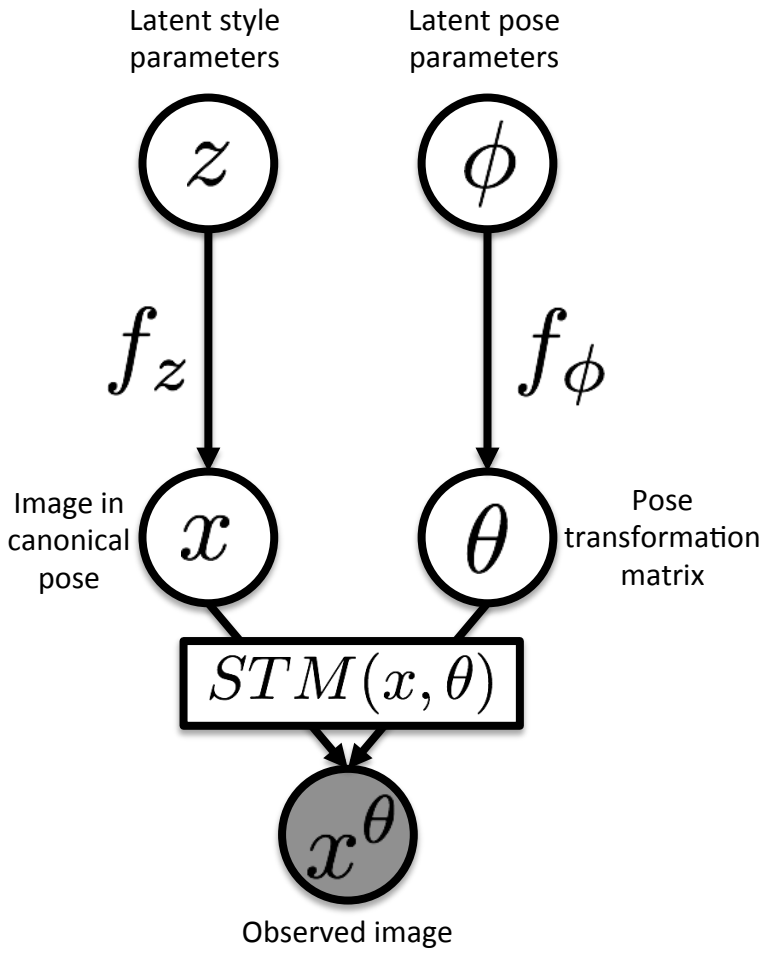
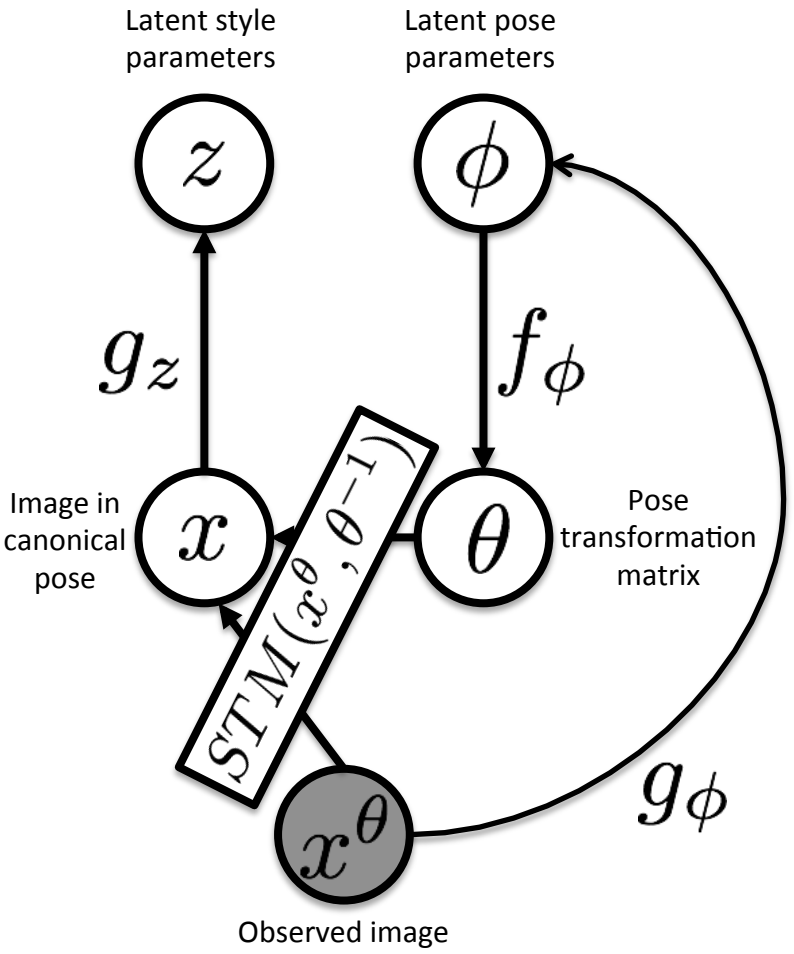
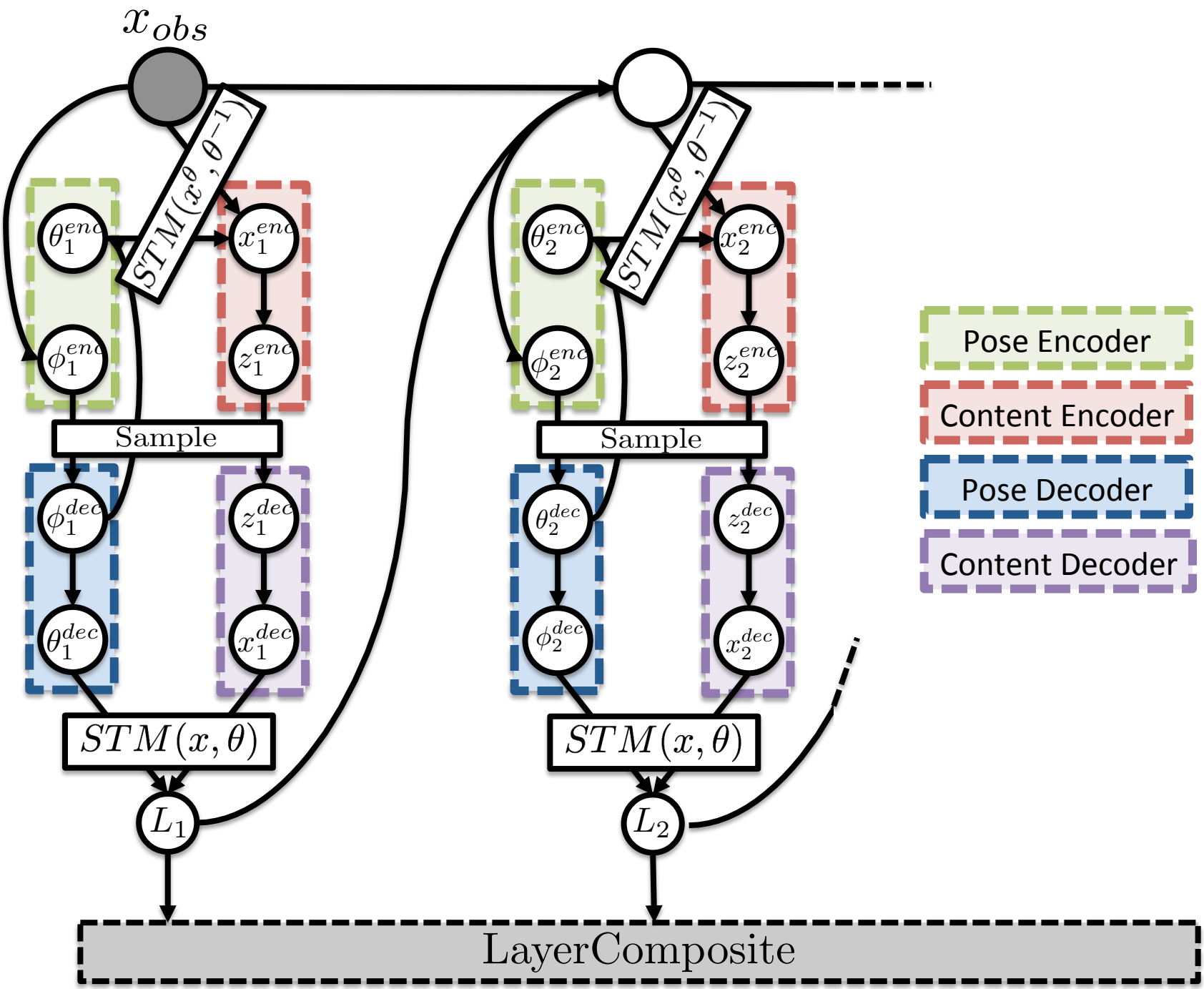
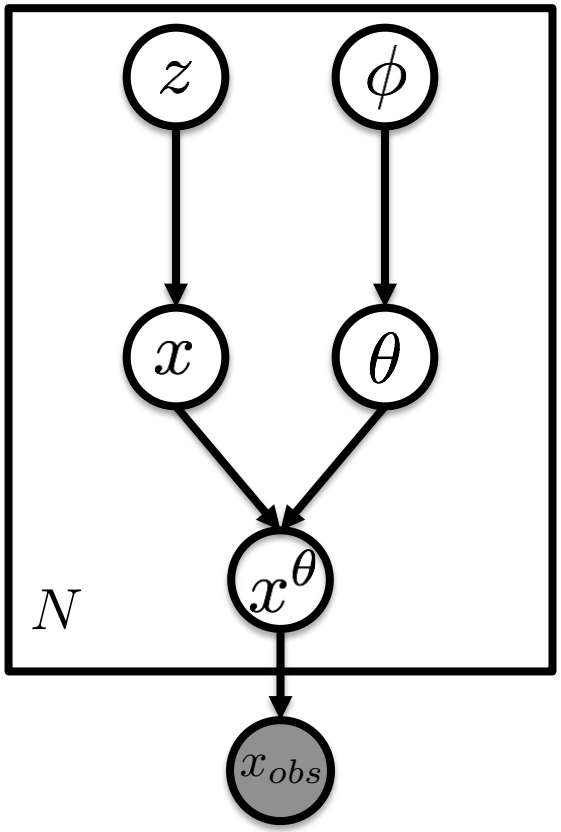
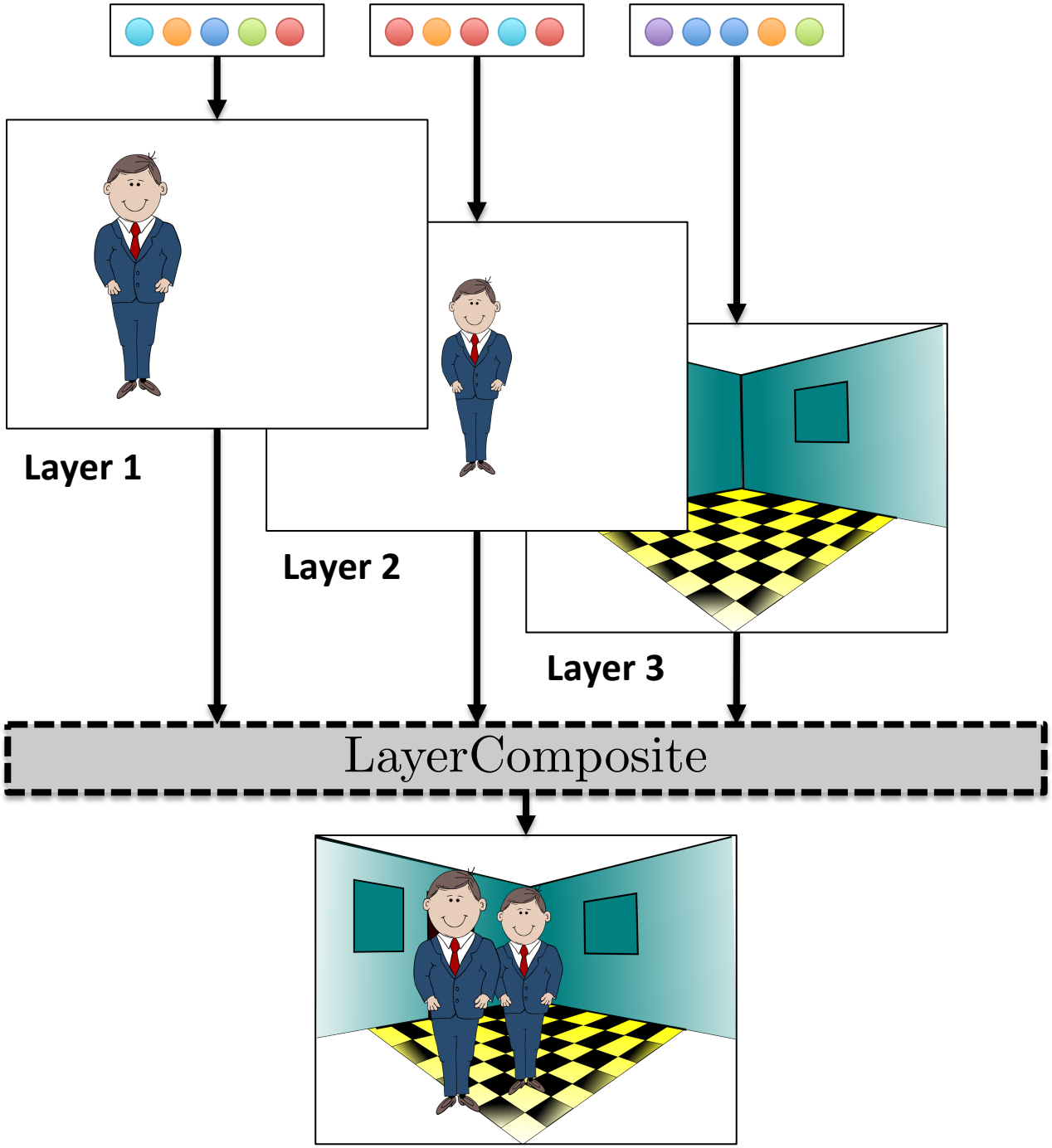
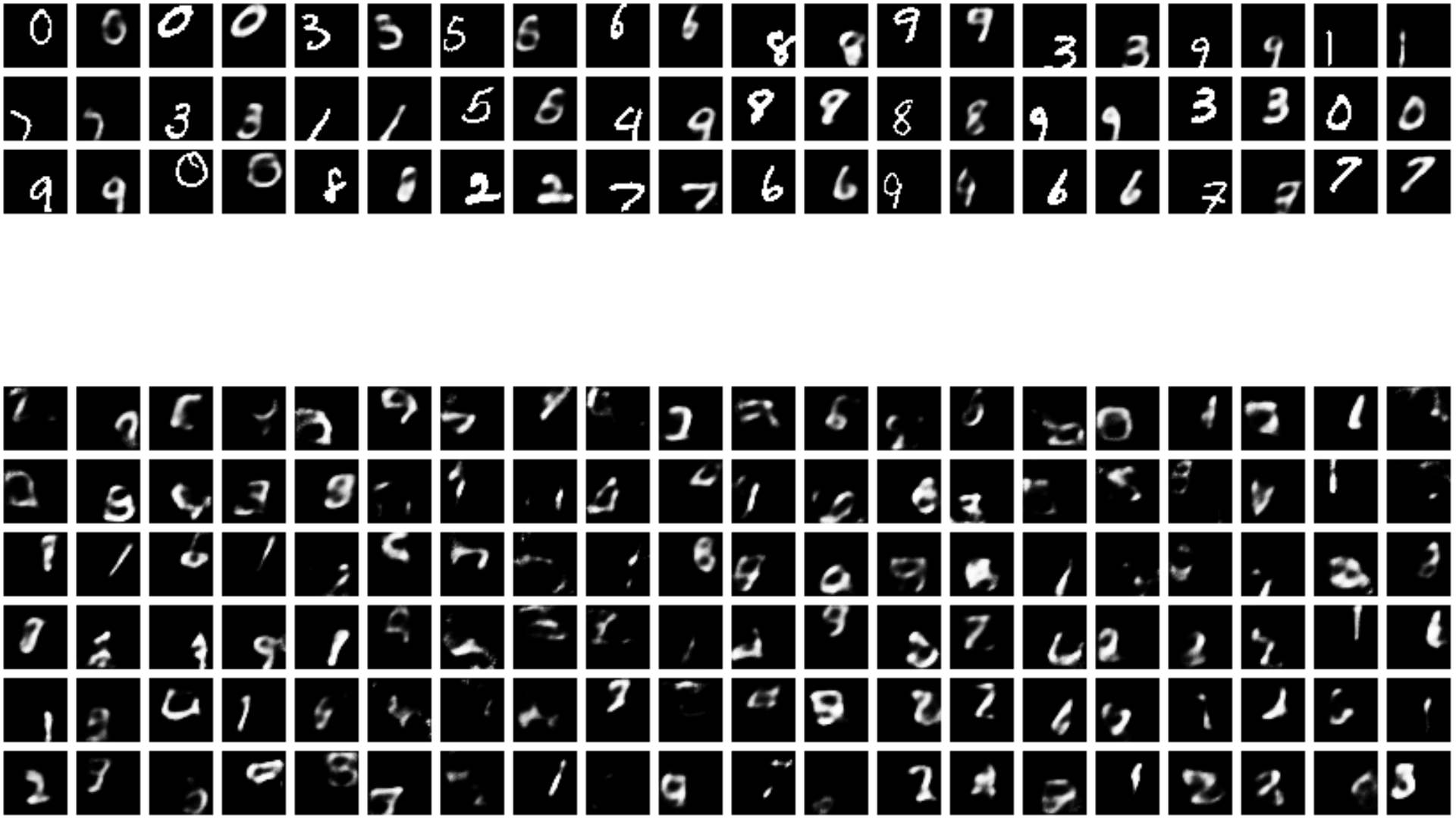

Latent style parameters
Latent pose parameters
z
ϕ
f_z
f_ϕ
Image in canonical pose
x
θ
Pose transformation matrix
$STM(x, \theta)$
x^θ
Observed image

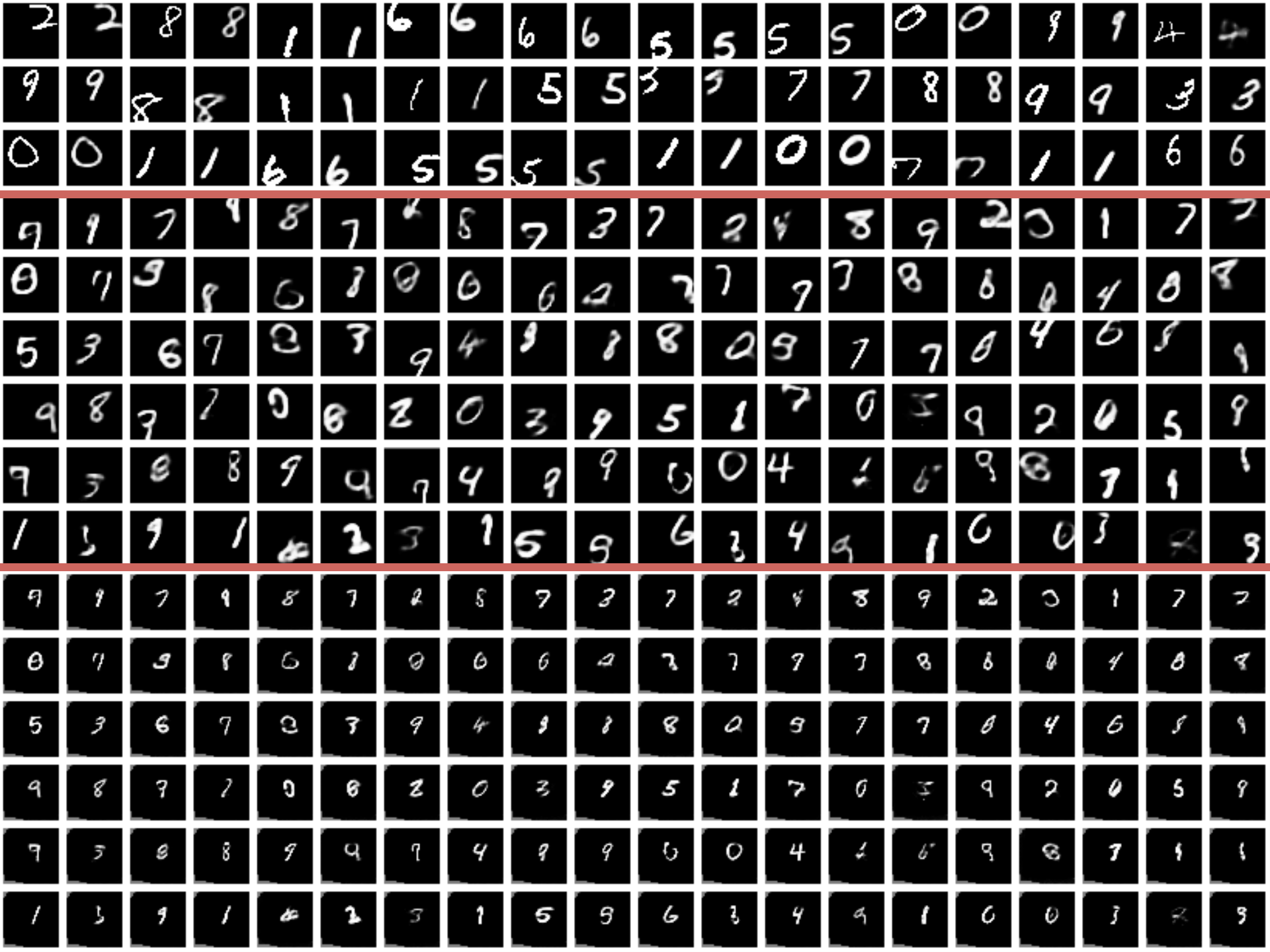

Latent style parameters
Latent pose parameters
z
ϕ
g_z
f_ϕ
Image in canonical pose
x
$STM(x^\theta, \theta^{-1})$
θ
Pose transformation matrix
g_ϕ
x^θ
Observed image


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

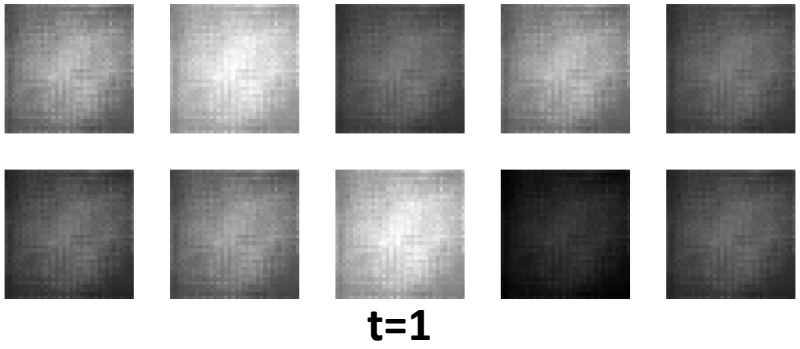
ST-VAE samples

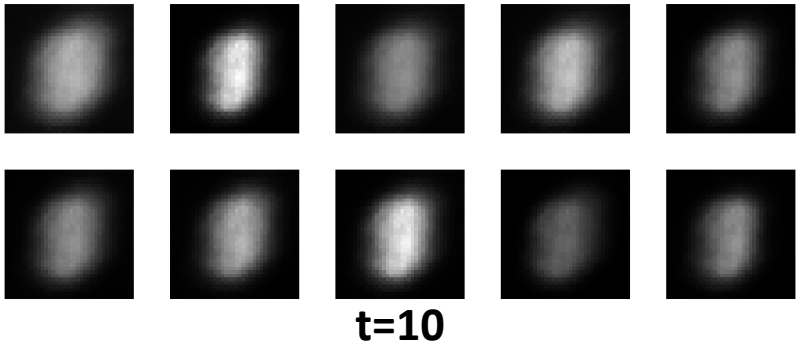
ST-VAE samples in canonical pose

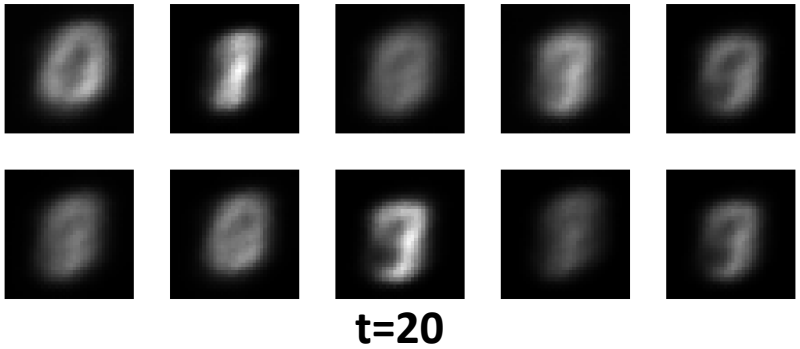
VAE samples

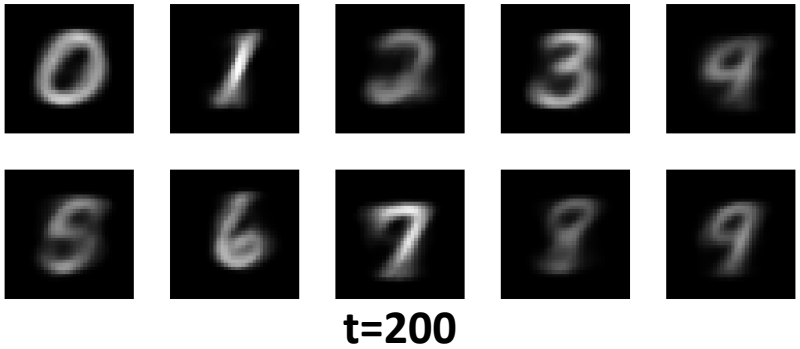
ST-VAE samples

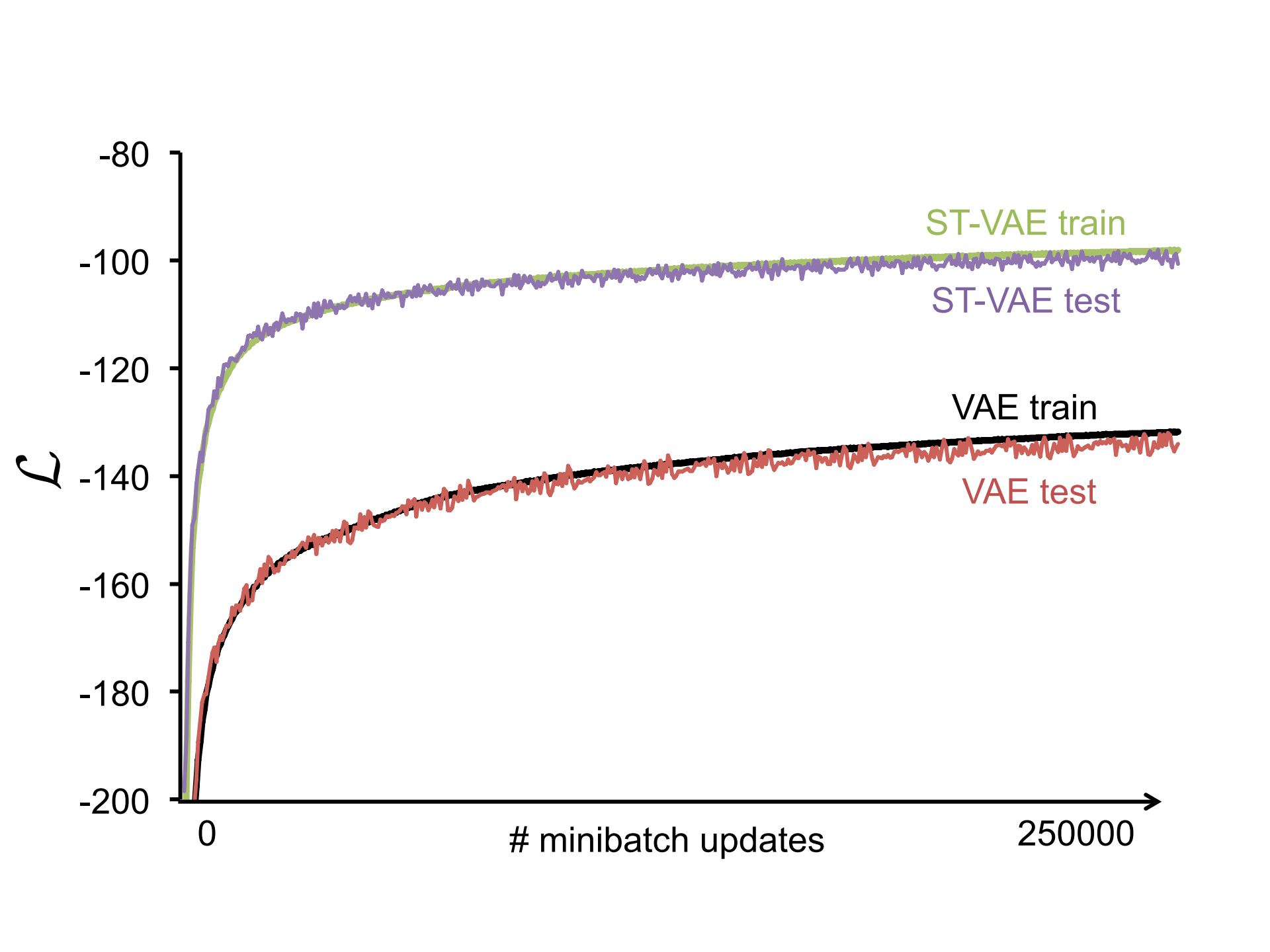
ST-VAE samples in canonical pose

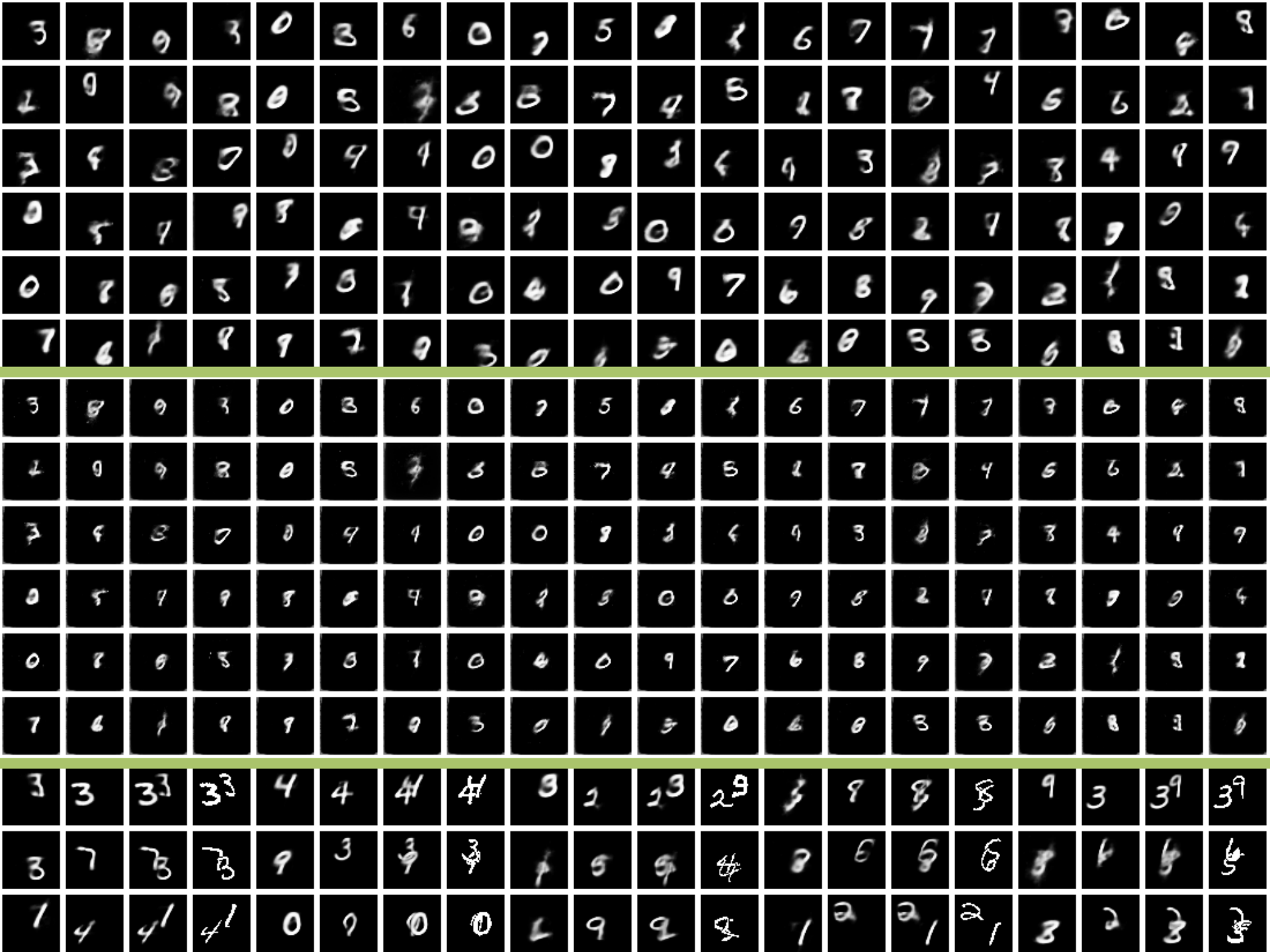

t=1

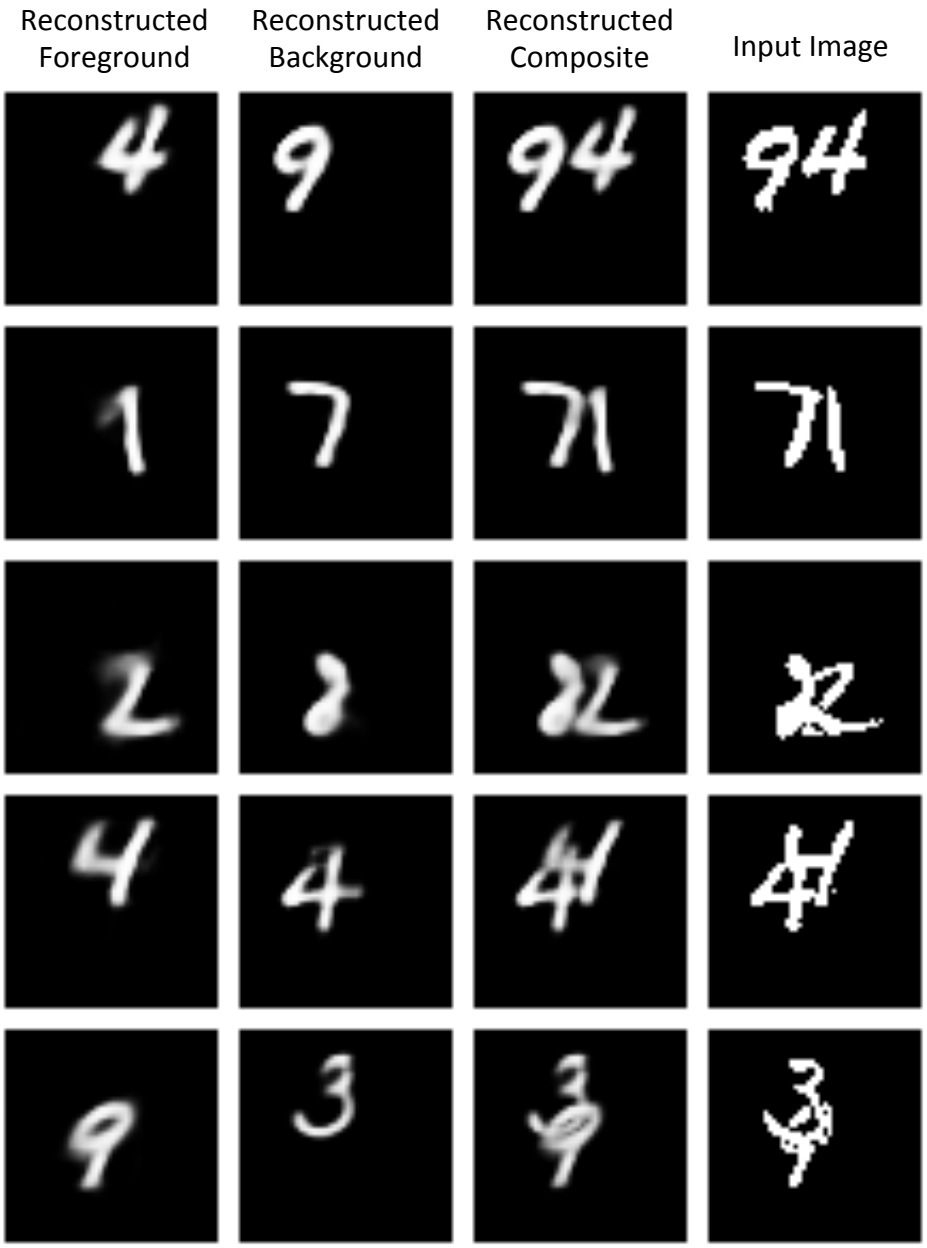

t=10


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