Image statistics for material perception

Nishida, Shin'ya

Available online at www.sciencedirect.com
ScienceDirect
Current Opinion in Behavioral Sciences

Image statistics for material perception

Shin'ya Nishida[1,2]

For estimation of material properties, inverse optics is generally too difficult to solve. Human material perception seems to rely on image features that are correlated with the material property under natural viewing environments. The critical features often take the form of image statistics, because many material properties can be characterized by how they optically modulate the natural image statistics. For instance, a critical image statistic for surface wetness perception is enhanced color saturations, while that for subresolution fineness perception is reduced luminance contrasts. There are optical reasons these image features vary in correlation with physical material properties, as well as psychophysical evidence that human material perception does respond to the features. That the shape (skewness) of the luminance histogram strongly affects surface material (gloss) perception, while not surface shape perception, suggests that material and shape perceptions may rely on independent image features — material (surface reflectance) perception relies on the magnitude of luminance gradient, while shape perception relies on the order of luminance gradient. I also discuss the merit and demerit of image statistics in relation to mid-level perceptual features, and deep neural network features.

Addresses
[1] Graduate School of Informatics, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan
[2] NTT Communication Science Laboratories, Nippon Telegraph and Telephone Corporation, 3-1 Morinosato Wakamiya, Atsugi-shi, Kanagawa 243-0198, Japan

Corresponding author: Nishida, Shin'ya (shinyanishida@mac.com)

Current Opinion in Behavioral Sciences 2019, **30**:94–99

This review comes from a themed issue on **Visual perception**

Edited by **Hannah E Smithson** and **John S Werner**

For a complete overview see the Issue and the Editorial

Available online 9th August 2019

https://doi.org/10.1016/j.cobeha.2019.07.003

Introduction

A retinal image is formed by an optical event in the scene. According to a traditional view [1], the task of vision is to estimate scene variables from the retinal image as an inverse-optics problem. For vision problems for which the underlying optics is relatively straightforward (e.g. stereo vision based on binocular disparity), one may be able to develop elegant computational inverse-optics theories. For complex problems, however, a strict inverse-optics approach may not be feasible. Material perception is such a complex problem. We humans can visually recognize not only the shape and color of an object, but also the reflectance property (e.g. gloss), material type (e.g. metal), and surface condition (e.g. wet) of the object. Physically, these material properties mainly reflect the surface reflectance properties represented by a complex function depending on the directions of illumination and viewer, known as Bidirectional Reflectance Distribution Function (BRDF), or even more complex functions including subsurface scattering and/or spatial variations. To correctly estimate the reflectance variables from a single object image is a very hard problem, because the image is produced as a result of complex optical interactions among surface reflectance property, surface geometry and lighting of the object, each of them being complicated and high dimensional. This ill-posed problem can be solved only under strong assumptions [2,3].

Instead of rigorously solving the inverse optics, the human visual system may use certain image features as clues for statistical inference of the physical material properties. Those features are correlated with the target physical property under most, but not all, natural scenes. In agreement with this view, human estimation of surface reflectance properties is not very accurate, being significantly affected by object's geometry [4,5] and lighting [6,7]. These findings also suggested the importance of low-level image statistics. For example, it was found that the shape of the histogram of image pixel intensity or that of specific subbands of spatial frequency/orientation of the image, characterized by second and third-order moments (standard deviation and skewness), is correlated well with gloss perception [8,9].

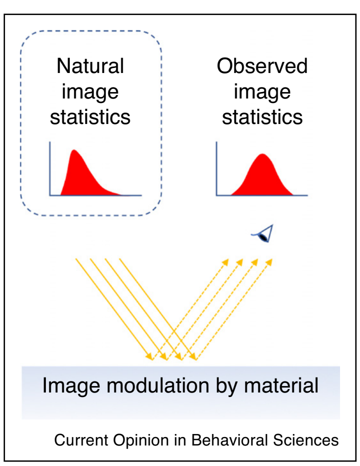
In this monograph, I first explain why image statistics is important, and how they should be used, in material perception research. I then introduce specific image statistics contributing to perception of wetness, fineness, and viscosity. I review the controversy over gloss perception, and discuss simultaneous estimation of material and shape. I also discuss the pros and cons of image statistics relative to mid-level perceptual features and deep neural network features.

Why image statistics?

In machine vision research, it has been shown that low-level image statistics are effective and robust features for a variety of tasks. In human vision research, investigation on image statistics has been made mainly in relation to texture perception, starting with Julesz's proposal that two textures with identical second-order statistics are not preattentively segregated [10]. Later, second-order statistics was substituted by subband statistics that is associated with the activation pattern of the bank of V1 Gabor filters [11,12]. More recent studies emphasized the importance of the joint statistics defined by the combinations of V1-filter outputs, represented beyond V1 [13–16]. Specifically, with perfect matching with regard to Portilla–Simoncelli (PS) statistics, which includes the joint statistics in addition to the intensity and subband statistics [17], observers cannot perceptually discriminate two textures with their peripheral vision.

Statistical features are expected to be particularly useful for material perception, since in many cases, material perception is based on the way the material modulates natural image statistics [18]. For instance, the surface reflectance property is specified from the way the reflection modulates the spatial pattern of illumination (Figure 1). If the pattern reflected at the surface of metallic object is blurred, the observer can infer that the surface is not well-polished, because the image of things in the reflected natural scene are originally sharp. As long as the observers know such rules of natural image statistics, they do not need to know the environment pattern *per se*. Similarly, under the situation where refraction by a transparent material optically deforms a background texture pattern, the observers can recognize the presence, and estimate the property, of the transparent material from the pattern of deviation from natural image statistics, without knowing the exact pattern of background texture.

Figure 1



In many cases, material perception is visual estimation of optical modulation of image statistics. Assuming natural image statistics of the environment, our visual system is able infer the modulation characteristics by surface reflectance from the observed image statistics.

Image statistics as perceptual cues

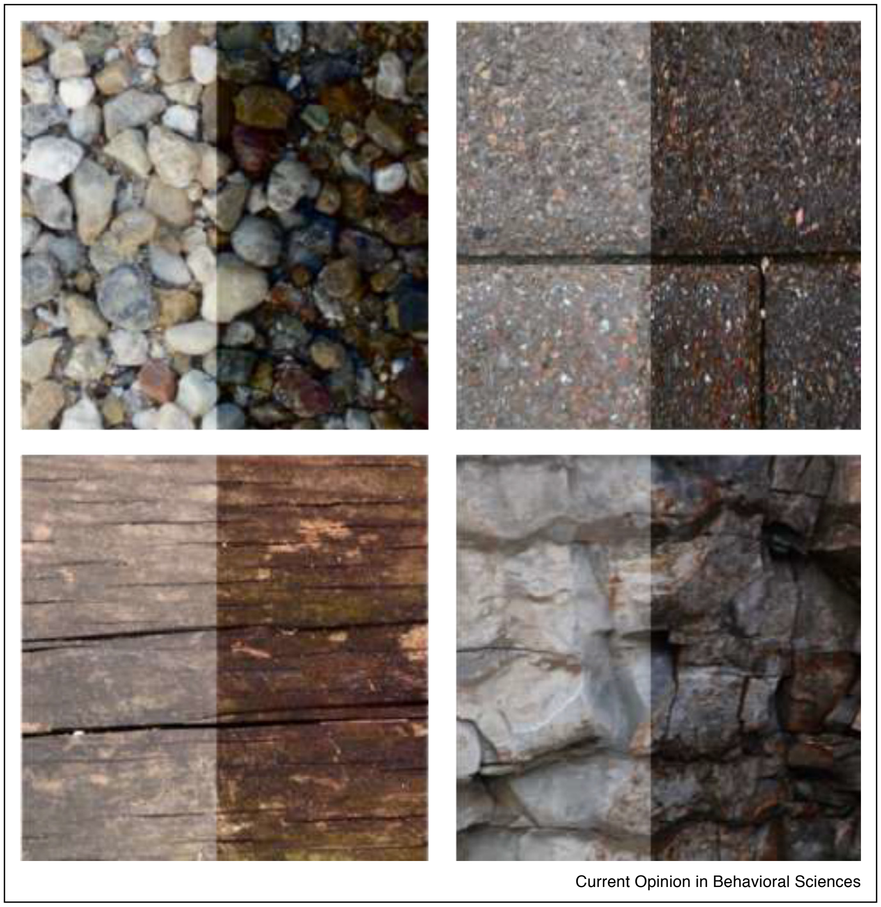
One should clarify several points to make a convincing claim that a given image statistic is a visual cue to perception of a particular material property. First, one needs to prove the 'perceptual' link between the image statistic (proximal stimulus) and material perception. A standard methodology for this is to collect human ratings of the material property for a number of images, and then find a small number of image features correlated well with the rating. This however only shows correlation. To prove the causal relationship, one should also demonstrate that changes in the candidate feature do affect the material perception in the predicted way. In addition, in the cases where the material perception can be regarded as the estimation of some physical variable/category (distal stimulus), one should establish the 'physical' link between the image statistic and the physical ground truth. The evidence for this link could be provided by theoretical analysis of the physical event, image rendering based on computer simulation, and real image measurements. It is also useful to see how well machine vision can estimate the ground truth from the candidate image statistic. In addition to clarifying the perceptual and physical links, it is important to clarify the conditions under which the image cues become effective. Just as certain monocular depth cues (e.g. relative size and aerial blur) are effective only under specific scene configurations, most of the image statistics, including all the cues described in this paper, are able to predict material perception only when other conditions are satisfied.

That a given image statistic is a visual cue to specific material perception does not necessarily imply that the brain explicitly computes it in an intermediate processing stage. The brain may analyze different stimulus aspects that are strongly correlated with the image statistic. Even in that case, specification of the key feature directly computable from the image provides clear insights into the visual processing for material perception.

Wetness, fineness, and viscosity

Let me explain some examples of image statistics as cues for material perception. Figure 2 shows an image transformation, named Wet Enhancing Transformation (WET), which is able to enhance the wetness impression of the scene [19•]. WET consists of two operations. One is to increase the luminance histogram skew, which makes the image look darker and glossy. The other is to enhance color saturation. Theoretical analysis as well as natural image analysis indicates that surface wetting indeed produces these image changes. The color change can be ascribed to an increase in the number of light-material interactions by wetting. Psychophysical experiments indicate that WET generally increases the wetness rating for natural texture images, but the effect magnitude varies across images. Sawayama *et al.* [19•] found that a critical factor to determine the effectiveness of WET is

Figure 2



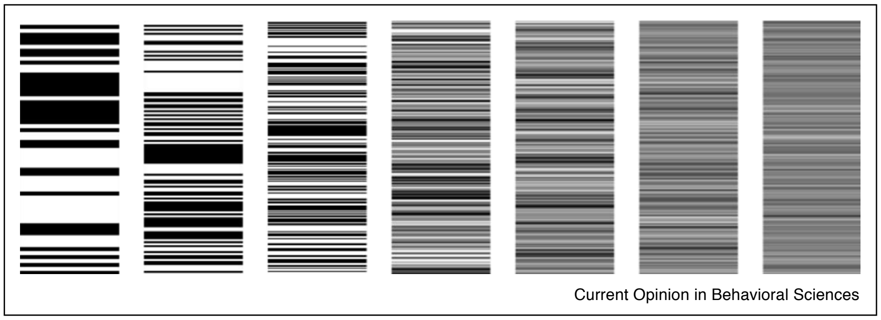
Wet Enhancing transform (overall enhancement of color saturation combined with an increase in luminance skewness) applied to the right side of each texture photograph. Reproduced from Ref. [19•].

color entropy of the image. Specifically, WET is more effective for images with a larger number of colors (high color entropy) than for those with a small number of colors (low color entropy). This could be interpreted as a result of Bayesian inference by the visual system about what physical factor is likely to cause of the image changes produced by WET. In general, high color saturation could be an inherent property of the object, rather than a result of wetting. However, if it is observed simultaneously at many locations in the scene, it is likely to be produced by global wetting.

Although surface geometry, surface material (optical property) and lighting are said to be three main factors of image rendering, they cannot be strictly separated from one another. Specifically, surface optical properties are affected by micro-scale and meso-scale surface geometry. Perception of superfine texture is an example of the effects of meso-scale surface geometry on material perception [20•]. Human observers are able to judge the relative fineness of line textures even when the width of each line element is narrower than the spatial resolution of the visual system (Figure 3). Thanks to this ability, we are able to estimate the fineness of hair and fabric even at a good distance. Image analysis of the superfine texture indicates that an increase in fineness is accompanied by a reduction in texture luminance contrast. This is because as each line elements becomes finer, the number of elements optically merged at each location (pixel) becomes larger. Consequently, luminance variation among spatial neighbors decreases according to the central limit theorem. Psychophysical test indicates that a reduction in luminance contrast indeed increases apparent fineness of artificial and natural textures. It is also found that for stimulus contrast of fine element textures to affect apparent fineness, the shape of luminance histogram should be close to a Gaussian distribution. This suggests that the human visual system implicitly knows a statistical rule following the central limit theorem.

Image statistics useful for material perception are not only about luminance and color, but also about motion flow.

Figure 3



Subresolution fineness perception based on luminance contrast reduction. Texture fineness appears to steadily increase from left to right, while the minimum line width does not physically change for the right four textures. Reproduced from Ref. [20•].

Perception of liquid viscosity, a mechanical material property, is significantly affected by local motion speed and flow smoothness [21]. When a static pattern is dynamically deformed at a specific range of spatiotemporal frequency, one can see a vivid perception of transparent liquid material [21].

Gloss

Concerning the visual processing for gloss perception, there was a controversy over the proposal that the skewness of intensity histogram is a low-level image cue to human gloss perception [8]. One criticism is that the absolute skewness value, which is affected by many other factors, does not always predict gloss perception [22]. Another criticism is that histogram skewness cannot explain spatial factors on gloss perception. Specifically, gloss perception is impaired when a highlight is placed at an incorrect surface location, or in an incorrect shape [23]. Apparent 3D layout of the surface also affects gloss perception [24].

It seems difficult to explain the difference between correct and incorrect highlights only in terms of image statistics, since the texture synthesis algorithm that equates PS statistics [17] is unable to synthesize images with proper highlights, and a pair of synthesized texture images, one with a correct highlight pattern, and the other with an incorrect highlight pattern, are perceptually similar and indistinguishable [25]. These observations suggest that image analysis beyond PS statistics is necessary to judge highlight spatial consistency. A more powerful texture synthesis algorithm based on image features developed in a deep neural network (DNN) [26] can reproduce patterns with highlight-like blobs, but only when features represented in high-level layers are used [27•].

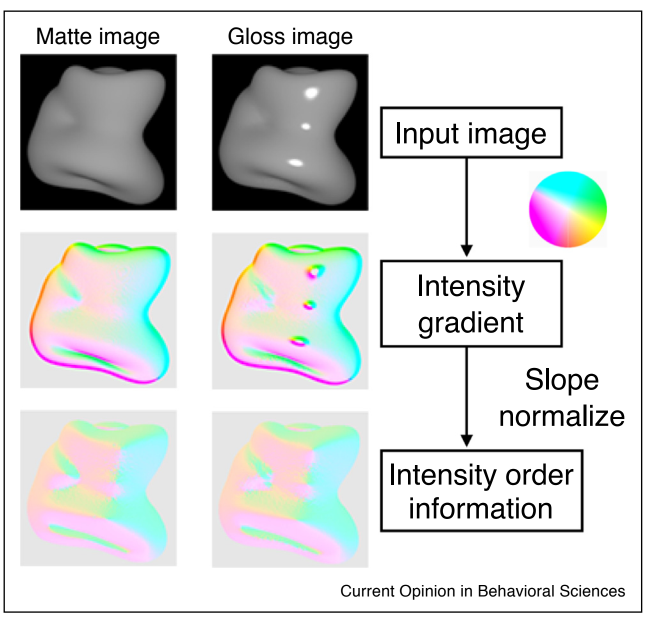
For better explanation of spatial effects on gloss perception, instead of low-level image statistics, mid-level perceptual features, such as the coverage of specular highlights, are proposed [22]. I do appreciate the merit of this approach, but a drawback of the mid-level features is that they are also outputs of visual processing, estimated from human judgments. To define the coverage of highlights, the visual system has to specify which image regions are highlights. To specify highlights, however, the visual system has to solve the core problem of gloss processing, that is, intrinsic image decomposition [28]. As a result, it is not very clear whether the mid-level perceptual cues cause gloss perception, or they are computed together with gloss. In contrast, low-level image statistics are directly computable from a region of interest of the image via relatively simple operations.

We recently proposed a hypothesis to reconcile the low-level histogram skew effect and the high-level spatial effect in gloss perception [29••]. Originally, the idea to associate luminance histogram skewness with gloss perception is based on a robust finding that an increase (or a decrease) in the histogram skew by histogram matching significantly enhances (reduces) apparent gloss. The test natural scenes consist mainly of dielectric materials having both diffuse and specular reflection components. In relation to this image manipulation, two additional observations are notable. First, changes in the luminance histogram do not affect the spatial structure of the images. More specifically, they preserve the luminance order relationship among pixels. Second, changes in the luminance histogram have little effect on apparent 3-dimentional shape. At least it preserves apparent qualitative shape. These observations led us to an idea that the visual system estimates geometry (shape) and material (reflectance) separately from two different aspects of image intensity gradients: shape from the order information of intensity gradients, while material from the magnitude information of intensity gradients. Physically, the intensity order (which is equivalent to isophote, that is, curves connecting points of equal intensities) with gradient direction information, is relatively stable against changes in material given the shape and illumination are kept the same (Figure 4). Therefore, luminance order is expected to provide robust information about shape estimation, although how to estimate (qualitative) depth only from intensity order information remains an outstanding problem. On the other hand, for non-textured objects, intensity gradient magnitude is dependent on surface material, being steep for glossy materials, while shallow for matte materials. Note also that intensity gradient magnitude has a direct relation with the shape of the intensity histogram, as well as the shape of the specular lobe in the surface's BRDF. It is therefore possible to make glossy materials look matte by simply removing steep gradients at high intensity zone (whereas it is not easy to make matte materials look glossy in the same way without knowing the spatial information about specular components) [29••]. For estimation of material, the slope of the intensity gradient should be evaluated relative to the slope of surface orientation gradient, and this is where 3D shape information interacts with material perception.

Deep neural network features

In the last several years, deep learning [30] has dramatically improved the performance of machine vision, including material perception [31–33]. It is being recognized that deep leaning is a useful tool not only for engineering purposes, but also for scientific understanding of human visual processing [34,35]. Specifically, to train a DNN for some material judgment task and then analyze the neural representation having developed in intermediate layers (deep features) seems to be a promising data-driven method to find the critical image features, along with neural processing, for the material judgment. Deep features are directly computable from the input image like low-level image features, and are able to predict perception as well as, or better than, mid-level perceptual features. Deep features can be also used to synthesize realistic materials [36]. However, it remains debatable how much similar the representations are between artificial neural networks and human neural networks. Moreover, unlike conventional image features, the deep features are not easily interpretable. Although many challenging problems remain, analyzing the relationship of the deep features with the low-level and mid-level cues will be a promising direction of future research on material perception.

Figure 4



Two objects with different materials, but with the same shape and illumination are similar in intensity-gradient orientation map when the gradient magnitude is normalized. This implies that material differences mainly affect intensity-gradient magnitude, but not in intensity order. Reproduced from Ref. [29••].

Conclusion

Visual cues in the form of images statistics are directly computable from image (unlike mid-level perceptual features) and computationally understandable (unlike deep features). Specifying such image statistics is not the final goal, but an important step to understand visual computation underlying material perception.

Conflict of interest statement

Nothing declared.

Acknowledgement

Funding: This work is supported by Grant-in-Aid for Scientific Research on Innovative Areas from Japan Society for the Promotion of Science (JSPS KAKENHI Grant Number JP15H05915).

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Marr D: *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: Freeman; 1982.
2. Oxholm G, Nishino K: **Shape and reflectance estimation in the wild**. *IEEE Trans Pattern Anal Mach Intell* 2015, **38**:376-389.
3. Lombardi S, Nishino K: **Reflectance and illumination recovery in the wild**. *IEEE Trans Pattern Anal Mach Intell* 2015, **38**:129-141.
4. Nishida S, Shinya M: **Use of image-based information in judgments of surface-reflectance properties**. *J Opt Soc Am A* 1998, **15**:2951-2965.
5. Vangorp P, Laurijssen J, Dutré P: **The influence of shape on the perception of material reflectance**. *ACM Trans Graph* 2007, **26**:77 1–9.
6. Fleming RW, Dror RO, Adelson EH: **Real-world illumination and the perception of surface reflectance properties**. *J Vis* 2003, **3**:347-368.
7. Motoyoshi I, Matoba H: **Variability in constancy of the perceived surface reflectance across different illumination statistics**. *Vis Res* 2012, **53**:30-39.

8. Motoyoshi I, Nishida S, Sharan L, Adelson EH: **Image statistics and the perception of surface qualities**. *Nature* 2007, **447**:206-209.

9. Sharan L, Li Y, Motoyoshi I, Nishida S, Adelson EH: **Image statistics for surface reflectance perception**. *J Opt Soc Am A* 2008, **25**:846-865.

10. Julesz B: **Visual pattern discrimination**. *IRE Trans Inf Theory* 1962, **8**:84-92.

11. Bergen JR, Adelson EH: **Early vision and texture perception**. *Nature* 1988, **333**:363-364.

12. Chubb C, Landy M: **Orthogonal distribution analysis: a new approach to the study of texture perception**. In *Computational Models of Visual Processing.* Edited by Landy MS, Movshon JA. MIT Press; 1991:291-301.

13. Freeman J, Simoncelli EP: **Metamers of the ventral stream**. *Nat Neurosci* 2011, **14**:1195-1201.

14. Ziemba CM, Heeger DJ, Simoncelli EP, Movshon JA, Freeman J: **A functional and perceptual signature of the second visual area in primates**. *Nat Neurosci* 2013:1-12.

15. Okazawa G, Tajima S, Komatsu H: **Image statistics underlying natural texture selectivity of neurons in macaque V4**. *Proc Natl Acad Sci U S A* 2015, **112**:E351-E360.

16. Rosenholtz R: **Capabilities and limitations of peripheral vision**. *Annu Rev Vis Sci* 2016:437-457.

17. Portilla J, Simoncelli E: **Texture modeling and synthesis using joint statistics of complex wavelet coefficients**. *Int J Comput Vis* 2000:49-70.

18. Adelson EH: **On seeing stuff: the perception of materials by humans and machines**. Edited by Rogowitz BE, Pappas TN. 2001:1-12. SPIE.

19. • Sawayama M, Adelson EH, Nishida S: **Visual wetness perception based on image color statistics**. *J Vis* 2017, **17**:7 1–24.

20. • Sawayama M, Nishida S, Shinya M: **Human perception of subresolution fineness of dense textures based on image intensity statistics**. *J Vis* 2017, **17**:8 1–18.

21. Kawabe T, Maruya K, Fleming RW, Nishida S: **Seeing liquids from visual motion**. *Vis Res* 2015, **109**:125-138.

22. Marlow PJ, Kim J, Anderson BL: **The perception and misperception of specular surface reflectance**. *Curr Biol* 2012, **22**:1909-1913.

23. Kim J, Marlow P, Anderson BL: **The perception of gloss depends on highlight congruence with surface shading**. *J Vis* 2011, **11**:4 1–19.

24. Marlow PJ, Todorović D, Anderson BL: **Coupled computations of three-dimensional shape and material**. *Curr Biol* 2015, **25**: R221-R222.

25. Wang Q, Motoyoshi I, Nishida S: **Characterization of high-level images features for surface gloss perception**. *J Vis* 2013, **13**:202 (Vision Sciences Society Meeting abstract).

26. Gatys LA, Ecker AS, Bethge M: **Texture synthesis using convolutional neural networks**. *Adv Neural Inf Process Syst* 2015, **28**.

27. • Wallis TSA, Funke CM, Ecker AS, Gatys LA, Wichmann FA, Bethge M: **A parametric texture model based on deep convolutional features closely matches texture appearance for humans**. *J Vis* 2017, **17**:5 1–29.

28. Barrow HG, Tenenbaum JM: **Recovering intrinsic scene characteristics from images**. *Computer Vision Systems*. Academic Press; 1978:3-26.

29. •• Sawayama M, Nishida S: **Material and shape perception based on two types of intensity gradient information**. *PLoS Comput Biol* 2018, **14**:e1006061.

30. Krizhevsky A, Sutskever I, Hinton GE: **ImageNet classification with deep convolutional neural networks**. *Adv Neural Inf Process Syst* 2012, **25**:1097-1105.

31. Cimpoi M, Maji S, Kokkinos I, Mohamed S, Vedaldi A: **Describing textures in the wild**. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2014.

32. Schwartz G, Nishino K: **Automatically discovering local visual material attributes**. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

33. Bell S, Upchurch P, Snavely N, Bala K: **Material recognition in the wild with the materials in context database**. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2015.

34. Yamins DLK, DiCarlo JJ: **Using goal-driven deep learning models to understand sensory cortex**. *Nat Neurosci* 2016, **19**:356-365.

35. Kriegeskorte N, Douglas PK: **Cognitive computational neuroscience**. *Nat Neurosci* 2018, **21**:1148-1160.

36. Gatys LA, Ecker AS, Bethge M: **Image style transfer using convolutional neural networks**. *Proc IEEE Conference on Computer Vision and Pattern Recognition*. 2016.