Unsupervised Learning of Disentangled 3D Representation from a Single Image

Junliang Lv, Haiyong Jiang and Jun Xiao*

University of Chinese Academy of Sciences, China *Email: xiaojun@ucas.ac.cn

Abstract

Learning 3D representation of a single image is challenging. Previous works either seek image annotation or 3D supervision to learn meaningful factors or employ a StyleGAN-like framework for image synthesis. We combine the advantages of both frameworks and propose an image disentanglement method based on 3D representation.

Introduction

Learning 3D representation of an image, e.g. the 3D shape and the texture, is important for many real-world applications in image synthesis and augmented reality. [1] creates high-quality 2D images by unsupervised learning but lacks the correspondence for property. [2] uses 3D models templates and key points in images to learn the disentanglement of different factors, while [3] and [4] learns 3D representation of an object by explicitly providing 3D supervisions.

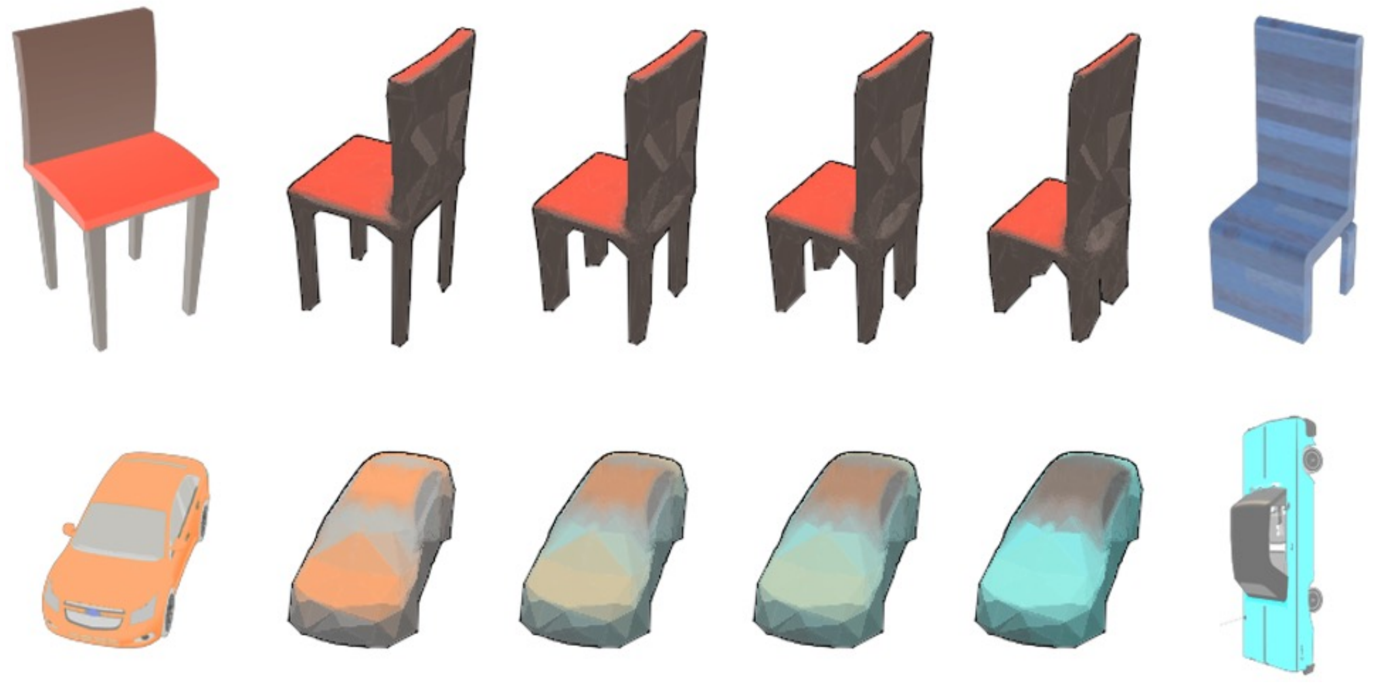

Figure 1: Interpolation of factors. The first column is the input image, the second column is the reconstructed image, and the subsequent columns represent the gradual change of a certain factor which ends with the image in the last column. The first row is for shape, while the second row is for texture.

Methods and Results

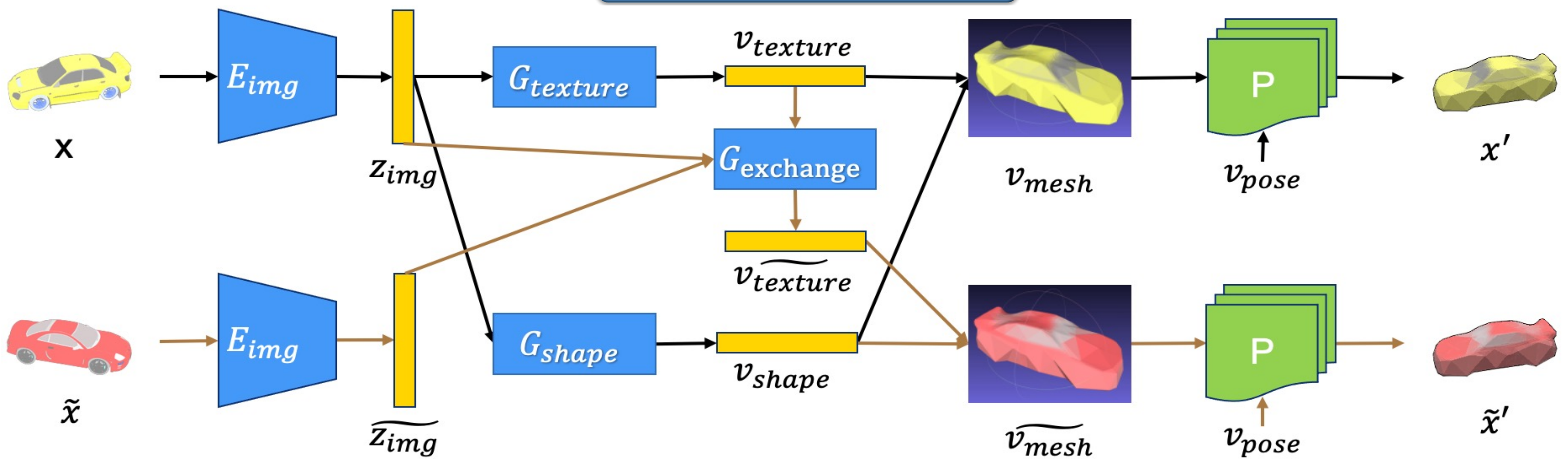

Figure 2: The pipeline of our framework. ***Black arrows*** *show the reconstruction branch, while* ***brown arrows*** *give the disentangling branch for shape and texture disentanglement.*

In Fig. 2, the reconstruction branch gets the deformed mesh and generate images in different views by [5]. The disentangling branch learns the transformation of texture. We learn the network through optimizing the reconstruction branch with the typical perceptual loss [6] and reconstruction loss, and the disentangling branch with the silhouette loss and the style loss [6]. By explicitly using a 3D mesh representation, our method can generate novel views with consistent geometry as shown in Fig. 3 (a). Results in Fig. 3 (b) demonstrate our method can achieve an disentanglement of textures and shapes even without any supervision. Fig. 1 shows that our method can achieve interpolation of factors.

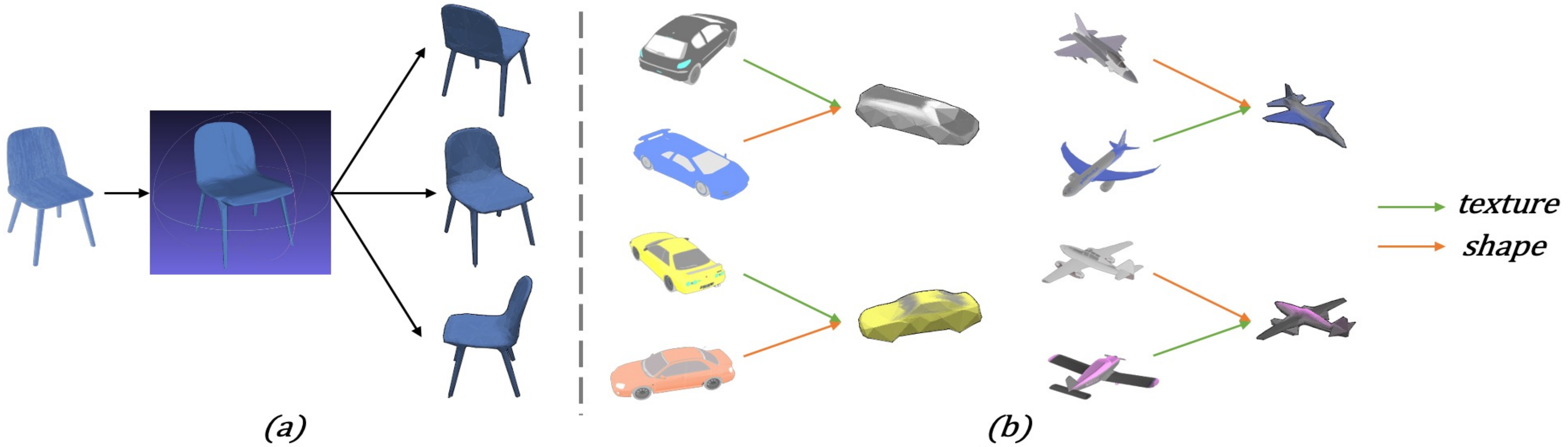

Figure 2*:Applications of our method.* ***(a)*** *Our method can reconstruct textured shape of an input image and use it to generate consistent novel view images.* ***(b)*** *Our method can help image editing by explicitly swapping the shape factors or texture factors so that images with novel textures and shapes can be generated.*

Conclusion

In this work, we present a novel approach to learn 3D representation from a single image. Our method can disentangle different factors of the 3D representation, which enables interesting applications, like texture remapping, pose editing and object replacement in an image. Results demonstrate our method can achieve plausible 3D factor disentanglement by exploring constraints between generated images as supervisions. In the future, we will deploy our method in more challenging scenarios, like novel image synthesis based on product images in shopping websites and room refurnishing based on a user image.

References

[1] KARRAS T., LAINE S., AILA T.: A style-based generator architecture for generative adversarial networks. CVPR (2019), pp. 4401–4410. 1

[2] KANAZAWA A., TULSIANI S., EFROS A. A., MALIK J.: Learning category-specific mesh reconstruction from image collections. ECCV (2018), pp. 371–386. 2

[3] CHEN X., COHEN-OR D., CHEN B., MITRA N. J.: Neural graphics pipeline for controllable image generation. arXiv:2006.10569 (2020). 2

[4] ZHU J.-Y., ZHANG Z., ZHANG C., WU J., TORRALBA A., TENENBAUM J. B., FREEMAN W. T.: Visual object networks: Im- age generation with disentangled 3d representation. arXiv:1812.02725 (2018). 2

[5] LIU S., LI T., CHEN W., LI H.: Soft rasterizer: A differentiable renderer for image-based 3d reasoning. ICCV (2019), pp. 7708–7717. 2

[6] JOHNSON J., ALAHI A., FEI-FEI L.: Perceptual losses for real-time style transfer and super-resolution. ECCV (2016), Springer, pp. 694–711. 2