The effect of shape and illumination on material perception: model and applications

ANA SERRANO, Max-Planck-Institut für Informatik, Germany and Centro Universitario de la Defensa, Spain
BIN CHEN, Max-Planck-Institut für Informatik, Germany
CHAO WANG, Max-Planck-Institut für Informatik, Germany
MICHAL PIOVARČI, Institute of Science and Technology Austria, Austria
HANS-PETER SEIDEL, Max-Planck-Institut für Informatik, Germany
PIOTR DIDYK, Università della Svizzera Italiana, Switzerland
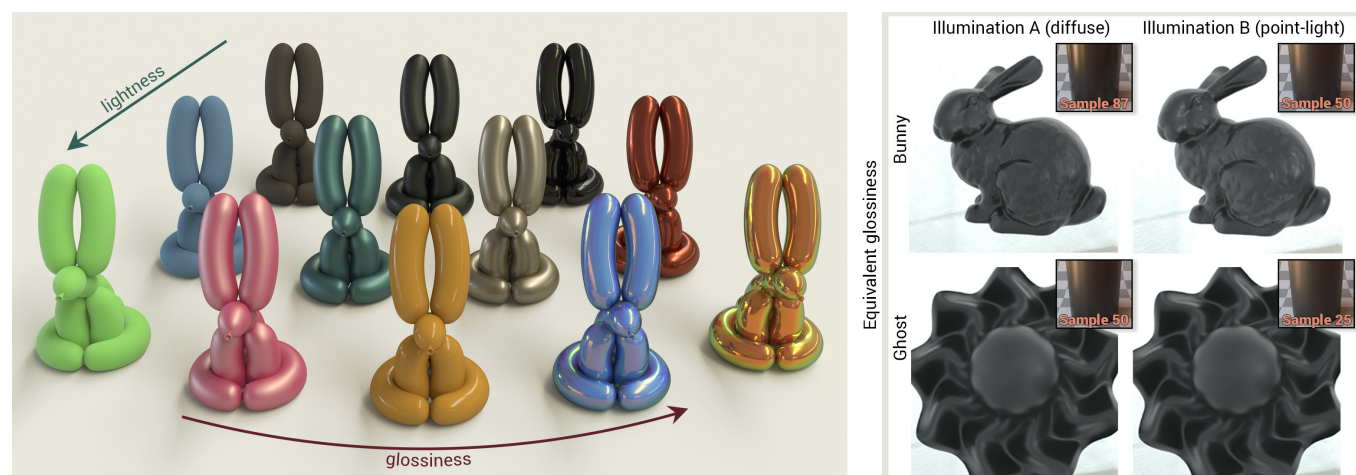KAROL MYSZKOWSKI, Max-Planck-Institut für Informatik, Germany

Fig. 1. We present a model that reckons with the effects of shape and illumination for predicting material appearance attributes that correlate with human judgments and show that it can be leveraged for several applications. Left: Our predictor can be used to sort material datasets according to desired properties such as perceived glossiness and lightness for target illuminations or shapes, assisting scene design. Right: We demonstrate the capabilities of our predictor for gloss reproduction in 3D printing. While applying the same varnish under different illuminations and shapes would yield different gloss perception, our predictor allows us to find optimal varnish mixtures (insets) for reproducing the desired equivalent gloss under different illuminations and geometries.

Material appearance hinges on material reflectance properties but also surface geometry and illumination. The unlimited number of potential combinations between these factors makes understanding and predicting material appearance a very challenging task. In this work, we collect a large-scale dataset of perceptual ratings of appearance attributes with more than 215,680 responses for 42,120 distinct combinations of material, shape, and illumination. The goal of this dataset is twofold. First, we analyze for the first time the effects of illumination and geometry in material perception across such a large collection of varied appearances. We connect our findings to those of the literature, discussing how previous knowledge generalizes across very diverse materials, shapes, and illuminations. Second, we use the collected dataset to train a deep learning architecture for predicting perceptual attributes that correlate with human judgments. We demonstrate the consistent and robust behavior of our predictor in various challenging scenarios, which, for the first time, enables estimating perceived material attributes from general 2D images. Since our predictor relies on the final appearance in an image, it can compare appearance properties across different geometries and illumination conditions. Finally, we demonstrate several applications that use our predictor, including appearance reproduction using 3D printing, BRDF editing by integrating our predictor in a differentiable renderer, illumination design, or material recommendations for scene design.

CCS Concepts: • **Computing methodologies** → *Appearance and texture representations*; *Perception*.

Additional Key Words and Phrases: Material appearance, perception

1 INTRODUCTION

Material appearance is determined not only by material reflectance, but also by surface geometry and illumination. The same material may have a different appearance for different combinations of lightings and shapes, or conversely, different materials may appear similar for specific illumination and shape choices [Ramanarayanan et al. 2007; Zhang et al. 2020a]. In the context of images such effects can be even more pronounced due to inherent ambiguities in the reflected radiance that depends on those factors.

Understanding the interactions of geometry and illumination in material appearance is a long-standing problem and the underlying perceptual processes that facilitate appearance perception are still not fully understood [Anderson 2011; Fleming 2014]. Despite a large body of work devoted to investigating the perceptual properties of materials [Anderson 2011; Fleming 2017; Fleming et al. 2015; Marlow et al. 2012], most of them focus on a single aspect of appearance, such as gloss [Chadwick and Kentridge 2015; Pellacini et al. 2000] or translucency [Fleming and Bülthoff 2005; Gkioulekas et al. 2015]. The impact of geometry and illumination is typically studied under a limited variety of appearances, geometries, and illuminations [Adams et al. 2018; Marlow and Anderson 2013; Motoyoshi et al. 2007; Olkkonen and Brainard 2010; Pont and te Pas 2006; Vangorp et al. 2007]. Further, these works often specialize in different subsets of these factors, making it difficult to compare them and generalize their insights. Some recent works study a wider variety of appearances but do not take into account the effect of illumination and geometry [Hu et al. 2020; Serrano et al. 2016]. While studying small subsets under rigidly controlled setups is a necessary and valuable step for understanding individual phenomena governing perception, we argue that analyzing a large-scale collection of very diverse appearances is crucial for generalizing existing knowledge and bringing derived insights closer to real applications.

In this work, we generate a diverse collection of 42,120 images with distinct combinations of isotropic and anisotropic materials, shapes, and illuminations, which covers a wide variety of appearances. Using a web-based online study, we crowdsource ratings for different reflectance attributes of these images, namely glossiness, sharpness of reflections, contrast of reflections, lightness, metallicness, and anisotropy. This constitutes the largest collection of subjective material ratings to date. We use this dataset for two main purposes. First, we perform an in-depth statistical analysis on the effects of illumination and geometry on perceived appearance. Using cumulative mixed models, we isolate and estimate the impact of each attribute. We then connect our insights to previous literature and provide an extensive discussion on how existing findings generalize for complex interactions of diverse shapes and illuminations. Second, inspired by recent work that shows a good correlation between deep features and complex encodings of the visual input explaining material perception [Delanoy et al. 2020; Fleming and Storrs 2019], we train a deep learning architecture

for predicting perceptual material properties from image features. For the first time, this prediction allows estimating perceived material attributes from general 2D images, which brings our insights closer to real applications. We thoroughly evaluate our model and demonstrate that its behavior is predictable and consistent with perceptual judgments under challenging scenarios, including real pictures, changes in illumination and geometry, and various tone mapping algorithms. Finally, we show different applications that directly benefit from our predictor, such as appearance reproduction using 3D printing, BRDF editing by integrating our predictor in a differentiable renderer, illumination design, and material recommendations for scene design. Our dataset (including renderings and perceptual data), data collection platform, trained predictor and code is available at https://mig.mpi-inf.mpg.de/.

2 RELATED WORK

The aim of this work is to understand and model the effect of geometry and illumination on perceived reflectance properties. Since surface reflectance is dominated by gloss [Chadwick and Kentridge 2015; Fleming 2017], we firstly focus our discussion on this property and its perceptual dimensions [Pellacini et al. 2000] as well as lightness, which is another critical property determining appearance [Toscani et al. 2020] (Sec. 2.1). Additionally, since our model estimates appearance properties, we also discuss previous work on perceptual embeddings of material appearance (Sec. 2.2), and image-based solutions for their estimation (Sec. 2.3).

2.1 The impact of geometry and illumination

Gloss. Previous work has shown that gloss perception depends on a number of factors in a very complex manner. It has been suggested that both the structure of bright highlights [Beck and Prazdny 1981; Berzhanskaya et al. 2005] and lowlights [Kim et al. 2012] play a strong role in gloss perception. Apparent gloss also increases with the brightness of highlights, their contrast, sharpness (distinctness) and coverage [Marlow and Anderson 2013; Marlow et al. 2012]. Likewise, those factors can be affected by the structure of environment maps and the orientation of dominant light sources [Marlow et al. 2012]. Surfaces appear more matte under slowly changing lighting, and glossiness increases under more directional, higher contrast lighting [Adams et al. 2018; Dror et al. 2004; Motoyoshi and Matoba 2012; Pont and te Pas 2006; Zhang et al. 2020a,b]. In general, gloss judgments are more consistent under natural illuminations [Fleming et al. 2003], and this consistency can improve further in the presence of additional cues such as color, motion, and disparity [Wendt et al. 2010]. Complex interactions between shape and illumination also play a role in the perception of gloss. Surfaces with strong curvatures lead to the spatial compression of highlights or lowlights, so that their intensity respectively increases or decreases [Kim et al. 2012]. Bumpier surfaces appear glossier, but this relation might be non-monotonic, as for higher bump magnitudes perceived gloss might decrease [Landy et al. 2008; Marlow et al. 2012]. Faul [2019] concludes that surface curvatures are also relevant for the case of gloss perceived from Fresnel effects. It has also been shown that the discrimination of subtle gloss changes is poor for tessellated surfaces, and might be lower for simple surfaces like a sphere than

for surfaces with more complex curvatures such as blobs [Vangorp et al. 2007]. Therefore, observers' ability to match glossy properties might be altered given different surface geometries [Havran et al. 2016; Nishida and Shinya 1998; Vangorp et al. 2007], but also under different lighting conditions [Doerschner et al. 2010; Leloup et al. 2010; Olkkonen and Brainard 2010; Pont and te Pas 2006].

Lightness. Matte materials typically maintain relatively constant lightness (albedo) perception [Olkkonen and Brainard 2010; Zhang et al. 2020a]; however, a positive correlation between lightness and the standard deviation (akin to contrast) of environment map luminance has been observed [Motoyoshi et al. 2007]. Additionally, lightness perception seems to be positively correlated to the highest percentiles of luminance values present in the environment [Toscani et al. 2017]. Geometry also plays a role in lightness perception [Toscani and Valsecchi 2019]: perceived lightness may increase when a visible surface region is directly exposed and perpendicularly oriented to dominant light sources. This increases the maximum brightness in the image that anchors the perception of lightness for less exposed regions [Gilchrist et al. 1999]. Lightness perception for glossy objects is far more complex [Toscani et al. 2017]. There is some evidence that lightness and gloss interact, so that darker surfaces are perceived as more glossy, and conversely glossy surfaces are perceived as darker [Chadwick and Kentridge 2015; Hunter and Harold 1987; Motoyoshi and Matoba 2012; Pellacini et al. 2000]. However, in other studies, only a weak impact of specularity has been found due to the apparent ability of the human visual system (HVS) to discount specular highlights [Olkkonen and Brainard 2010; Todd et al. 2004; Toscani et al. 2017].

These works are usually targeted to carefully test a particular effect, and therefore the experimental stimuli are reduced to a few controlled materials, typically rendered with basic analytic reflectance models. Further, different works often use different subsets of stimuli, making it difficult to compare them and extrapolate their insights to general scenarios. As discussed in recent works [Hansmann-Roth and Mamassian 2017; Landy et al. 2008], experiments on simultaneous interactions between multiple factors are relatively sparse. In this work, we perform a large-scale crowdsourced study, in which we include a wide variety of materials (520), as well as a carefully selected set of illuminations (9) and geometries (9). We measure different properties of perceived reflectance, perform an in-depth analysis of our collected data, and discuss the effects of illumination and geometry on perceived appearance with respect to existing works, providing a unified view of such effects.

2.2 Perceptual material embeddings

A step forward in understanding perceived appearance is to analyze the relationships between different materials and derive low-dimensional perceptual embeddings useful for many applications, such as BRDF editing or similarity metrics for BRDF compression, fitting, and gamut mapping. A majority of existing works focus on gloss perception.

Pellacini et al. [2000] employ a multi-dimensional scaling analysis and show that gloss appearance can be explained by two perceptual dimensions that roughly correspond to contrast gloss and distinctness-of-image gloss [Hunter 1937; Hunter and Harold 1987].

They correlate these dimensions with the parameters of the Ward reflectance model [1992] to make them perceptually uniform. In the same spirit, Wills et al. [2009] propose another 2D embedding for measured BRDFs [Matusik et al. 2003] and correlate the perceptual dimensions with parameters of multiple reflectance models. Recently, Vangorp et al. [2017] focused on analyzing BRDF materials with two specular components and showed the importance of the haze gloss dimension [Hunter 1937]. Further, Toscani et al. [2020] suggests that there are three main perceptual dimensions representative of reflectance: gloss, lightness, and metallicness. Going beyond gloss-related parameters, Serrano et al. [2016] proposed an intuitive control space for material appearance, which was later employed for material gamut mapping [Sun et al. 2017]. In a similar spirit, Hu et al. [2020] proposed to use deep representations for manipulating measured materials.

The above works are limited to a fixed geometry (sphere [Pellacini et al. 2000; Serrano et al. 2016], bunny [Wills et al. 2009], and blob [Vangorp et al. 2017]) and few lighting scenarios. Therefore, it is unclear how the proposed perceptual embeddings generalize for departures from perceived gloss constancy as discussed in Sec. 2.1.

2.3 Image-based perception modeling

Classic methods. Initial efforts suggested that simple image statistics derived from luminance histograms such as the standard deviation, kurtosis, skewness [Motoyoshi et al. 2007; Sharan et al. 2008], or image sub-band histograms [Motoyoshi and Matoba 2012] correlate well with surface glossiness and lightness. More complex heuristics combining multiple cues such as highlight coverage, sharpness, and contrast have been also proposed to explain perceived gloss [Marlow and Anderson 2013]. However, follow-up work [Anderson and Kim 2009; Chadwick and Kentridge 2015; Fleming 2014; Olkkonen and Brainard 2010] has shown that these methods do not generalize across existing perceptual data. Recent research suggests that the HVS cannot perform inverse optics computations that would revert the complex image formation process so that surface reflectance properties are disentangled [Anderson 2011]. Therefore, highly non-linear encodings of the visual input may be the key to explain material perception processes [Fleming 2014].

Deep features. Fleming [2014] advocated for data-driven solutions for appearance modeling instead of computing physical parameters. It has been recently argued that complex representations of the visual input might be needed for explaining perceptual processes [Delanoy et al. 2020; Fleming and Storrs 2019; Lagunas et al. 2021]. Deep learning offers a framework for computationally modeling such processes. In a concurrent work, Storrs and Fleming [2020] use rendered images of a flat surface with variable bump structures, illuminated by different environment maps for unsupervised training of a variable autoencoder (PixelVAE). Although the network is not trained with perceptual data, its latent vectors correlate well with human gloss ratings when observing such images. Schwartz and Nishino [Schwartz and Nishino 2019] also make use of deep features to discover a set of representative material traits by training material classifiers. The closest work to ours is that of Lagunas et

al. [2019]: They trained a deep architecture (ResNet) that can effectively estimate appearance similarity distances between different materials in images.

Following this trend, we leverage deep features by training a deep architecture with our extensive set of perceptual data. We show that this model can effectively predict perceptual reflectance attributes, and demonstrate that its behavior is robust in challenging scenarios. We believe our work is complementary to that of Lagunas et al. While their work focuses on similarity as a whole, we focus on particular material properties that may drive this similarity. Further, in their work they only collect subjective data for a single geometry and illumination, hence our data is better suited for analyzing the interactions between these two factors. We discuss how our two perceptual spaces relate in Sec. 5.4.

3 CROWDSOURCING PERCEPTUAL DATA

Our goal is to characterize the perceived appearance of a wide variety of materials under different geometries and illuminations, for which we have rendered 42,120 different images covering different materials, illuminations, and geometries. We summarize in this section our stimuli and procedure for crowdsourcing perceptual data, please refer to the supplementary material for more details.

3.1 Stimuli

Materials. We include 520 measured BRDFs, focusing on isotropic appearances (389 BRDFs). We also include (131) anisotropic materials for further analysis in Sec. 4.2. We use isotropic materials from the MERL [Matusik 2003] and RGL [Dupuy and Jakob 2018] datasets, and complement our selection with edits over MERL BRDFs [Sun et al. 2018]. Anisotropic materials are obtained from the UTIA [Filip and Vávra 2014] and RGL datasets. From the UTIA dataset, we select 50 materials with high anisotropy effects and compute their isotropic equivalents [Filip 2015], in order to allow for comparisons between perceived attributes with and without anisotropy (Sec. 4.2).

Illuminations. We focus on real-world natural illuminations [Fleming et al. 2003] and include nine illumination maps selected to produce different appearances. We compute the high-frequency content [Brossier et al. 2004] of a set of 412 high-quality HDR environment maps[1] and select nine of them, aiming for variety, since this feature has been suggested to have an effect on material recognition [Dror et al. 2004; Lagunas et al. 2021] and appearance [Zhang et al. 2020a]. In this selection, we also ensure a variety of indoor/outdoor environments, and different types of light sources. We normalize the exposure of the environment maps such that the integral of the luminance is the same for all illuminations [Fleming et al. 2003]. Fig. 2, first row shows the selected environment maps.

Geometries. We select nine geometries with varied features. We include geometries widely used in the graphics community (*bunny*, *buddha*, *cylinder*, *teapot*) and in material perception research. The *blob* shape is reported as one of the best shapes for material discrimination [Vangorp et al. 2007], and the *ghost* shape is optimized for single image material comparison [Havran et al. 2016]. We add

the *dragon* and the *statuette* as geometries with steep gradients and higher complexity. Fig. 2, third row shows the selected geometries.

3.2 Methods

Attributes. The work of Toscani et al. [2020] suggests that three main perceptual dimensions are representative of diffuse and specular reflection: gloss, lightness, and metallicness. These three dimensions are also present in the work of Serrano et al. [2016]. Consequently, we chose them as our main attributes. For gloss in particular, previous works [Pellacini et al. 2000; Wills et al. 2009] have suggested that the two main dimensions of perceived gloss are related to distinctness-of-image gloss (or sharpness) and contrast gloss [Hunter 1937; Hunter and Harold 1987], so we also consider these two attributes. Finally, since we are including anisotropic materials in our dataset, we add this attribute to our list. Our final list consists of six attributes: *glossiness, sharpness of reflections, contrast of reflections, metallicness, lightness,* and *anisotropy*.

Participants. A total of 3,217 participants (37% female, average age 38.1, $\sigma = 11.96$ years old) participated in our perceptual survey. From these participants, 19% claimed above intermediate experience in computer graphics, 17% claimed to have experience with design or modeling software, and 16% claimed artistic knowledge. The study was mostly carried on computer displays (92.5%), although a few users used other devices (such as tablets).

Procedure. We implemented a web interface for performing our perceptual survey (refer to the supplementary material for more details). We perform our study in a crowdsourcing platform (Amazon Mechanical Turks), which has been proven an effective tool for studies involving graphical perception experiments [Heer and Bostock 2010], and material perception studies [Bousseau et al. 2013; Lavoué et al. 2021; Serrano et al. 2016]. We then follow a rigorous procedure to deal with noise and improve robustness of the data, including a tutorial, training with examples [Doroudi et al. 2016], and control images for anchoring scales and discarding unreliable users [Cunningham and Wallraven 2011]. Before starting the survey, the task and different attributes were explained to the participants (detailed explanations and examples, highlighting relevant areas in the images). To allow them to familiarize themselves with the interface and to minimize worker unreliability [Welinder et al. 2010] a short training session with obvious examples (not part of our tested stimuli) was presented. Participants were asked to rate each of the six attributes for each given image, focusing on the main object rather than on the background. Similar to previous works [Serrano et al. 2016; Zhang et al. 2020a] we chose a Likert rating task, which offers a good trade-off between the amount of trials and difficulty of the task. We chose to use a seven-point Likert scale since it offers a good balance between granularity and complexity [Nunnally 1994]. The extremes of the scales were labeled as "not at all" (low rating) and "a lot" (high rating). Each perceptual survey consisted of 20 images selected randomly from our pool and 3 control images to detect inattentive or malicious participants. We rejected 11% of the participants, and finally obtained 215,680 valid responses, from which we ensure that each image has at least 5 views.

---

[1]https://hdrihaven.com/

4 DATA ANALYSIS

In this section we first perform a statistical analysis on the effects of illumination and geometry on perceived reflectance properties (glossiness, lightness, contrast of reflections, sharpness of reflections, and metallicness). Then, based on our observations, we select a subset of geometries and illuminations to analyze their pairwise interactions. Later, in Sec. 4.2 we analyze anisotropy effects focusing on the subset of materials included for this purpose (anisotropic-isotropic pairs). Additionally, in the supplementary material we provide an analysis of image statistics computed over the final images and show that they can not fully predict all observed variations in our perceptual attributes. During our analysis, we focus on connecting our insights to previous knowledge in the literature (usually scattered and limited to different subsets of materials, illuminations, and shapes), and discussing how existing findings generalize for complex interactions of diverse shapes and illuminations.

4.1 Statistical analysis

We investigate whether geometry and illumination have a significant impact on users' ratings. Our dependent variables (attributes) are 7-point Likert items and the collected observations are not normally distributed ($p < 0.05$ for the Kolmogorov-Smirnov test). Therefore in order to analyze our data we use a cumulative link mixed model (CLMM) for ordinal-scale observations [Agresti 2003; Christensen 2018; McCullagh 1980]. Cumulative link models are tailored to handle the ordered nature of Likert items, which avoids treating these variables as quantitative. In these models, for each level of the ordinal response, the cumulative probability of being in such level or lower is modeled in the form of cut-off points that separate the levels of the ordinal response. We include in the supplementary material the parameters of the fitted models for each attribute as well as the predicted rating probabilities by geometry and illumination. For post-hoc analysis we use Estimated Marginal Means with Bonferroni correction for multiple comparisons. We include the *illumination* and *geometry* as fixed factors, and the *user* and *material* as mixed effects. This allows us to effectively account for residual differences due to user diversity and variations in experimental setup (e.g., different displays for visualization). In all our tests, we fix the significance level α to 0.05. Fig. 2 summarizes the estimated effects for geometry and illumination.

4.1.1 Effect of illumination. There is a significant effect of the environment maps in the perceived attributes ($p < 0.001$), Fig. 2 shows this effect in the mean response for each attribute, adjusted by other variables in the model. Note that the estimated marginal mean effects show relative changes with respect to the average model: they account for the estimated effects of illumination in isolation, across all variety of geometries and materials. Based on the categorization we used for selecting our environment maps, we can see that for the selected illuminations, in general increasing the high frequency content of the environment map increases the perceived attributes. This is the case for all gloss-related attributes (glossiness, contrast gloss, sharpness of reflections, and metallicness). Interestingly, for most illuminations, different features still lead to similar gloss perception: environment maps from (A) to (F) contain different frequency content and dynamic range, indoors and outdoors environments, or

presence and absence of point-like light sources, yet the changes in perceived glossiness only differ in 0.5 points in the scale, which translates to 6% of gloss variation. Interestingly, metallicness is less affected by lighting variations. For perceived lightness, the environment maps (A, B, C, F), and (G) lead to very similar ratings: these environments, in general, have very large area illuminations, and very few point-like lights, if any. High lightness seems to be associated very strong localized light sources (H, I), while low lightness ratings seem related to back-lighting (environment map D).

4.1.2 Effect of geometry. There is a clear effect in the perceived attributes for changes in geometry ($p < 0.01$). Geometry plays a key role in how a surface reflects the incoming light and, therefore, it has a strong impact on the final appearance of the material. Fig. 2 shows the effect of geometry in the mean response for each attribute, adjusted by other variables in the model. The trend is similar for all attributes. In general, the *dragon* produces lower ratings of gloss-related attributes, followed by a cluster composed of *statuette*, *cylinder*, *buddha*, *bunny* and *teapot*, followed by *sphere*, then *blob*, and then *ghost*. These results suggest that smooth surfaces tend to produce higher ratings, while very complex geometries with steep gradients tend to produce lower ratings for reflectance-related attributes. For perceived lightness and metallicness, a similar trend can be observed, although the influence of geometry is less pronounced. An interesting effect that we observe is that the confidence intervals are larger for glossiness than for lightness. A potential explanation for this is that the perception of lightness is based primarily on diffusely reflected light [Todd et al. 2004], leading to relatively stable lightness judgmenets [Zhang et al. 2020a].

4.1.3 Interactions of geometry and illumination. For exploring the interaction of geometry and illumination, computing all possible pairwise post-hoc tests would yield 81 combinations (9 illuminations × 9 geometries), hindering the convergence of our CLMM statistical model. Therefore, we work with a subset of our data and show our results in Fig. 4. To support our discussion, we show in Fig. 3 different statistics computed for the illumination and the geometry against the glossiness, lightness, and metallicness estimated marginal effects. For environment maps, we compute the following statistics: standard deviation (I_{std}), skewness (I_{skew}), kurtosis (I_{kurt}), light-source coverage (I_{area}, where small values correspond to point-like light sources), contrast between light-sources and the rest of the image (I_{range}), and high-frequency content (I_{hfc}) [Brossier et al. 2004]. For geometries, since users only could see the shape from the rendered point of view, we compute the following statistics over the normal maps (relative to the rendered viewpoint): sum of gradients (G_{grad}), size of the largest smooth area (G_{area}), and high-frequency content (G_{hfc}). Since gloss-related attributes share similar behaviors, we focus on glossiness and lightness, which are arguably the most prominent attributes discussed in the literature [Chadwick and Kentridge 2015; Fleming 2017].

Glossiness. The pattern of gloss perception purely due to lighting changes for a fixed geometry is quite similar across different geometries (Fig. 4, bottom left). Two exceptions can be noticed: (1) for the combination of highly directional lighting (I) and the *blob* a smaller glossiness boost is observed for this particular lighting
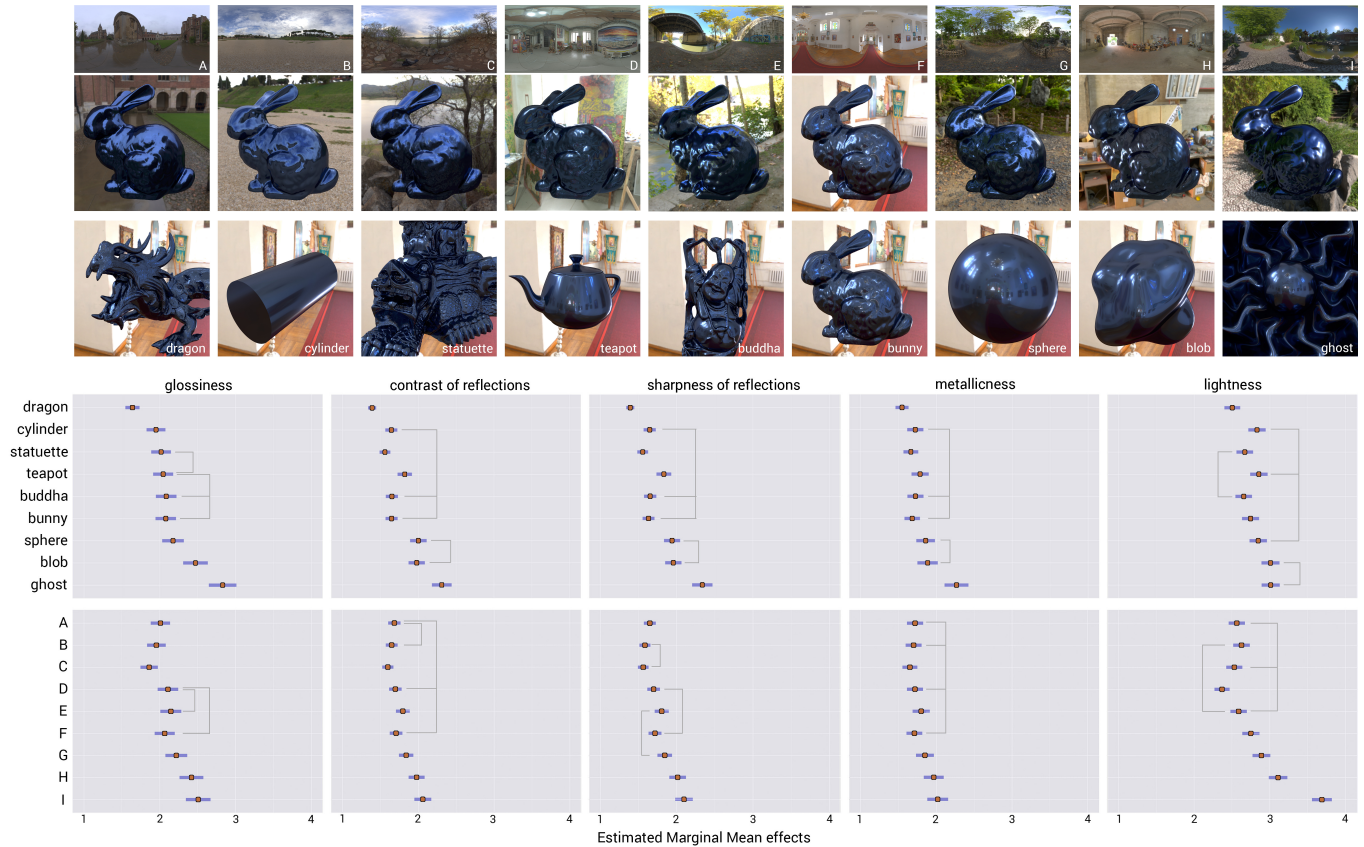
Fig. 2. Top: Examples of our nine illuminations (first and second row) and geometries (third row) with the material *blue-metallic-paint2* from the MERL dataset. Bottom: Effect of geometry (fourth row) and illumination (fifth row) in perceived attributes. Estimated Marginal Means approximate the mean response for each factor, adjusted by other variables in the model. The orange points mark the mean value, while blue bars indicate a 95% confidence interval. We mark with gray lines clusters of factors that do not yield statistically significant differences.

when compared to other geometries, and (2) for the combination of area lighting (G) and the *ghost* an inverse trend holds. Our data confirms literature findings showing that for any geometry gloss perception is reduced for slowly changing lighting (low I_{hfc}), and it becomes stronger for more directional and higher intensity light sources (low I_{area}) [Dror et al. 2004; Pont and te Pas 2006; Zhang et al. 2020a,b, 2015].

The perceived glossiness in Fig. 4, top left shows a strong impact of the surface geometry, where the highest gloss is always perceived for the *ghost* which is a combination of a smooth spherical shape (high G_{area}) with a wavy region roughly perpendicular to the viewpoint and strongest illuminants. This results in a large highlight coverage: The spherical shape reproduces highlight sharpness, while the wavy part strongly contributes to the highlight contrast. The *blob* also features a smooth surface, but the highlight coverage and its contrast is reduced, decreasing the overall perceived gloss. Gloss is further reduced for less smooth geometry featuring small surface details as in the case of *buddha*. The *dragon* is the least glossy due to its tessellated-like surface that is poor in reproducing the discussed highlight features. This is in agreement with observations made by

Vangorp et al. [2008; 2007] who reported reduced glossy material discrimination for tessellated surfaces.

Lightness. A pattern of lightness rating (Fig. 4, bottom right) can be seen across different environment maps that roughly repeats across all objects. Environment map (I) that features a high-intensity light source (high I_{range}) of small spatial extents (low I_{area}) results in the highest lightness. This is in agreement with observations made by Toscani et al. [2017] who claimed that lightness perception is strongly correlated to the highest percentiles of luminance values in environment maps. This environment map also features the highest contrast, which has been also reported to impact perceived lightness [Motoyoshi et al. 2007]. The spatial extents of key illuminants (I_{area}) gradually increase in (G), (E) and then in (A), which results in reduced lightness [Toscani et al. 2017].

The lightness ratings in Fig. 4, top right make a clear pattern across different geometries that roughly repeats for all environment maps. The *blob* shape, due to its slowly changing curvature, and the *ghost* shape, due to its perpendicular positioning with respect to the view direction, are likely to be directly exposed to the highest intensity regions in the environment maps. This in turn results in
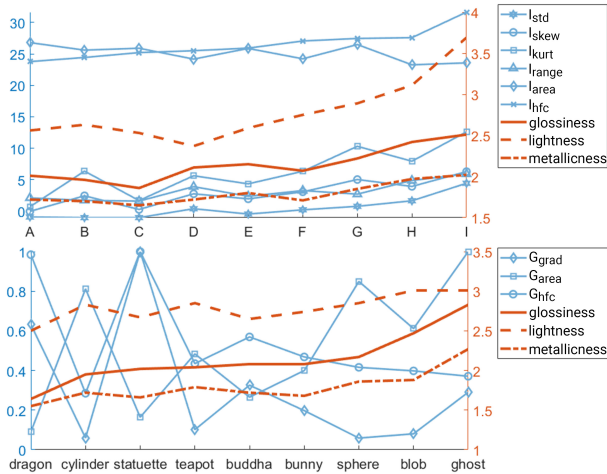
Fig. 3. Estimated marginal effects for glossiness, metallicness, and lightness for different illuminations (top) seem to be tied mostly to environment map statistical moments, as well as the range (I_{range}) and high frequency content (I_{hfc}). Changes for different geometries (bottom) seem to be tied to smooth areas (G_{area}), gradients (G_{grad}), and frequency content (G_{hfc}).
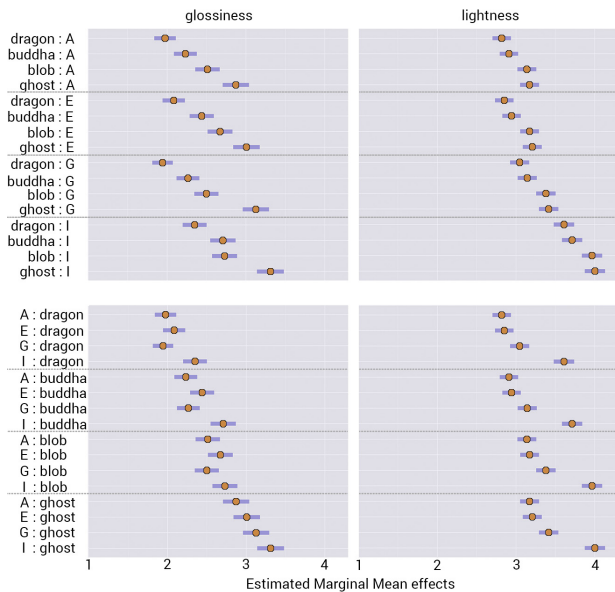

Fig. 4. Estimated Marginal Means effects of glossiness and lightness for the interaction of illumination and geometry in a subset of our data.

bright coherent regions in the image, which are the most informative for lightness perception [Gilchrist et al. 1999; Toscani and Valsecchi 2019; Toscani et al. 2017]. The availability of such regions and their brightness might be reduced for the *dragon* and the *buddha* due self-shadowing, and sparser, more fragmented regions that are perpendicularly exposed to the highest intensity regions in environment maps [Toscani and Valsecchi 2019]. This results in the overall lower lightness for such geometries.

4.2 Anisotropy

Due to the limited amount of measured anisotropic materials in existing datasets, our dataset is biased towards isotropic materials. Therefore, in order to analyze anisotropy, we select 50 anisotropic materials from the UTIA dataset, and their isotropic counterparts. First, we want to analyze the effect of illumination and geometry in the perception of anisotropy. Second, we want to analyze if the presence of anisotropy affects the perception of gloss, lightness, and metallicness. Similarly to the previous section, we make use of cumulative link mixed models. In this case, we additionally include as a fixed factor a binary variable that encodes whether each image contains an anisotropic material, and the interactions of this variable with illumination and geometry. For the materials tested, we found that anisotropy ratings are only significantly different for isotropic-anisotropic pairs for the *sphere* (p < 0.001), and for the illuminations (H) (p = 0.0012) and (I) (p = 0.0039). Our results are consistent with those of Filip et al. [2015]: sensitivity to anisotropy decreases for complex shapes and illuminations. Interestingly, we found that some environment maps (H and I), even though complex, allow for consistently identifying anisotropy. These two environment maps correspond to the lowest I_{area}, and the highest I_{range} (Fig. 3), i.e., they contain point-like light sources: We show that these two statistics are highly correlated to the perception of anisotropy. We did not find significant differences in glossiness, metallicness, and lightness under the presence of anisotropy.

After these results, we consider the anisotropy attribute only for the presented analysis and we do not include it for prediction in the following section. Unfortunately, current datasets of measured materials do not contain enough variety of anisotropic materials, and therefore, our data is biased towards isotropic materials. Additionally, most of these anisotropic materials show very subtle effects, and are hard to see for most geometries even for human observers, as we have shown in this section. This produces unreliable results for anisotropy estimation (see Fig. 6).

5 PREDICTING PERCEPTUAL ATTRIBUTES

In this section we describe our approach for predicting material appearance attributes from images. Note that, as discussed in Sec. 4.2, we do not consider anisotropy as an attribute for prediction. To our knowledge, our work is the first that allows for estimating perceptual material appearance attributes, such as glossiness and lightness, taking into account the final appearance including the joint effects of geometry and illumination. We then show that our predictor is robust under different scenarios and discuss failure cases of our predictor. Finally, we discuss our perceptual space of attributes in the context of existing perceptual embeddings.

5.1 Implementation Details

We use as input for training our collection of rendered images, together with the median of the collected ratings for each image as a robust statistic (recall that each image has been viewed at least five times). For training, we fully remove from the dataset all images containing the *bunny* and all images containing the (A) illumination, leading to 33,280 images (520 BRDFs x 8 geometries x 8 illuminations), augmented to more than 3,328,000 images using crops, flips,

Table 1. Evaluation results with the *bunny* dataset (set A) and extrapolation results with the additional validation dataset (set B). Mean absolute error ± standard deviation for our two tested architectures and loss functions.

	Evaluation (Set A)		Extrapolation (Set B)	
Network structure	MSE loss	MSLE loss	MSE loss	MSLE loss
ResNet52	0.1069 ± 0.0866	0.1021 ± 0.0840	0.1718 ± 0.0978	0.1688 ± 0.0931
VGG16	0.0957 ± 0.0818	0.0955 ± 0.0831	0.1672 ± 0.0984	0.1694 ± 0.1046

Table 2. Evaluation (set A) and extrapolation (set B) results for each of our five predicted attributes with our selected architecture (VGG16, MSE loss). Mean absolute error ± standard deviation for each attribute.

	Glossiness	Contrast of reflections	Sharpness of reflections	Metallicness	Lightness	All
Evaluation (Set A)	0.0914 ± 0.1122	0.0736 ± 0.1068	0.0770 ± 0.1103	0.08568 ± 0.1118	0.1329 ± 0.1177	0.0957 ± 0.0818
Extrapolation (Set B)	0.1905 ± 0.1564	0.1647 ± 0.1375	0.1591 ± 0.1422	0.1846 ± 0.1420	0.1380 ± 0.0982	0.1672 ± 0.0984

shifts, rotations, scaling, and adding Gaussian and Poisson noise. We select the *bunny* and the illumination (A) because they show certain redundancy with the *buddha* and the ilumination (B) respectively: this way we make sure that we are not missing interesting effects in the training data. When performing augmentation techniques, we ensure that at least 80% of the geometry remains visible, so that the ratings are still representative of the images. We train for 30 epochs with batch-size 4 and input resolution 512×512. The optimization is performed using the Adam optimizer [Kingma and Ba 2014] with an initial learning rate set to 10^{-5}. Since we are interested in regression and not classification, we remove the last layer of the network and introduce a fully-connected layer which outputs a 5-dimensional (attribute) vector. We test two different architectures widely used for extracting image features: VGG16 [Simonyan and Zisserman 2014; Xu et al. 2018], which stacks convolutions with non-linearities, and ResNet52 [He et al. 2016; Lagunas et al. 2019], which includes a residual block allowing to train very deep networks with outstanding performance. Since we are interested in regression predictive modeling (predicting continuous attributes from image features), we use the Mean Squared Error (MSE) as a loss function (Eq. 1):

$$L_{MSE} = \frac{1}{N} \sum_{i=1}^{N} (\widehat{a}_i - a_i)^2 \tag{1}$$

where, $\widehat{a}$ is the 5-dimensional vector containing the predicted attribute values, a is the vector of ground truth user ratings, and N is the number of attributes. Additionally, we test the Mean Squared Logarithmic Error (MSLE), which has been widely used in regression problems with skewed distributions [Dawson et al. 2014] (Eq. 2):

$$L_{MSLE} = \frac{1}{N} \sum_{i=1}^{N} (\log(\widehat{a}_i + 1) - \log(a_i + 1))^2 \tag{2}$$

5.2 Validation

We evaluate our model with the set of images containing the *bunny* geometry and the (A) illumination (set A). Note that, since we left out of training all images containing either of these, our network has never seen the *bunny*, nor any image with this illumination. Since our goal is that the predicted attributes are as close as possible

Fig. 5. Prediction results (predicted / ground truth) for the additional validation dataset for glossiness (G), lightness (L), sharpness of reflections (S), contrast gloss (C), and metallicness (M). From left to right: examples of images from Lagunas et al., analytic BRDFs, and real materials. Note that in all these cases, the predictor has never seen the tested materials, geometries, nor illuminations.

to the measured attributes, and in order to evaluate fairly our two different loss functions, we use the Mean Absolute Error (MAE) to evaluate our results (Eq. 3):

$$L_{MAE} = \|\widehat{a}_i - a_i\| \tag{3}$$

Tab. 1 shows the evaluation results for our two tested architectures and loss functions: all configurations yield similar results for our test set, with VGG16 slightly outperforming ResNet.

In order to test the extrapolation abilities of our predictor, we perform an additional evaluation with challenging cases, for which we collect a new set of data (set B). We include: 20 new edited BRDFs from MERL [Sun et al. 2018] and 20 analytic BRDFs (none of these 40 materials has been previously seen by our predictor) under five new environment maps and five new geometries; 20 real pictures of materials from the Flickr Material Dataset (fabric, metal, plastic, and leather categories) [Sharan et al. 2014]; and 18

rendered images containing new geometries and illuminations from Lagunas et al. [2019]. For testing the consistency of our predictor, we also include 20 images containing four scenes with different geometries and illuminations in which we rotate the geometry for testing different viewpoints (5 rotations for each configuration). For collecting ground truth data for this new validation dataset, we follow the same procedure described in Sec. 3. Tab. 1 shows the results for our two tested architectures and loss functions with the MAE metric, and Fig 5 and 7, top show examples of our predictions for this dataset (please, see Sec. 2 in the supplementary material for more examples, including different attributes). We observe that the VGG16 architecture with MSE loss performs slightly better for this task, therefore, we choose this configuration as our final predictor. Finally, we show in Tab. 2 the prediction error with our selected architecture for each of our five attributes.

Failure cases. In this additional validation dataset, we include several challenging cases, including real pictures of patterned cloth, and new materials. In general, our predictor performs well in most cases, including the challenging case of real pictures with diverse materials such as cloth, metals, or plastics. We found two clear cases in which our predictor fails: images with strong patterns and no visual cues of reflectance (e.g., total absence of highlights), and extremely glossy surfaces in which the resulting reflections are almost a perfect mirror. Examples of these are shown in Fig. 6, top. For the former, we believe that the gradients produced by the cloth patterns interfere with the prediction, while for the latter, we hypothesize that the mirror-like reflections are interpreted as background irrelevant features (as if light were passing through), instead of as glossy features. We also include in Fig. 6, bottom examples of unreliable anisotropy prediction due to the insufficient amount of available anisotropic materials, and the subtlety of their effects. Although the predictor is able to detect the anisotropic materials (center, right), it also produces high values for anisotropy for similar but isotropic materials (left).

5.3 Consistency evaluation

We further test our predictor to analyze its behavior under changes in geometry, illumination, point of view, and image features. For conciseness, we discuss examples with the most relevant attributes in each case, please refer to the supplementary material for more results. Note that the geometries and illuminations used in this section (except the sphere, which we use as baseline for comparisons) were never seen during training.

Changes in point of view. In Fig. 7, top we show different views for the same material in the same geometry under the same illumination, and our predictions for glossiness and lightness. While the predictions for lightness remain constant, glossiness predictions slightly vary while the geometry rotates. This is to be expected, since the point of view also plays a role in reflectance perception, i.e., different orientations of the geometry interact with the illumination, producing different amount and strength of reflections as the geometry rotates, specially around the armadillo's face and nose.

Geometry complexity. Fig. 7, center shows that the predicted glossiness and lightness increases as the complexity of the geometry
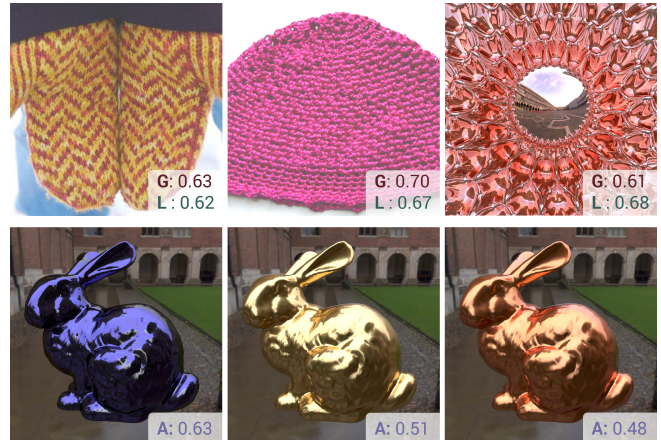

Fig. 6. Top: Three examples of failure cases in which our predictor does not accurately predict glossiness (G) and lightness (L). Images in which patterns are more relevant than reflectance cues (left, center), and almost perfect mirror reflections (right). Bottom: Due to the limited amount of available anisotropic materials and their subtle effects, our predictor is not able to fully learn the image features that characterize anisotropy: while it does detect the two anisotropic materials (center, right), it shows similar anisotropy predictions for a similar but isotropic material (left).
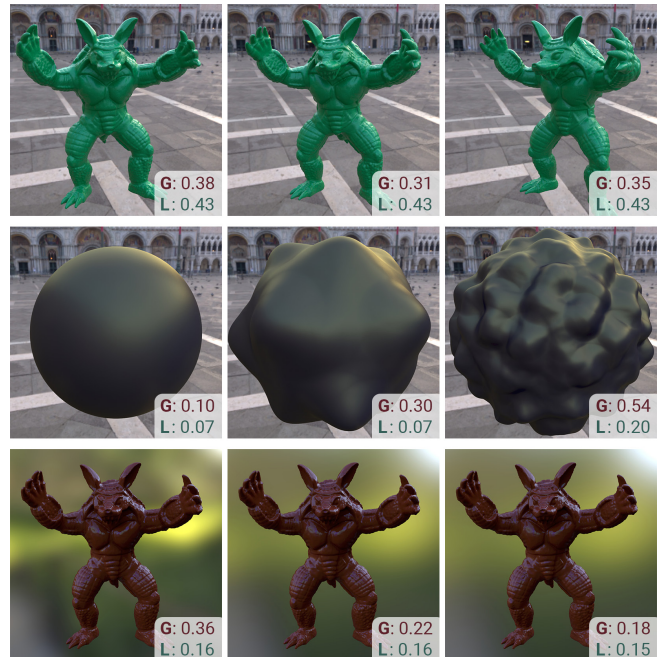

Fig. 7. Changes in glossiness and lightness predictions for different points of view (top), increasing surface relief (center), and increasingly blurry environment map (bottom). Our predictor is in agreement with visual changes and the insights derived in our analysis.

increases from a simple sphere, to an increasingly bumpy sphere. We can observe that this behavior is consistent with our collected data: the sphere was also rated with lower values of glossiness and

Fig. 8. Influence of the tone mapping operator in glossiness predictions. Our predictor is consistent with visual changes in the image: different operators produce different contrast in the final image, which in turn, affects the perception of gloss [Wiebel et al. 2015].

lightness than the blob (which is similar to the bumpy spheres in terms of geometry features). The results of our predictor are also in agreement with previous work on gloss perception for a similar scene setup [Marlow and Anderson 2013, Figure 6].

Illumination frequencies. In Sec. 4 we discuss how different illumination characteristics may affect the perception of glossiness and lightness: as in previous literature [Adams et al. 2018; Zhang et al. 2020a], we observe that high frequencies seem to be related to higher ratings for glossiness. We show in Fig. 7, bottom the predictions for these attributes under and increasingly blurry environment map. Similarly to our data, the prediction for glossiness decreases as the high frequency content of the environment map decreases.

Influence of tone mapping. Previous work has shown that the choice of particular tone mapping operator has a relatively small impact on human judgments of gloss when applied to an object with visible background, although it may have a larger effect when images are shown without any additional context [Adams et al. 2018]. In Fig. 8 we show glossiness predictions for the same image, tone mapped with two different exposures (gamma-exposure TMO), and with the Reinhard [2002] and Durand [2002] TMOs. Our predictor is consistent with visual changes induced by the different TMOs. This small variations are expected also for human observers, since different TMOs produce different contrast, and this effectively affects gloss perception [Wiebel et al. 2015].

Real Images. We validate our predictor with ground truth annotations for 20 real images from the Flickr Material Dataset in Sec. 5.2. In this section, we further illustrate the performance of our model in real pictures with a subset of 40 images from the MINC-2500 dataset [Bell et al. 2015] (10 for each of four categories: fabric, leather, plastic, and metal) illustrating both success and representative failure cases. Predictions for the attributes glossiness and lightness are shown in Fig. 9. We mark with red squares failure cases that we discuss in the following. In general, our predictor performs well for leather materials, since usually they do not exhibit strong patterns, they show uniform colors, and reflectance cues are very apparent. The same applies for metallic materials, with the exception of a metal with very smoothly varying reflections, which our predictor scores with relatively low glossiness, almost on par with the fabrics. Due to the particular lighting direction and the high roughness of this material, we believe that reflectance cues are wrongly interpreted as smooth color variations, producing unreliable glossiness predictions. For the case of fabrics, strong fine

Fig. 9. Predictions for lightness (y-axis) and glossiness (x-axis) for 40 example real images (four categories: fabric, leather, plastic, and metal) from the MINC-2500 dataset. We mark failure cases in red, please refer to the text for an extended discussion on these cases.

patterns may overrule reflectance cues, producing unreliable predictions as also discussed in in Sec. 5.2. This is the case of the yellow fabric with fine white pattern marked in red in the figure: this result suggests that the predictor is interpreting the white pattern as highlights, thus incorrectly producing a high glossiness score. For the case of plastics, we show as failure case a plastic figurine for which our predictor scores a relatively low glossiness. While most of the image shows a matte plastic, the plastic piece corresponding to the hair is highly glossy. This is a complex case, since different plastics with different glossiness are present in the image. We believe our predictor performs reasonably, since a larger portion of the image displays a matte plastic, however, predictions in these cases with several materials within the same image are challenging. In these cases, a potential approach could be to mask out irrelevant materials for the prediction, and leave only visible the materials of interest.

In general, real pictures displaying uniform materials without strong patterns are best suited for our predictor, since the images used during training were rendered with uniform BRDFs. Images containing several materials, strong patterned fabrics, or very complex geometries may produce unreliable results. However, as shown in Fig. 9, our model is overall able to handle real images, including some challenging cases with complex geometries.

5.4 Discussion on other perceptual embeddings

In the last decade a number of works have proposed perceptual reparameterizations of material appearance, performed either directly on BRDF space (or a low-dimensional representation of the
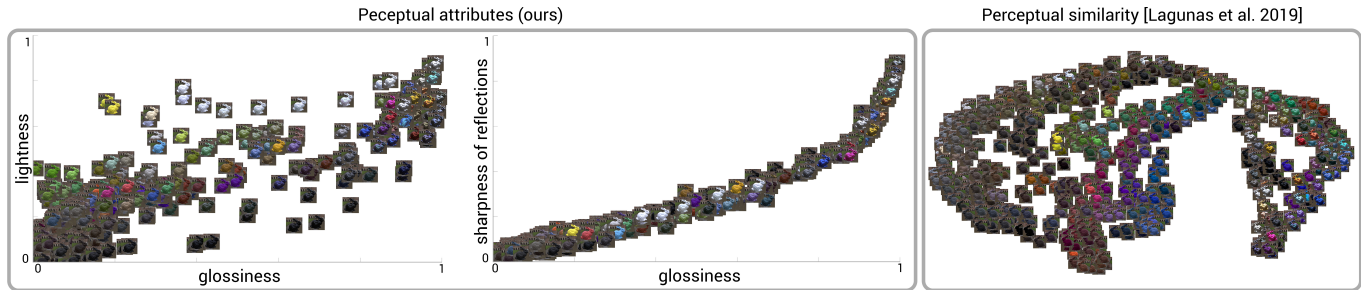
Fig. 10. Left: Perceptual embeddings generated with our predictor: lightness – glossiness, and glossiness – sharpness of reflections. Several interesting effects can be observed: First, even though glossiness and lightness are not completely orthogonal, both are very distinctive material properties. Second, while glossiness and sharpness of reflections start growing together (one cannot perceive sharp reflections if there is not sufficient glossiness), once sufficient glossiness is achieved, the relation between these two attributes becomes non-linear: glossiness tends to saturate, and then sharpness of reflections becomes the distinctive factor between different materials. Right: perceptual similarity embedding proposed by Lagunas et al. [2019]. Their embedding (UMAP non-linear principal dimensions) shows a space organized first by reflectance, and second by color as a secondary grouping factor. Note that in both cases, the tested images were not used during training.

BRDF) [Fores et al. 2012; Guarnera et al. 2018; Hu et al. 2020; Serrano et al. 2016], or in image space [Bieron and Peers 2020; Lagunas et al. 2019; Ngan et al. 2006; Pereira and Rusinkiewicz 2012; Sztrajman et al. 2019]. These reparameterizations usually take the form of perceptual embeddings, or similarity metrics. Embeddings in BRDF space are computed directly from measured BRDFs or parameters in analytical models, and do not take into account the final appearance of the material, which may substantially change due to the effects of geometry, illumination, and point of view. We show an example of this in Fig. 11: even though the final appearance of the *blue-metallic-paint* BRDF (MERL) looks very different under the two environment maps (left and right), BRDF metrics cannot account for this difference. Instead, embeddings computed in image space allow for accounting for the final appearance of materials after interacting with geometry and illumination. Ngan et al. [2006] compute L2 distances between the cube root of linear RGB rendered images. Pereira and Rusinkiewicz [2012] propose the use of L4 distances in Lab space in the context of gamut mapping. Bieron et al. [2020] recently introduced a new metric for optimizing BRDF fitting by using the CSSIM color metric. These works yield very good perceptual approximations when targeting a given environment map and geometry, however, since they are based on pixel-wise distances, they do not allow for comparing perceptual changes under different geometries and illuminations (see Fig. 11). The closest methodology to ours has been recently proposed by Lagunas et al. [2019]. In their work, Lagunas et al. learn a perceptual similarity metric by training a deep network with similarity triplets, which allows for comparing perceived material similarity between two images. In Fig. 10 we compare their perceptual space to two embeddings created using our predictor: lightness – glossiness, and glossiness – sharpness of reflections (refer to the supplementary material for more embeddings including different combinations of attributes, as well as more illuminations and geometries). As the authors discuss in their paper, their embedding shows a "*clear gradient in reflectance, followed by color as a secondary, softer grouping factor*": the effect of color and reflectance cannot be easily disentangled. This can be seen in Fig. 11, the two test images are deemed almost at the

same distance to the reference, in one case due simply to different color, and in the other due to reflectance changes. There are two fundamental differences of our work with respect to Lagunas et al.'s work. First, our embeddings allow for perceptual comparison between attributes, for example, when computing differences a single attribute (e.g., glossiness) can be targeted. In other words, it is possible to identify the properties that make two materials similar or different. This can be useful for many applications, for example, for appearance reproduction, as we will show in Sec. 6.1. Second, we can predict meaningful properties without a reference to compare (i.e., we can predict glossiness, or metallicness for a given image), whilst their similarity metric is most useful for comparing between two images. This enables a straight-forward application our predictor to applications such as perceptual editing, as we will demonstrate in Sec. 6.2. Nevertheless, we believe these two embeddings can be complementary. While the work of Lagunas et al. provides a unique measurement of similarity as a whole, our work is useful for dealing with isolated material properties.

6 APPLICATIONS

We have shown in Sec. 5.4 that our predictor can be used to generate perceptual embeddings and also as a distance metric. In this section we illustrate further applications of our attribute predictor.

6.1 Reproducing appearance

Faithfully reproducing objects' appearance is a challenging problem with applications ranging from product design, through the preservation of cultural heritage, to the creation of life-like prosthetic. A vital enabler of these applications is novel printing devices capable of depositing various materials at a high resolution [Elkhuizen et al. 2019; Luongo et al. 2019; Piovarči et al. 2017; Rouiller et al. 2013; Weyrich et al. 2009]. Such materials then mix in the human visual system, enabling to reproduce any BRDF within the convex hull of available materials [Matusik et al. 2009]. Varnishes are materials of particular interest, since they modify the surface reflectance without a change in the subsurface color [Baar et al. 2014; Samadzadegan et al. 2015].

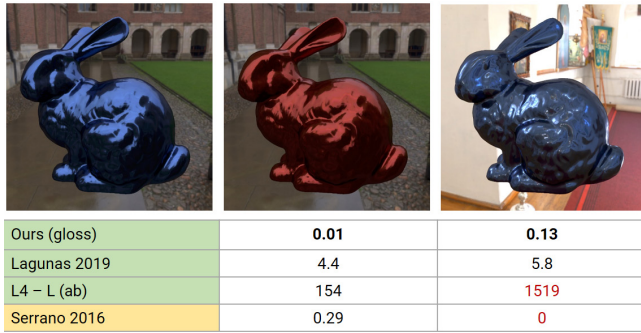| Ours (gloss)	**0.01**	**0.13**
Lagunas 2019	4.4	5.8
L4 − L (ab)	154	1519
Serrano 2016	0.29	0

Fig. 11. Measuring pairwise distances between a reference material (left), and the same material with different color (center), or differently illuminated (right). Our predictor can be used to compute differences between specific attributes, e.g., gloss as in this example. Simpler image-based metrics (green) based on pixel-wise distances, such as the L4 on Lab space (Pereira and Rusinkiewicz [2012]) tend to produce extremely large distances under drastic changes in illumination and/or geometry. The similarity embedding generated by Lagunas et al. [2019] tends to rely on both reflectance and color, therefore, it is not easy to disambiguate these two factors. Metrics in BRDF space (yellow), such as the one proposed by Serrano et al. [2016] simply cannot take into account perceptual variations due to geometry and illumination.

We take inspiration from the recent work of Piovarci et al. [2020] that proposed a novel hardware setup capable of processing off-the-shelf varnishes. Our goal is to use our predictor to enable perceptual gloss reproduction in printed samples, therefore we use a similar approach to fabricate our samples. Note that our predictor is not constrained to a single manufacturing method but rather acts as a distance metric suitable for optimization, therefore other printing approaches are possible. First, we start by 3D printing a test macroscale geometry (a cylinder) on a resin FormLabs printer. Then, we mix two acrylic varnishes (*Amsterdam Matte 115* and *Amsterdam High Gloss 113*) in various ratios to generate different reflectances, and we apply them to the test cylinders. We measure the reflectance of the different varnish mixtures in a setup similar to the one described by Piovarci et al., and fit these measurements to a Cook-Torrance model with GGX distribution [Trowbridge and Reitz 1975]. This yields a collection of available measured varnish mixtures $\mathbf{v} = \{v_i; i = 1..N\}$ (see Fig. 12, top) that we can render under different illuminations and geometries. In this case we have selected a black resin for all our samples, nevertheless other resins can be considered simply by rendering the varnishes applied over other resin colors (i.e., changing the diffuse color in the Cook-Torrance model).

We can then use our predictor to find the varnish mixture that better approximates the glossiness of a target image I_g under a new desired illumination (L_t) and geometry (G_t) by finding the varnish mixture from our collection yielding the closest glossiness to the target gloss (v_{opt}):

$$v_{opt} = \arg\min_{v_i \in \mathbf{v}} \|A(R(v_i; L_t, G_t)) - A(I_g)\|_2 \tag{4}$$

where $A(I)$ means applying our predictor (in this case, we focus on the glossiness attribute) for a given image, and R denotes a
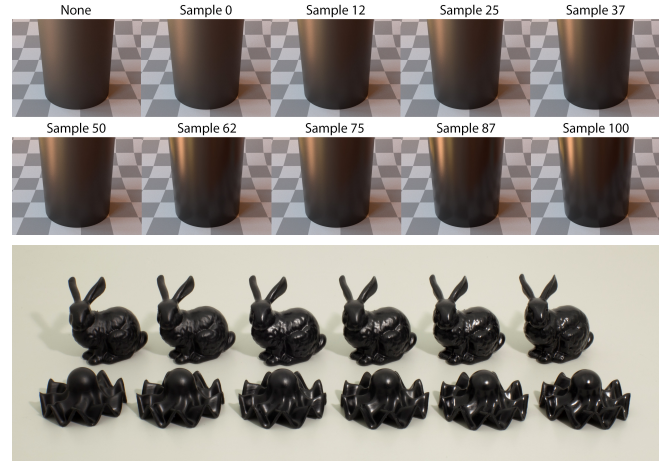

Fig. 12. Top: Collection of varnish mixtures fitted to a Cook-Torrance model, rendered under the same illumination. The labeled value corresponds to the percentage of *Amsterdam High Gloss 113* present in the mixture. Bottom: Examples of different varnish mixtures (increasing glossiness from left to right) applied over our printed samples for the *bunny* (top) and *ghost2* (bottom) geometries.

Fig. 13. Optimized varnishes for a target specular illumination and the dragon geometry (bottom) yielding the desired glossiness for sample materials from the Flickr Material Database (top).

rendering operation under the target illumination (L_t) and geometry (G_t). Since we do not depend on pixel-wise distances, we can use any given image as target gloss: Fig. 13 shows examples of this application with the Flickr Material Database [Sharan et al. 2008]. Our predictor also enables finding the optimal varnish mixture for reproducing the desired gloss under different illuminations and geometries: Fig. 1, right shows how different mixtures selected using perceptual gloss distances according to our predictor result in equivalent glossiness across illuminations and shapes.

Validation in real samples. We validate our equivalent glossiness predictions on real samples. We print two different geometries (*bunny* and *ghost2*) and capture two environment maps in a controlled setup (*diffuse* and *specular*). Fig. 12, bottom shows examples

Table 3. Results of our confirmatory study for optimized varnish selection under different geometries and illuminations. Almost all users responded that the optimized varnish mixture was closer to the reference in terms of perceived glossiness (votes are shown in parentheses).

	Diffuse illumination	Specular illumination	
	Reference	Same mixture	Optimized
bunny	Sample 87	Sample 87 (0)	Sample 50 (5)
	Sample 100	Sample 100 (0)	Sample 62 (5)
ghost2	Sample 50	Sample 50 (1)	Sample 25 (4)
	Sample 87	Sample 87 (1)	Sample 50 (4)

of these samples. We then perform a small confirmatory study with five users. They were presented with a reference mixture under the diffuse light, and two test mixtures under the specular light: the same mixture as in the reference, and a new optimized mixture using our predictor for better reproducing the reference gloss. Tab. 3 shows the results, almost all users determined the optimized varnish mixture to be closer in terms of glossiness to the reference.

6.2 Appearance editing

Integrating our predictor into a differentiable renderer allows to perform material editing using intuitive, perceptually-relevant sliders. We illustrate this application by using our predictor together with the differentiable renderer *Mitsuba 2* [Nimier-David et al. 2019]. To illustrate this application, we use the analytical built-in surface scattering model *roughconductor* ($\mathbf{x} = (\alpha, eta, k, specular\_reflectance)$) with a GGX distribution [Walter et al. 2007]. We then select a set of target glossiness values and fix all the parameters except for α, which represents the roughness of the surface microgeometry. We implant our attribute predictor A into the differentiable rendering pipeline for predicting attributes from the rendered images. This prediction is then used as the perceptual loss to be minimized such that:

$$\mathcal{L} = \|A(R(\mathbf{x}; L_t, G_t)) - T_g\|^2, \tag{5}$$

where R is the differentiable rendering operation in the target illumination (L_t) and geometry (G_t), $\mathbf{x}$ contains the BRDF analytic parameters as described above (including the surface roughness α to be optimized), and T_g is the target gloss. Fig. 14 shows the edited results obtained when adjusting the target glossiness to 0.1, 0.3, 0.5 and 0.9 from left to right (the framed image shows the fixed starting parameter at 0.72 glossiness). Fig. 15 shows editing results with target gloss set to 0.9 for a different illumination map (center) and geometry (right). Note that the optimized α value changes for each of these configurations to adapt for the effects of shape and illumination according to our predictor. More results are available in the supplementary material.

6.3 Material recommendations for scene design

Performing material selections for complex scenes with many items is an onerous process in which users typically resort to trial and error [Zsolnai-Fehér et al. 2018]. For the case of large unlabeled databases, this process may require browsing through hundreds of

Fig. 14. Material appearance editing results. From left to right edited with target glossiness: 0.1, 0.3, 0.5, 0.72 (fixed starting parameter), and 0.9.

Fig. 15. Example of BRDF editing using our predictor under a different illumination (center) and geometry (right). We set the target glossiness to 0.9. Our predictor takes into account changes in illumination and geometry and updates α accordingly.

materials [Lagunas et al. 2019]. In order to assist this process our predictor can be used to sort materials according to desired properties, such as glossiness and lightness, for target illuminations or shapes. This is illustrated in Fig. 1. Alternatively, the user can directly provide our predictor with a query value for a set of attributes so that suggestions of different materials with the desired properties are presented. Fig. 16 shows an example for high metallicness queries with low and high lightness. Note how our suggestions adapt to the different geometry and illumination based on our predictor, e.g., for the ganesha-ennis set overall brightness is very high, therefore the only plausible choices for high metallicness and low lightness are almost fully black materials. Conversely, the centaur-grace set is overall very dark, therefore the plausible choices for light materials are very bright, colorful ones.

6.4 Illumination design

Our predictor can be leveraged to estimate and visualize the effect that a given illumination may have over a range of materials and geometries. This may be useful for applications such as planning illuminations for interior design or exhibitions. We show an example of this in Fig 17: we show the glossiness distribution under three different illuminations for a set of five materials and two geometries. The *Konzerthaus* and *Autumn park* illuminations produce respectively lower and higher glossiness appearances, while the *Birchwood* illumination produces a larger range of glossiness values, boosting material differences.

7 CONCLUSION

In this work we have presented a model that reckons with the effects of shape and illumination on material perception. In particular, we have focused on six crucial dimensions of appearance according to the literature: glossiness, sharpness of reflections, contrast of reflections [Pellacini et al. 2000], lightness, metallicness [Toscani et al. 2020], and anisotropy. We have presented a large-scale dataset
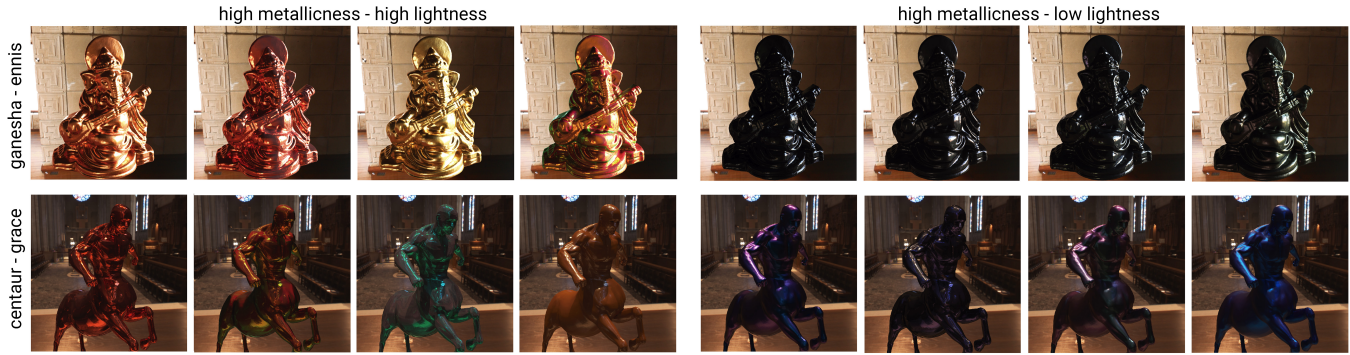
Fig. 16. Material recommendations for scene design. For different geometries and illuminations, we show suggested materials for high metallicness with high lightness (left) and low lightness (right). Our predictor is able to suggest plausible choices even for very extreme environment maps.
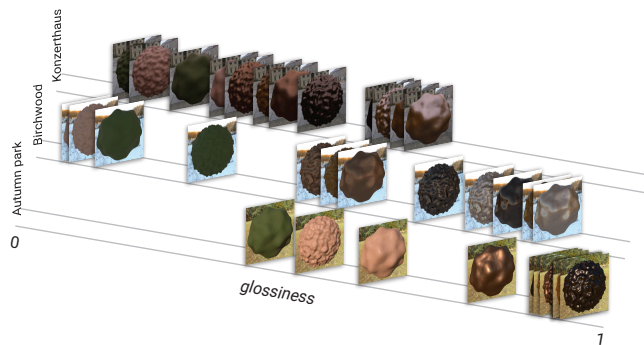

Fig. 17. Our predictor can be used to visualize the effect of a given illumination over a range of materials and geometries.

of 42,120 images featuring a variety of (520) materials, (9) geometries, and (9) illuminations, for which we collect subjective ratings for our attributes through a web-based online study. From this data, we have presented an extensive discussion on the effects of shape and illumination in material appearance, discussing our findings with those of the literature. Our data enables for the first time a comprehensive discussion on this topic taking into account a wide variety of appearances. We have analyzed several aspects of our data and focused on quantifying general trends that hold across the sampled factors, nevertheless, we believe this data is still not exhausted and it can foster new research on this topic.

Material reflectance, shape, and illumination interact in highly complex manners, and fully understanding their interactions is an open problem. For instance, further research could focus on analyzing a subset of our data: matte, semi-glossy, or fully glossy materials may interact in different ways with shape and illumination [Schmid et al. 2020]. We have shown and discussed illumination and geometry statistics that are directly linked to changes in attribute perception. However, similarly to previous literature [Fleming and Storrs 2019; Lagunas et al. 2021; Olkkonen and Brainard 2010], we observe that image statistics are not enough for fully explaining changes in attribute perception. Therefore, we have resorted to deep features and trained a simple network architecture for predicting

material appearance attributes given an input image, and we have thoroughly tested its behavior and discussed its limitations in challenging scenarios. We have shown that all the tested combinations of loss function and network architecture yield very similar performance; future network architectures that may outperform our choice could be easily adopted.

We included anisotropy as one of our attributes for analysis, however, we found the amount of measured anisotropic materials available in existing datasets [Dupuy and Jakob 2018; Filip 2015] not sufficient to reliably predict anisotropy in static images under complex geometries and illuminations. Most of the materials have very subtle effects, that can be better visualized under viewpoint changes rather than static images. Larger datasets of measured materials could translate into improvements of our predictor, alternatively, generating anisotropic samples with analytic models could be considered.

We have shown that our predictor enables interesting applications. In particular, we have shown that it can be integrated in a differentiable rendering pipeline [Nimier-David et al. 2019] for editing applications. We have focused on optimizing material parameters; an interesting avenue for future work is to use this framework for optimizing geometry or illumination for material perception [Bousseau et al. 2011; Havran et al. 2016]. Finally, we hope that our work and dataset can motivate additional research on this topic as well as new applications.

ACKNOWLEDGMENTS

REFERENCES

Wendy J Adams, Gizem Kucukoglu, Michael S Landy, and Rafał K Mantiuk. 2018. Naturally glossy: Gloss perception, illumination statistics, and tone mapping. *Journal of Vision* 18, 13 (2018), 4–4.

Alan Agresti. 2003. *Categorical data analysis*. Vol. 482. John Wiley & Sons.

Barton L Anderson. 2011. Visual perception of materials and surfaces. *Current Biology* 21, 24 (2011), R978–R983.

Barton L Anderson and Juno Kim. 2009. Image statistics do not explain the perception of gloss and lightness. *Journal of Vision* 9, 11 (2009), 10–10.

Teun Baar, Sepideh Samadzadegan, Hans Brettel, Philipp Urban, and Maria V Ortiz Segovia. 2014. Printing gloss effects in a 2.5 D system. In *Measuring, Modeling, and Reproducing Material Appearance*, Vol. 9018. International Society for Optics and Photonics, 90180M.

Jacob Beck and Slava Prazdny. 1981. Highlights and the perception of glossiness. *Perception & Psychophysics* 30, 4 (1981), 407–410.

Sean Bell, Paul Upchurch, Noah Snavely, and Kavita Bala. 2015. Material Recognition in the Wild with the Materials in Context Database. *Computer Vision and Pattern Recognition* (2015).

Julia Berzhanskaya, Gurumurthy Swaminathan, Jacob Beck, and Ennio Mingolla. 2005. Remote Effects of Highlights on Gloss Perception. *Perception* 34, 5 (2005), 565–575.

J Bieron and P Peers. 2020. An adaptive brdf fitting metric. In *Computer Graphics Forum*, Vol. 39. 59–74.

Adrien Bousseau, Emmanuelle Chapoulie, Ravi Ramamoorthi, and Maneesh Agrawala. 2011. Optimizing environment maps for material depiction. In *Computer Graphics Forum*, Vol. 30. 1171–1180.

Adrien Bousseau, James P O'shea, Frédo Durand, Ravi Ramamoorthi, and Maneesh Agrawala. 2013. Gloss perception in painterly and cartoon rendering. *ACM Trans. Graph.* 32, 2 (2013), 1–13.

Paul Brossier, Juan Pablo Bello, and Mark D Plumbley. 2004. Real-time temporal segmentation of note objects in music signals. In *Proceedings of ICMC 2004, the 30th Annual International Computer Music Conference*.

Alice C Chadwick and RW Kentridge. 2015. The perception of gloss: A review. *Vision Research* 109 (2015), 221–235.

Rune Haubo B Christensen. 2018. Cumulative link models for ordinal regression with the R package ordinal. *Submitted in J. Stat. Software* (2018).

Douglas W Cunningham and Christian Wallraven. 2011. *Experimental design: From user studies to psychophysics*. CRC Press.

CW Dawson, Nick J Mount, Robert J Abrahart, and John Louis. 2014. Sensitivity analysis for comparison, validation and physical legitimacy of neural network-based hydrological models. *Journal of Hydroinformatics* 16, 2 (2014), 407–424.

Johanna Delanoy, Manuel Lagunas, Ignacio Galve, Diego Gutierrez, Ana Serrano, Roland Fleming, and Belen Masia. 2020. The Role of Objective and Subjective Measures in Material Similarity Learning. In *ACM SIGGRAPH Posters*. Article 51, 2 pages.

Katja Doerschner, Laurence T. Maloney, and Huseyin Boyaci. 2010. Perceived glossiness in high dynamic range scenes. *Journal of Vision* 10, 9, Article 11 (2010).

Shayan Doroudi, Ece Kamar, Emma Brunskill, and Eric Horvitz. 2016. Toward a learning science for complex crowdsourcing tasks. In *Human Factors in Computing Systems*. 2623–2634.

Ron O. Dror, Alan S. Willsky, and Edward H. Adelson. 2004. Statistical characterization of real-world illumination. *Journal of Vision* 4 (2004), 821–837.

Jonathan Dupuy and Wenzel Jakob. 2018. An adaptive parameterization for efficient material acquisition and rendering. *ACM Trans. Graph.* 37, 6 (2018), 1–14.

Frédo Durand and Julie Dorsey. 2002. Fast bilateral filtering for the display of high-dynamic-range images. In *Proc. ACM SIGGRAPH*. 257–266.

Willemijn Elkhuizen, Tessa Essers, Yu Song, Jo Geraedts, Clemens Weijkamp, Joris Dik, and Sylvia Pont. 2019. Gloss, Color, and Topography Scanning for Reproducing a Painting's Appearance Using 3D Printing. *Journal on Computing and Cultural Heritage (JOCCH)* 12, 4 (2019), 1–22.

Franz Faul. 2019. The influence of Fresnel effects on gloss perception. *Journal of Vision* 19, 13 (2019), 1–39.

Jiří Filip. 2015. Analyzing and predicting anisotropic effects of BRDFs. In *Proc. ACM Symposium on Applied Perception*. 25–32.

J. Filip and R. Vávra. 2014. Template-Based Sampling of Anisotropic BRDFs. *Computer Graphics Forum* 33, 7 (2014), 91–99.

Roland W Fleming. 2014. Visual perception of materials and their properties. *Vision Research* 94 (2014), 62–75.

Roland W Fleming. 2017. Material perception. *Annual Review of Vision Science* 3 (2017), 365–388.

Roland W. Fleming and Heinrich H. Bülthoff. 2005. Low-Level Image Cues in the Perception of Translucent Materials. *ACM Trans. Appl. Percept.* 2, 3 (2005), 346–382.

Roland W. Fleming, Ron O. Dror, and Edward H. Adelson. 2003. Real-world illumination and the perception of surface reflectance properties. *Journal of Vision* 3, 5 (2003), 3–3.

Roland W Fleming, Shin'ya Nishida, and Karl R Gegenfurtner. 2015. Perception of material properties. *Vision Research* 115 (2015), 157–62.

Roland W Fleming and Katherine R Storrs. 2019. Learning to see stuff. *Current Opinion in Behavioral Sciences* 30 (2019), 100–108.

Adria Fores, James Ferwerda, and Jinwei Gu. 2012. Toward a perceptually based metric for BRDF modeling. In *Proc. Color and Imaging Conference*. 142–148.

A. Gilchrist, Kossyfidis C., Bonato F., and et al. 1999. An anchoring theory of lightness perception. *Psychol Rev.* 106, 4 (1999), 795–834.

Ioannis Gkioulekas, Bruce Walter, Edward H Adelson, Kavita Bala, and Todd Zickler. 2015. On the appearance of translucent edges. In *Computer Vision and Pattern Recognition*. 5528–5536.

Dar'ya Guarnera, Giuseppe Claudio Guarnera, Matteo Toscani, Mashhuda Glencross, Baihua Li, Jon Yngve Hardeberg, and Karl R Gegenfurtner. 2018. Perceptually validated cross-renderer analytical BRDF parameter remapping. *IEEE Transactions on Visualization and Computer Graphics* 26, 6 (2018), 2258–2272.

Sabrina Hansmann-Roth and Pascal Mamassian. 2017. A Glossy Simultaneous Contrast: Conjoint Measurements of Gloss and Lightness. *i-Perception* 8, 1 (2017).

Vlastimil Havran, Jiri Filip, and Karol Myszkowski. 2016. Perceptually motivated BRDF comparison using single image. In *Computer Graphics Forum*, Vol. 35. 1–12.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Computer Vision and Pattern Recognition*. 770–778.

Jeffrey Heer and Michael Bostock. 2010. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. In *Human Factors in Computing Systems*. 203–212.

Bingyang Hu, Jie Guo, Yanjun Chen, Mengtian Li, and Yanwen Guo. 2020. DeepBRDF: A Deep Representation for Manipulating Measured BRDF. In *Computer Graphics Forum*, Vol. 39. 157–166.

R.S. Hunter. 1937. Methods of Determining Gloss. *Part of Journal of Research of the National Bureau of Standards* 18 (1937), 19–39.

R.S. Hunter and R.W. Harold. 1987. *The measurement of appearance* (2nd ed.). Wiley, New York.

Juno Kim, Phillip J. Marlow, and Barton L. Anderson. 2012. The dark side of gloss. *Nature Neuroscience* 15, 11 (2012), 1590–1595.

Diederik P Kingma and Jimmy Ba. 2014. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014).

Manuel Lagunas, Sandra Malpica, Ana Serrano, Elena Garces, Diego Gutierrez, and Belen Masia. 2019. A Similarity Measure for Material Appearance. *ACM Trans. Graph.* 38, 4 (2019).

Manuel Lagunas, Ana Serrano, Diego Gutierrez, and Belen Masia. 2021. The joint role of geometry and illumination on material recognition. *Journal of Vision* 21, 2 (2021). https://doi.org/10.1167/jov.21.2.2

Y-X. Landy, M.S. Landy, and L.T. Maloney. 2008. Conjoint measurement of gloss and surface texture. *Psychological Science* 19, 2 (2008), 196–204.

Guillaume Lavoué, Nicolas Bonneel, Jean-Philippe Farrugia, and Cyril Soler. 2021. Perceptual quality of BRDF approximations: dataset and metrics. *Computer Graphics Forum* 40, 2 (2021).

Frédéric B Leloup, Michael R Pointer, Philip Dutré, and Peter Hanselaer. 2010. Geometry of illumination, luminance contrast, and gloss perception. *J. Opt. Soc. Am. A* 27, 9 (2010), 2046–2054.

A Luongo, V Falster, MB Doest, MM Ribo, ER Eiriksson, DB Pedersen, and JR Frisvad. 2019. Microstructure Control in 3D Printing with Digital Light Processing. In *Computer Graphics Forum*, Vol. 39. 347–359.

Phillip J Marlow and Barton L Anderson. 2013. Generative constraints on image cues for perceived gloss. *Journal of Vision* 13, 14 (2013), 2–2.

Phillip J Marlow, Juno Kim, and Barton L Anderson. 2012. The perception and misperception of specular surface reflectance. *Current Biology* 22, 20 (2012), 1909–1913.

Wojciech Matusik. 2003. *A data-driven reflectance model*. Ph.D. Dissertation. Massachusetts Institute of Technology.

Wojciech Matusik, Boris Ajdin, Jinwei Gu, Jason Lawrence, Hendrik P. A. Lensch, Fabio Pellacini, and Szymon Rusinkiewicz. 2009. Printing Spatially-Varying Reflectance. *ACM Trans. Graph.* 28, 5 (2009), 1–9.

Wojciech Matusik, Hanspeter Pfister, Matt Brand, and Leonard McMillan. 2003. A Data-Driven Reflectance Model. *ACM Trans. Graph.* 22, 3 (2003), 759–769.

Peter McCullagh. 1980. Regression models for ordinal data. *Journal of the Royal Statistical Society: Series B (Methodological)* 42, 2 (1980), 109–127.

Isamu Motoyoshi and Hiroaki Matoba. 2012. Variability in constancy of the perceived surface reflectance across different illumination statistics. *Vision Research* 53, 1 (2012), 30–39.

Isamu Motoyoshi, Shin'ya Nishida, Lavanya Sharan, and Edward H Adelson. 2007. Image statistics and the perception of surface qualities. *Nature* 447, 7141 (2007), 206–209.

Addy Ngan, Frédo Durand, and Wojciech Matusik. 2006. Image-driven Navigation of Analytical BRDF Models. *Rendering Techniques* (2006), 399–407.

Merlin Nimier-David, Delio Vicini, Tizian Zeltner, and Wenzel Jakob. 2019. Mitsuba 2: A retargetable forward and inverse renderer. *ACM Trans. Graph.* 38, 6 (2019), 1–17.

Shin'ya Nishida and Mikio Shinya. 1998. Use of image-based information in judgments of surface-reflectance properties. *J. Opt. Soc. Am. A* 15, 12 (1998), 2951–2965.

Jum C Nunnally. 1994. *Psychometric theory 3E*. Tata McGraw-hill education.

Maria Olkkonen and David H Brainard. 2010. Perceived glossiness and lightness under real-world illumination. *Journal of Vision* 10, 9 (2010), 5–5.

Fabio Pellacini, James A Ferwerda, and Donald P Greenberg. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proc. ACM SIGGRAPH*. 55–64.

Thiago Pereira and Szymon Rusinkiewicz. 2012. Gamut mapping spatially varying reflectance with an improved BRDF similarity metric. In *Computer Graphics Forum*, Vol. 31. 1557–1566.

Michal Piovarči, Michael Wessely, Michał Jagielski, Marc Alexa, Wojciech Matusik, and Piotr Didyk. 2017. Directional screens. In *Proceedings of the 1st Annual ACM Symposium on Computational Fabrication*. ACM, 1.

Michal Piovarči, Michael Foshey, Vahid Babaei, Szymon Rusinkiewicz, Wojciech Matusik, and Piotr Didyk. 2020. Towards Spatially Varying Gloss Reproduction for 3D Printing. *ACM Trans. Graph.* 39, 6, Article 206 (2020).

Sylvia C Pont and Susan F te Pas. 2006. Material — Illumination Ambiguities and the Perception of Solid Objects. *Perception* 35, 10 (2006), 1331–1350.

Ganesh Ramanarayanan, James Ferwerda, Bruce Walter, and Kavita Bala. 2007. Visual equivalence: towards a new standard for image fidelity. *ACM Trans. Graph.* 26, 3 (2007), 76–es.

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic tone reproduction for digital images. In *Proc. ACM SIGGRAPH*. 267–276.

Olivier Rouiller, Bernd Bickel, Jan Kautz, Wojciech Matusik, and Marc Alexa. 2013. 3D-printing spatially varying BRDFs. *IEEE Computer Graphics and Applications* 33, 6 (2013), 48–57.

Sepideh Samadzadegan, Teun Baar, Philipp Urban, Maria V Ortiz Segovia, and Jana Blahová. 2015. Controlling colour-printed gloss by varnish-halftones. In *Measuring, Modeling, and Reproducing Material Appearance 2015*, Vol. 9398. International Society for Optics and Photonics, 93980V.

A. C. Schmid, P. Barla, and K. Doerschner. 2020. Material category determined by specular reflection structure mediates the processing of image features for perceived gloss. *bioRxiv* (2020).

Gabriel Schwartz and Ko Nishino. 2019. Recognizing material properties from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42, 8 (2019), 1981–1995.

Ana Serrano, Diego Gutierrez, Karol Myszkowski, Hans-Peter Seidel, and Belen Masia. 2016. An intuitive control space for material appearance. *ACM Trans. Graph.* 35, 6 (2016).

Lavanya Sharan, Yuanzhen Li, Isamu Motoyoshi, Shin'ya Nishida, and Edward H Adelson. 2008. Image statistics for surface reflectance perception. *J. Opt. Soc. Am. A* 25, 4 (2008), 846–865.

Lavanya Sharan, Ruth Rosenholtz, and Edward H Adelson. 2014. Accuracy and speed of material categorization in real-world images. *Journal of Vision* 14, 9 (2014), 12–12.

Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014).

Katherine R Storrs and Roland W Fleming. 2020. Unsupervised Learning Predicts Human Perception and Misperception of Specular Surface Reflectance. *bioRxiv* (2020).

Tiancheng Sun, Henrik Wann Jensen, and Ravi Ramamoorthi. 2018. Connecting measured BRDFs to analytic brdfs by data-driven diffuse-specular separation. *ACM Trans. Graph.* 37, 6 (2018), 1–15.

Tiancheng Sun, Ana Serrano, Diego Gutierrez, and Belen Masia. 2017. Attribute-preserving gamut mapping of measured BRDFs. In *Computer Graphics Forum*, Vol. 36. 47–54.

Alejandro Sztrajman, Jaroslav Křivánek, Alexander Wilkie, and Tim Weyrich. 2019. Image-based Remapping of Spatially-Varying Material Appearance. *Journal of Computer Graphics Techniques (JCGT)* 8, 4 (2019), 1–30.

James T. Todd, J. Farley Norman, and Ennio Mingolla. 2004. Lightness Constancy in the Presence of Specular Highlights. *Psychological Science* 15, 1 (2004), 33–39.

Matteo Toscani, Dar'ya Guarnera, Giuseppe Claudio Guarnera, Jon Yngve Hardeberg, and Karl R Gegenfurtner. 2020. Three perceptual dimensions for specular and diffuse reflection. *ACM Transactions on Applied Perception (TAP)* 17, 2 (2020), 1–26.

Matteo Toscani and Matteo Valsecchi. 2019. Lightness Discrimination Depends More on Bright Rather Than Shaded Regions of Three-Dimensional Objects. *i-Perception* 10, 6 (2019), 1–10.

Matteo Toscani, Matteo Valsecchi, and Karl R Gegenfurtner. 2017. Lightness perception for matte and glossy complex shapes. *Vision Research* 131 (2017), 82–95.

TS Trowbridge and Karl P Reitz. 1975. Average irregularity representation of a rough surface for ray reflection. *J. Opt. Soc. Am. A* 65, 5 (1975), 531–536.

Peter Vangorp, Pascal Barla, and Roland W Fleming. 2017. The perception of hazy gloss. *Journal of Vision* 17, 5 (2017), 19–19.

Peter Vangorp and Philip Dutré. 2008. Shape-dependent gloss correction. In *Proc. Applied Perception in Graphics and Visualization*. 123–130.

Peter Vangorp, Jurgen Laurijssen, and Philip Dutré. 2007. The influence of shape on the perception of material reflectance. In *Proc. ACM SIGGRAPH*. 77:1–77:9.

Bruce Walter, Stephen R Marschner, Hongsong Li, and Kenneth E Torrance. 2007. Microfacet Models for Refraction through Rough Surfaces. *Rendering Techniques* (2007), 195–206.

Gregory J. Ward. 1992. Measuring and Modeling Anisotropic Reflection. *Proc. ACM SIGGRAPH* 26, 2 (1992), 265–272.

Peter Welinder, Steve Branson, Pietro Perona, and Serge Belongie. 2010. The multidimensional wisdom of crowds. *Advances in Neural Information Processing Systems* 23 (2010), 2424–2432.

G. Wendt, F. Faul, V. Ekroll, and R. Mausfeld. 2010. Disparity, motion, and color information improve gloss constancy performance. *Journal of Vision* 10, 9 (2010).

Tim Weyrich, Pieter Peers, Wojciech Matusik, and Szymon Rusinkiewicz. 2009. Fabricating microgeometry for custom surface reflectance. *ACM Trans. Graph.* 28, 3 (2009), 32.

Christiane B Wiebel, Matteo Toscani, and Karl R Gegenfurtner. 2015. Statistical correlates of perceived gloss in natural images. *Vision Research* 115 (2015), 175–187.

Josh Wills, Sameer Agarwal, David Kriegman, and Serge Belongie. 2009. Toward a perceptual space for gloss. *ACM Trans. Graph.* 28, 4 (2009), 1–15.

Haoyu Xu, Zhenqi Han, Songlin Feng, Han Zhou, and Yuchun Fang. 2018. Foreign object debris material recognition based on convolutional neural networks. *Eurasip Journal on Image and Video Processing* 2018, 1 (2018), 1–10.

Fan Zhang, Huib de Ridder, Pascal Barla, and Sylvia Pont. 2020a. Effects of light map orientation and shape on the visual perception of canonical materials. *Journal of Vision* 20, 4, Article 13 (2020), 18 pages.

Fan Zhang, Huib de Ridder, Pascal Barla, and Sylvia Pont. 2020b. A systematic approach to testing and predicting light-material interactions. *Journal of Vision* 19, 4, Article 11 (2020), 22 pages.

Fan Zhang, Huib de Ridder, and Sylvia Pont. 2015. The influence of lighting on visual perception of material qualities. In *Human Vision and Electronic Imaging*, Vol. SPIE 9394. 239–248.

Károly Zsolnai-Fehér, Peter Wonka, and Michael Wimmer. 2018. Gaussian Material Synthesis. *ACM Trans. Graph.* 37, 4, Article 76 (2018).