# Novel View Synthesis and Style Transfer via 3D feature embeddings

**Shubham Agrawal Somendra Tripathi Najim Yaqubie Iddo Drori**

## Abstract

The advent of conditional adversarial models has given rise to very powerful image synthesis. However, despite the rich depth in 2D image models, mathematical models of 3D environments have not been explored thoroughly. We propose to apply a hybrid generational model for style transfer upon an underlying 3D scene structure.

## 1. Introduction

Recent years have seen progress in applying machine learning techniques to develop a 3D representation of a particular scene from 2D images. With the introduction of efficient 3D representations such as DeepVoxels (Sitzmann et al., 2019a), Scene Representation Networks (SRNs) (Sitzmann et al., 2019b) and Neural Meshes (Kato et al., 2018), these neural networks are able to generate novel views of an object learned from a set of 2D images.

However, little work has been done in transforming these underlying 3D structures in meaningful ways. Can we, for instance, apply a new floral pattern to a 3D model of shape? We aim to explore the style transfer of a 2D image upon latent 3D models for novel view synthesis.

## 2. Previous Work

There are a set of deep neural networks proposed to solve novel view synthesis that are multi-view consistent. Many techniques make use of various underlying structures such as voxel grids (Sitzmann et al., 2019a), continuous ray marching functions (Sitzmann et al., 2019b), or approximate gradients (Kato et al., 2018). Our aim is to explore these underlying structures to determine which is extensible for style transfer.

Deep models for 2D image synthesis and manipulation have shown promising results in generating photorealistic images, especially using adversarial generative models (Karras et al., 2019). In other cases, deep neural networks separate and recombine content and style of arbitrary images to create artistic views (Gatys et al., 2016). We aim to follow previous methods of style transfer to synthesize novel views with combined content.

## 3. Methodology

We propose collecting two datasets for our approach. First, we will extract 2D poses, similar to the DeepVoxel dataset, from Shapenet (Chang et al., 2015) as our target objects. Using a native 3D object from Shapenet allows us to examine methods that require 3D objects vs 2D images. Second, we will collect high quality fine art images as our style to transfer onto these 3D objects.

We envision multiple baselines to help guide our approach. First, we will unfold a 3D object into a 2D image (Massarwi et al., 2007) (Futurologist, 2019) and apply 2D image-to-image style transfer. We will then refold the result back into a 3D representation. Second, we will explore using neural meshes to render novel stylized views (Kato et al., 2018). Lastly, we will explore applying a style transfer network on the input images to a pre-trained DeepVoxel network.

As our main approach, we will use a pre-trained DeepVoxel network along with a convolutional network for style transfer. To ensure applying style transfer has not degraded novel view synthesis that is multi-view consistent, we will rely on the content loss measure used by the DeepVoxel network to backpropogate. To measure style loss, we will propagate the generated image to a style-transfer-network (Gatys et al., 2016). Combining these two losses should mitigate 3D object deformation and encourage multi-view consistent stylized images.

## 4. Evaluation Criteria

Quantitatively expressing the success of combining the content of one image and the style of another image is a difficult problem and not in the scope of this project. Prior work has also yet to systematically measure the success of style transfer (Gatys et al., 2016). However, we will qualitatively measure whether a novel view retains the 3D object while also ”looking like” the style art image provided. We hope to improve upon the 2D-to-3D style transfer as compared to the neural mesh renderer where the generated shape’s style is discontinuous across views (Kato et al., 2018).

## References

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., Su, H., et al. Shapenet: An information-rich 3d model repository. *arXiv preprint arXiv:1512.03012*, 2015.

Futurologist. Unfold a 3d mesh to a 2d plane, 2019. URL `https://computergraphics.stackexchange.com/a/8785`.

Gatys, L. A., Ecker, A. S., and Bethge, M. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2414–2423, 2016.

Karras, T., Laine, S., and Aila, T. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4401–4410, 2019.

Kato, H., Ushiku, Y., and Harada, T. Neural 3d mesh renderer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3907–3916, 2018.

Massarwi, F., Gotsman, C., and Elber, G. Papercraft models using generalized cylinders. In *15th Pacific Conference on Computer Graphics and Applications (PG'07)*, pp. 148–157. IEEE, 2007.

Sitzmann, V., Thies, J., Heide, F., Nießner, M., Wetzstein, G., and Zollhofer, M. Deepvoxels: Learning persistent 3d feature embeddings. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2437–2446, 2019a.

Sitzmann, V., Zollhöfer, M., and Wetzstein, G. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *arXiv preprint arXiv:1906.01618*, 2019b.