Unsupervised Style-based Explicit 3D Face Reconstruction from Single Image Supplementary Material

Heng Yu Zoltán Á. Milacski László A. Jeni
Robotics Institute, Carnegie Mellon University
{hengyu, zmilacsk}@andrew.cmu.edu laszlojeni@cmu.edu

1. Architecture Detail

We show the architectures of our method in this section: generator network G in Table 3, discriminator network D in Table 1 and style network S in Table 2. 'Type' indicates whether this layer is shared between different branches and all layers in encoder/decoder are shared. 'Style', 'Light' and 'View' branch output style code s, light l and viewpoint w respectively and they are directly append to the encoder. 'Depth' and 'Albedo' branches output depth map d and albedo a. They share the whole encoder-decoder layers. 'Num' represents repeating numbers of specific layers and we also utilize leaky ReLU [1] in discriminator network D.

2. Additional Results

More qualitative results can be found in https://ue3dst.github.io/.

Table 1. Discriminator Architecture

Num	Layer	Resample	Dim
1×	Conv 1×1	–	$64 \times 64 \times 64$
1×	ResBlk	AvgPool	$32 \times 32 \times 128$
1×	ResBlk	AvgPool	$16 \times 16 \times 256$
1×	ResBlk	AvgPool	$8 \times 8 \times 512$
1×	ResBlk	AvgPool	$4 \times 4 \times 512$
1×	Leaky ReLU	–	$4 \times 4 \times 512$
1×	Conv 4×4	–	$1 \times 1 \times 512$
1×	Leaky ReLU	–	$1 \times 1 \times 512$
1×	Reshape	–	512
1×	Linear * N	–	K * N

Table 2. Style Network Architecture

Num	Type	Layer	Activation	Dim
4×	Shared	Linear	ReLU	512
3×	Unshared	Linear	ReLU	512
1×	Unshared	Linear	–	64

References

[1] Andrew L Maas, Awni Y Hannun, Andrew Y Ng, et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30, page 3. Citeseer, 2013. 1

Table 3. Generator Architecture

Name	Type	Layer	Stride	Activation	Norm	Dim
Encoder	Shared	Conv 4×4	2	LeakyReLU	GroupNorm	$32 \times 32 \times 64$
	Shared	Conv 4×4	2	LeakyReLU	GroupNorm	$16 \times 16 \times 128$
	Shared	Conv 4×4	2	LeakyReLU	GroupNorm	$8 \times 8 \times 256$
	Shared	Conv 4×4	2	LeakyReLU	GroupNorm	$4 \times 4 \times 512$
	Shared	Conv 4×4	1	LeakyReLU	GroupNorm	$1 \times 1 \times 256$
Decoder	Shared	ConvTranspose 4×4	1	ReLU	–	$4 \times 4 \times 512$
	Shared	Conv 3×3	1	ReLU	–	$4 \times 4 \times 512$
	Shared	ConvTranspose 4×4	2	ReLU	AdaIN	$8 \times 8 \times 256$
	Shared	Conv 3×3	1	ReLU	AdaIN	$8 \times 8 \times 256$
	Shared	ConvTranspose 4×4	2	ReLU	AdaIN	$16 \times 16 \times 128$
	Shared	Conv 3×3	1	ReLU	AdaIN	$16 \times 16 \times 128$
	Shared	ConvTranspose 4×4	2	ReLU	AdaIN	$32 \times 32 \times 64$
	Shared	Conv 3×3	1	ReLU	AdaIN	$32 \times 32 \times 64$
	Shared	Upsample 2×2	–	–	–	$32 \times 32 \times 64$
	Shared	Conv 3×3	1	ReLU	AdaIN	$64 \times 64 \times 64$
Style	Unshared	Linear	–	–	–	64
	Unshared	Linear	–	–	–	64
Depth	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 5×5	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 5×5	1	–	–	$64 \times 64 \times 1$
Albedo	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 5×5	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 5×5	1	Tanh	–	$64 \times 64 \times 3$
Light	Unshared	Conv 1×1	1	ReLU	–	$64 \times 64 \times 256$
	Unshared	Conv 1×1	1	ReLU	–	$64 \times 64 \times 256$
	Unshared	Conv 1×1	1	Tanh	–	$64 \times 64 \times 4$
View	Unshared	Conv 1×1	1	ReLU	–	$64 \times 64 \times 256$
	Unshared	Conv 1×1	1	ReLU	–	$64 \times 64 \times 256$
	Unshared	Conv 1×1	1	Tanh	–	$64 \times 64 \times 6$
Conf	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 3×3	1	ReLU	GroupNorm	$64 \times 64 \times 64$
	Unshared	Conv 5×5	1	Softplus	GroupNorm	$64 \times 64 \times 2$