

2	2	8	8	1	1	6	6	6	6	5	5	5	5	0	0	9	9	4	4
9	9	8	8	1	1	1	1	5	5	3	3	7	7	8	8	9	9	3	3
0	0	1	1	6	6	5	5	5	5	1	1	0	0	7	7	1	1	6	6
9	9	7	9	8	7	8	7	3	7	2	4	8	9	2	0	1	7	7	
0	7	9	8	6	8	0	0	6	4	2	7	7	8	6	4	4	0	8	
5	3	6	7	3	3	9	4	9	8	8	0	9	7	7	0	4	6	5	9
9	8	7	9	0	8	2	0	3	9	5	1	7	0	5	9	2	0	5	9
7	3	8	8	9	4	7	4	9	9	0	0	4	4	6	9	8	7	1	1
1	3	9	1	4	2	3	1	5	9	6	3	4	9	1	0	0	3	2	9
9	9	7	9	8	7	8	7	3	7	2	4	8	9	2	0	1	7	7	
0	7	9	8	6	8	0	0	6	4	2	7	7	7	8	6	4	4	0	8
5	3	6	7	3	3	9	4	9	8	8	0	9	7	7	0	4	6	5	9
9	8	7	9	0	8	2	0	3	9	5	1	7	0	5	9	2	0	5	9
7	3	8	8	9	4	7	4	9	9	0	0	4	4	6	9	8	7	1	1
1	3	9	1	4	2	3	1	5	9	6	3	4	9	1	0	0	3	2	9

VAE
samples



ST-VAE
samples



ST-VAE samples
in canonical pose



VAE
samples

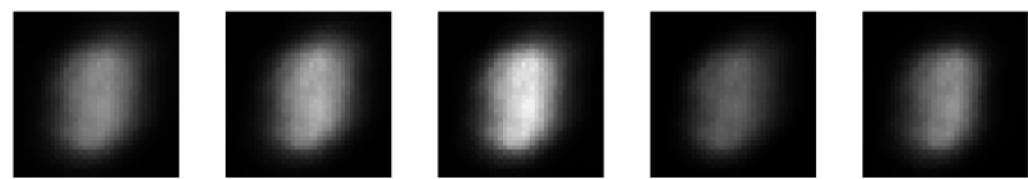
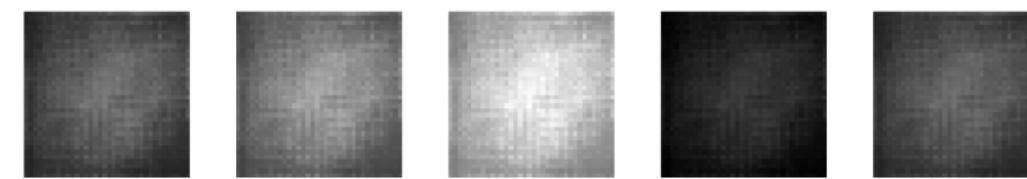
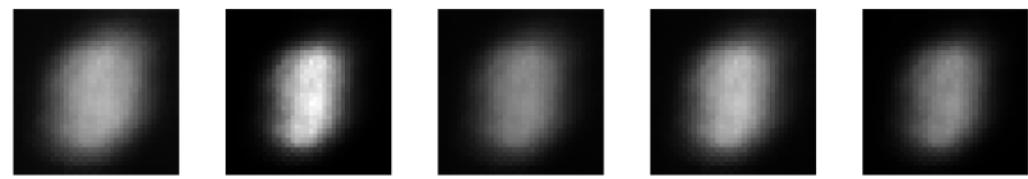
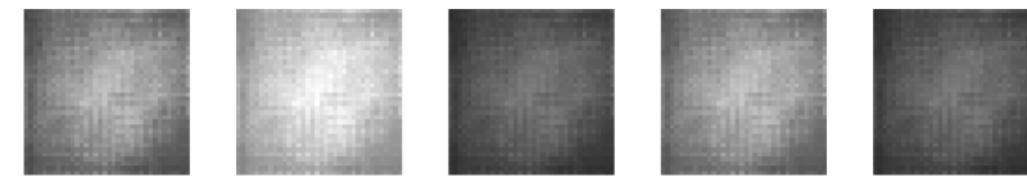


ST-VAE
samples



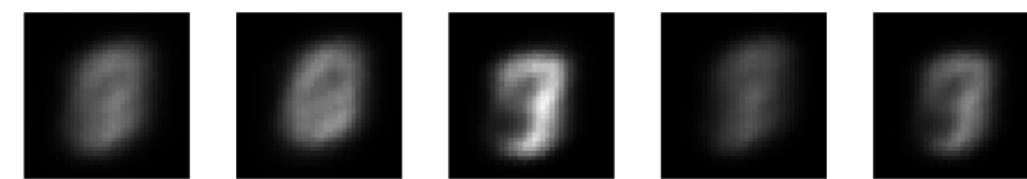
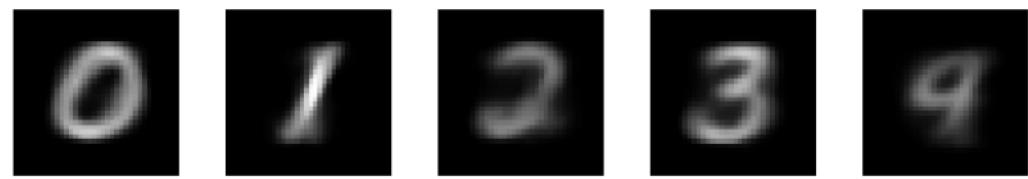
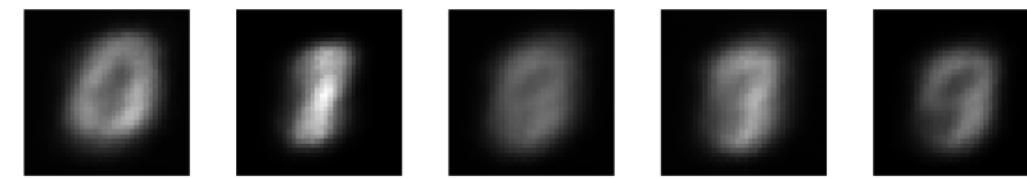
ST-VAE samples
in canonical pose





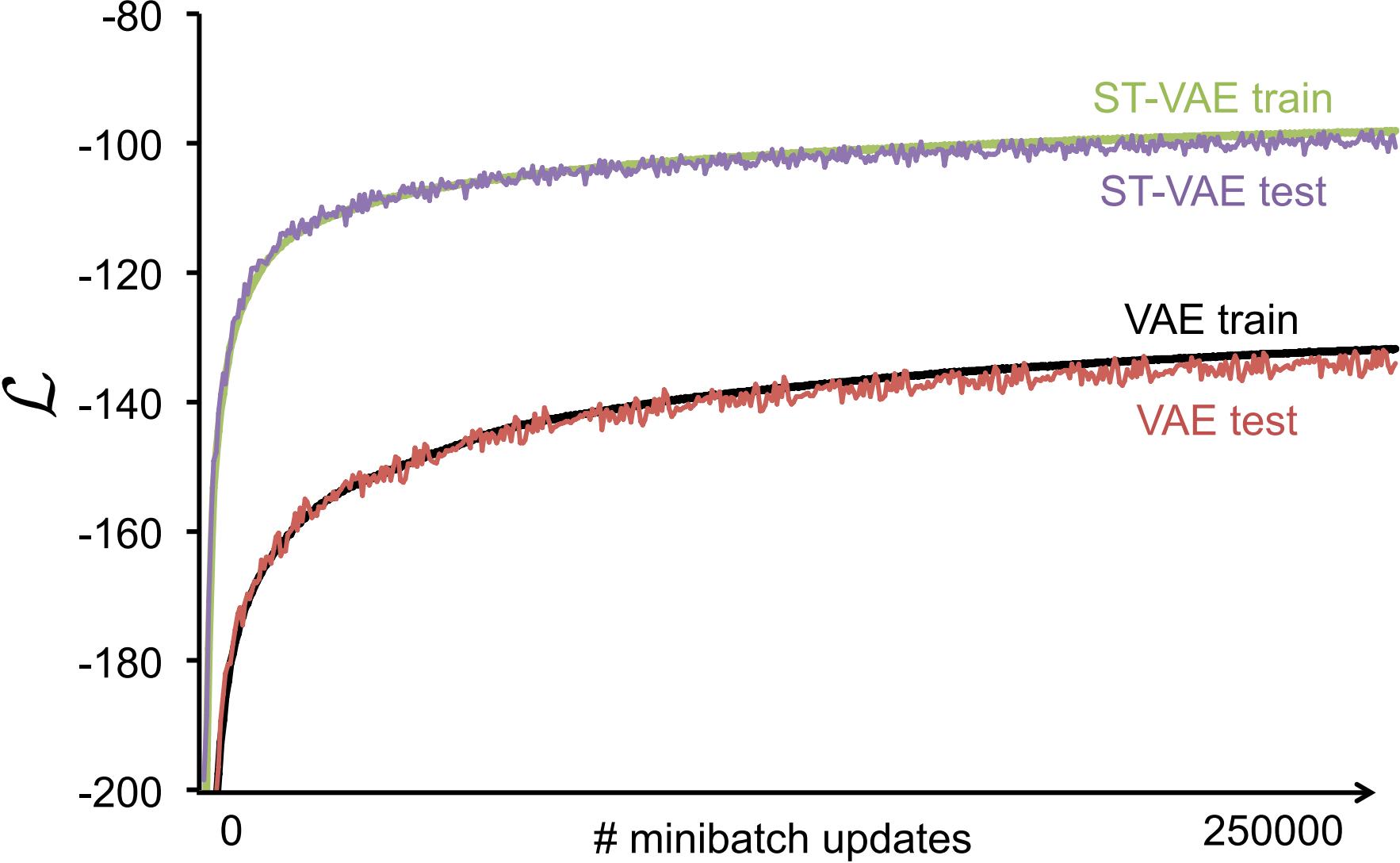
$t=1$

$t=10$



$t=20$

$t=200$



3	8	9	3	0	3	6	0	9	5	0	3	6	7	7	7	3	6	8	8
4	9	9	8	0	5	3	3	8	7	4	5	1	7	0	4	6	6	2	1
3	4	3	0	0	9	4	0	0	8	3	4	9	3	8	7	3	4	9	9
0	5	7	9	3	6	4	9	2	3	0	0	9	8	2	9	7	9	0	4
0	8	0	3	3	0	7	0	4	0	9	7	6	8	9	7	2	4	8	2
7	6	4	8	9	2	0	3	0	4	5	0	6	0	3	8	6	8	1	6

3	8	9	3	0	3	6	0	9	5	0	3	6	7	7	7	3	6	8	8
4	9	9	8	0	5	3	3	8	7	4	5	1	7	0	4	6	6	2	1
3	4	3	0	0	9	4	0	0	8	3	4	9	3	8	7	3	4	9	9
0	5	7	9	3	6	4	9	2	3	0	0	9	8	2	9	7	9	0	4
0	8	0	3	3	0	7	0	4	0	9	7	6	8	9	7	2	4	8	2
7	6	4	8	9	2	0	3	0	4	5	0	6	0	3	8	6	8	1	6

3	3	3 ³	3 ³	4	4	4 ¹	4 ¹	3	2	2 ³	2 ³	3	9	3	3 ⁸	9	3	3 ⁹	3 ⁹
3	7	7 ³	7 ³	9	3	3 ³	3 ³	4	5	5 ³	4 ³	8	6	6	6	8	6	6	6
1	4	4 ¹	4 ¹	0	0	0	0	1	9	9	9	8	1	2	2 ¹	2 ¹	3	2	3

Observed images



Inferred explanations

Reconstruction



1st layer



2nd layer



$$f_C(\cdot;\theta_C)$$

$$f_T(\cdot;\theta_T)$$

$$z_i^C$$

$$C_i$$

$$z_i^T$$

$$T_i$$

$$z^T$$

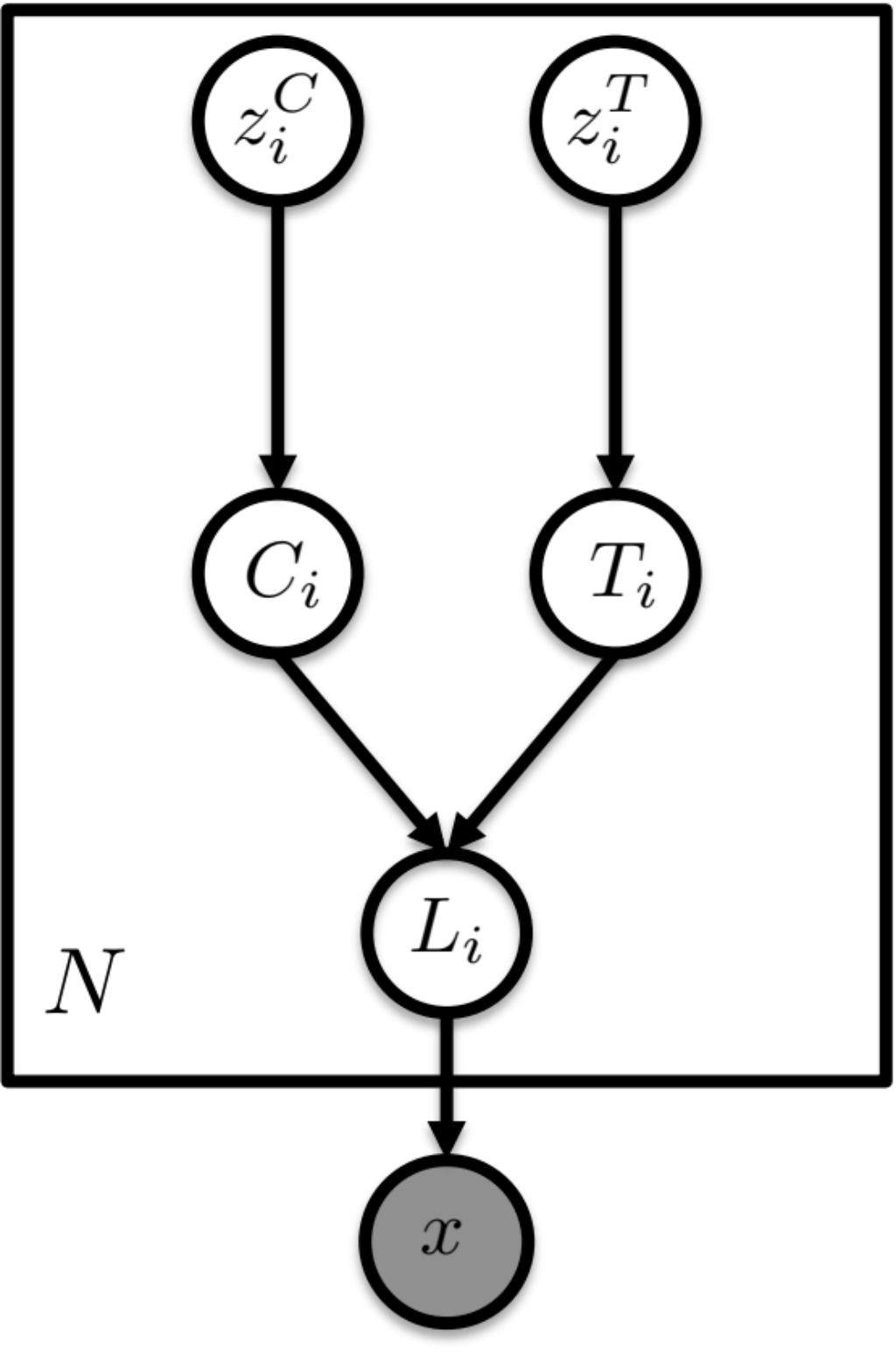
$$\boldsymbol{x}$$

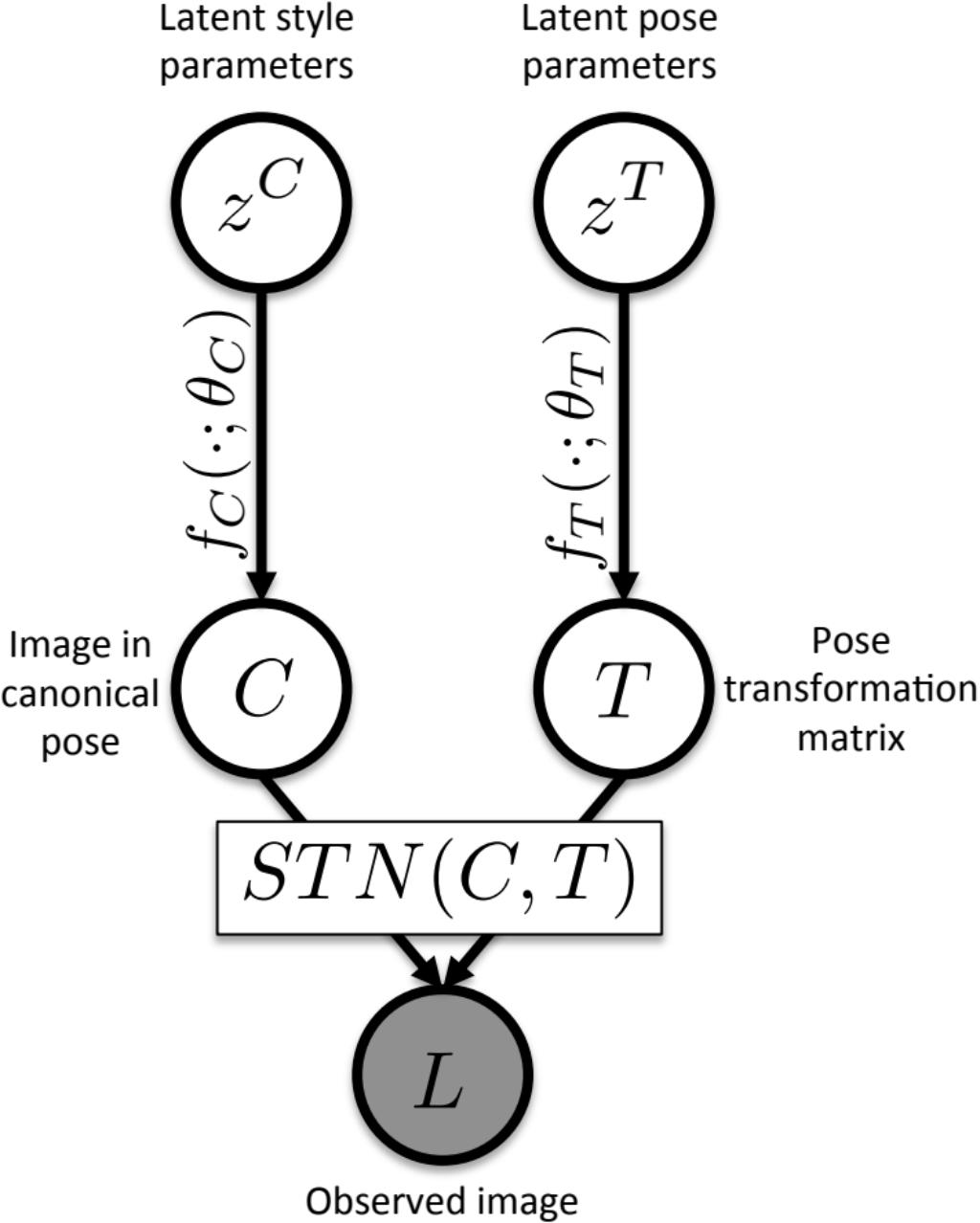
$$L_i$$

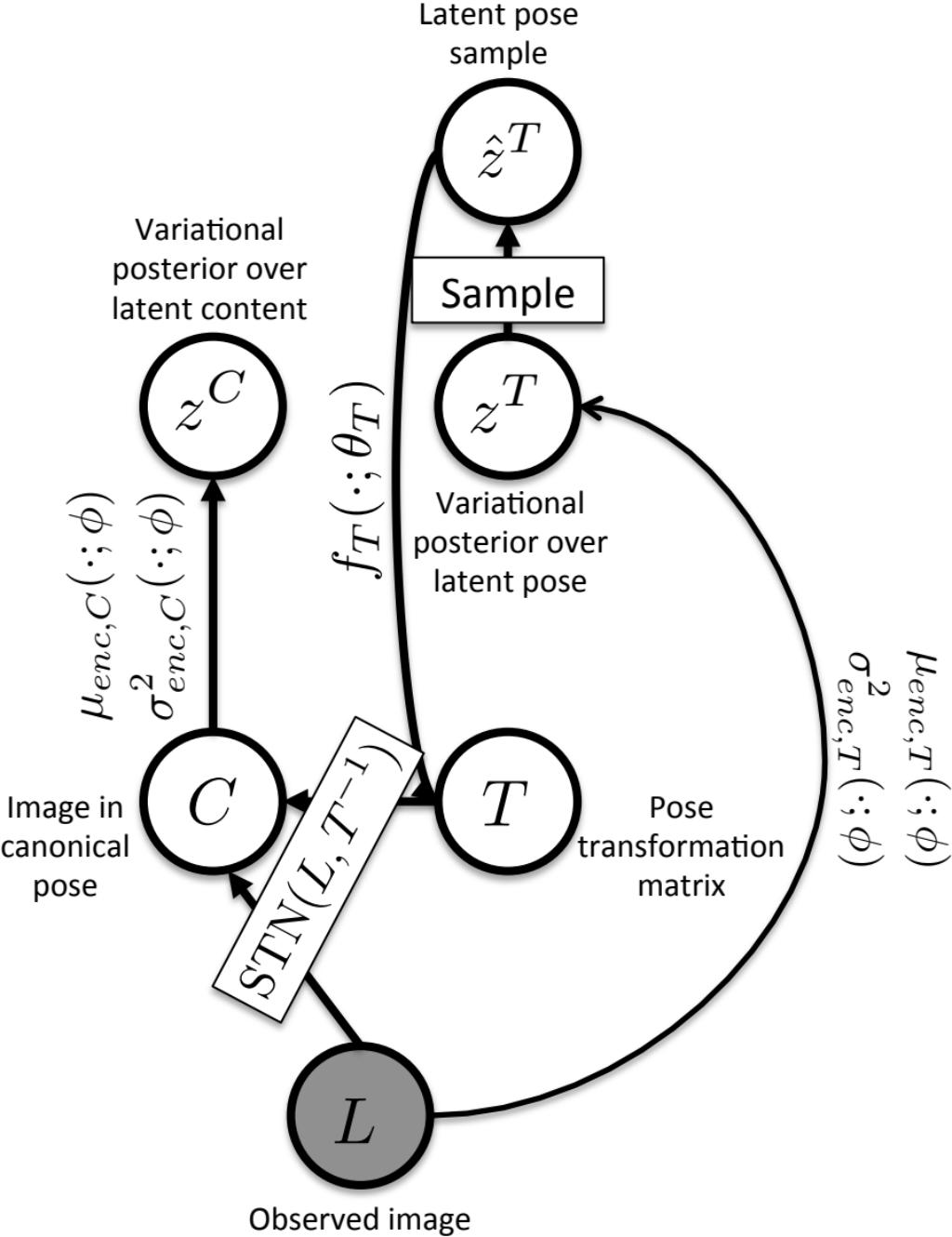
$$STN(C,T)$$

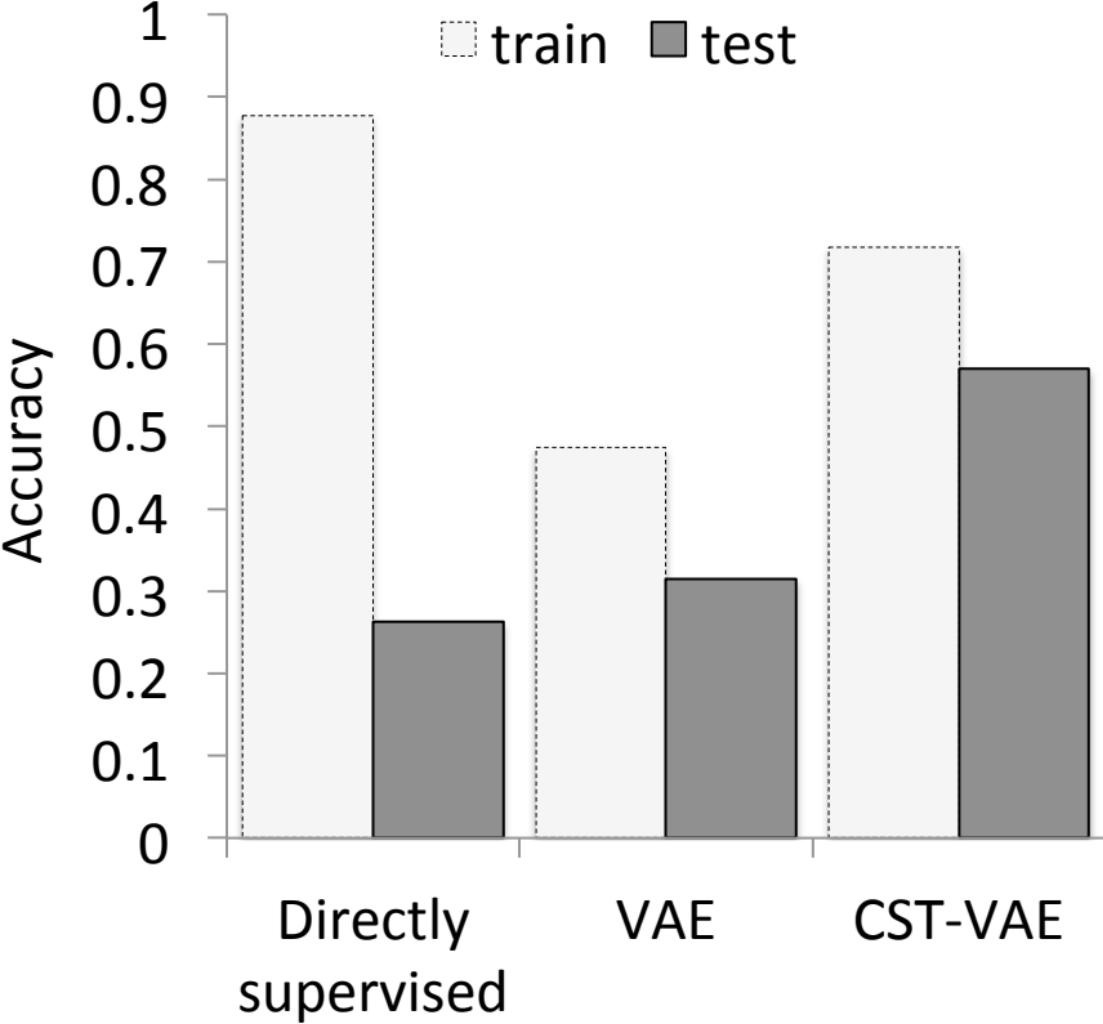
$$\begin{matrix} z^C \\ C \\ L \end{matrix}$$

$$\boldsymbol{T}$$









↓

