Unsupervised Learning of Disentangled and Interpretable Representations of Material Appearance

Santiago Jiménez, Julia Guerrero-Viu, Belen Masia

Graphics and Imaging Lab

MOTIVATION

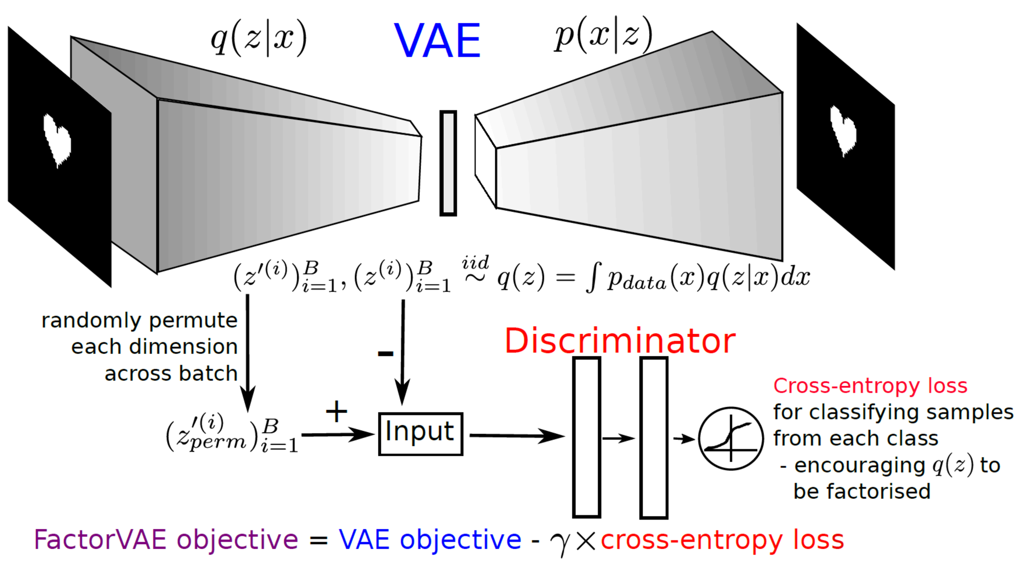
The final **perception** we get of an object's **appearance** depends not only on the **material** that was used to craft it, but also on the **interaction between external confounding factors** such as the lighting conditions, geometry, or point of view. This interaction is not yet fully understood, and the final perception ultimately depends on the **subjectivity** of the viewer, which could be further biased by their background and previous knowledge.

Designing and creating content is a transversal task that ranges from designers or artists to novice users. When creating a new material, design systems usually ask users to input a set of **complex** and **non-intuitive parameters**, which hinder the overall process of creating such material. Ideally, the making of a new material would just involve defining a **reduced** set of parameters that are **intuitive** to modify, and lead to **predictable** changes in the final appearance.

We present a **learning-based** algorithm that effectively **disentangles perceptual features** of rendered **images** in an **unsupervised** way, enabling intuitive **representation** and **editing** of materials.

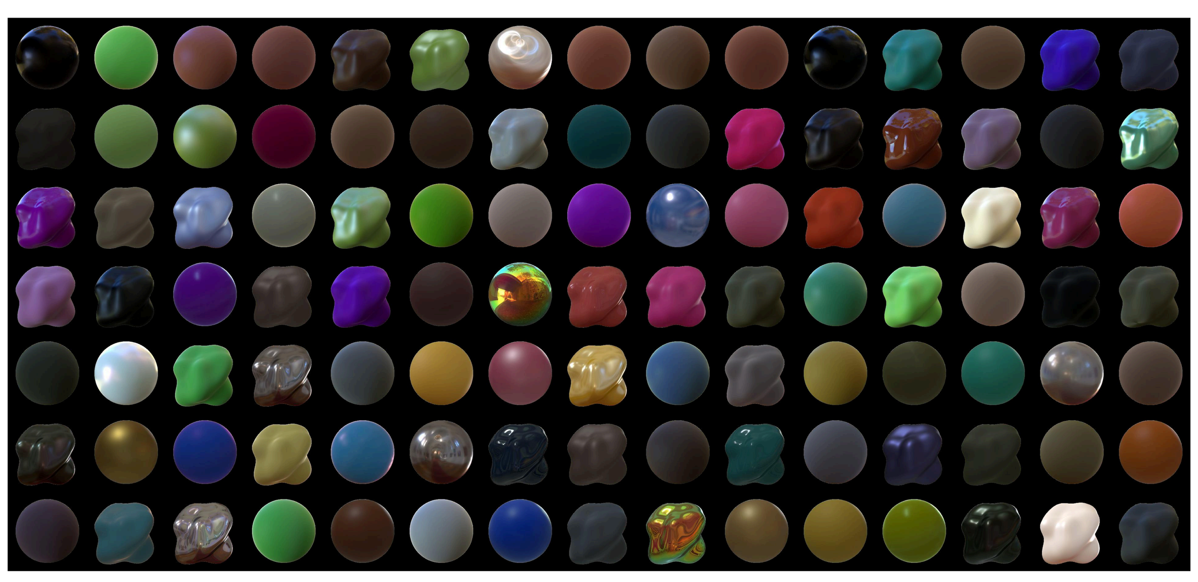
APPROACH

We chose to use a modification of a Variational Autoencoder (VAE) [1] called **FactorVAE** [2] due to its good balance in terms of disentanglement, regularization, and reconstruction quality.



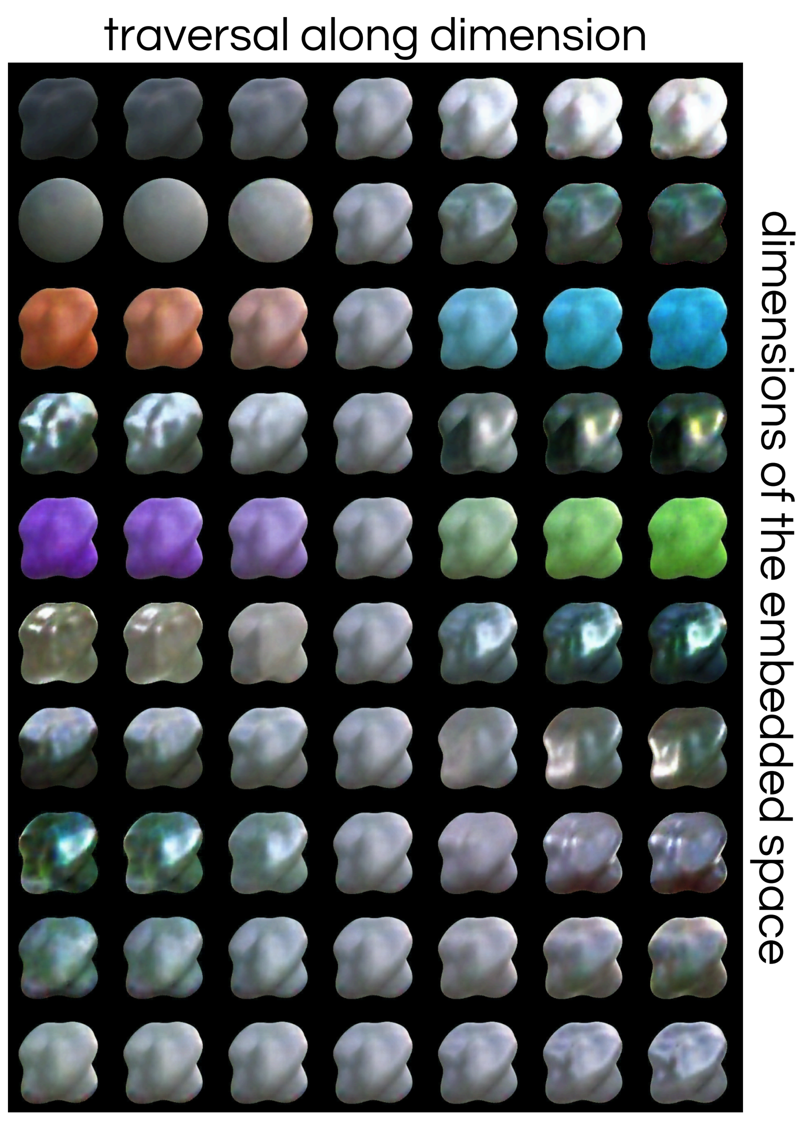
DATA

To train our models, we utilize a modified version of the Serrano Dataset [3], which includes renderings depicting a diverse range of **materials**, **lighting conditions**, and **geometries**.

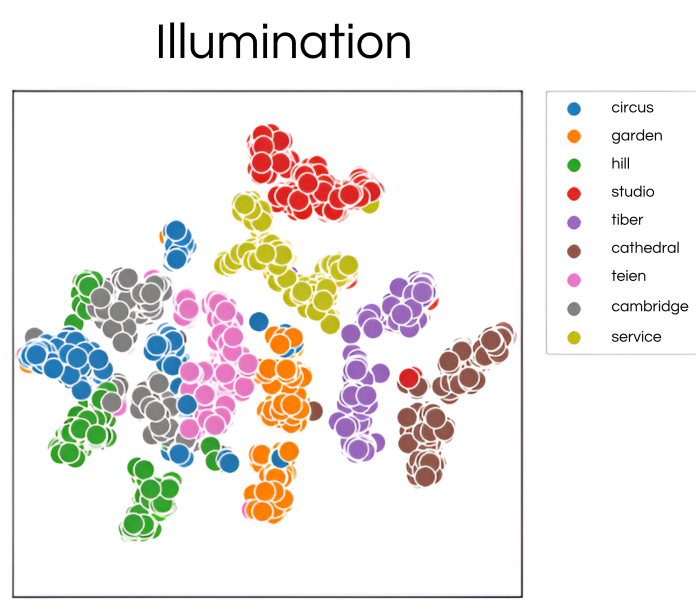


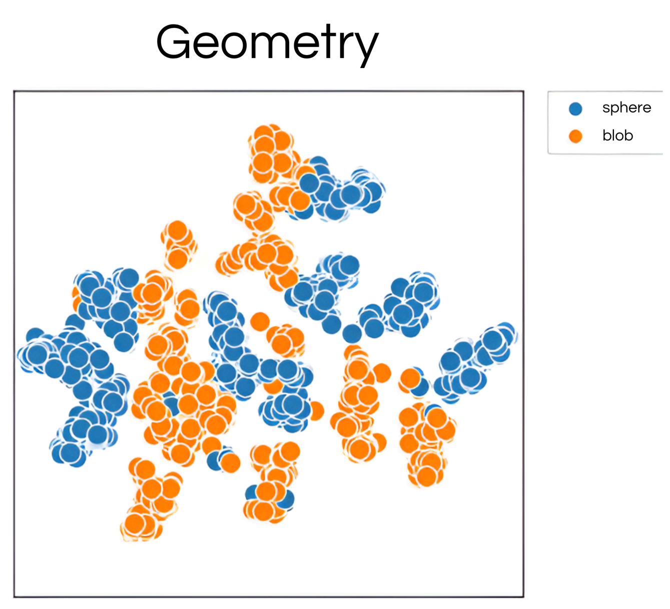
RESULTS

Our experiments reveal an embedded space that is **highly disentangled** in terms of independence of dimensions. We can visually assess how it successfully identified certain **perceptual features** of the training data such as **lightness**, **geometry**, **glossiness**, or **lighting**. ▷



By visualizing the distribution of this space, we notice how the model initially **groups** the input data based on its **lighting**, and then based on its **shape**. ▽





ACKNOWLEDGEMENTS

We acknowledge funding from the Spanish Agencia Estatal de Investigación (PID2019-105004GB-I00, PID2022-141539NB-I00), and the European Union's Horizon 2020 MSCA (grant No. 956585). Julia Guerrero-Viu was supported by the FPU20/02340 grant and Santiago Jimenez by the TFM Fellowship Program of I3A.

REFERENCES

[1]. KINGMA, Diederik P.; WELLING, Max. Auto-Encoding Variational Bayes. stat, 2014.
[2]. KIM, Hyunjik; MNIH, Andriy. Disentangling by factorising. In International conference on machine learning. PMLR, 2018. p. 2649-2658.
[3]. SERRANO, Ana, et al. The effect of shape and illumination on material perception: model and applications. ACM Transactions on Graphics (TOG), 2021, vol. 40, no 4, p. 1-16.
[4]. STORRS, Katherine R., et al. Unsupervised learning predicts human perception and misperception of gloss. Nature Human Behaviour, 2021, vol. 5, no 10, p. 1402-1417.
[5]. GUERRERO-VIU, Julia, et al. Predicting Perceived Gloss: Do Weak Labels Suffice?. In Computer Graphics Forum. 2024. p. e15037.

MINISTERIO DE CIENCIA, INNOVACIÓN Y UNIVERSIDADES
Cofinanciado por la Unión Europea
AGENCIA ESTATAL DE INVESTIGACIÓN
PRIME