Table 1. A detailed description of G1, the generator in Stage 1 (the shape stage). Layer 2 and B are joined in the channel dimension.

	actuned descrip	on or or, in	e generator	in stage I (the shap	c stage). Lay	or 2 und 1	o are join	ou in th	e chamer anner	151011.	
	Layer Index (By Depth)	Join from	Network Type	# Input Feature Channels	# Output Feature Channels	Kernel Size	Stride	Pad	Followed by BN and Activation?	output s	siee
ĺ	Layer 1	$\{\mathbf{z}_S, \mathbf{d}\}$	Deconv	150	1024	4	4	0	BN / ReLU	4x4	
	Layer 2	Layer 1	Deconv	1024	512	4	2	1	BN / ReLU	8×8	
	Layer A	$\downarrow m(S_0)$	Conv	4	64	3	1	1	BN / ReLU	3×3	
	Layer B	Layer A	Conv	64	128	3	1	1	BN / ReLU	3×3 3×8	
	Layer 3	Layer 2, B	Deconv	640 (512 + 128)	256	4	2	1	BN / ReLU	loxic	
	Layer 4	Layer 3	Deconv	256	128	4	2	1	BN / ReLU	35,435	
	Layer 5	Layer 4	Deconv	128	64	4	2	1	BN / ReLU	D4x 6ce	
	Layer 6	Layer 5	Deconv	64	7	4	2	1	SoftMax	128 x15 8	

Table 2. A detailed description of G2, the generator in Stage 2 (the rendering stage). Layer 3 and C are joined in the channel dimension.

	icianeu uescrij	puon or G2, in	e generator	m stage 2 (the fend	ering stage).	Layer 5 a	na C are	joineu i	ii tile chamilei u	inclisi
	Layer	Join	Network	# Input	# Output	Kernel			Followed by	
	Index			Feature	Feature	Size	Stride	Pad	BN and	
	(By Depth)	from	Type	Channels	Channels	Size			Activation?	
ĺ	Layer 1	$\{\mathbf{z}_I, \mathbf{d}\}$	Deconv	150	1024	4	4	0	BN / ReLU	40
Ì	Layer 2	Layer 1	Deconv	1024	512	4	2	1	BN / ReLU	δ×
Ì	Layer 3	Layer 2	Deconv	512	256	4	2	1	BN / ReLU	16
	Layer A	Š Į ČĖX	Conv Conv	7	64	4	2	1	BN / ReLU	64
Ì	Layer B	Layer A	Conv	64	128	4	2	1	BN / ReLU	32
	Layer C	Layer B	Conv	128	256	4	2	1	BN / ReLU	16
- [	Layer 4	Layer 3, C	Deconv	512 (256 + 256)	128	4	2	1	BN / ReLU	32
	Layer 5	Layer 4	Deconv	128	64	4	2	1	BN / ReLU	64
ĺ	Layer 6	Layer 5	Deconv	64	3	4	2	1	Tanh	13
	1-4-11-1 1	+! C D 1 +1		-+ ! C+ 1 (+1	-1	T 4	1 D	- 1 - 1	1 111	11

Table 3. A detailed description of D1, the discriminator in Stage 1 (the shape stage). Layer 4 and B are joined in the channel dimension.  $\mathbf{d}_{\times 4}$  denotes the four-time replicated version of  $\mathbf{d}$  in the two spatial dimensions. It is joined with Layer 5 in the channel dimension.

Layer Index	Join from	Network	# Input Feature	# Output Feature	Kernel Size	Stride	Pad	Followed by BN and
(By Depth)	Hom	Type	Channels	Channels	Size			Activation?
Layer 1	G1 Output	Conv	7	64	4	2	1	LeakyReLU(0.2)
Layer 2	Layer 1	Conv	64	128	4	2	1	BN / LeakyReLU(0.2)
Layer 3	Layer 2	Conv	128	256	4	2	1	BN / LeakyReLU(0.2)
Layer 4	Layer 3	Conv	256	512	4	2	1	BN / LeakyReLU(0.2)
Layer A	$\downarrow m(S_0)$ for	Conv	4	64	3	1	1	BN / LeakyReLU(0.2)
Layer B	Layer A	Conv	64	128	3	1	1	BN / LeakyReLU(0.2)
Layer 5	Layer 4, B	Conv	640 (512 + 128)	1024	4	2	1	BN / LeakyReLU(0.2)
Layer 6	Layer 5, $\mathbf{d}_{\times 4}$	Conv	1074 (1024 + 50)	1024	1	1	0	BN / LeakyReLU(0.2)
Layer 7	Layer 6	Conv	1024	1	4	4	0	BN / LeakyReLU(0.2)

Table 4. A detailed description of D2, the discriminator in Stage 2 (the rendering stage). Layer 3 and C are joined in the channel dimension.

 $\mathbf{d}_{\times 4}$  denotes the four-time replicated version of  $\mathbf{d}$  in the two spatial dimensions. It is joined with Layer 5 in the channel dimension.

4 denotes the four-time replicated version of <b>d</b> in the two spatial dimensions. It is joined with Layer 3 in the channel dimension.									
ſ	Layer	Join	Network Type	# Input	# Output	Kernel Size	Stride	Pad	Followed by
	Index	Index from		Feature	Feature				BN and
	(By Depth)	Hom		Channels	Channels				Activation?
ĺ	Layer 1	G2 Output	Conv	3	64	4	2	1	LeakyReLU(0.2)
ĺ	Layer 2	Layer 1	Conv	64	128	4	2	1	BN / LeakyReLU(0.2)
[	Layer 3	Layer 2	Conv	128	256	4	2	1	BN / LeakyReLU(0.2)
	Layer A	Š171/123	Conv	7	64	4	2	1	BN / LeakyReLU(0.2)
ſ	Layer B	Layer A	Conv	64	128	4	2	1	BN / LeakyReLU(0.2)
[	Layer C	Layer B	Conv	128	256	4	2	1	BN / LeakyReLU(0.2)
ĺ	Layer 4	Layer 3, C	Conv	512 (256 + 256)	512	4	2	1	BN / LeakyReLU(0.2)
Ì	Layer 5	Layer 4	Conv	512	1024	4	2	1	BN / LeakyReLU(0.2)
ĺ	Layer 6	Layer 5, $\mathbf{d}_{\times 4}$	Conv	1074 (1024 + 50)	1024	1	1	0	BN / LeakyReLU(0.2)
ĺ	Layer 7	Layer 6	Conv	1024	1	4	4	0	BN / LeakyReLU(0.2)