Network Architectures and Hyperparameters

Operation	Kernel	Strides	Feature Maps	BN	Nonlinearity
G(z,c) - 1 x 1 x 28 input					
Dense			3136	Yes	ELU
Reshape - 7 x 7 x 64					
Transposed Convolution	4 x 4	2 x 2	128	Yes	ELU
Transposed Convolution	4 x 4	1 x 1	64	Yes	ELU
Transposed Convolution	4 x 4	2 x 2	1	No	Sigmoid
E(x) - 28 x 28 x 1 input					
Convolution	3 x 3	1 x 1	32	Yes	ELU
Convolution	3 x 3	2 x 2	64	Yes	ELU
Convolution	3 x 3	2 x 2	128	Yes	ELU
Dense			1024	Yes	ELU
Output z: Dense			1	No	Tanh
Output c_{disc}: Dense			10	No	Softmax
Output c_{cont}: Dense			2	No	Linear
D(x) - 28 x 28 x 1 input					
Convolution	3 x 3	2 x 2	64	Yes	ELU
Convolution	3 x 3	2 x 2	128	Yes	ELU
Dense			512	Yes	ELU
D(Z) - 1 x 1 x 28 input					
Convolution	1 x 1	1 x 1	64	Yes	ELU
Convolution	1 x 1	1 x 1	128	Yes	ELU
Dense			512	Yes	ELU
D(x,Z) - 1 x 1 x 1024 input					
concatenate D(x) and D(Z) along channel axis					
Dense			1024	Yes	ELU
Dense			1	No	Sigmoid
Optimizer D:	Adam ($\alpha = 10^{-4}, \beta_1 = 0.5$)				
Optimizer G and E:	Adam ($\alpha = 5 \times 10^{-4}, \beta_1 = 0.5$)				
Learning Rate Reduction	Polynomial with final LR one fifth of the original				
Batch Size	128				
Iterations	30 000				
G: weight, bias init	Truncated Normal ($\mu = 0, \sigma = 0.02$)				
D and E: weight, bias init	$\mathcal{N}(0, \sqrt{inp/2}^{-1})$ with inp number of inputs to a unit				

Table 1: MNIST Model Hyperparameters

Operation	Kernel	Strides	Feature Maps	BN	Nonlinearity
G(z,c) - 1 x 1 x 98 input					
Dense			2048	Yes	ELU
Reshape - 2 x 2 x 512					
Transposed Convolution	5 x 5	2 x 2	512	Yes	ELU
Transposed Convolution	5 x 5	2 x 2	256	Yes	ELU
Transposed Convolution	5 x 5	1 x 1	128	Yes	ELU
Transposed Convolution	5 x 5	1 x 1	64	Yes	ELU
Transposed Convolution	5 x 5	1 x 1	32	Yes	ELU
Transposed Convolution	4 x 4	1 x 1	3	No	Sigmoid
E(x) - 32 x 32 x 3 input					
Convolution	3 x 3	1 x 1	32	Yes	ELU
Convolution	3 x 3	2 x 2	64	Yes	ELU
Convolution	3 x 3	1 x 1	128	Yes	ELU
Convolution	3 x 3	2 x 2	256	Yes	ELU
Convolution	3 x 3	2 x 2	512	Yes	ELU
Dense			1024	Yes	ELU
Output z: Dense			1	No	Tanh
Output c_{disc}: Dense			4 x 10	No	Softmax
Output c_{cont}: Dense			4	No	Linear
D(x) - 32 x 32 x 3 input					
Convolution	4 x 4	2 x 2	64	Yes	ELU
Convolution	4 x 4	2 x 2	128	Yes	ELU
Convolution	4 x 4	2 x 2	256	Yes	ELU
Dense			1024	Yes	ELU
D(Z) - 1 x 1 x 28 input					
Convolution	1 x 1	1 x 1	64	Yes	ELU
Convolution	1 x 1	1 x 1	128	Yes	ELU
Convolution	1 x 1	1 x 1	256	Yes	ELU
Dense			1024	Yes	ELU
D(x,Z) - 1 x 1 x 2048 input					
concatenate D(x) and D(Z) along channel axis					
Dense			1024	Yes	ELU
Dense			1	No	Sigmoid
Optimizer D, G, and E:	Adam ($\alpha = 10^{-4}, \beta_1 = 0.5$)				
Learning Rate Reduction	Polynomial with final LR one fifth of the original				
Batch Size	32				
Iterations	300 000 (SVHN); 500 000 (CelebA				
G: weight, bias init	Truncated Normal ($\mu = 0, \sigma = 0.02$)				
D and E: weight, bias init	$\mathcal{N}(0, \sqrt{inp/2}^{-1})$ with inp number of inputs to a unit				

Table 2: CelebA / SVHN Model Hyperparameters