In-the-wild Material Appearance Editing using Perceptual Attributes

J. Daniel Subias[1], M. Lagunas[2]

[1] Graphics & Imaging Lab
Instituto de Investigación en Ingeniería de Aragón (I3A)
Universidad de Zaragoza, Mariano Esquillor s/n, 50018, Zaragoza, Spain.
Tel. +34-976762707, e-mail: dsubias@unizar.es
[2]Amazon, Spain

Abstract

We present a single-image appearance editing framework that allows to intuitively modify the material appearance of an object by increasing or decreasing high-level perceptual attributes describing appearance, without supplementary information of the scene.

Introduction

Humans are visual creatures, most of our input information comes just from sight. However untangling the processes that happen in our visual system, and finding a direct relationship between our subjective impression and the physical parameters that govern the interaction between light and matter is an open challenge. Moreover, this problem is further aggravated since the mathematical models that simulate the interaction between light and matter are usually not intuitive nor predictable, highly dimensional, and only understood by professionals. We present an image-based framework that does not rely on any physically-based rendering but instead modifies directly the material appearance in the image pixels.

Our Framework

Our approach relies on a STGAN [1] architecture taking as input taking a certain image of an object as input and modifies the appearance based on varying the value of the desired high-level perceptual attribute describing appearance (e.g., glossy or metallic). Our goal is to maintain the high-freuquency geometrical details from the input image whitout supplementary information of the scene.

Architecture

Our framework is composed of a generator *G* and a discriminator *D* with the same number of convolutional layers. The generator *G* takes as input a high-resolution image of a single object **x** together with a target high-level perceptual attribute (e.g glossy, metallic, ...) **att** in the range [0, 1]. Then the encoder module G_{enc} compresses the input image in a latent code **z**. After that, the decoder module G_{dec} reconstructs an edited image y by adding the target high-level perceptual attribute **att** in **z**. The features maps sent from the G_{enc} to G_{dec} are processed by the Selective Transfer Units [1] (STUs) to remove residual information.

Training

We adopt the adversarial training proposed by *G* ulrajani et al. [2] and introduce a GAN model where the discriminator *D* has two branches D_{adv} and D_{att} . D_{adv} consists of five convolution layers to predict whether an image is fake (edited) or real and the perceptual attribute value from the input image. Our loss function is described as follows:

$$\mathrm{Loss}_D = -\ L_{Dadv} + \lambda_1 L_{Datt}$$

$$\mathrm{Loss}_G = -\ L_{Gadv} + \lambda_2 L_{Gatt} + \lambda_3 L_{rec}$$

We also intruduced a reconstruction loss function L_{rec} to give feedback to generator *G* on the quality of its generated images. To optimize these losses, we use $\beta1 = 0.5$ and $\beta2 = 0.999$ for the Adam optimizer. The learning rate is 2 × 10−4 for both *G* and *D*; and does not decay at training. The tradeoff parameters are $\lambda_1 = 10$, $\lambda_2 = 100$ and $\lambda_3 = 1000$. We leverage the training dataset of Delanoy et. al [3], designed for material appearance perception tasks. This dataset contains renderings of 13 different geometries, illuminated by 7 catured real-world illuminations. Renders have been made using 100 different BRDFs For each combination of material × shape × illumination, 5 different images with slight variations in the viewpoint have been rendered, as. The dataset has 45,500 images

Results

In Figure 1 we show high-quality edits of material attributes such as glossy or metallic, while preserving the geometrical structure and details. Our approach learns to edit perceptual cues properly while objects' shape remains unchanged. We compare our results against the state-of the-art-method of Delanoy et al. [3]. Table 1 shows we outperform the state of the art as a result of introducing STU cells in each skip-connection. As illustrated in Figure 2 our method keeps high-frequency details from the input image without the need of supplementary information of the object's surface as the input. In Figure 3 we can see a comparison between the edited images by our method and the one by Delanoy et al. [3].

Conclusions

We have presented a framework for intuitive material appearance editing using single in-the-wild images. We relied on a large set of images to train our framework, and use a generative neural network and devise a loss function that allows us to learn how to edit material appearance based on such high-level attributes, without any pairs of original and edited images. Our results show that the presented method can achieve realistic results, almost on par with real photographs. However, our method is not free of limitations, as when using input photographs that exhibit highly specular highlights, although they are capable of conveying the appearance of the target high-level perceptual attribute, our framework may have difficulty editing them. Instead of taking into account the original albedo to perform the edit of the reflections when the glossiness decreases.

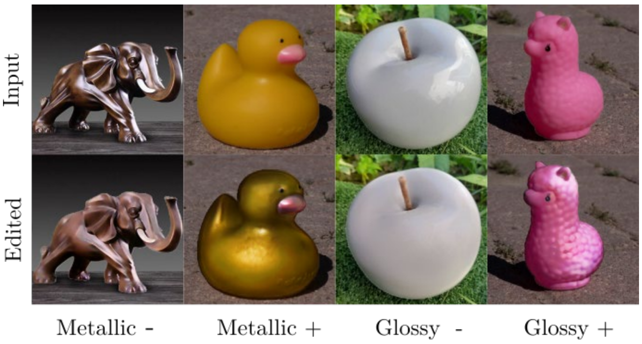

Figure 1: A sample of the real photographs edited by our framework without suplementary information of the scene.

Tabla 1. Average PSNR, SSIM, MSE and MAE reconstructing the input image.

Method	PSNR	SSIM	MAE
Delanoy	15,989	0,761	0,031
Ours	**27,388**	**0,967**	**0,023**

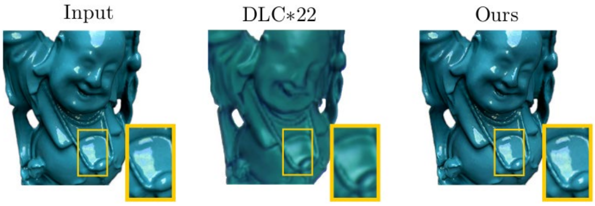

Figure 2: A demonstration of the reconstruction quality.

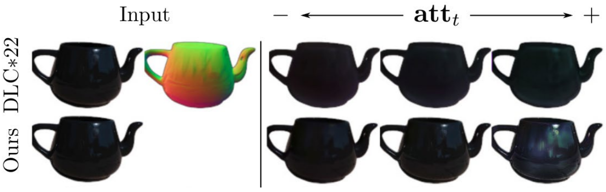

Figure 3: Comparison editing the glossy attribute using the method of Delanoy et al. [3] and our framework.

REFERENCIAS

[1]. LIU M., DING Y., XIA M., LIU X., DING E., ZUO W.,WEN S.: Stgan: A unified selective transfer network for arbitrary image attribute editing. In Proc. Computer Vision and Pattern Recognition (CVPR) (June 2019).

[2]. GULRAJANI I., AHMED F., ARJOVSKY M., DUMOULIN V., COURVILLE A. C.: Improved training of wasserstein gans. In Advances in Neural Information Processing Systems (2017), Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., (Eds.), vol. 30, Curran Associates, Inc.

[3]. DELANOY J., LAGUNAS M., CONDOR J., GUTIERREZ D., MASIA B.: A generative framework for image-based editing of material appearance using perceptual attributes. Computer Graphics Forum 41, 1(2022), 453–464.