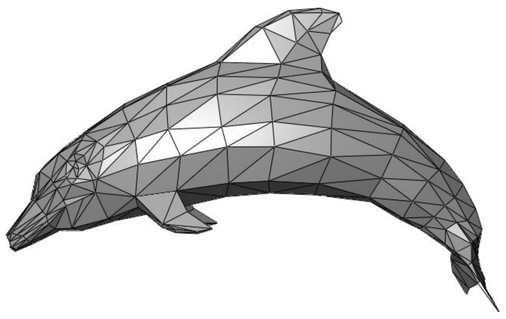
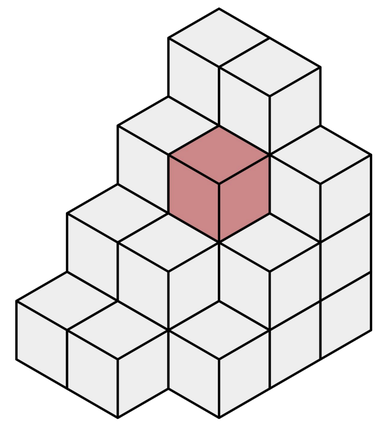
Novel View Synthesis with Style Transfer via 3D feature embeddings



Shubham Agrawal, Somendra Tripathi, Najim Yaqubie

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

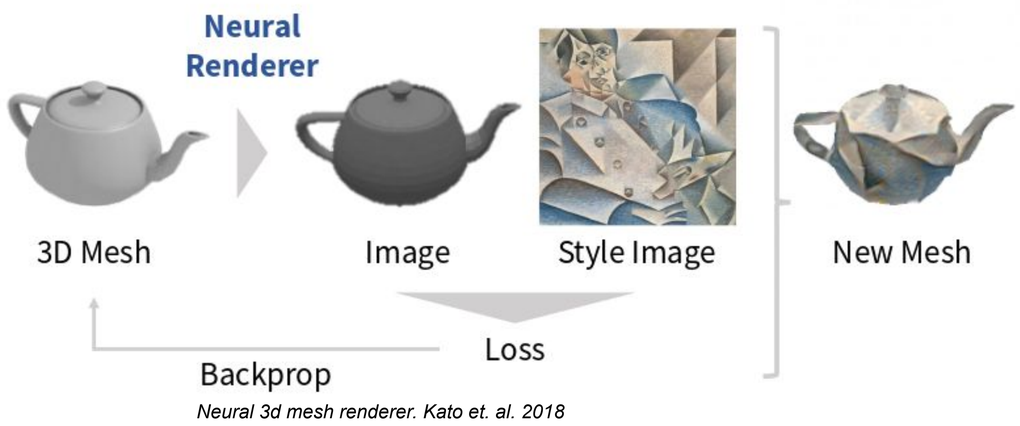
Data, Models, and Code: github.com/najerama/3dstyletransfer



Introduction and Methods

2D Style Transfer is a solved problem, applying style transfer to latent 3D scene environments is unexplored.
We compare polygon mesh and voxel grid underlying 3D representations with respect to transformations such as style transfer.
We use Neural Renderer and DeepVoxels to apply multi-view consistent style transfer onto novel views generated from 3D models.
Key contributions:

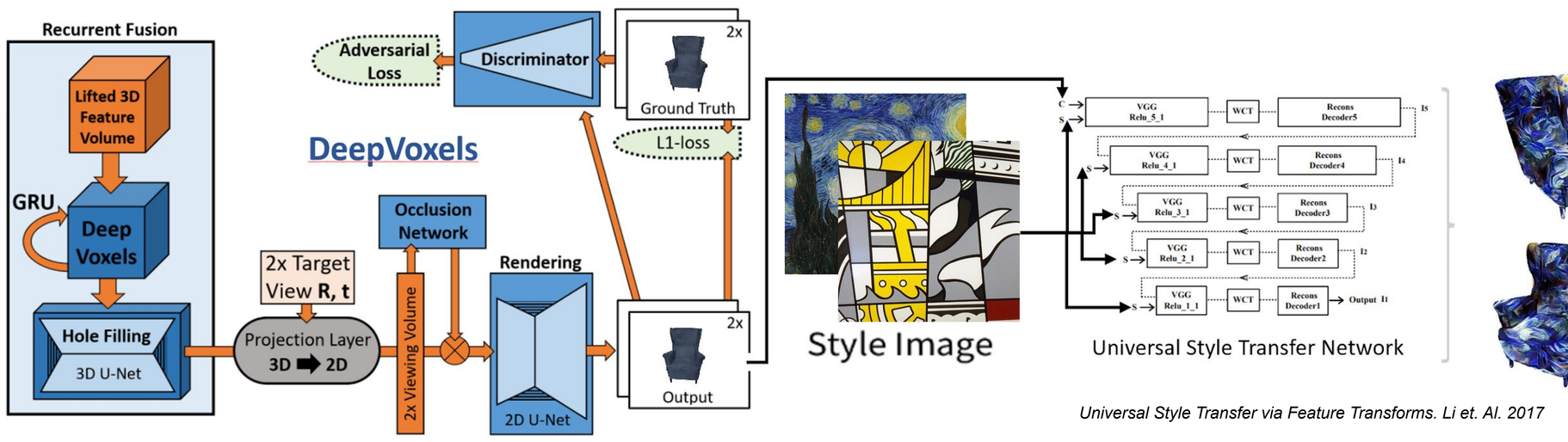
- Hardness scale for 3D Style Transfer for a dataset of 50 images
- Extended Neural Meshes to ShapeNet's objects for future standardized research.
- Quantified combined content and style loss with 3D style transfer.
- Extended our approach to DeepVoxels and demonstrated applying the network to new datasets and style transfer.
- Publicly available source code for 3D style transfer networks.



Neural 3d mesh renderer. Kato et. al. 2018

$$L = \lambda_c * \sum_{\{v_i, v_i^c \in (m, m^c)\}} |v_i - v_i^c|^2 + \lambda_s * |M(f_s(R(m, \phi))) - M(f_s(x_s))|_F^2 + \lambda_t * \sum_{\{\boldsymbol{p}_a, \boldsymbol{p}_b\} \in P} |\boldsymbol{p}_a - \boldsymbol{p}_b|_2^2$$

Neural Mesh Style Transfer: We use this network with trained meshes to backpropogate style loss onto the mesh and generate stylized novel views (left). Overall loss is a combination of content loss and style loss (right).

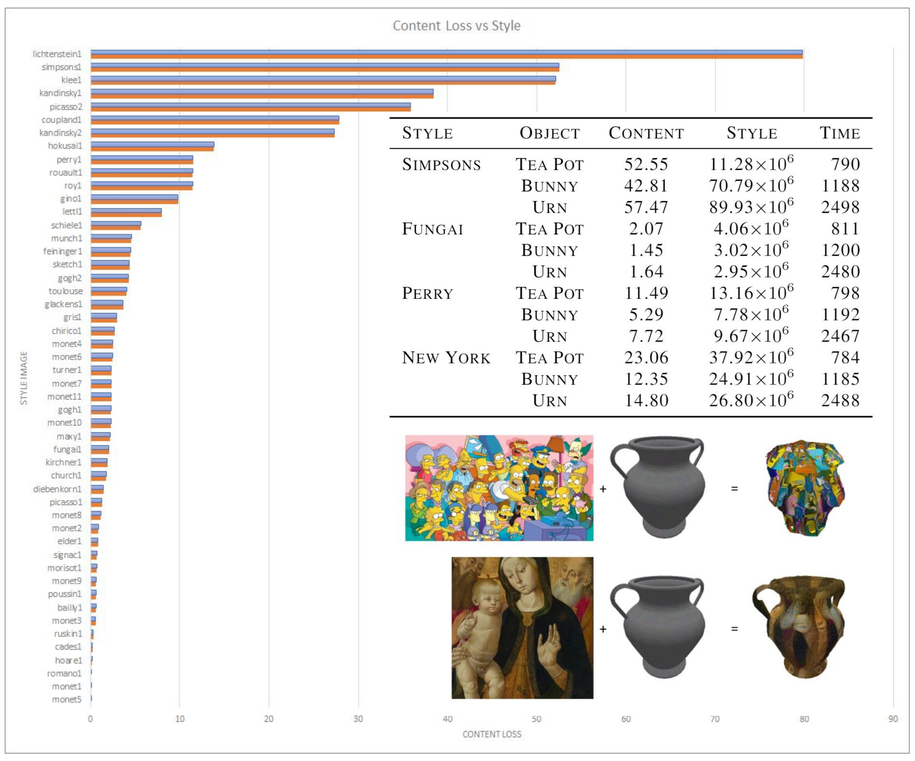


Deep voxels: Learning persistent 3d feature embeddings Sitzmann et. Al. 2019

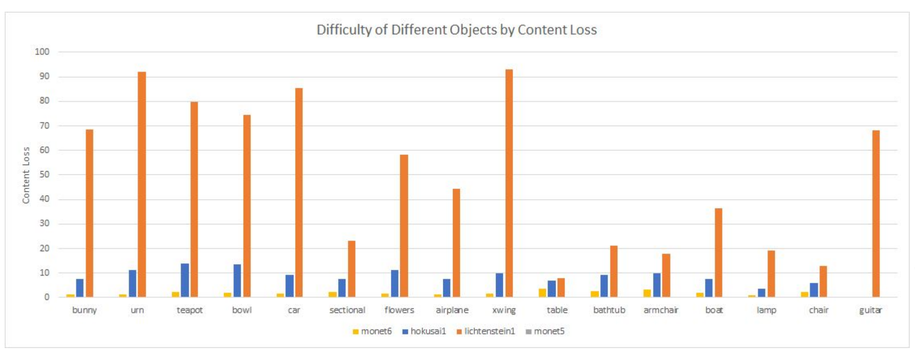
Universal Style Transfer via Feature Transforms. Li et. Al. 2017

DeepVoxels Style Transfer: we used a pretrained multi-layer Universal Style Network that uses a series of pretrained VGG-19 networks while rendering randomized novel views via DeepVoxels for style loss.

Results



Style	Object	Content	Style	Time
Simpsons	Tea Pot	52.55	11.28×10^6	790
	Bunny	42.81	70.79×10^6	1188
	Urn	57.47	89.93×10^6	2498
Fungai	Tea Pot	2.07	4.06×10^6	811
	Bunny	1.45	3.02×10^6	1200
	Urn	1.64	2.95×10^6	2480
Perry	Tea Pot	11.49	13.16×10^6	798
	Bunny	5.29	7.78×10^6	1192
	Urn	7.72	9.67×10^6	2467
New York	Tea Pot	23.06	37.92×10^6	784
	Bunny	12.35	24.91×10^6	1185
	Urn	14.80	26.80×10^6	2488

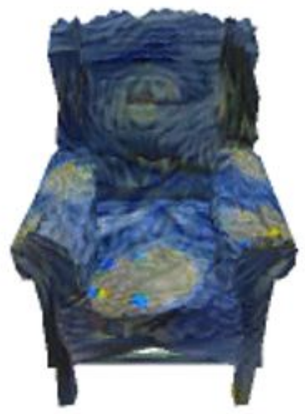


(Left) Styles and objects have varying levels of difficulty in terms of applying style transfer onto 3D models. Styles that are more like textures are much easier to apply when compared to vector graphics, and have consistent ranking of difficulty. Time for training is a factor of object and not of style. Content loss results in deformation of novel views.

(Right) We can directly compare the style loss across both novel views rendered by each model. Neural Meshes resulted in **1.81e-2** loss while DeepVoxels performed better at **7.25e-3** loss. DeepVoxels outperforms Neural Meshes when applying Van Gogh's Starry Night to an armchair. We do see deformation in both models, though, it is not quantitatively comparable by content loss functions.

Style Image

Neural Meshes



Original Model

DeepVoxels

Acknowledgements: We thank Deep Learning professor Iddo Drori and TA Benedikt Schifferer for guidance and feedback during our project.