - *General shape objects*



Augmented Reality Application: Material Cloning

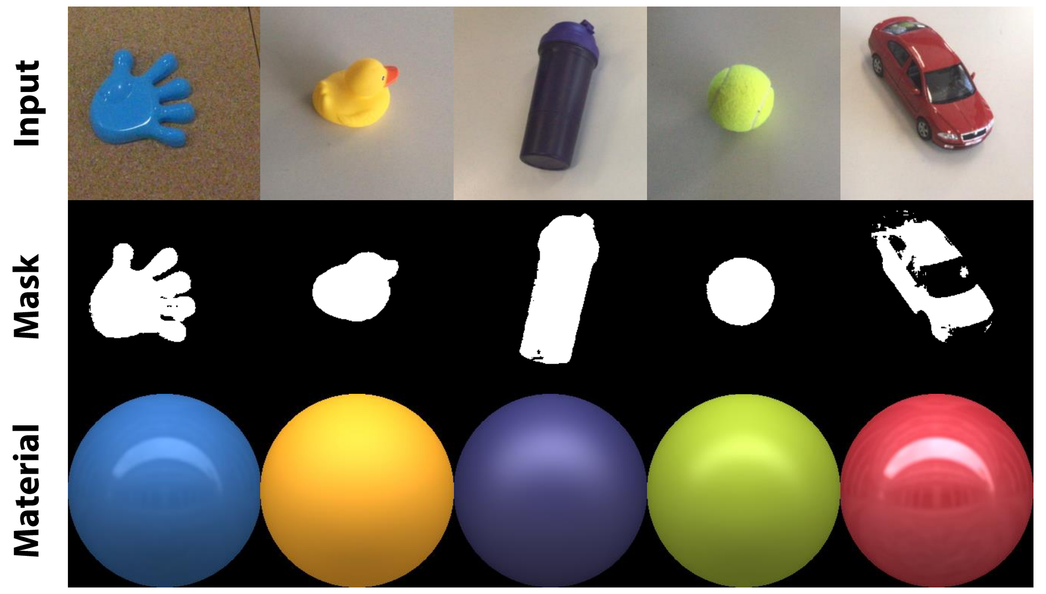
4. Training Data

- Synthetically rendered training dataset with 100,000 samples
- 55 synthetic 3D models augmented with random scaling, orientation and position

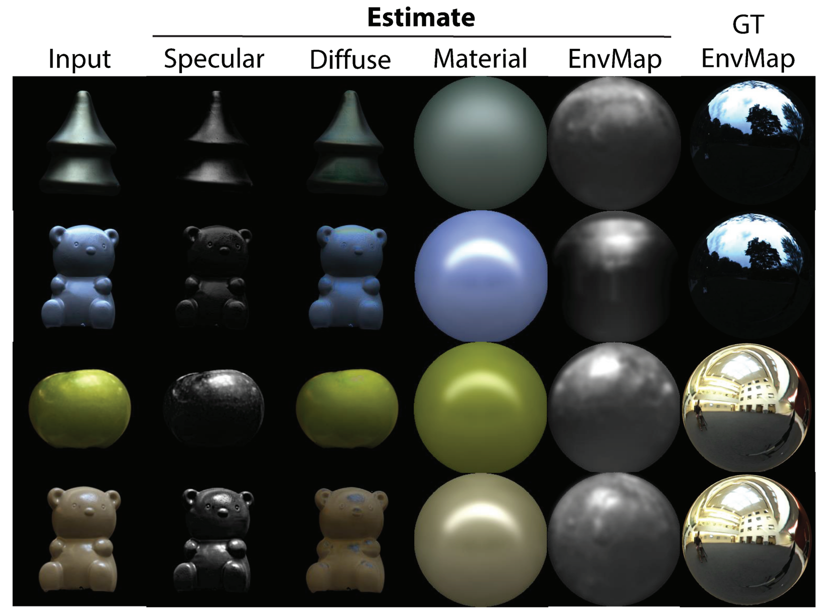


- Rendered with uniformly sampled albedo parameters from YUV color space
- 45 indoor environment maps captured in varying lighting conditions - homes, offices, classrooms, auditoriums – randomly rotated before rendering
- Input images augmented with gaussian noise and random background textures
- Entire training dataset available on our website – **gvv.mpi-inf.mpg.de/projects/LIME**

5. Results & Evaluation



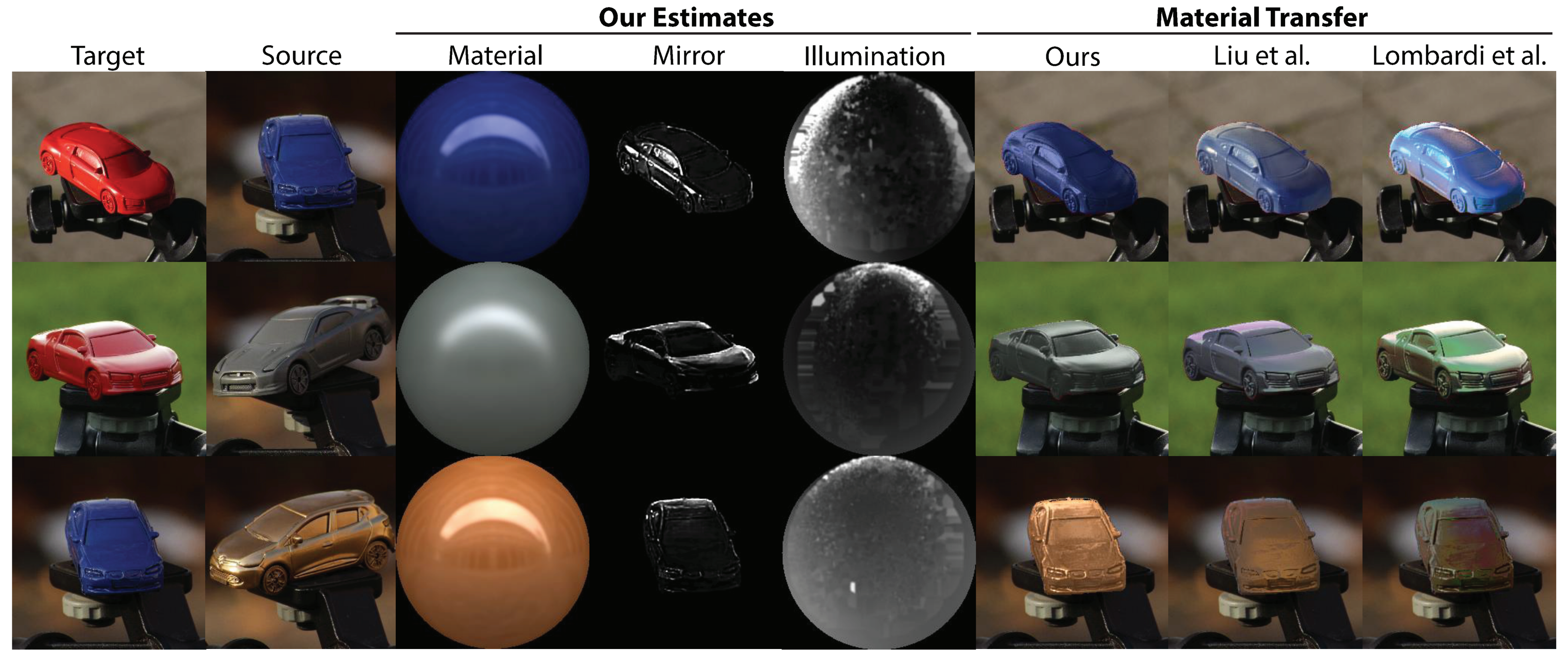
Object segmentation and material estimation results on objects of increasing specularity



Diffuse-Specular decomposition and Material & Environment Map estimation results

	Shininess Exponent (correct bin + adjacent bin)	Average Error: Diffuse Albedo	Average Error: Specular Albedo	Average Error: Shininess ($\log_{10}$)
Our full approach	45.07% + 40.12%	0.0674	0.2158	0.3073
without perceptual loss	45.15% + 40.96%	0.1406	0.2368	0.3038
without MirrorNet	36.29% + 40.28%	0.0759	0.2449	0.3913
with exponent regression (log10)	44.09% + 41.28%	0.0683	0.2723	0.2974
Reflectance Map Based Estimation	13.57% + 25.29%	0.0408	0.1758	0.7243

Qualitative Evaluation: In the ablation study our full method achieves best performance. In comparison to a reflectance-map-based approach , our albedo estimation is slightly worse but the shininess exponent is much more accurate.



Material retargeting results - We estimate materials from source objects and transfer them to target shapes with consistent lighting by estimating a high-frequency environment map for target shapes from our mirror image layer. We also compare against other state-of-the-art methods and obtain more accurate retargeting results.

[1] Liu et al., "Material Editing Using a Physically Based Rendering Network" , ICCV 2017
[2] Lombardi and Nishino., "Reflectance and illumination recovery in the wild", IEEE TPAMI 2016

2. Model

- We decompose the input image into its physical constituents, inspired by the rendering equation

$$L(x,\boldsymbol{\omega_0}) = \int_{\Omega} f(\boldsymbol{x},\boldsymbol{\omega_i},\boldsymbol{\omega_0})(\boldsymbol{\omega_i}.\boldsymbol{n})E(\boldsymbol{\omega_i})d\boldsymbol{\omega_i}$$

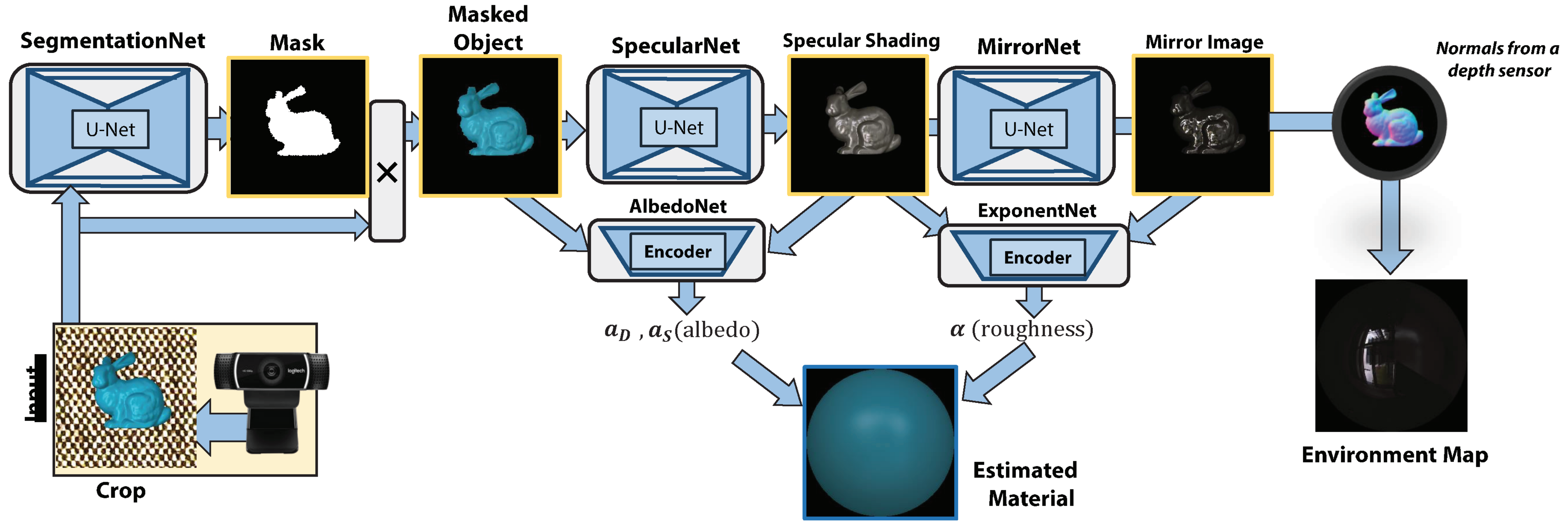
- Surface radiance (L) is given by the integral over the product of surface reflectance (f) and the incident illumination (E)
- The blinn-phong reflection model is used to simplify the rendering equation
- The incident lighting is parameterized by an environment map

$$\mathbf{I} = \underbrace{\mathbf{a_D} \bullet \left(\sum_{i\,\epsilon\,\Omega}[\hat{\mathbf{n}}.\hat{\boldsymbol{l}}_i]E_i\right)}_{\text{Diffuse Albedo x Diffuse Shading}} + \underbrace{\mathrm{a_S} \bullet \left(\sum_{i\,\epsilon\,\Omega}[\hat{\mathbf{n}}.\hat{\mathbf{h}}]^{\alpha}E_i\right)}_{\text{Specular Albedo x Specular Shading}}$$

Blinn-Phong Reflection Model

- Here '**I**' is the input RGB image, '**n**' is surface normal, 'l' is light direction, '**h**' is the Blinn-Phong half vector and 'α' is the specular shininess
- The Blinn-Phong model allows us to decompose the input image into linear components of diffuse and specular Albedos and Shadings

3. Pipeline



- We use 5 CNN's that are tailored to perform specific sub-tasks on the input image
- All the networks are trained jointly in an end-to-end fashion by combining a ground-truth loss with a ***novel perceptual rendering loss***:

$$\|\,(\mathbf{a_D}\cdot\text{Diffuse Shading} + \mathrm{a_S}\cdot\text{Specular Shading}) - \mathbf{I}\,\|_2^2$$

- Novel strategy to estimate the non-linear shininess exponent – our proposed 'mirror image' representation (perfectly specular version of the object), acts as an absolute reference for the specular shading image, allowing for a more accurate estimation
- When object normals are available from a depth sensor, the 'mirror image' can be unwrapped to also estimate the environment map

Graphics Vision & Video
erc European Research Council

This work was supported by EPSRC grant CAMERA (EP/M023281/1), ERC Starting Grant CapReal (335545) and the Max Planck Center for Visual Computing and Communication (MPC-VCC)