https://doi.org/10.2352/CIC.2022.30.1.39

An intrinsic image network with properties of human lightness perception

Richard F. Murray[1]***, David H. Brainard***[2]***, Jaykishan Y. Patel***[1]***, Ethan Weiss***[1]***, and Khushbu Y. Patel***[1]
[1] Department of Psychology and Centre for Vision Research, York University
[2] Department of Psychology, University of Pennsylvania

Abstract

Research on human lightness perception has revealed important principles of how we perceive achromatic surface color, but has resulted in few image-computable models. Here we examine the performance of a recent artificial neural network architecture in a lightness matching task. We find similarities between the network's behaviour and human perception. The network has human-like levels of partial lightness constancy, and its log reflectance matches are an approximately linear function of log illuminance, as is the case with human observers. We also find that previous computational models of lightness perception have much weaker lightness constancy than is typical of human observers. We briefly discuss some challenges and possible future directions for using artificial neural networks as a starting point for models of human lightness perception.

Introduction

Visual perception emerged through evolution because it helps us to survive. For survival, it is often useful to perceive properties of creatures and things in the environment, rather than properties of our retinal images of them. Vision enables us to perceive the distance, shape, and identity of objects, among many other characteristics. One fundamental property is *surface color*, which is determined by the proportion of light that a surface reflects at various wavelengths, regardless of what spectrum of light happens to be illuminating it. Surface color is a multi-dimensional property, and vision researchers often study perception of achromatic surfaces viewed under achromatic illumination as a simplified problem that nevertheless retains several of the features that make color perception difficult and interesting. *Reflectance* is a scalar property that represents the proportion of incident light reflected by a surface. Black surfaces have reflectances near zero, white surfaces have reflectances near one, and grey surfaces have values in between. Lightness has different definitions in different research literatures, and in the present work we follow the convention that lightness is the perceived reflectance of achromatic surfaces seen under achromatic illumination [1]. Thus for our purposes, lightness perception is the visual perception of black, white, and grey surface color.

Lightness perception has been an active research area since the beginnings of experimental psychology, and many studies have shown the importance of features such as shadow boundaries and transparency cues for lightness [1]. Mathematical and computational models of lightness perception in realistic scenes have been more difficult to formulate, though, and there are few such models.

Recently, work on artificial neural networks (ANNs) has made progress on estimating surface reflectance in realistic scenes [2, 3, 4]. These networks were not developed as models of human vision. However, human lightness perception often relies on implicit knowledge about natural scenes to make rational inferences from retinal images [5, 6, 7]. This suggests that ANN approaches to reflectance estimation may be useful starting points for normative models of human lightness perception, as ANNs also exploit regularities in natural scenes to overcome the intrinsic ambiguity of 2D images.

In the computer vision literature, ANNs that estimate reflectance are often evaluated using summary statistics, such as the mean-squared error of reflectance estimates, averaged over pixels and images. This is a useful way of quantifying network performance, but it provides little insight into how similar networks are to human vision. Here we examine the performance of a recent ANN architecture [4] on a reflectance estimation task, and in addition to summary statistics of performance, we also report how the network behaves in a lightness matching experiment of a kind that has often been used to study human lightness perception. To provide context for our results, we use the same approach to evaluate other relevant models [8, 9, 10].

Related work

Land and McCann's [11] retinex model is one of the few image-computable models of human lightness perception. It estimates reflectance by integrating the derivative of luminance along paths through an image, discarding small derivatives (assumed to be due to illumination gradients) and retaining large derivatives (assumed to be due to reflectance edges). Retinex is an impressively long-lived model, and it has been developed further in more recent variants [12]. It does have limitations, such as assuming that lighting changes only gradually throughout a scene.

There are many image-computable models of *brightness*, defined as perceived luminance [13]. In lightness research, brightness models are sometimes evaluated as well [14, 6], given the scarcity of image-computable models of lightness, and the weakness of our current understanding of the relationship between lightness and brightness. Here we consider one such model, the oriented difference of Gaussians (ODOG) model, which operates by normalizing contrast energy in orientation and spatial frequency passbands [10, 15],

Dakin and Bex [9] developed an image-computable lightness model that normalizes the Fourier power spectrum of a stimulus image to match the $1/f^{\alpha}$ spectrum that is typical of natural images. There are many spatial filtering models of brightness, and Dakin and Bex's model is an interesting case where filtering is the basis of a lightness model (though also see [16]). Dakin and Bex

equivocate on whether their model predicts lightness or brightness, and here we evaluate it as a model of lightness.

Reflectance estimation is a classic problem in computer vision, where it is part of the more general problem of 'intrinsic image decomposition', in which reflectance, shading, and sometimes shape are estimated from a single image. For example, Barron and Malik [17] developed a Bayesian algorithm for estimating surface color, shape, and lighting, guided by statistical knowledge about natural scenes. Recently there have been several ANN approaches to intrinsic image decomposition. The network we use below is based on Yu and Smith's InverseRenderNet [4], which used simple priors and self-supervised learning to train a network using web-crawled images. The network estimated reflectance, surface normals, and global lighting. Interestingly, this network first estimated reflectance and local surface orientation, and then used these estimates to infer global lighting; this is the opposite of the order in many models of human vision [5]. Other ANN approaches to reflectance estimation include [2, 3]. Flachot et al. [18] have evaluated ANNs as models of color constancy, and Storrs et al. [19] have examined similar approaches to material perception.

Methods

Network architecture. We used PyTorch to implement the first stage of InverseRenderNet, which maps a luminance image to an equally sized color image [4]. Instead of an $n \times n \times 3$ output layer representing surface color, we used an $n \times n \times 1$ layer representing achromatic reflectance . The resulting network was a 30-layer convolutional neural network (CNN) in an hourglass architecture with skip connections. To indicate that the network is derived from but also different from InverseRenderNet, we call it IRNet.

We adapted this 30-layer network from InverseRenderNet, and it is an interesting question whether so many layers are necessary. In future work, we will explore alternative architectures.

Training data. We used Blender [20], an open-source rendering package, to render 100,000 training images, 5,000 validation images, and 5,000 test images. We used images of simple geometric objects, as a first step in exploring the hypothesis that many properties of lightness perception are due to generic features of 3D scenes, such as cast shadows and occlusion, rather than to more subtle properties of genuine natural scenes. Each scene contained 20 randomly positioned, greyscale geometric objects (spheres, cubes, and tori; Figure 1a). Each object had probability 0.5 of being colored solid grey, with reflectance uniformly drawn from the interval [0.1, 0.9], and probability 0.5 of having a greyscale Voronoi texture. The background consisted of three planes, intersecting at randomly chosen angles between 80° and 100°, and each independently assigned a randomly chosen reflectance from [0.1, 0.9]. Lighting consisted of an ambient source and a directional (i.e., infinitely distant) source. The direction and intensity of the directional source were randomized across scenes, as was camera position. All surfaces were Lambertian, and rendering did not include interreflections. We rendered a luminance image and a reflectance image for each scene. The size of each image was 256 × 256 pixels.

Training. We trained IRNet to infer reflectance images from luminance images. We used the Adam optimizer [21] with a mean-squared error criterion, and a batch size of five images. Batches were randomly sampled without replacement from the 100,000 training images, and training continued for seven epochs, by which time the error on the 5,000 validation images had asymptoted.

Evaluation. We measured the root-mean-square (RMS) error of IRNet's reflectance estimates on the 5,000 test images, which were not used during training.

Using separately rendered stimuli (described below), we also measured the network's Thouless ratio, which is a measure of perceptual constancy [22]. In a lightness matching task, the Thouless ratio is

$$\tau = \frac{\log r_M - \log r_0}{\log r_R - \log r_0} \tag{1}$$

Here r_M is the match reflectance chosen by the observer, r_R is the reference reflectance, and r_0 is the match reflectance that would be chosen by an observer who has no lightness constancy and simply matches image luminance. This ratio measures the extent to which perceived reflectance is independent of illumination. If an observer is completely lightness constant ($r_M = r_R$), and reflectance estimates are independent of illumination, then the Thouless ratio is one. If an observer completely confounds reflectance and image luminance ($r_M = r_0$), so that for example doubling illumination (and hence image luminance) also doubles perceived reflectance, then the Thouless ratio is zero. Values between zero and one represent degrees of partial lightness constancy.

To measure Thouless ratios, we used Blender to render images of two probe cubes and a sphere (Figure 2a). The reference cube was located in the shadow of the sphere, and the test cube was located outside the shadow. We measured Thouless ratios for reference cube reflectances 0.1, 0.2, and 0.4. For each reference cube reflectance, we rendered scenes where the test cube had reflectance 0.0, 0.1, 0.2, 0.4, 0.8, or 1.0, and the directional light produced illuminance 0.0, 0.2, 0.4, 0.6, 0.8, 1.0, or 1.5 lux on a frontoparallel surface (hence $6 \times 7 = 42$ scenes). The ambient light had fixed luminance 0.5 cd/m^2 in all directions, and so generated illuminance $\pi/2$ lux on unobstructed surfaces. The illuminance at the upper surface of the reference cube (in shadow) was determined by the ambient light intensity, whereas the illuminance at the upper surface of the test cube included contributions from both the directional light and the ambient light. We inferred the illuminances by dividing the rendered image luminance of the upper faces of the reference and test cubes by their known reflectances (and multiplying by the factor of π required by the definitions of photometric units). The background panel reflectances were the same in all scenes (floor 0.25; left wall 0.45; right wall 0.55). Under each directional light intensity, we recorded the network's mean output over the pixels of the upward-facing surface of the reference cube (reflectance fixed) and the test cube (reflectance varied across scenes). We then used interpolation to find the test cube reflectance required for the network to have the same output at the reference and test cubes. We took this to be the network's 'match' setting, indicating the actual test cube reflectance for which the network assigned the same perceived reflectance to the reference and test cubes. We found this match reflectance under all seven directional lighting intensities. Equation (1) implies that if the Thouless ratio τ is constant across changes in illuminance, then log match reflectance is an affine function of log illuminance at the test patch, with slope $m = \tau - 1$ (see Appendix).

We found IRNet's match reflectance for each of the seven directional lighting intensities, fitted a least-squares regression line to log match reflectance versus log illuminance at the test patch, and used the inverse relationship $\tau = m + 1$ to find the Thouless ratio. We repeated this procedure for each reference cube reflectance (0.1, 0.2, and 0.4), resulting in three Thouless ratios.

Model comparisons. We used the same methods to evaluate three additional models: a variant of retinex [8], Dakin and Bex's band-pass normalization model [9], and the ODOG brightness model [10, 15].

Results

Figure 1 shows examples of luminance images, ground truth reflectance images, and IRNet's response to the luminance images. Images of the network's response show that it largely removed cast shadows from the luminance image, though with residual mottling in some places. It also largely removed *shading*, the variation in luminance due to variations in surface orientation relative to the light source. Figure 1d shows scatterplots of estimated reflectance versus true reflectance at 1,000 randomly chosen pixels in each image. Estimated reflectance was approximately proportional to true reflectance. Across 5,000 test images like the ones shown, the median RMS reflectance error was 0.063. For comparison, we found that a model that simply maps each luminance image to fill the reflectance range [0.1, 0.9] via an affine transformation had a median RMS error of 0.22.

Figure 3 shows corresponding results for retinex, the Dakin-Bex model, and ODOG. Here shading and cast shadows remain clearly visible in the model outputs. Scatterplots show that model output increased as a function of true reflectance, but the scatter was much greater than for IRNet. Furthermore, these model outputs were in arbitrary units. We converted model outputs to reflectance estimates by applying the affine transform that minimized the resulting RMS error between reflectance estimates and true reflectance. After this transformation, the median RMS reflectance error across the 5,000 test images was 0.16 for all three models. This is substantially higher than the error found with IRNet, and only moderately better than the model described above that simply applies an affine transformation to luminance. However, it is also true that IRNet was trained specifically on this stimulus set, unlike the three other models. In future work, it will be important to test how well all four models perform outside the stimulus set used here.

Figure 2 shows results for IRNet in the Thouless ratio experiment. Like human observers, the network had partial lightness constancy, and test patches in higher illumination were judged to have slightly higher reflectance. Furthermore, log match reflectance declined approximately linearly as a function of log illuminance. Thouless ratios for reference reflectances 0.1, 0.2, and 0.4 were 0.69, 0.75, and 0.69, respectively. Human observers have been found to have Thouless ratios in the range 0.35 to 0.75 in sparse scenes like the ones used here [1, p. 31], and higher values in richer scenes [23]. Thus although we have not measured Thouless ratios for human observers in these scenes, the values we found for IRNet are broadly similar to what we might expect from humans, and the approximately linear relationship between log match reflectance and log illumination is also typical of human observers [23]. Thouless ratios for human observers also sometimes show moderate variations from one reference reflectance to another [23].

Figure 4 shows corresponding results for the other three models. In all cases, lightness constancy was weak. The Thouless ratios given at the right of each panel show that in most cases, the models' behaviour was close to luminance matching ($\tau = 0$). This is much weaker lightness constancy than is typical of human observers, even in simple scenes, but in future work we will establish a point of comparison by measuring human lightness constancy in the same scenes used to test the computational models.

Discussion

Practically all research to date on human lightness perception has been guided by parametric models that describe how lightness depends on selected properties (e.g., surface orientation) in a limited range of scenes, or by qualitative observations about how various image features affect lightness. Without a doubt, these are valuable approaches, and they have revealed much about how we perceive lightness. At the same time, another goal of research on lightness should be to formulate image-computable models that predict what people see in complex, realistic scenes. Progress toward this goal has been limited. Our examination of a recent ANN architecture suggests that such networks may be useful points of departure for a new class of models of human lightness perception. For example, we find that IRNet's log reflectance matches are an approximately linear function of log illuminance (Figure 2), as is the case for human observers. We also find that IRNet has Thouless ratios comparable to those often found with human observers. To continue this approach to modelling lightness, there are certainly more tests that should be done to compare ANNs to human observers. Are ANNs highly tolerant of lighting inconsistencies, as human observers are [24]? Are they susceptible to the same lightness illusions as human observers? Do they generalize well beyond the training stimuli? Furthermore, black-box algorithms like the one used here are not very enlightening when considered as models, and more interpretable architectures are needed. Despite these challenges, we find it highly appealing to have a large new class of image-computable, data-driven models that predict lightness in complex scenes, and we suggest that this is a promising way forward for modelling lightness.

Appendix

We assume that all surfaces are Lambertian. Let the reference cube have log reflectance r_R, log illuminance i_R, and log luminance $\ell_R = r_R + i_R - \log \pi$. (In SI units, the luminance (cd/m^2) of a Lambertian surface equals reflectance (unitless) times illuminance (lux) divided by the constant π cd / m^2 · lux.) Let the observer's match setting at the test cube be log reflectance r_M, under log illuminance i_M, and thus have luminance $\ell_M = r_M + i_M - \log \pi$. If the observer has no lightness constancy and simply matches luminance at the reference and test cubes, then their match setting r_0 satisfies $r_0 + i_M = r_R + i_R$, or $r_0 = r_R + i_R - i_M$. If we substitute this expression for r_0 into equation (1) and solve for r_M, we find $r_M = (\tau - 1)(i_M - i_R) + r_R$, which is an affine function of i_M with slope $m = \tau - 1$.

References

[1] A L Gilchrist. *Seeing black and white*. Oxford University Press, 2006. 1, 3

[2] M Janner, J Wu, T D Kulkarni, I Yildirim, and J Tenenbaum.

Self-supervised intrinsic image decomposition. In *Advances in Neural Information Processing Systems 30*, pages 5396–5946, 2017. 1, 2

[3] Z Li, A Xu, R Ramamoorthi, K Sunkavalli, and M Chandraker. Learning to reconstruct shape and spatially-varying reflectance from a single image. *ACM Transactions on Graphics*, 37(6):269:1–11, 2018. 1, 2

[4] Y Yu and W A P Smith. InverseRenderNet: Learning single image inverse rendering. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3155–3163. IEEE, 2019. 1, 2

[5] D H Brainard and L T Maloney. Surface color perception and equivalent illumination models. *Journal of Vision*, 11(5):1, 2011. 1, 2

[6] R F Murray. A model of lightness perception guided by probabilistic assumptions about lighting and reflectance. *Journal of Vision*, 20(7):28:1–22, 2020. 1

[7] R F Murray. Lightness perception in complex scenes. *Annual Review of Vision Science*, 7:417–436, 2021. 1

[8] J J McCann. Lessons learned from Mondrians applied to real images and color gamuts. In *Proceedings of the IS&T/SID Seventh Color Imaging Conference*, pages 1–8. Scottsdale, AZ, November 1999, 1999. 1, 3

[9] S C Dakin and P J Bex. Natural image statistics mediate brightness 'filling in'. *Proceedings of the Royal Society B*, 270:2341–2348, 2003. 1, 3

[10] B Blakeslee and M E McCourt. A multiscale spatial filtering account of the White effect, simultaneous brightness contrast and grating induction. *Vision Research*, 39:4361–4377, 1999. 1, 3

[11] E H Land and J J McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971. 1

[12] J J McCann and A Rizzi. *The art and science of HDR imaging*. New York: John Wiley & Sons, Ltd., 2012. 1

[13] F A A Kingdom. Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research*, 51(13):652–673, 2011. 1

[14] T Betz, R Shapley, F A Wichmann, and M Maertens. Noise masking of White's illusion exposes the weakness of current spatial filtering models of lightness perception. *Journal of Vision*, 15(14):1:1–17, 2015. 1

[15] B Blakeslee and M E McCourt. When is spatial filtering enough? Investigation of brightness and lightness perception in stimuli containing a visible illumination component. *Vision Research*, 60:40–50, 2012. 1, 3

[16] E L Dixon and A G Shapiro. Spatial filtering, color constancy, and the color-changing dress. *Journal of Vision*, 17(3):7:1–20, 2017. 1

[17] J T Barron and J Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015. 2

[18] A Flachot, A Akbarina, H H Schütt, R W Fleming, F A Wichmann, and K R Gegenfurtner. Deep neural models for color classification and color constancy. *Journal of Vision*, 22(4):17:1–24, 2022. 2

[19] K R Storrs, B L Anderson, and R W Fleming. Unsupervised learning predicts human perception and misperception of gloss. *Nature Human Behaviour*, 5:1402–1417, 2021. 2

[20] Blender Online Community. Blender, version 2.92.0, 2021. 2

[21] D P Kingma and J L Ba. Adam: a method for stochastic optimization. In *Proceedings of the Third International Conference on Learning Representations*, 2015. 2

[22] R H Thouless. Phenomenal regression to the real object. I. *British Journal of Psychology*, 21(4):339–359, 1931. 2

[23] K Y Patel, A P Munasinghe, and R F Murray. Lightness matching and perceptual similarity. *Journal of Vision*, 18(5):1:1–13, 2018. 3

[24] J D Wilder, W J Adams, and R F Murray. Shape from shading under inconsistent illumination. *Journal of Vision*, 19(6):2, 2019. 3

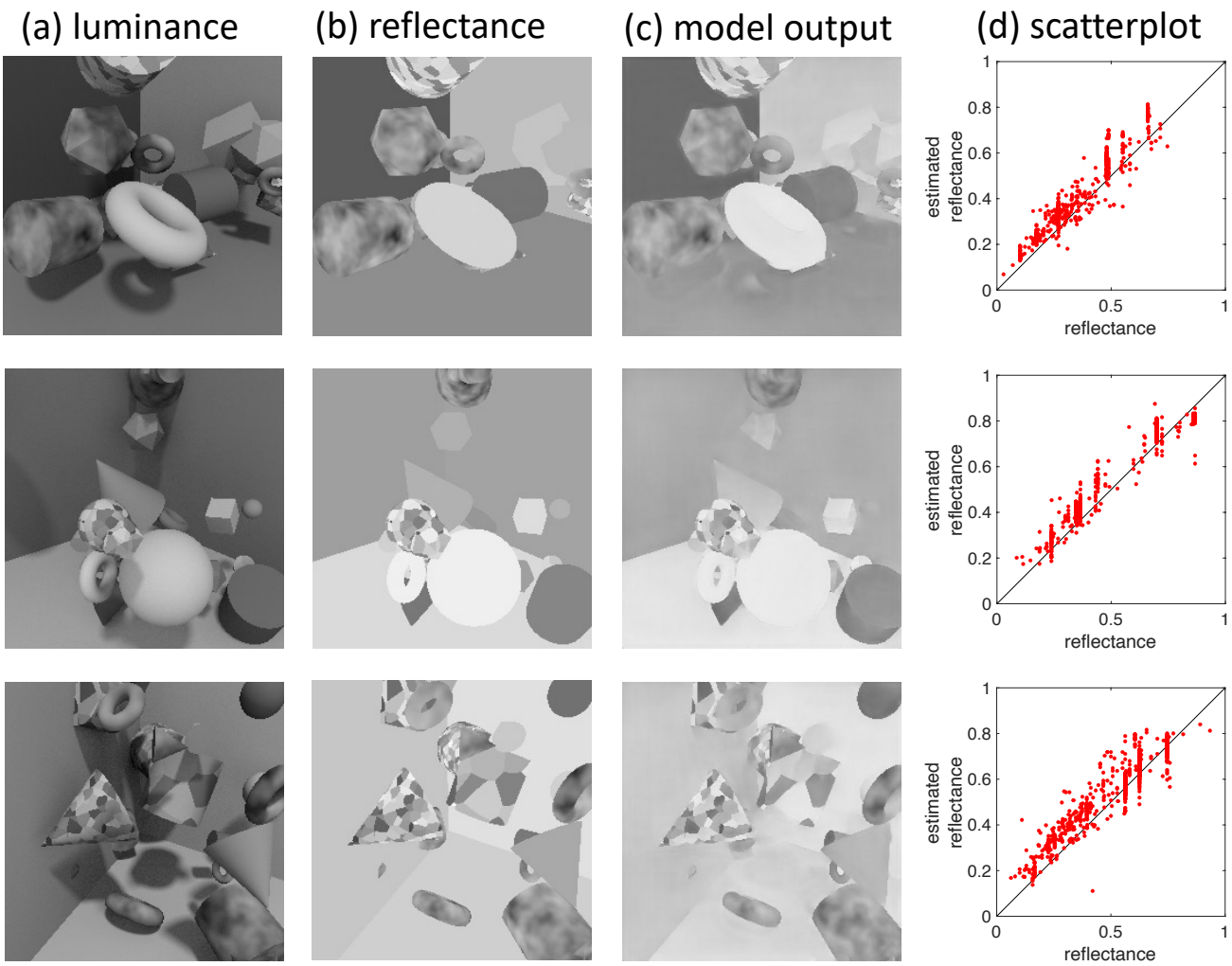

Figure 1. *Sample stimuli and IRNet responses. (a) Luminance images. (b) Ground truth reflectance images. (c) IRNet's response to luminance images. (d) Scatterplot of IRNet's pixelwise response versus true reflectance.*

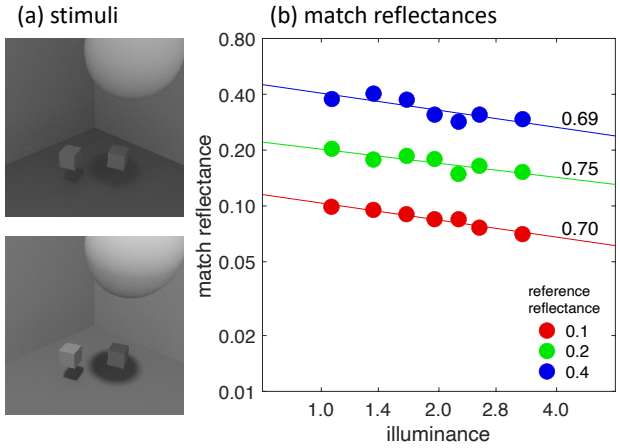

Figure 2. *Thouless ratio experiment. (a) Two examples of stimuli with different directional lighting intensities. The right cube (in shadow) is the reference cube, and the left cube is the test cube. (b) IRNet's match reflectance as a function of illuminance at the test cube, for three different reference reflectances. The numbers at the right of the panel, above each line, are Thouless ratios for the three reference reflectances.*

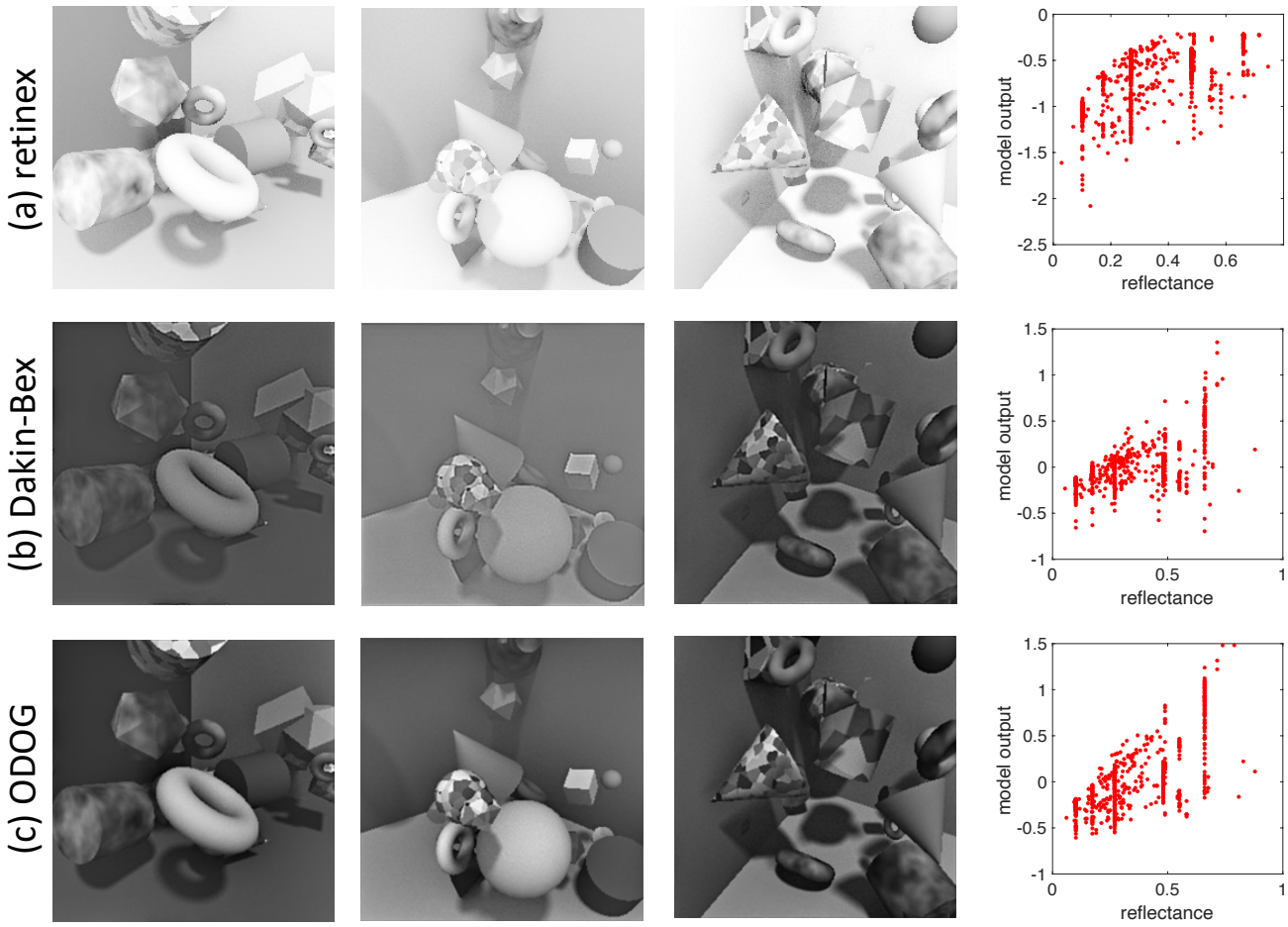

Figure 3. *Responses of three additional models. The first three columns show model responses to the three luminance images shown in Figure 1a. The fourth column shows scatterplots of pixelwise model response versus true reflectance, for the image in the first column.*

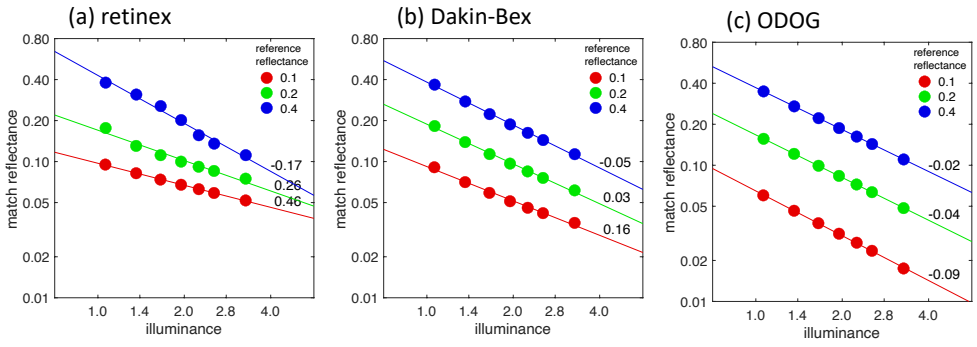

Figure 4. *Thouless ratio experiment for the three additional models. The numbers at the right, above each line, are Thouless ratios for the three reference reflectances.*