LIME: Live Intrinsic Material Estimation
— Supplemental Document —

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] MPI Informatics [2] Saarland Informatics Campus [3] Stanford University [4] University of Bath

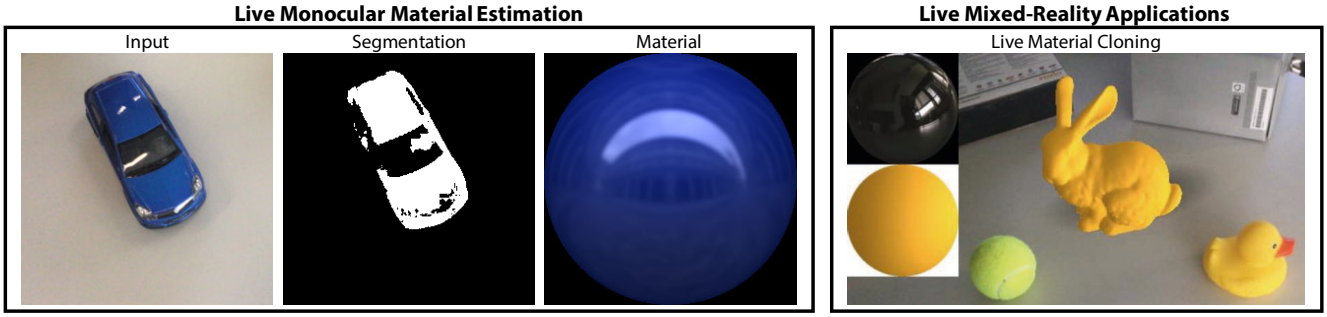

Figure 1. Our approach enables the real-time estimation of the material of general objects (left) from just a single monocular color image. This enables exciting live mixed-reality applications (right), such as for example cloning a real-world material onto a virtual object.

In this supplemental document, we show additional results, perform more evaluations, and we further justify the design choices made in our approach. More specifically, we show more results on real images (see Figure 3), more material transfer results (see Figure 4), more results on the data of Georgoulis et al. [2] (see Figure 6) and Lombardi and Nishino [4] (see Figure 7).

1. Shininess Classification versus Regression

As described in the main paper, our approach uses classification for recovering the specular exponent. For this classification task, we segmented the range of exponents by appearance into eight bins, which are illustrated in Figure 5. We opted for classification over regression, since regression was often overestimating the specular shininess of objects. As an example we show the material estimation results for a diffuse rubber duck in Figure 2. As is evident, the regression-based approach overestimates the specularity of the object, whereas classification is correctly able to estimate its diffuse nature.

We decided to train on synthetic training data and to sample the bins uniformly, instead of using a real-world dataset like MERL [5]. The MERL dataset consists of 100 materials. Model fits for various parametric BRDF models are available for these materials, including the Blinn–Phong model we use. The Blinn–Phong fit for these materials shows that more than 50 of the 100 material are highly specular and fall into the most specular bin in our classification system, or even higher. Sampling materials for our training data from the MERL dataset would result in our networks being heavily biased towards highly specular exponent estimation. Although a dataset such as MERL is very useful in studying how real materials span the large 4D space of BRDFs, it does not approximate well the distribution of BRDFs of everyday objects. For this reason, we chose to uniformly sample exponent values from a perceptually segmented classification system with 8 bins. For images of resolution 256×256, it is difficult to differentiate more levels of shininess, and using more bins does not provide much greater value perceptually.

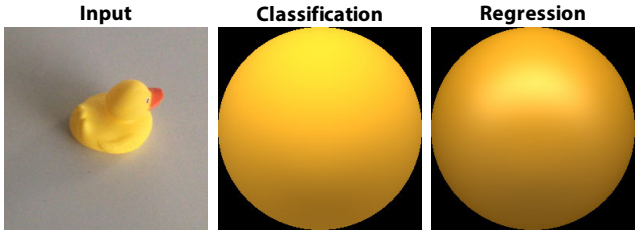

Figure 2. Regression often overestimates the shininess of objects. Therefore, we choose a classification-based approach.

2. Network Architecture Details

The network architecture of the five deep neural networks that make up our approach are detailed in Figures 8 to 11. SegmentationNet, SpecularNet and MirrorNet are based on the U-Net architecture [6], and AlbedoNet and ExponentNet are standard feed-forward networks for regression and classification, respectively.

References

[1] P. Debevec. Rendering synthetic objects into real scenes: bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 189–198, 1998.

[2] S. Georgoulis, K. Rematas, T. Ritschel, E. Gavves, M. Fritz, L. V. Gool, and T. Tuytelaars. Reflectance and natural illumination from single-material specular objects using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, preprints, 2017.

[3] G. Liu, D. Ceylan, E. Yumer, J. Yang, and J.-M. Lien. Material editing using a physically based rendering network. In *Proceedings of the International Conference on Computer Vision (ICCV)*, pages 2280–2288, 2017.

[4] S. Lombardi and K. Nishino. Reflectance and illumination recovery in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):129–141, 2016.

[5] W. Matusik, H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Transactions on Graphics (Proceedings of SIGGRAPH)*, 22(3):759–769, 2003.

[6] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 234–241, 2015.

Figure 3. Real-world material estimation results based on a single color input image. Our approach produces high-quality results for a large variety of objects and materials, without manual interaction. Note the high quality of the jointly computed binary segmentation masks.

Figure 4. Material estimation and transfer. Our method obtains convincing material estimates (second column) and material transfer results (column three to ten) on the data of Liu et al. [3]. We use the normal map estimated by Liu et al. [3] as basis for our environment map estimation.

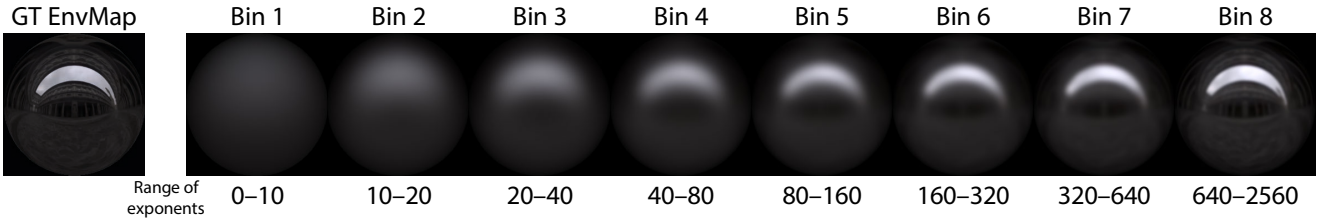

Figure 5. The 8 bins used for shininess exponent estimation in our approach, ranging from most diffuse (bin 1) to most shiny (bin 8). The visualization uses a material with a diffuse albedo of zero, a specular albedo of one, and shininess set to the mean value of each exponent bin. The materials are shown under the 'Uffizi' environment map Debevec [1].

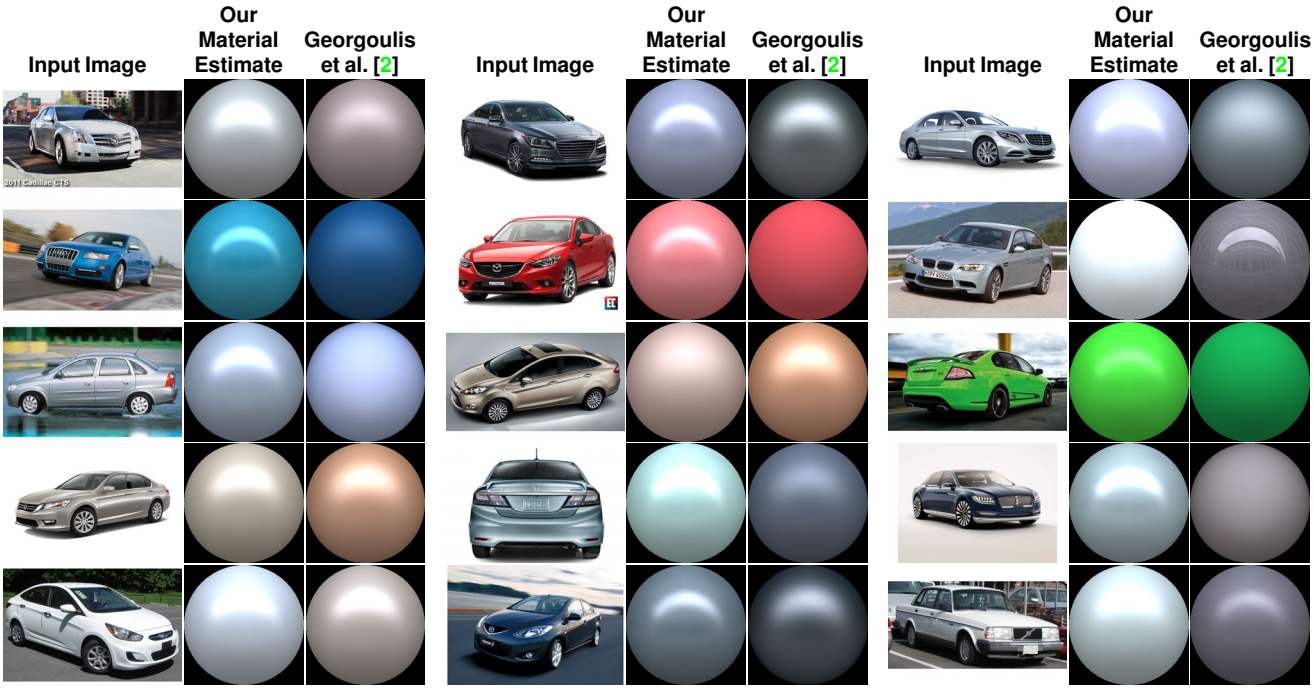

Figure 6. Comparison to the approach of Georgoulis et al. [2]. Note that their approach is specifically trained for the outdoor estimation scenario, while our approach is trained for the indoor setting. Nonetheless, our approach obtains results of similar or higher quality.

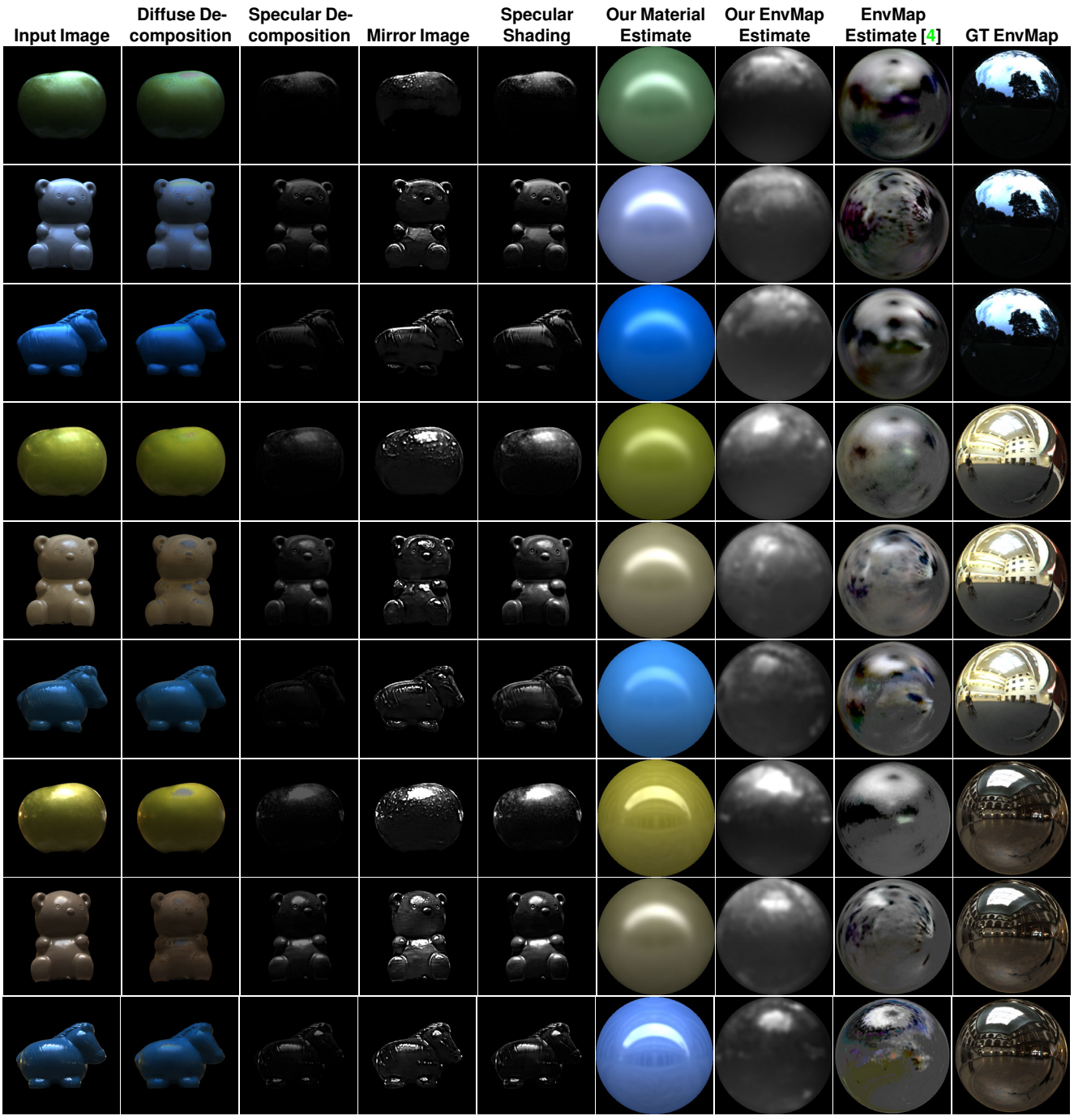

Figure 7. Our approach estimates the diffuse–specular decomposition from a single color image. The specular decomposition is obtained by multiplying specular albedo and specular shading layer, and the diffuse decomposition is obtained by subtracting the specular decomposition layer from the input image. Input images and ground-truth environment maps from Lombardi and Nishino [4]. We tone-mapped their images to process them with our approach.

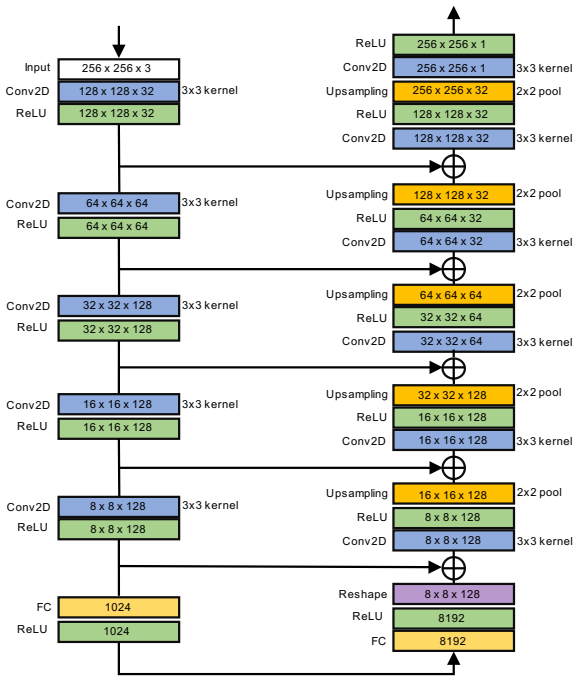

Figure 8. The architecture of our *SegmentationNet*, which learns a binary segmentation mask from a color input image. The numbers in each box denote *width*×*height*×*channels* of the layer's output, and a plus in circle represents concatenation of feature maps.

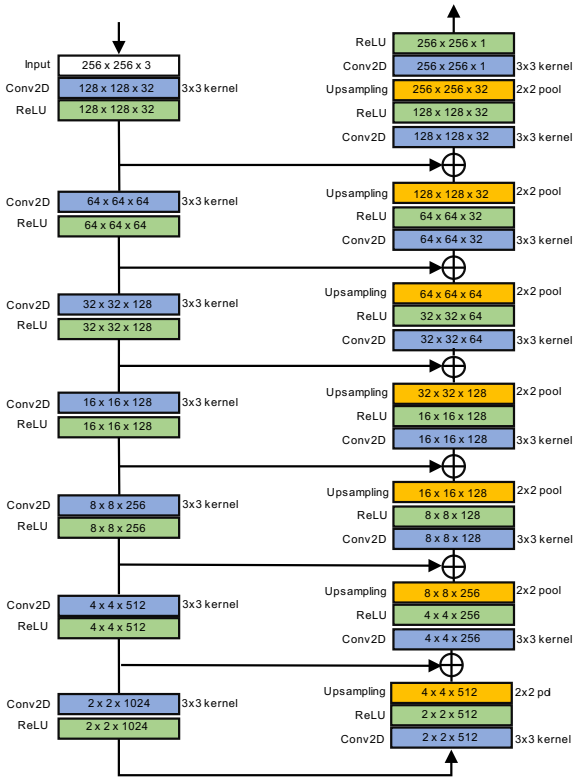

Figure 9. The architecture of our *SpecularNet*, which learns the grayscale specular decomposition from a masked color input image.

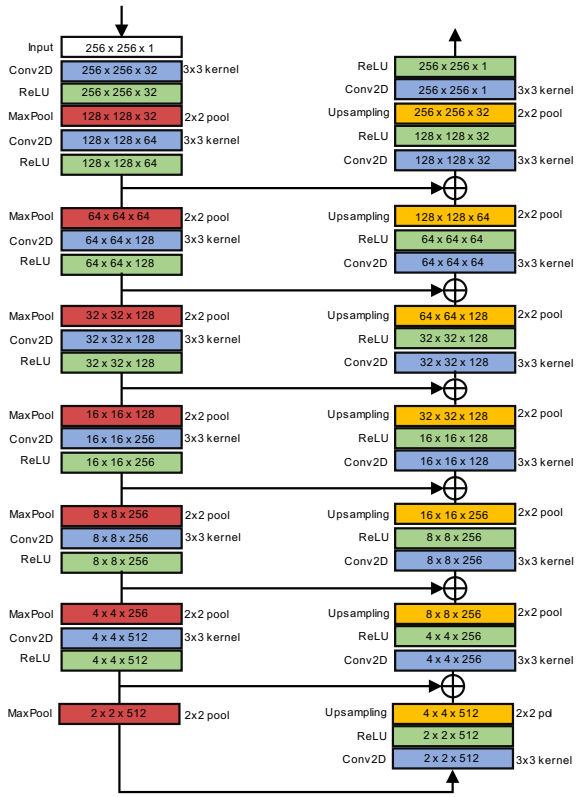

Figure 10. The architecture of our *MirrorNet*, which learns a grayscale mirror image from a grayscale specular shading image.

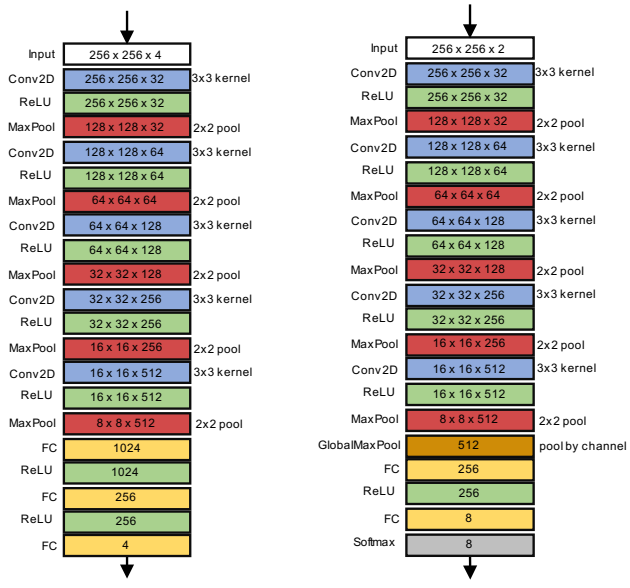

Figure 11. **Left:** The architecture of our *AlbedoNet*, which learns diffuse albedo in color (3 parameters) and grayscale specular albedo (1 parameter) from the masked color input image (3 color channels) concatenated with the grayscale specular image (1 color channel). **Right:** The architecture of our *ExponentNet*, which learns the shininess exponent using classification into 8 bins from the concatenation of the specular and mirror images (both grayscale).