

1      **Image editing using intrinsic image decomposition into shading and albedo.**

2  
3      RICHA MISHRA, Carnegie Mellon University, USA



13  
14      Fig. 1. From left: the input images, the predicted shading, predicted albedo, Recolored image where the green paprika was changed  
15      to cyan.

16      Intrinsic decomposition of an image is the process of separating an image into albedo and shading. These decompositions can  
17      further be used for image editing tasks like recoloring, relighting, retexturing. In this work, we explore evaluate the recent work by  
18      [3]. Their method uses a data-driven approach to decompose an image into shading and albedo. We will discuss their formulation and  
19      test it in images to identify the merits and shortcomings of the method and suggest a few ideas to overcome them.

20  
21      CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your  
22      Paper*; *Generate the Correct Terms for Your Paper*; *Generate the Correct Terms for Your Paper*.

23  
24      Additional Key Words and Phrases: Computational photography, Intrinsic decomposition, Image editing

25  
26      **ACM Reference Format:**

27      Richa Mishra. 2018. Image editing using intrinsic image decomposition into shading and albedo.. In *Woodstock '18: ACM Symposium on  
28      Neural Gaze Detection, June 03–05, 2018, Woodstock, NY*. ACM, New York, NY, USA, 8 pages. <https://doi.org/XXXXXX.XXXXXXX>

29  
30      **1 INTRODUCTION**

31      Intrinsic decomposition is an important computer vision problem that can be used for many downstream computational  
32      photography pipelines. It attempts to separate the illumination-based effects from fundamental color values. This  
33      process involves breaking down an observed image into its fundamental constituents: albedo, representing the inherent  
34      reflectance properties of surfaces, and shading, which accounts for the effects of lighting and surface orientation. The  
35      goal of intrinsic decomposition is to separate these two factors to gain insights into the underlying scene structure and  
36      lighting conditions. Intrinsic decomposition is a fundamental step in various computer vision applications, including  
37      object recognition, scene understanding, and 3D reconstruction. By isolating albedo and shading, computer vision  
38      systems can focus on the intrinsic properties of objects and scenes, facilitating better recognition and interpretation.

39  
40  
41  
42  
43      
$$I = A * S \quad (1)$$

44  
45      Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not  
46      made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components  
47      of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on  
48      servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

49      © 2018 Copyright held by the owner/author(s). Publication rights licensed to ACM.

50      Manuscript submitted to ACM

53 However, this task is non-trivial. As we can see from Equation 1, this problem is inherently unconstrained. This  
 54 ambiguity is caused by variations in lighting conditions, shadows, and specular reflections. Different surfaces may  
 55 have similar shading under certain lighting conditions, making it difficult to uniquely determine albedo and shading  
 56 components. Moreover, this problem is also scale invariant as we can see from Equation 1. For a given  $A$  and  $S$ , there  
 57 exists  $\frac{1}{c}A$  and  $cS$  which would also satisfy Equation 1.  
 58

59 Traditional methods proposed to do intrinsic decomposition often leverage assumptions about scene properties  
 60 and lighting conditions and lighting directions to separate albedo and shading. To get this decomposition from  
 61 in-the-wild images we can use data-driven methods. However it is hard to formulate this problem to leverage neural  
 62 networks. The main challenge comes from predicting a reliable shading from images accurately. Shading prediction is  
 63 inherently complex due to the intricate interplay between scene geometry, surface materials, and lighting conditions.  
 64 The intrinsic decomposition of an image into albedo and shading is an ill-posed problem. There can be multiple  
 65 plausible decomposition for a given image, making it challenging for data-driven methods to learn a unique mapping.  
 66 Ambiguities arise from factors such as specular reflections, shadows, and global illumination effects, which may not be  
 67 well-represented in the training data. We will explore a recent work [3], in which the authors have formulated this  
 68 problem to overcome these problems and predict reliable shading. We will also look at the cases where the methods  
 69 fails.  
 70

## 73 2 RELATED WORKS

74 Traditional methods focus on developing low-level priors for shading and albedo. But with increased availability of  
 75 training data, the focus of the field first shifted to sparse ordinal representations and then to direct regression of the  
 76 continuous-valued shading. We will briefly discuss prior works on ordinal shading and data-driven approaches.  
 77

### 78 2.1 Ordinal shading

79 The first large-scale dataset came with the sparse ordinal annotations on real-world images called Intrinsic Images  
 80 in-the-Wild (IIW) [1]. This lead to multiple works focusing on using the given annotations to train a network to predict  
 81 ordinal relationships between albedo values of pixel pairs. The sparse representation can then be used to estimate the  
 82 dense ordinal values [9]. However, in the work we use [3], the authors used the ordinal shading formulation but did not  
 83 regress dense shading values from sparse values. They leveraged both global and local image information to directly  
 84 estimate the dense decomposition. This is done in two steps. To generate the global constraints they resize the image to  
 85 fit the receptive field of the network and for local constraints they estimate the shading at higher resolution.  
 86

### 87 2.2 Data-driven approaches

88 The main bottleneck for any data-driven method is the availability of reliable ground-truth data. With faster and more  
 89 efficient physically-based rendering techniques available, it has become feasible to generate large-scale dataset and train  
 90 intrinsic decomposition networks. ShapeNet [4], MPI Sintel [2] and MPI Dataset [5] are few examples of large scale  
 91 datasets collected and used widely to train networks to do intrinsic decomposition. These datasets can be directly used  
 92 to supervise the estimation of shading and albedo. However, direct supervision has limitations for this task. Shading is  
 93 heavily influenced by the complex nature of lighting in a scene, including direct and indirect illumination, varying  
 94 light sources, and shadows. Predicting shading accurately requires a model to capture the nuances of different lighting  
 95 scenarios, which can be a daunting task for data-driven methods, especially when the training data does not sufficiently  
 96 cover the diversity of real-world lighting conditions.  
 97

Several other methods use two networks to predict shading and albedo separately and use a reconstruction loss. However, these methods do not incorporate the scale-invariant nature of the problem and often do not perform well on novel scenes which hinders their usage in downstream image editing tasks. The authors of [3] calculate the albedo from the predicted shading and the groundtruth input image.

### 2.3 Rendered vs real-world dataset

As already discussed, due to improvement in physically-based rendering techniques, it is possible to render large-scale datasets for directly supervising the intrinsic decomposition. However, training on rendered datasets usually leave a domain gap between training and in-the-wild photographs. One way to overcome these issues is to collect image sequences in varying illumination to enforce the reluctance consistency across multiple illuminations. In this work, the authors generated pseudo-ground-truth intrinsic components from multi-illumination data using Multi Illuminations Dataset [7].

## 3 METHOD

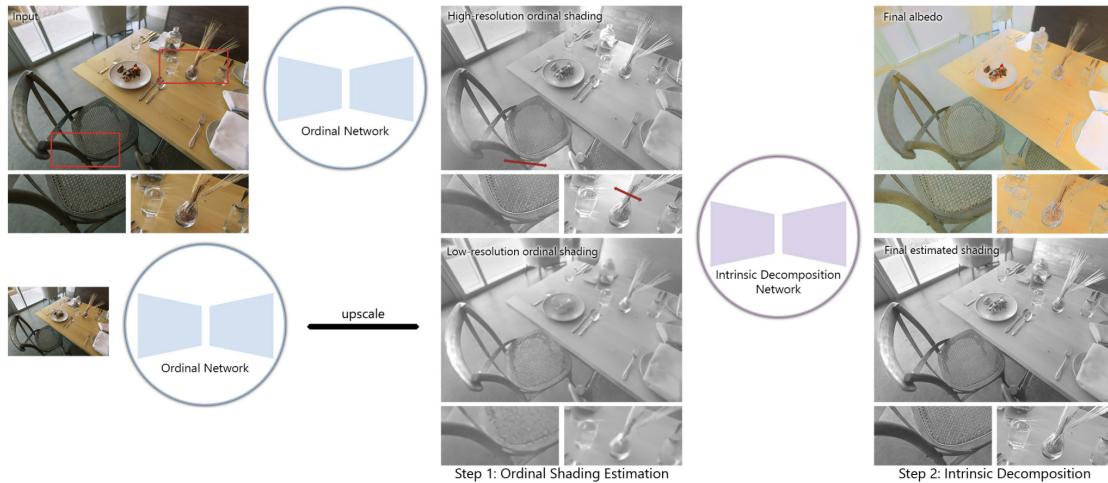


Fig. 2. The method for [3]. The ordinal network predicts the shading in low and high resolution. This is then used as an input to the intrinsic decomposition network. The image is taken directly from the paper.

The method used in [3] uses local as well as global constraints to achieve accurate shading values. In this method, the authors define a relaxed loss that enforces correct ordering of shading values. This relaxation incorporates the inherent scale-invariant nature of the problem. By ensuring that the scale values are always positive the ordering remains the same and we have two main observations:

- Under the relaxed formulation, at low resolutions, the predictions are globally coherent, but they lack details.
- At high resolutions, the estimated shading values have global inconsistencies but they have detailed local shading discontinuities. This is because the image resolution is bigger than the receptive field of the network.

Therefore the low-resolution image has information about the global ordering of the shading values of the entire image. The high-resolution image has detailed discontinuities in the local neighbourhood of the pixel. They use both the

157 images and feed it to another network to generate the final shading. This network can then predict highly detailed,  
 158 consistent shading values. We will define a few key steps in detail.  
 159

### 160 3.1 Inverse shading representation

161 Estimating shading using neural networks is not straightforward. The challenge arises from the fact that the illumination  
 162 can take on a large set of values due to specular objects present in the scene. This results in a long-tailed distribution of  
 163 values. One solution to this problem is to formulate the shading in the logarithmic scale. This prevents the long-tailed  
 164 distribution and is more uniform. But the logarithmic domain lacks contrast and a well-defined range. Using inverse  
 165 shading solves these two issues. The contrast is preserved and the shading range is between [0, 1] :  
 166

$$167 D = \frac{1}{S + 1} \quad (2)$$

168 where S is the shading value in linear scale. Moreover, since it ranges between [0, 1] ;, it is easier for a neural network  
 169 to be trained on this well-defined range. The inverse shading formulation also preserves the ordinal relationships in the  
 170 shading domain.  $D_i < D_j$  for  $S_i > S_j$  for all pixel pairs (i, j).  
 171

### 172 3.2 Ordinal shading estimation

173 We have to incorporate the scale-invariant nature of the Equation 1 in our model. Since it is hard to directly estimate  
 174 dense shading values, the authors used a relaxed formulation of direct shading estimation. This means that the network  
 175 is not tasked to predict the final shading values. Instead the network is only tasked to learn the relative ordering between  
 176 the shading values. This is induced by ensuring that  $f(S)$  where  $f(\cdot)$  is a monotonically increasing function is also a  
 177 valid solution.  
 178

179 We define  $D \in [0, 1]$ , which preserves the ordinality in shading. The relaxed loss function is formulated as follows:  
 180

$$181 L_{ord} = \frac{1}{N} \sum_i^N (f(O_i) - D_i^*)^2 \quad (3)$$

182 where  $O$  is estimated ordinal shading,  $D^*$  is the groundtruth inverse shading and  $f(\cdot)$  is a monotonically increasing  
 183 function defined as  
 184

$$185 f(x) = ax + b \quad (4)$$

$$186 (a, b) = argmin \sum_i (f(O_i) - D_i^*)^2, a > 0 \quad (5)$$

187 Under this formulation, we allow the network to predict shifted and scaled versions of the shading as it preserves the  
 188 relative ordering. The ordinal shading value  $O$  is not the global value. It is an unknown scale and shift away from the  
 189 ground-truth. The intrinsic model in Equation 1 is not satisfied here. This network learns to correctly order the shading  
 190 values for each pixel. This simplification allows the network to generate ordinal results with more high frequency  
 191 details when compared to direct shading estimation.  
 192

### 193 3.3 Ordinal Shading Network

194 The method uses ResNet101 encoder-decoder structure. The activation layers are replaced from ReLU to sigmoid to  
 195 bound in between [0, 1]. The network is briefly described in the figure 2.  
 196

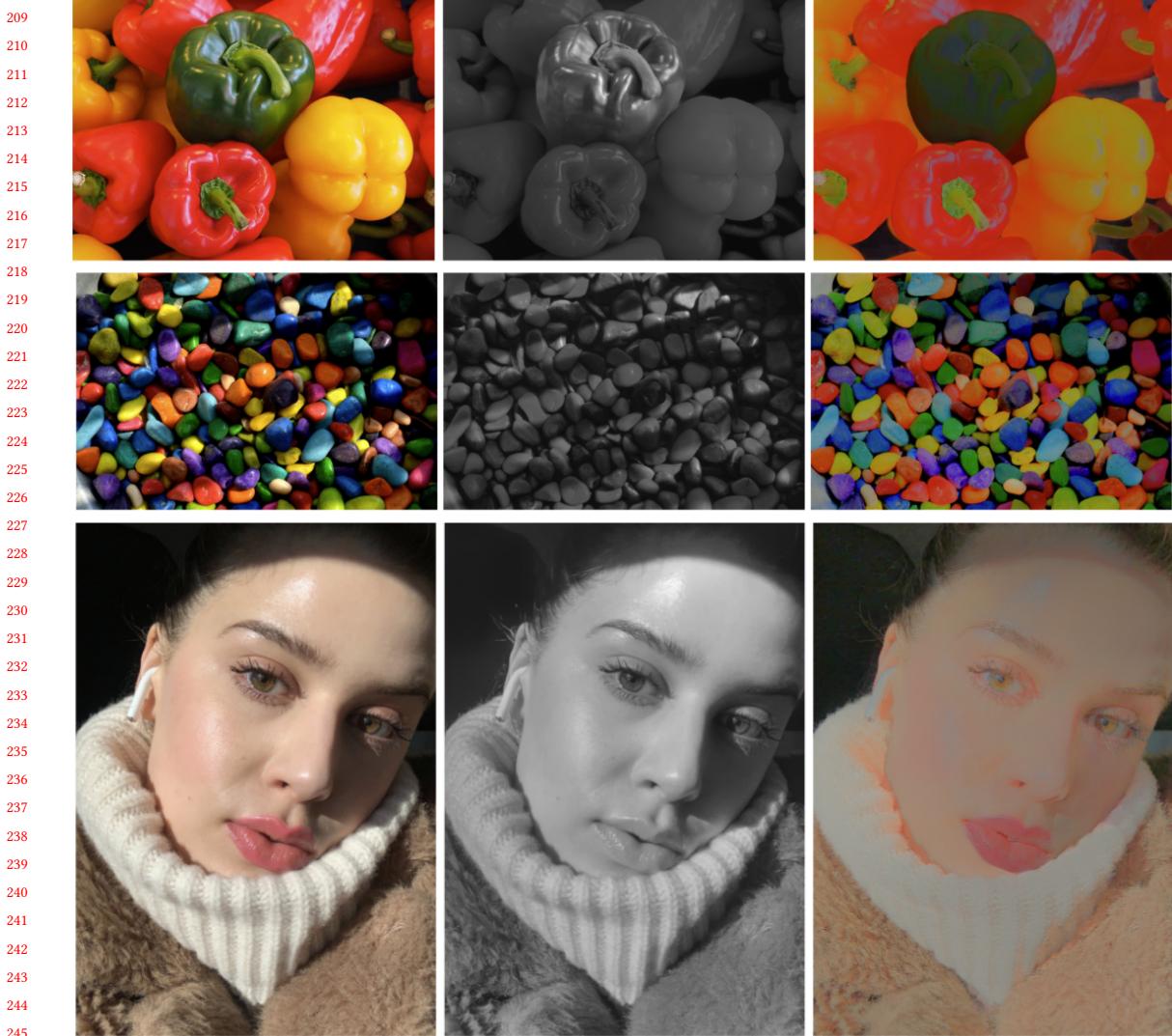


Fig. 3. Examples of the results from [3]. Left to right: Input image, shading and albedo. Images from [8]

### 3.4 Intrinsic Decomposition Network

The authors define a two-step approach to solve intrinsic decomposition problem. the first ordinal estimation, $O_L$  is generated at the receptive field resolution of the Ordinal shading network. Since it is at the receptive field resolution, there is global coherency. The ordinal shading network is able to provide accurate ordering of the shading values across the entire image.  $O_L$  provide a good starting point point for the final dense shading estimation. The final prediction of  $O_H$  is generated at a higher resolution. This lacks global coherency but contains local detailed discontinuities which provide the high frequency details to the second intrinsic decomposition network. The task of our intrinsic decomposition network is simplified with these three inputs. The network does not have to reason about the global



Fig. 4. 1907 Franklin Model D roadster. Photograph by Harris & Ewing, Inc. [Public domain], via Wikimedia Commons. (<https://goo.gl/VLCRBB>).

structure like illumination direction and geometry as it comes from  $O_L$ . The network does not have to reason whether a sharp change in RGB near edges is a change in shading or albedo as this information is provided in the form of  $O_H$ . The network simply has to adjust the values of  $O_L$  such that it satisfies intrinsic model of Equation 1 while also including the details from  $O_H$

### 3.5 Output formulation

The final output of the intrinsic decomposition network is a single channel inverse shading value. We then compute the shading and albedo as follows

$$S = \frac{1 - D}{D} \quad (6)$$

$$A = \frac{I}{S} = \frac{I * D}{1 - D} \quad (7)$$

313 where I and D represent the input image and estimated inverse shading respectively  
 314

### 315 3.6 Scale Invariance

316 The globally consistent  $O_L$  provides a point of reference for global scaling. The authors use a least squares fit between  
 317 the ground-truth and predicted shading. They use the low-resolution input to set the arbitrary scale in the ground-truth:  
 318

$$320 \quad c = \operatorname{argmin} \sum_i (xA^{**} - \hat{A}_L)^2, \quad (8)$$

$$322 \quad \hat{A}_L = \frac{I}{\hat{S}_L} \quad (9)$$

$$324 \quad \hat{S}_L = \frac{1 - O_L}{O_L} \quad (10)$$

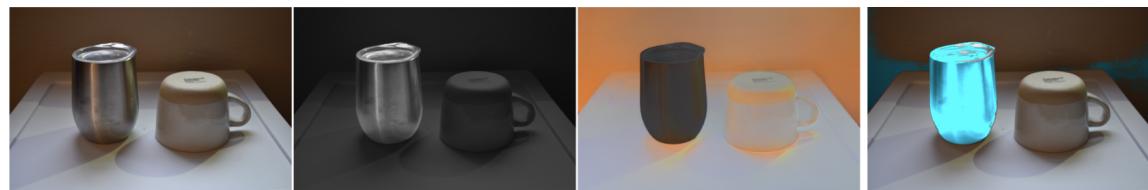
326 where  $A^{**}$  represents the ground-truth albedo at arbitrary scale. They then use a fixed scale to define the ground-truth  
 327 albedo and shading.  
 328

$$330 \quad A^* = c * A^{**} \quad (11)$$

$$333 \quad S^* = \frac{I}{A^*} \quad (12)$$

$$336 \quad D^* = \frac{1}{S^* + 1} \quad (13)$$

## 338 4 EXPERIMENTS



348 Fig. 5. Input image, shading, albedo, recolored image  
 349

351 We used the above pipeline on several images. Some results of the intrinsic decomposition is shown in figure 3. As  
 352 we can see from the figure, the method is able to produce decent albedo and shading from the images. However, the  
 353 results lack in the regions where there is high specularity and shadows. We can see from the figure 3 that there is  
 354 some issue where there are a lot of shadows and where the surface is highly glossy I decide to use my own scene with  
 355 highly specular metallic surfaces to check the limits of the method. I wanted to check for two main areas of interest.  
 356 Specular surfaces and shadows The results are shown in figure 6. I also tried to recolor the albedo and generate the  
 357 recolored images by generating a mask using Segment Anything Mask [6]. One example shown in 1 and another in 5.  
 358 We observe that the specular images and metallic surface are not decomposed effectively. This could be due to the fact  
 359 the dataset used to train the network has not seen this material. Or the network does not perform well with highly  
 360 specular images.  
 361



Fig. 6. Albedo and shading estimated by

## 5 FUTURE WORK

I want to introduce another step in the training network that first removes the specular effects from the image before ordinal shading. I have been following [?] to change the model as shown in figure 2 to remove specular effects. This should improve the shading and albedo prediction and we can add the specular mask again after recoloring.

## REFERENCES

- [1] Sean Bell, Kavita Bala, and Noah Snavely. 2014. Intrinsic Images in the Wild. *ACM Trans. on Graphics (SIGGRAPH)* 33, 4 (2014).
- [2] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. 2012. A naturalistic open source movie for optical flow evaluation. In *European Conf. on Computer Vision (ECCV) (Part IV, LNCS 7577)*, A. Fitzgibbon et al. (Eds.) (Ed.). Springer-Verlag, 611–625.
- [3] Chris Careaga and Yağız Aksoy. 2023. Intrinsic Image Decomposition via Ordinal Shading. *ACM Trans. Graph.* (2023).
- [4] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. 2015. *ShapeNet: An Information-Rich 3D Model Repository*. Technical Report arXiv:1512.03012 [cs.GR]. Stanford University – Princeton University – Toyota Technological Institute at Chicago.
- [5] Roger Grosse, Micah K. Johnson, Edward H. Adelson, and William T. Freeman. 2009. Ground-truth dataset and baseline evaluations for intrinsic image algorithms. In *International Conference on Computer Vision*. 2335–2342. <https://doi.org/10.1109/ICCV.2009.5459428>
- [6] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. 2023. Segment Anything. arXiv:2304.02643 [cs.CV]
- [7] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. 2019. A Dataset of Multi-Illumination Images in the Wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.
- [8] Sumit Shekhar, Max Reimann, Maximilian Mayer, Amir Semmo, Sebastian Pasewaldt, Jürgen Döllner, and Matthias Trapp. 2021. Interactive Photo Editing on Smartphones via Intrinsic Decomposition. *Computer Graphics Forum* 40, 2 (2021).
- [9] Tinghui Zhou, Philipp Krähenbühl, and Alexei A. Efros. 2015. Learning Data-driven Reflectance Priors for Intrinsic Image Decomposition. arXiv:1510.02413 [cs.CV]