[1] UNIVERSITY OF CAMBRIDGE
[2] TOYOTA

Style2NeRF: An Unsupervised One-Shot NeRF for Semantic 3D Reconstruction

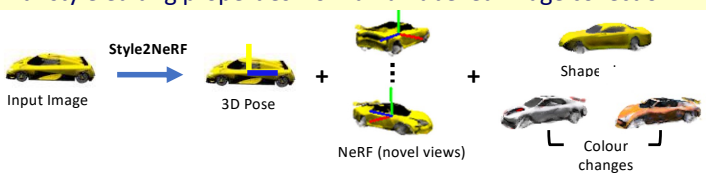
James Charles[1], Wim Abbeloos[2], Daniel Olemda Reino[2] and Roberto Cipolla[1]

Style2NeRF is an unsupervised model for recovery of 3D pose, shape and appearance of symmetric objects from a single image. The model can be trained from unlabelled single-view image collections of object instances.

No 2D or 3D annotation is required. We demonstrate the method on cars and show how our model can also work on car image collections where the underlying pose distribution is unknown.

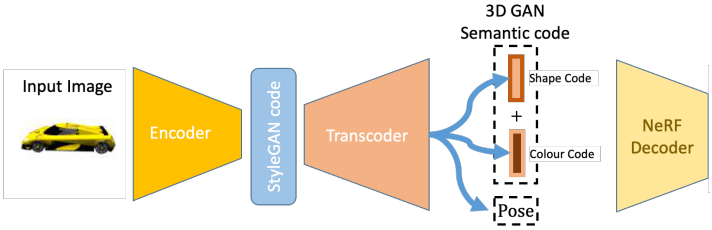
Background / Objective

Goal: Create a model for 3D pose, shape and appearance recovery with style editing properties from an unlabelled image collection



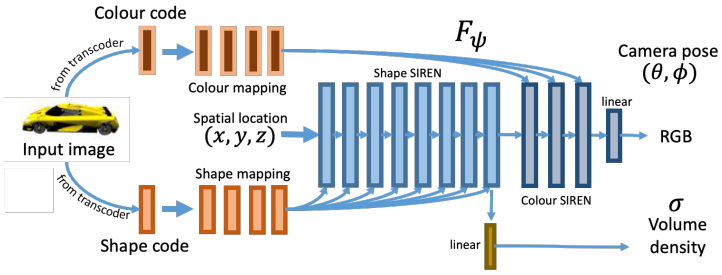
Trainable from unlabelled image collections e.g.:

Model Overview

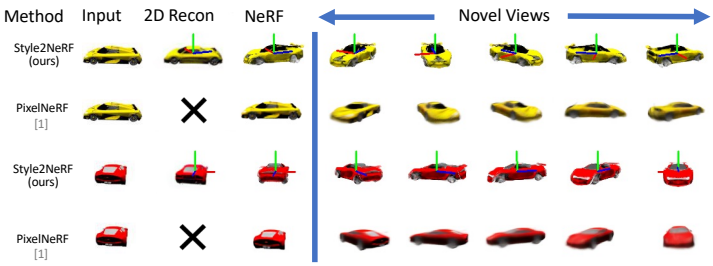


Example renderings from NeRF Decoder

NeRF Decoder



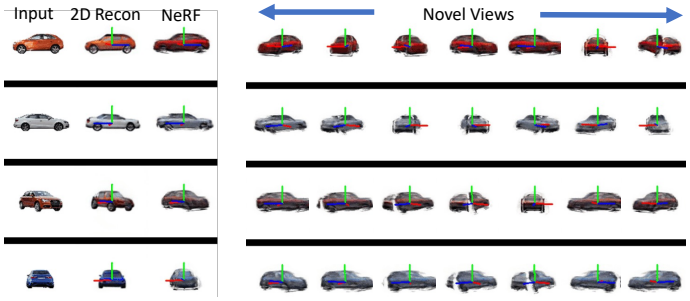
Qualitative Results (SRNCars Dataset)



More structured output than state of the art with improved reconstruction from unobserved view point

[1] Yu, Alex, et. al., pixelnerf: Neural radiance fields from one or few images, CVPR 2021

Qualitative Results (RealCars Dataset)



Trained without knowing underling pose distribution

Quantitative Results

Method	64 × 64 FID ↓	64 × 64 KID ↓	64 × 64 IS ↑	128 × 128 FID ↓	128 × 128 KID ↓	128 × 128 IS ↑
HoloGAN [25]	134	9.70	-	67.5	3.95	3.52
GRAF [31]	30	0.91	-	41.7	2.43	3.70
π-GAN [6]	13.59	0.34	3.85	29.2	1.36	4.27
Pix2NeRF unconditional	10.54	0.37	3.95	27.23	1.43	4.38
Pix2NeRF conditional	12.06	0.44	3.81	38.51	2.37	3.89
Style2NeRF (ours) unconditional	**9.29**	**0.33**	**3.98**	**21.93**	**1.03**	**4.57**
Style2NeRF (ours) conditional	11.03	0.43	3.82	35.84	2.00	3.86

Performs better than state of the art on CARLA Cars dataset

SRNCars - Rotation error (degrees)				RealCars - Rotation error (degrees)			
Pose regressor	Elev.	Azi.	min-Azi.	Pose regressor	Elev.	Azi.	min-Azi.
Baseline (ResNet34)	14	48	13	Baseline (ResNet34)	**1**	77	41
Style2NeRF w/o con.	10	85	16	Style2NeRF w/o con	6	114	32
Style2NeRF (ours)	**9**	**45**	**8**	Style2NeRF (ours)	3	**70**	**19**

Pose estimates are **better than fully supervised Baseline** model

Summary

1. Style2NeRF - unsupervised one-shot NeRF
2. Performs better than a fully supervised pose estimator
3. Semantically editable.
4. Transfers to image collection with unknown pose prior

Future Work

1. Improve speed of training using more efficient NeRF
2. Address articulated objects e.g. people