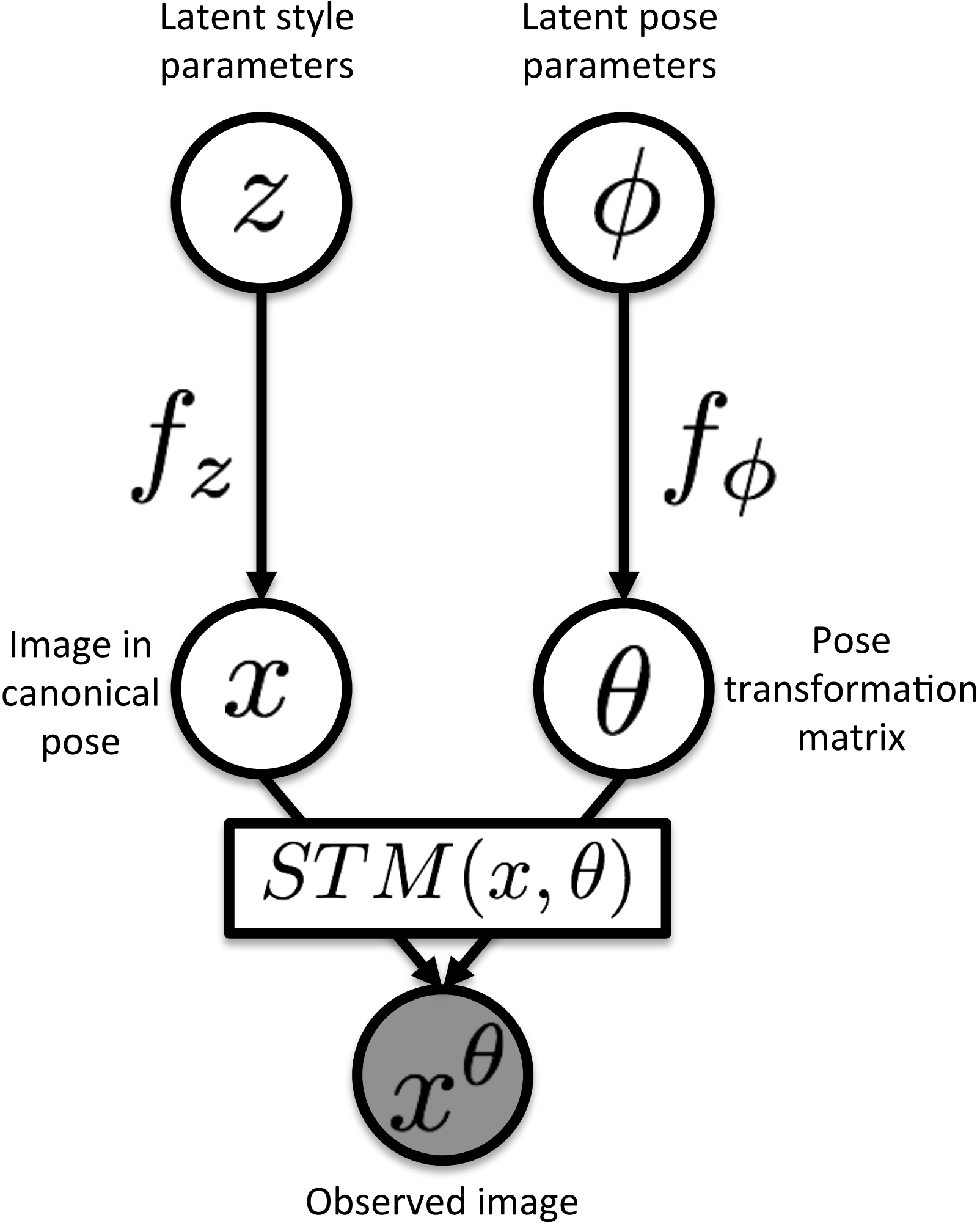
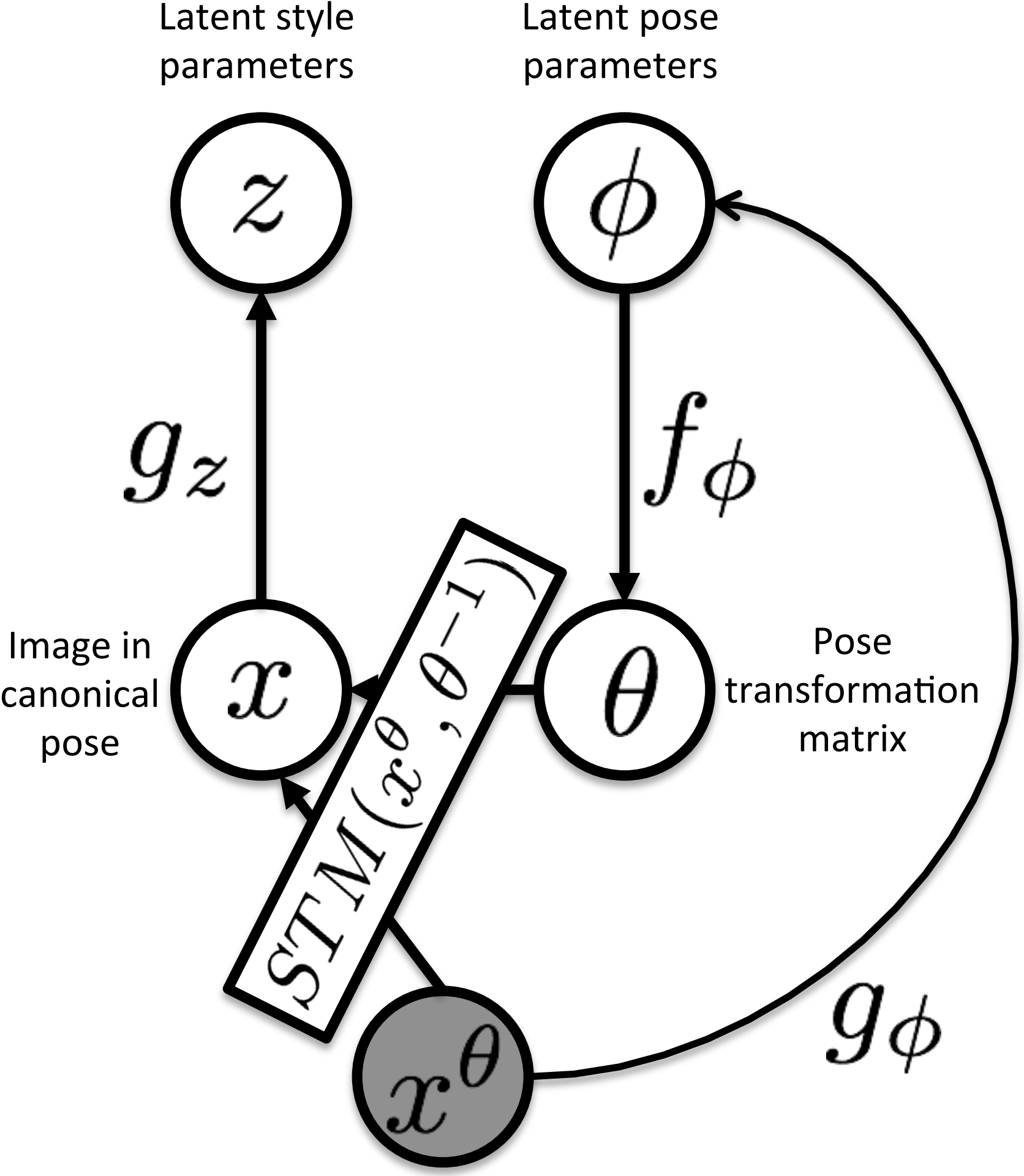
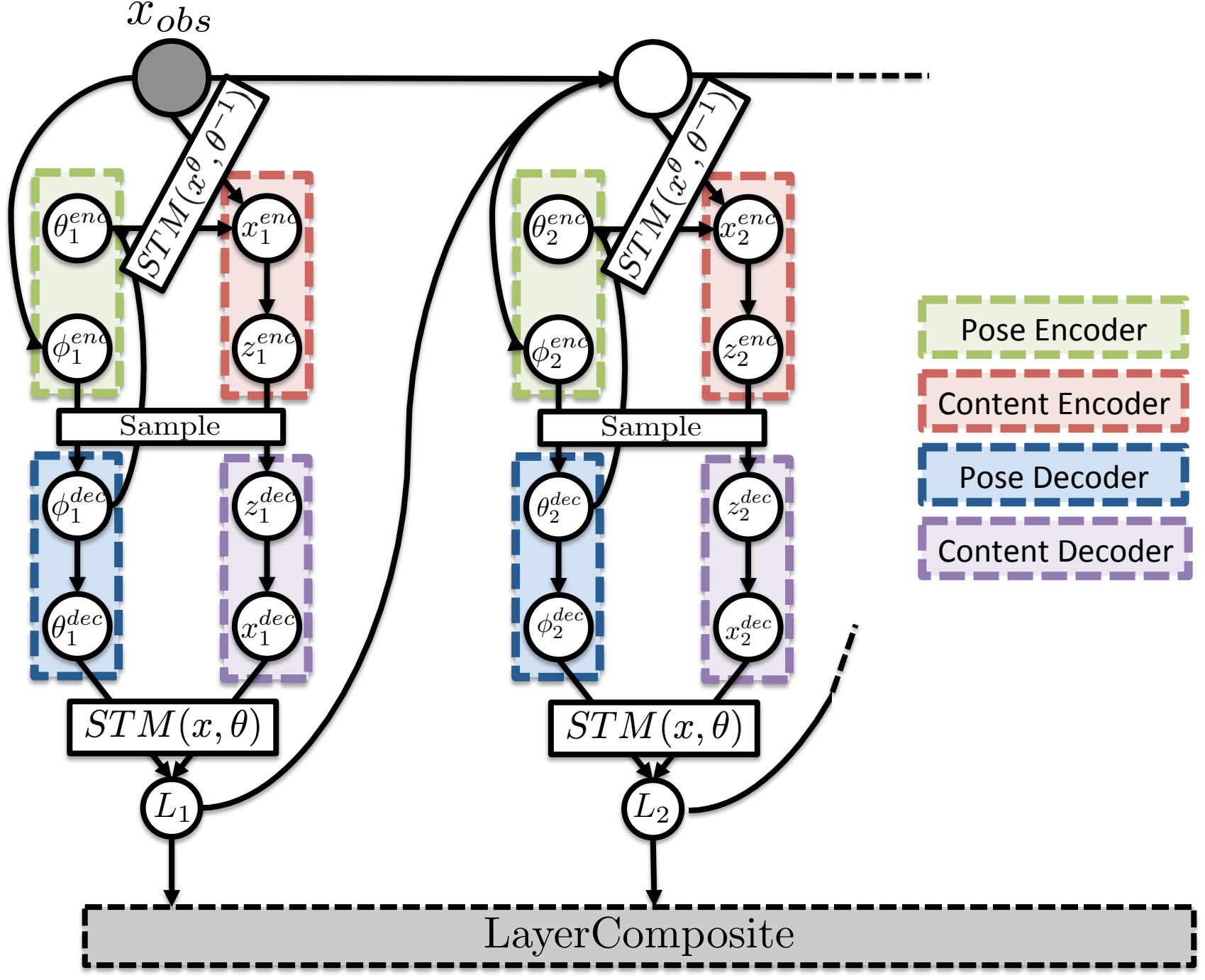
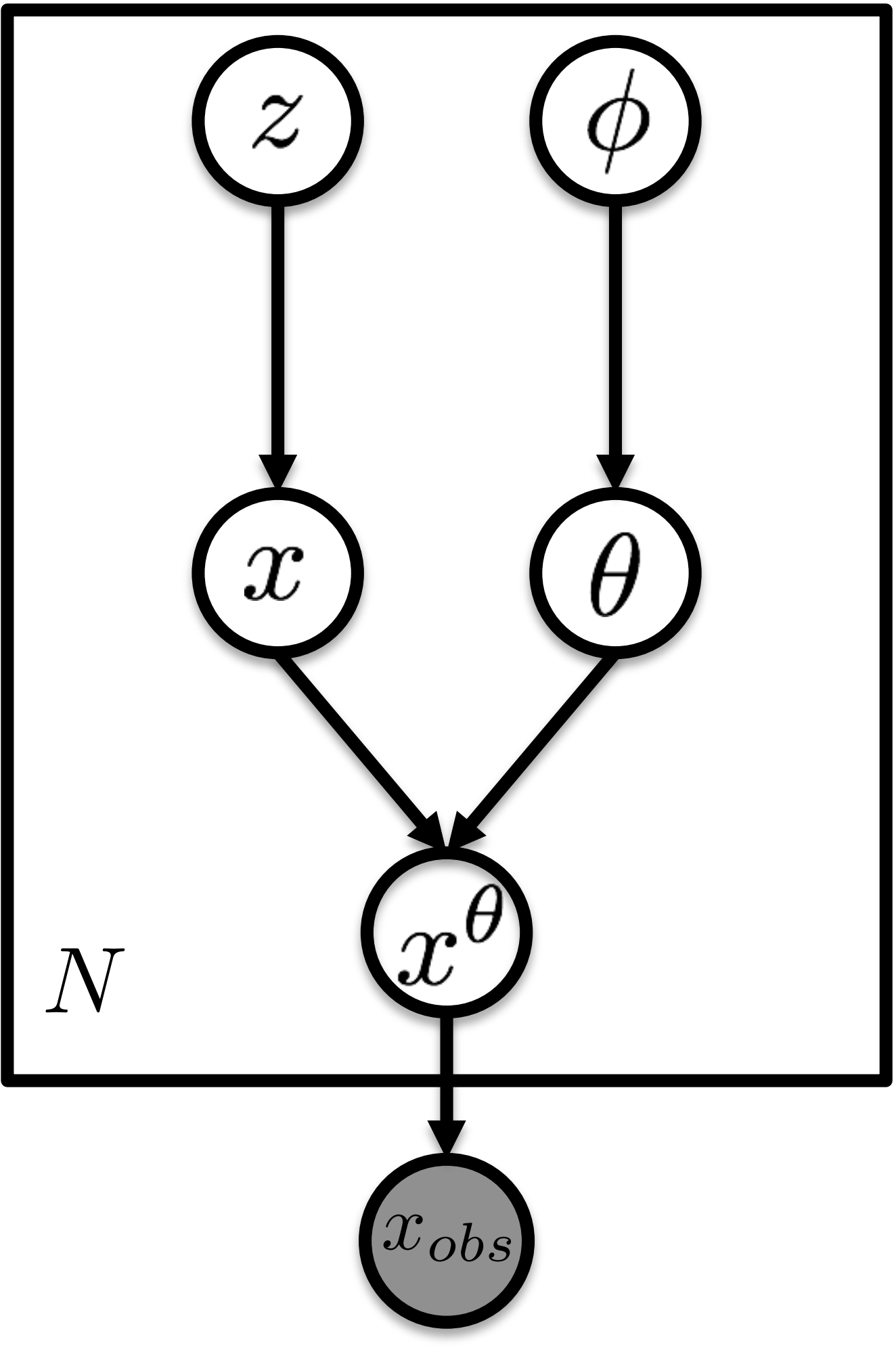
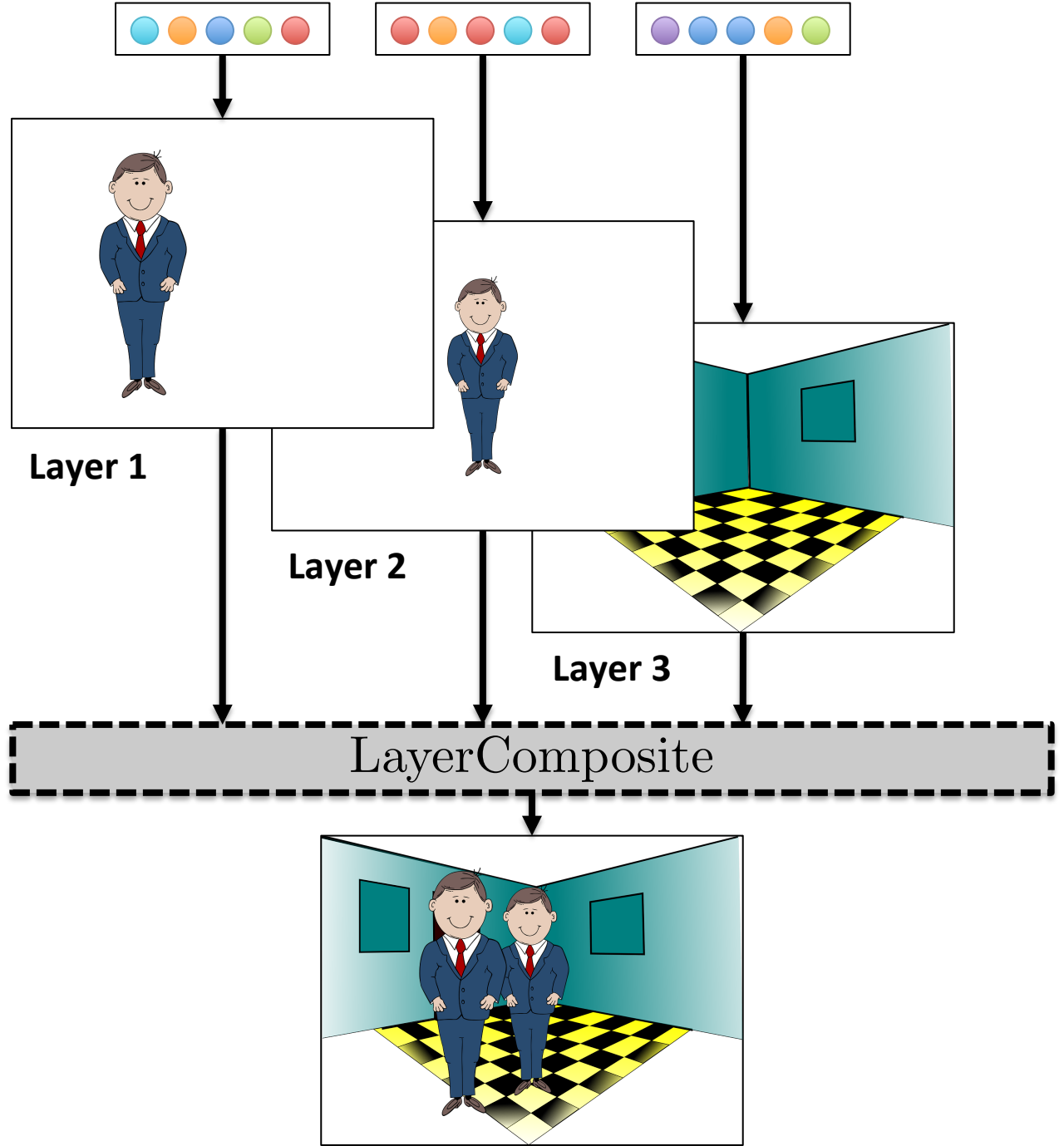
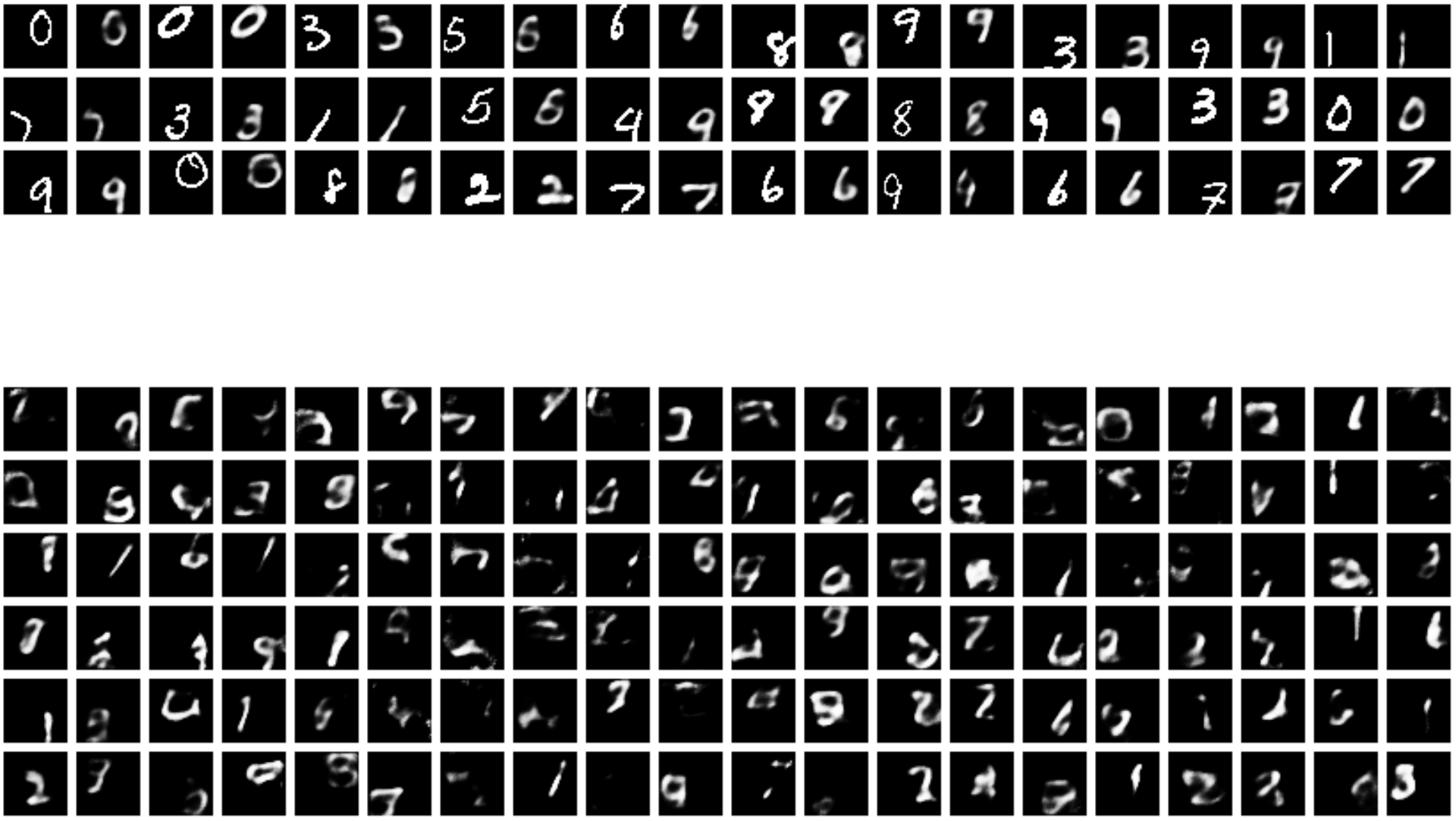

Latent style parameters
Latent pose parameters
z
ϕ
f_z
f_ϕ
Image in canonical pose
x
θ
Pose transformation matrix
$STM(x, \theta)$
x^θ
Observed image

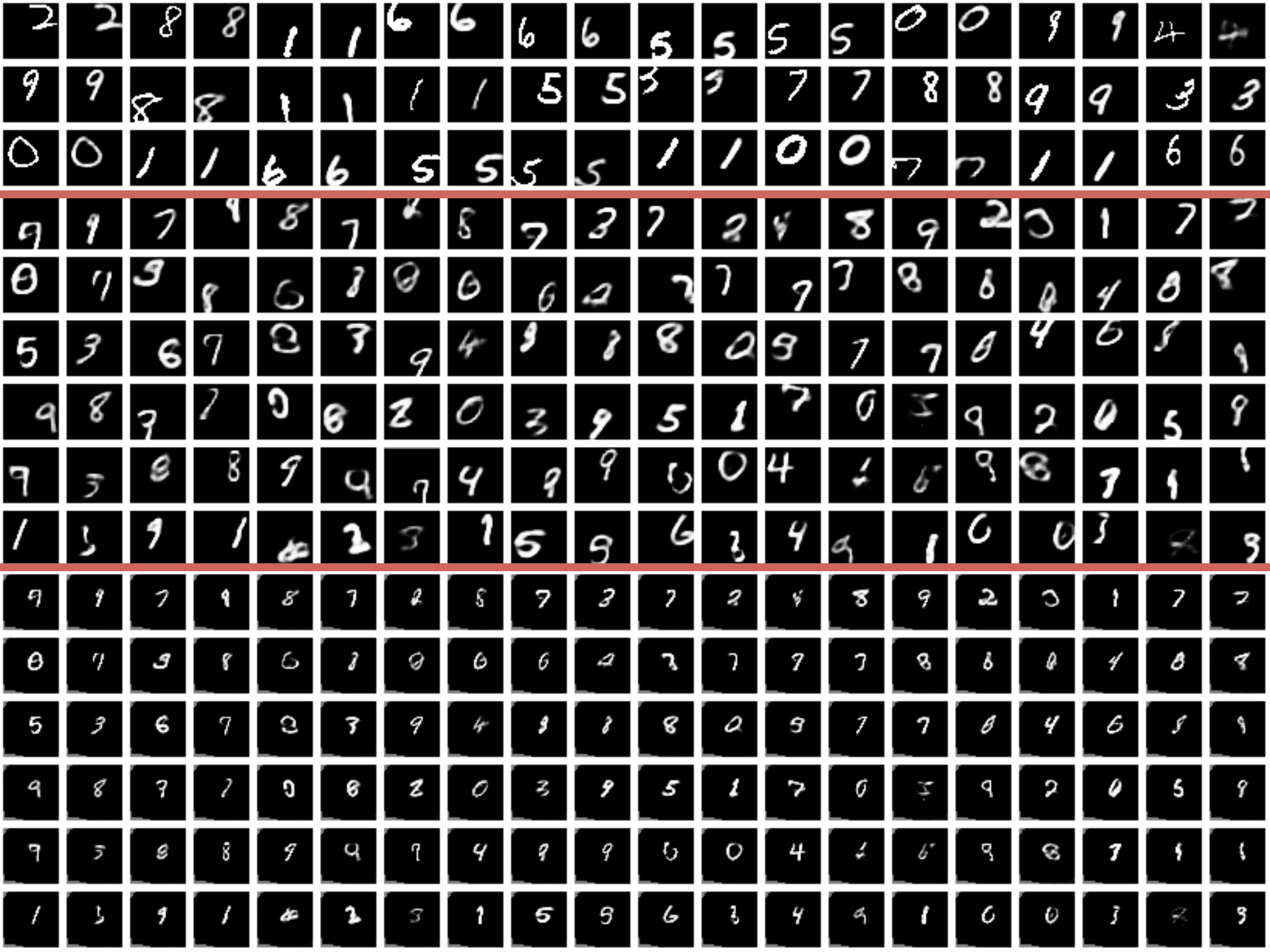

Latent style parameters
Latent pose parameters
z
ϕ
g_z
f_ϕ
Image in canonical pose
x
$STM(x^\theta, \theta^{-1})$
θ
Pose transformation matrix
x^θ
g_ϕ
Observed image


x_{obs}
$STM(x^{\theta}, \theta^{-1})$
θ_1^{enc}
x_1^{enc}
ϕ_1^{enc}
z_1^{enc}
Sample
ϕ_1^{dec}
z_1^{dec}
θ_1^{dec}
x_1^{dec}
$STM(x, \theta)$
L_1
$STM(x^{\theta}, \theta^{-1})$
θ_2^{enc}
x_2^{enc}
ϕ_2^{enc}
z_2^{enc}
Sample
θ_2^{dec}
z_2^{dec}
ϕ_2^{dec}
x_2^{dec}
$STM(x, \theta)$
L_2
LayerComposite
Pose Encoder
Content Encoder
Pose Decoder
Content Decoder

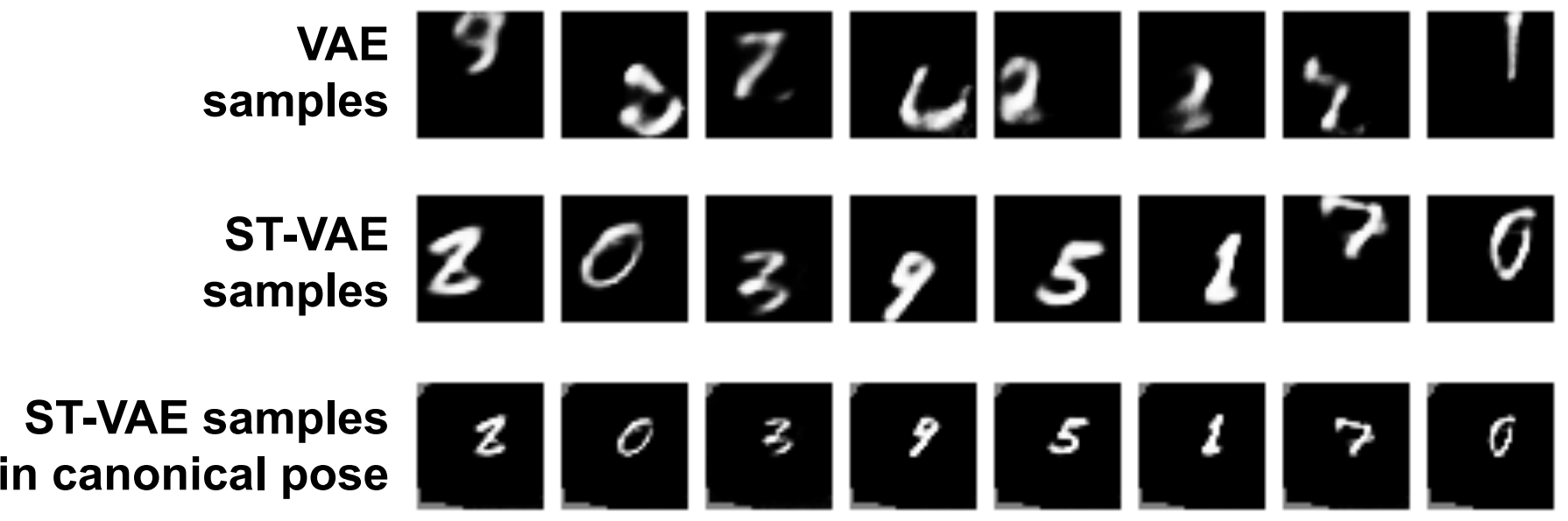

z
ϕ
x
θ
x^{θ}
N
x_{obs}

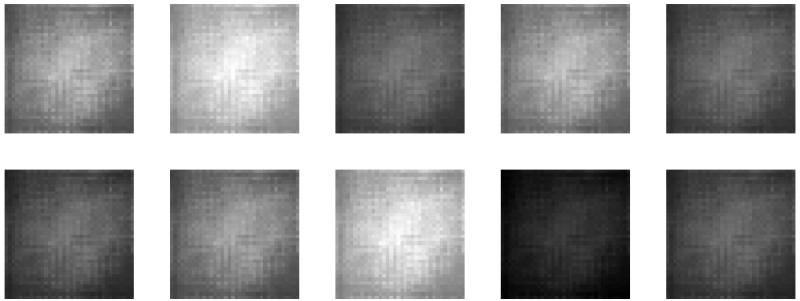

Layer 1
Layer 2
Layer 3
LayerComposite



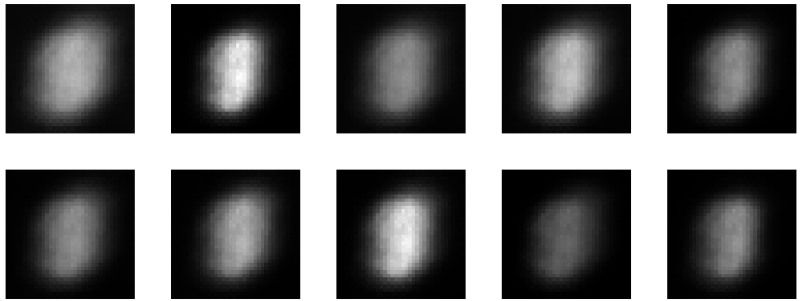



VAE
samples
ST-VAE
samples
ST-VAE samples
in canonical pose

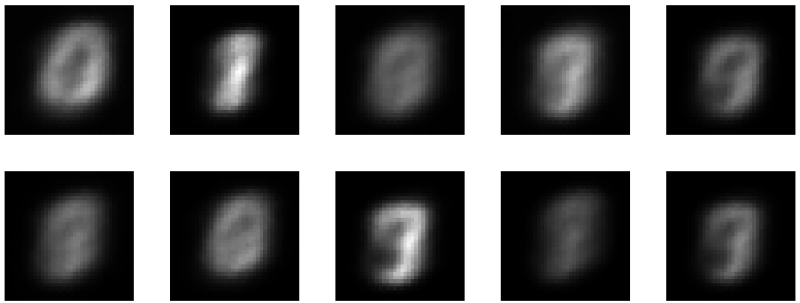

VAE
samples
ST-VAE
samples
ST-VAE samples
in canonical pose



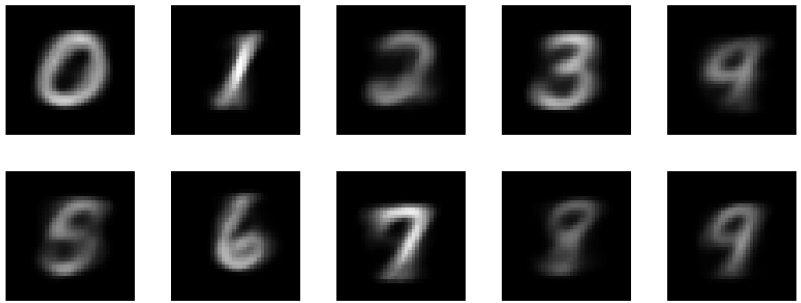
t=1



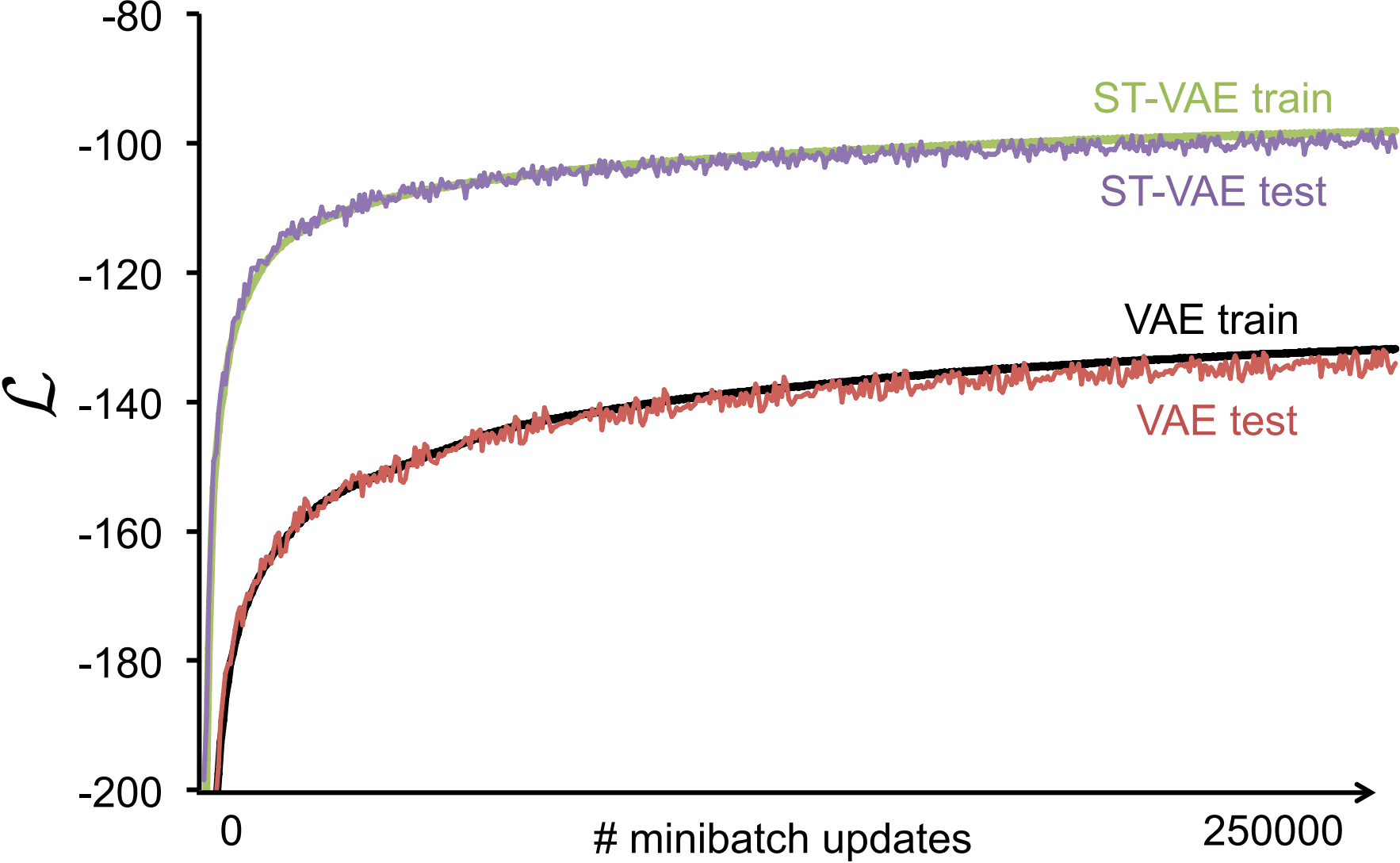
t=10

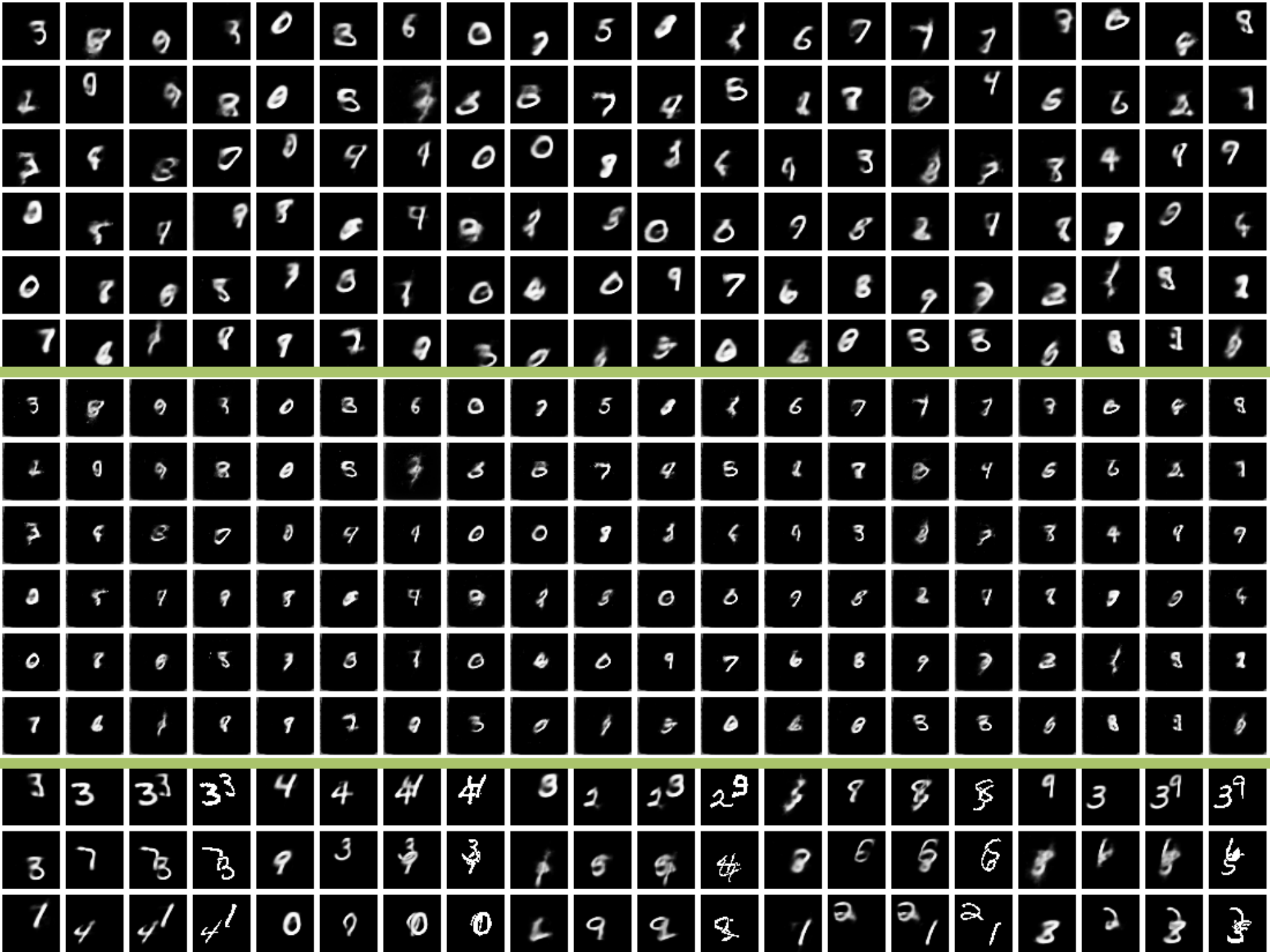


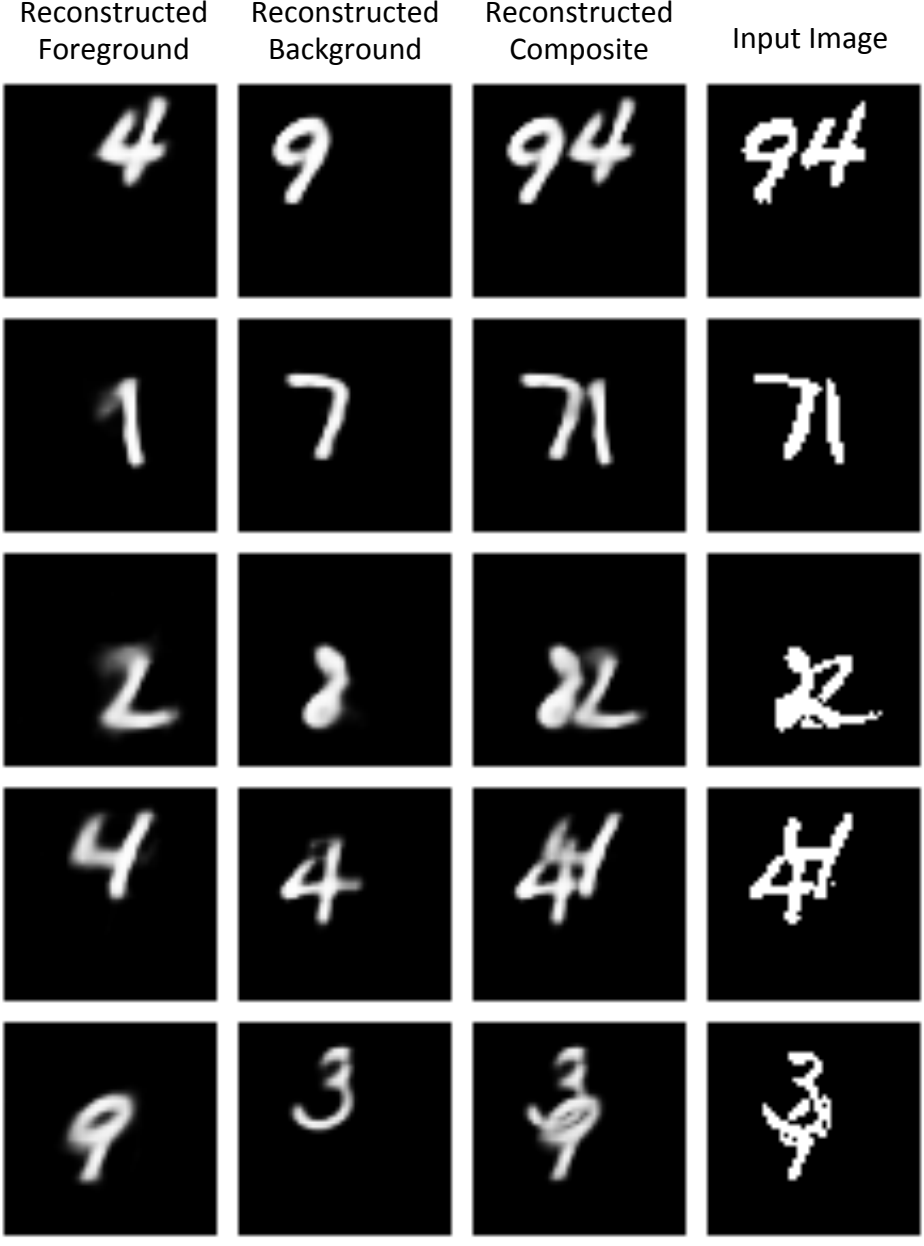
t=20

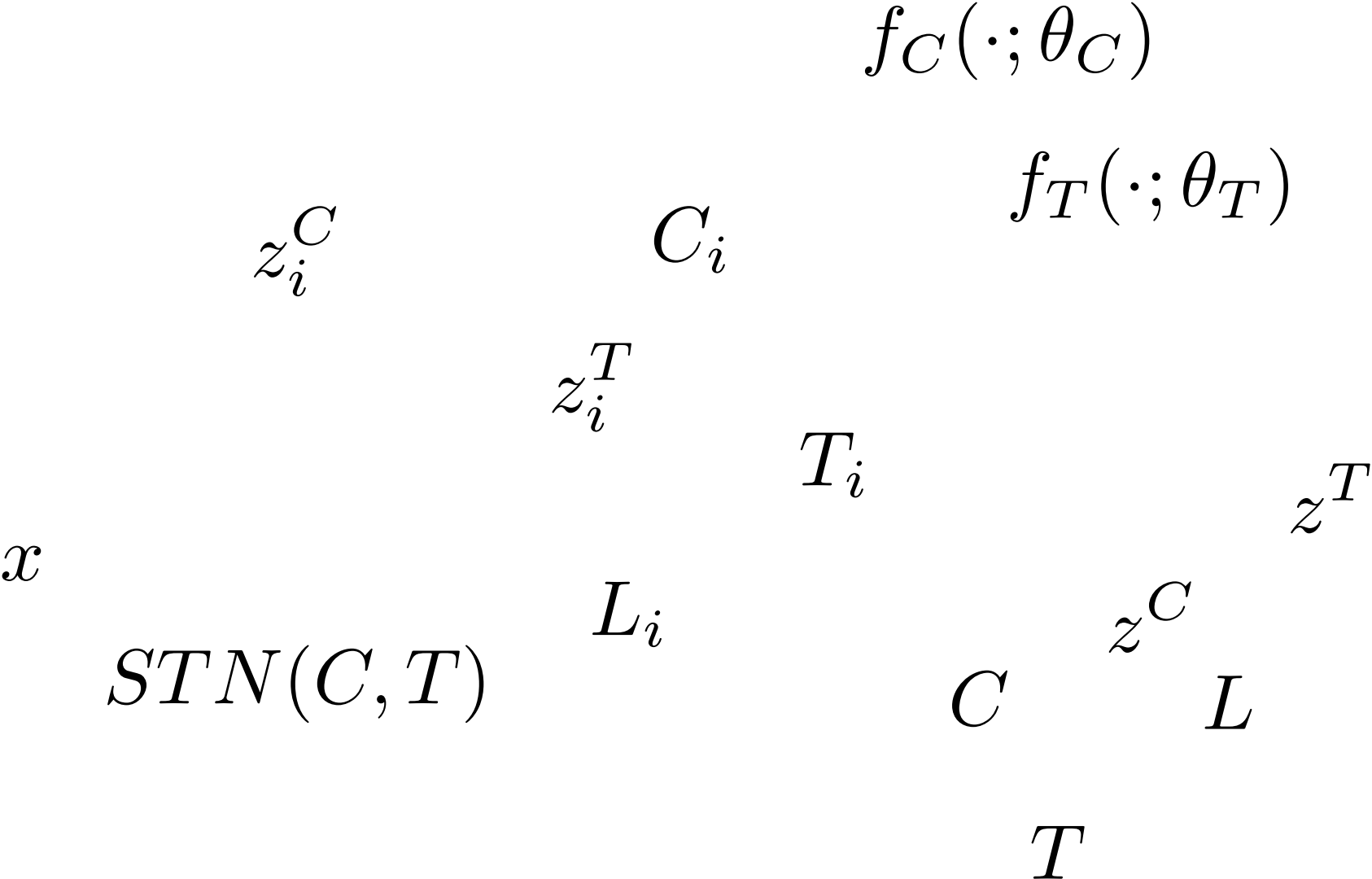


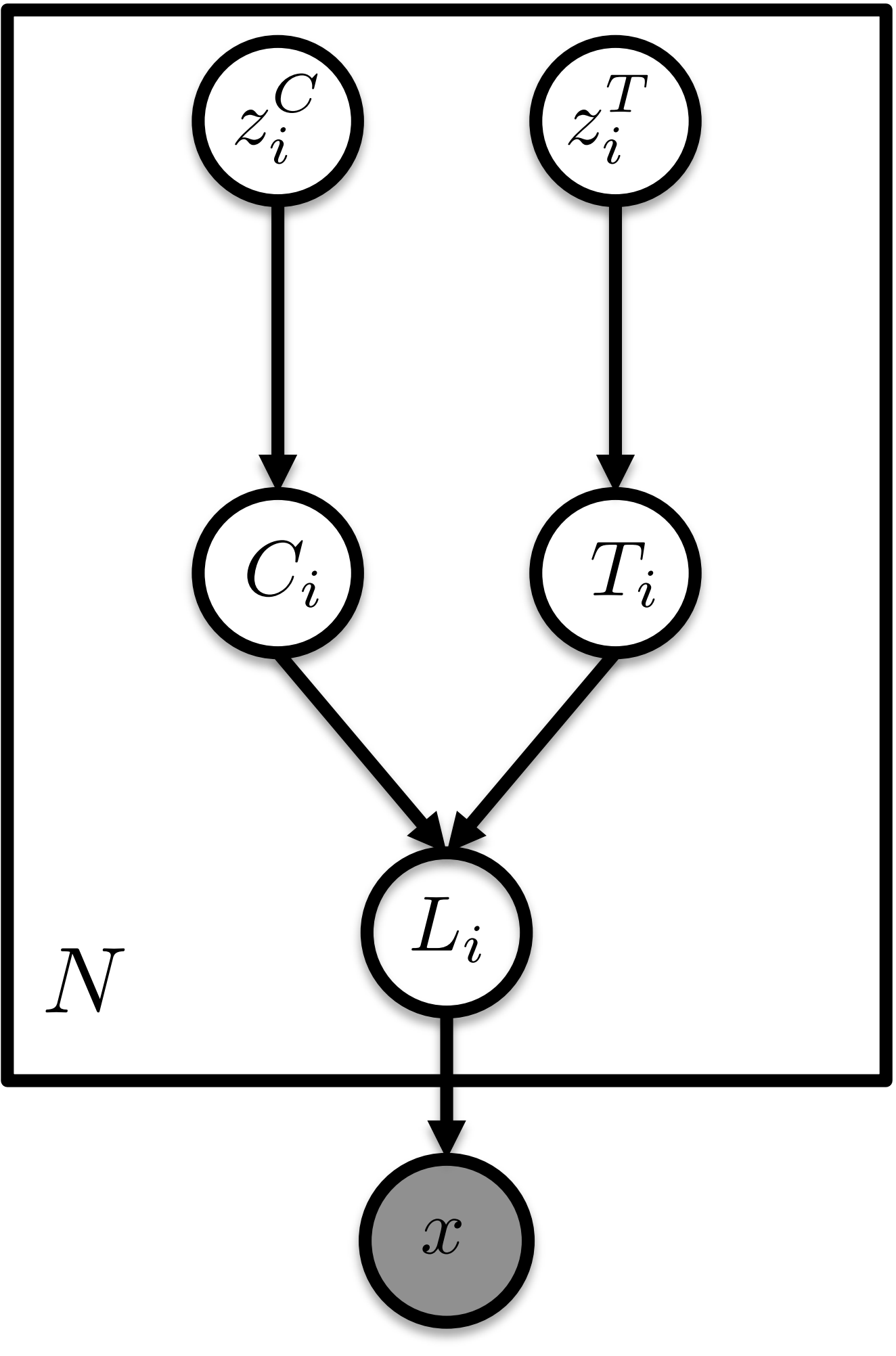
t=200

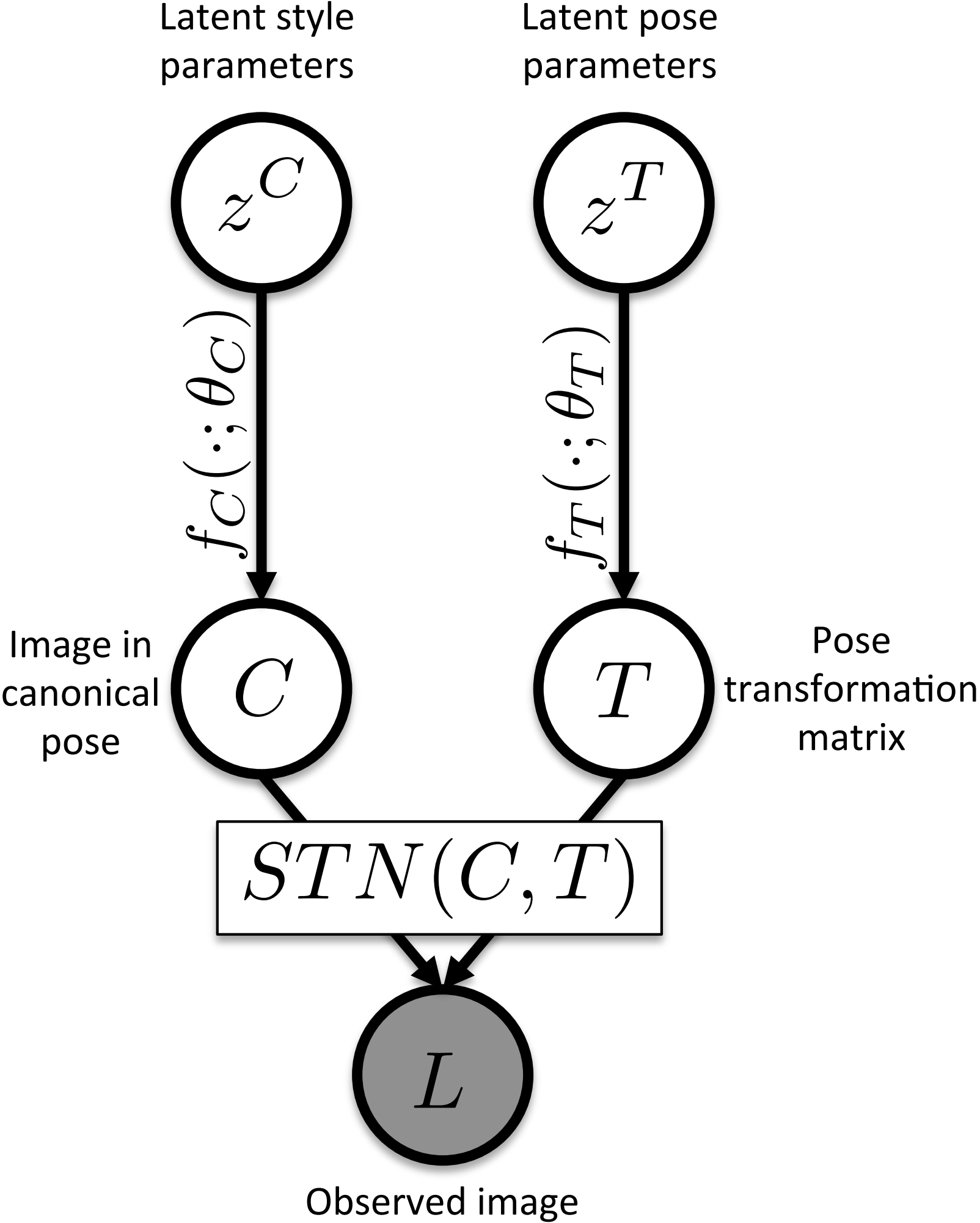

-80
-100
-120
-140
-160
-180
-200
$\mathcal{L}$
ST-VAE train
ST-VAE test
VAE train
VAE test
0
minibatch updates
250000

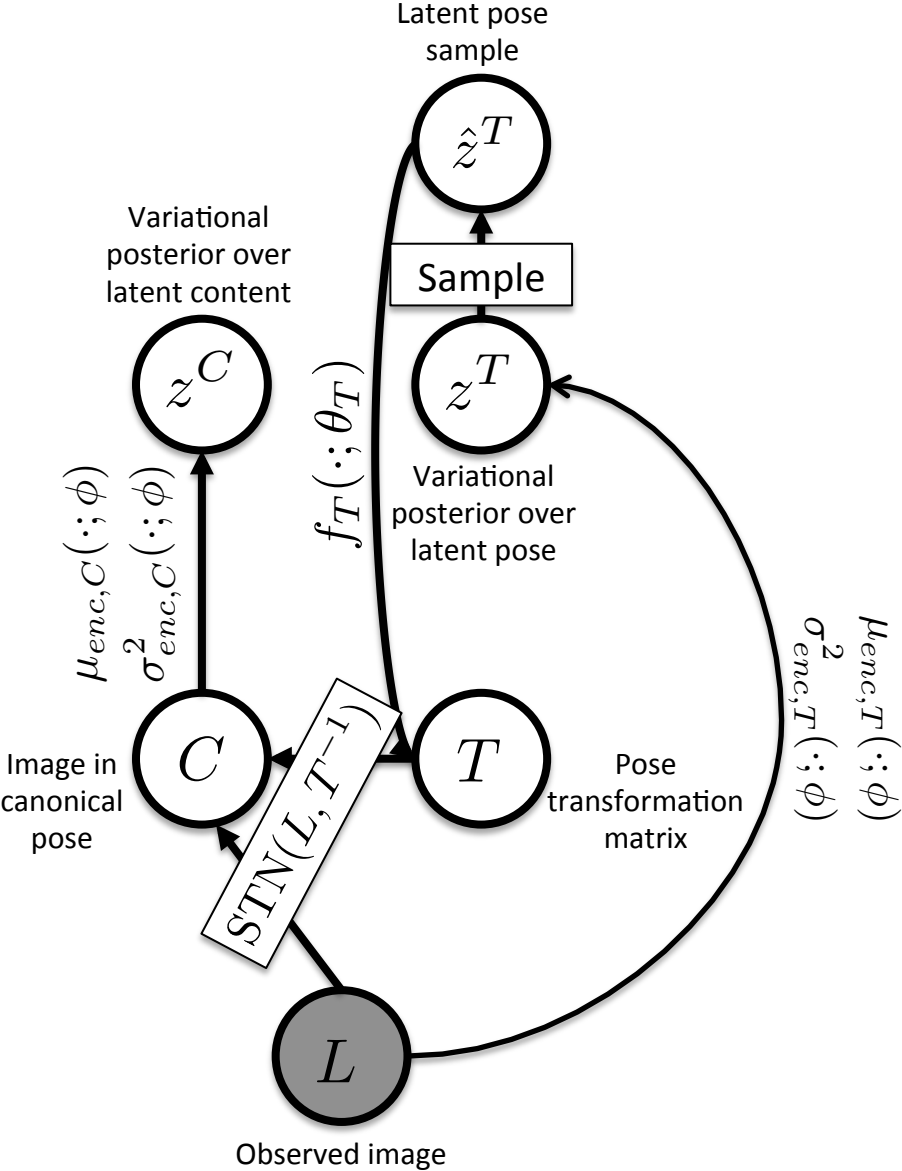



Reconstructed Foreground
Reconstructed Background
Reconstructed Composite
Input Image


$f_C(\cdot;\theta_C)$
$f_T(\cdot;\theta_T)$
z_i^C
C_i
z_i^T
T_i
z^T
x
L_i
z^C
$STN(C,T)$
C
L
T


z_i^C
z_i^T
C_i
T_i
L_i
N
x


Latent style parameters
Latent pose parameters
z^C
z^T
$f_C(\cdot;\theta_C)$
$f_T(\cdot;\theta_T)$
Image in canonical pose
C
T
Pose transformation matrix
$STN(C,T)$
L
Observed image


Latent pose sample
$\hat{z}^T$
Variational posterior over latent content
Sample
z^C
z^T
$f_T(\cdot;\theta_T)$
Variational posterior over latent pose
$\mu_{enc,C}(\cdot;\phi)$
$\sigma^2_{enc,C}(\cdot;\phi)$
$\mu_{enc,T}(\cdot;\phi)$
$\sigma^2_{enc,T}(\cdot;\phi)$
Image in canonical pose
C
T
Pose transformation matrix
$STN(L, T^{-1})$
L
Observed image