Structure-Guided Deep Video Inpainting

Chaoqun Wang, Xuejin Chen, *Member, IEEE*, Shaobo Min, Jiaping Wang, *Member, IEEE*, and Zheng-Jun Zha *Member, IEEE*

TABLE I
ENET ARCHITECTURE DETAILS.

Network	Arch	Operation	Kernel size	Output channels
G^E	Encoder	Conv3D	(3,7,7)	[64]
		Conv3D	(3,4,4)	[128,256]
	ResnetBlock	DilatedConv2D	(3,3)	[256]*8
	Decoder	ConvTranspose3D	(3,4,4)	[128,64]
		ConvTranspose3D	(3,7,7)	[1]
D^E		Conv2D	(4,4)	[64,128,256,512,1]

TABLE II
TEXNET ARCHITECTURE DETAILS. WE USE GATED CONVOLUTION FOR THE WHOLE NETWORK, WHICH IS OMITTED FOR SPACE SAVING. DECONVBLOCK DENOTES DECONV+CONV.

Network	Arch	Operation	Kernel size	Output channels
G^T	Coarse	Conv3D	(3,5,5)	[16]
		Conv3D	(3,3,3)	[32,64,128]
		DilatedConv3D	(3,3,3)	[256,256,256]
		Conv3D	(3,3,3)	[128]
		DeConvBlock3D	(3,3,3)	[32,3]
	Refine	Conv2D	(5,5)	[64]
		Conv2D	(3,3)	[128,128,256,256]
		DilatedConv2D	(3,3)	[256,256,256,256]
		Conv2D	(3,3)	[256]
		SAM	-	[256]
		Conv2D	(3,3)	[256]
		DeConvBlock2D	(3,3)	[128,32]
		Conv2D	(3,3)	[3]
D^T		Conv2D	(4,4)	[64,128,256,512,1]

C. Wang, X. Chen, S. Min, and Z. Zha are with the National Engineering Laboratory for Brain-inspired Intelligence Technology and Application, University of Science and Technology of China, Hefei, 230026, China.

J. Wang is with the Peng Cheng Laboratory, Shenzhen, 518000, China.