

Latent style  
parameters

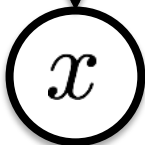
Latent pose  
parameters



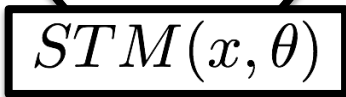
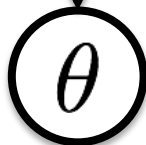
$f_z$

$f_\phi$

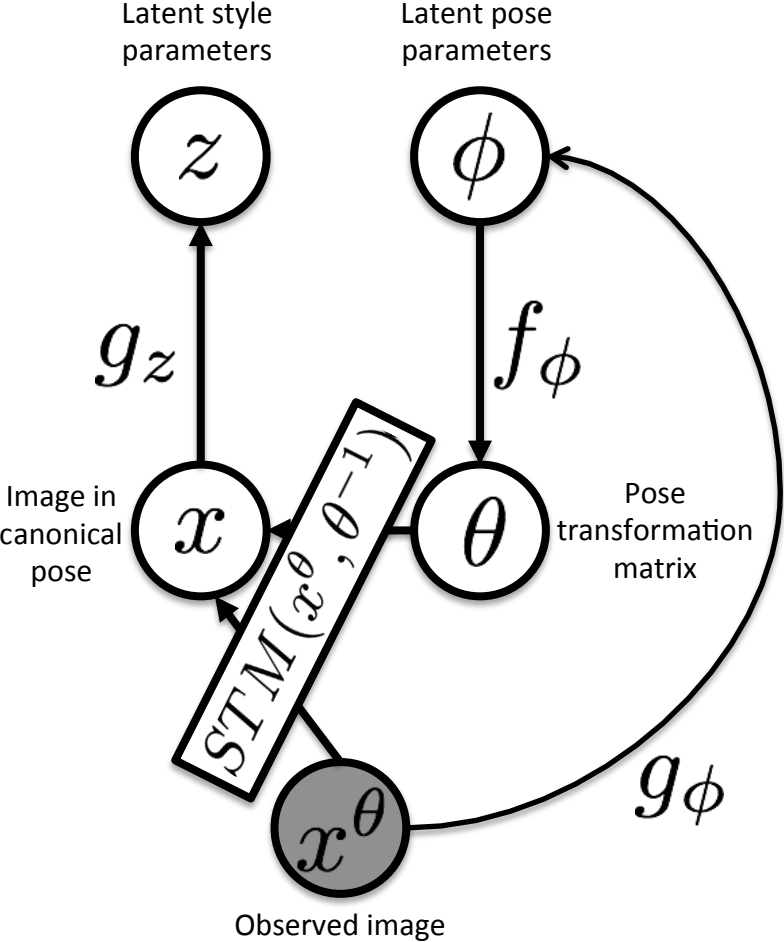
Image in  
canonical  
pose

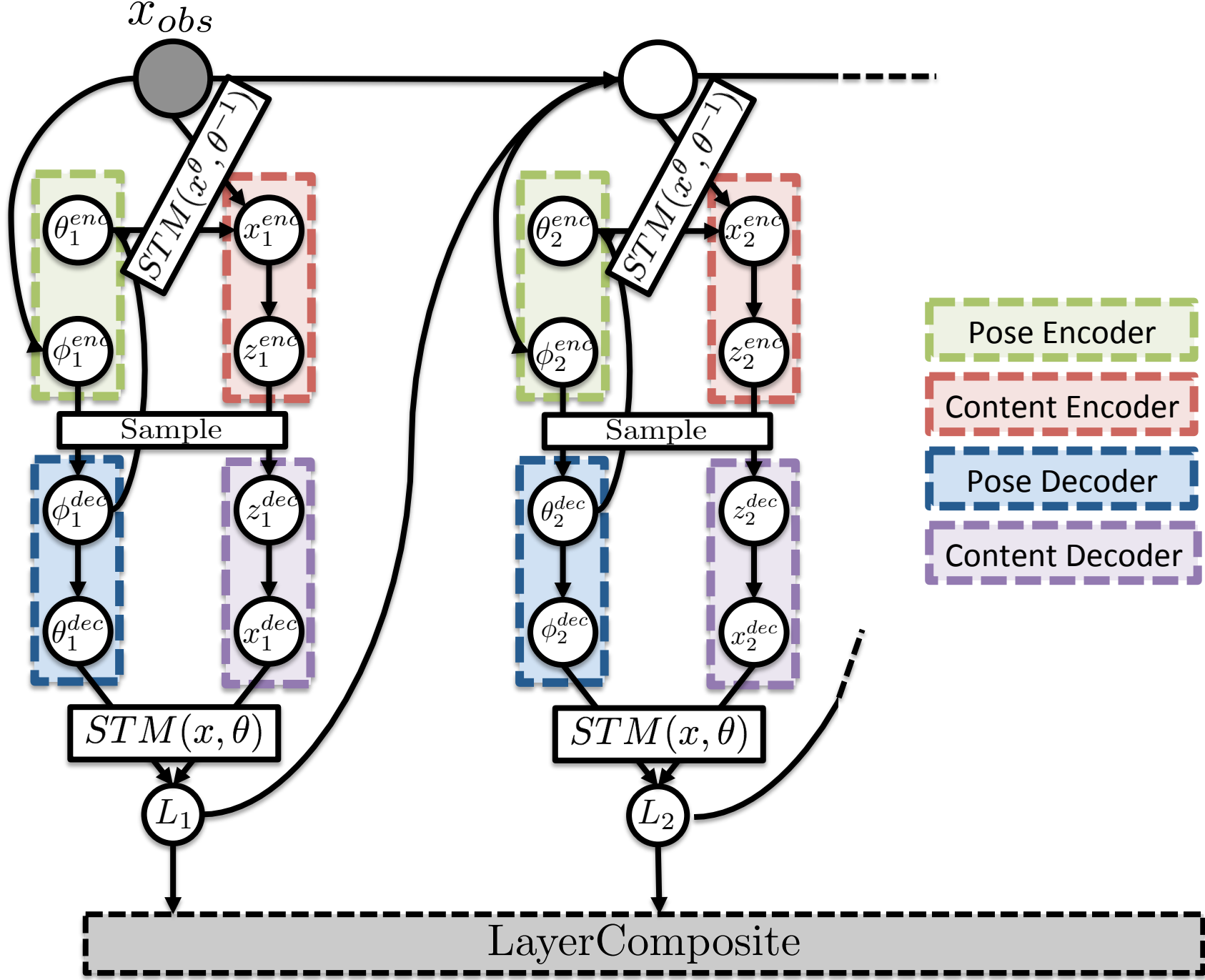


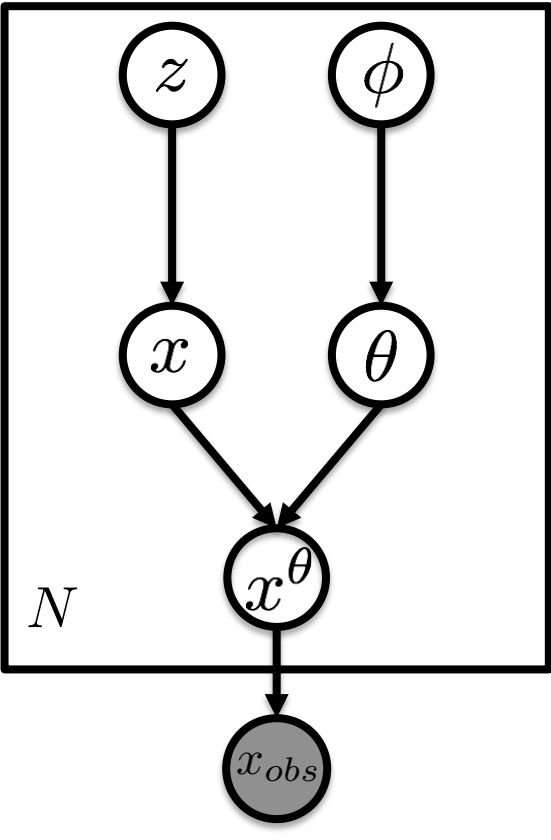
Pose  
transformation  
matrix

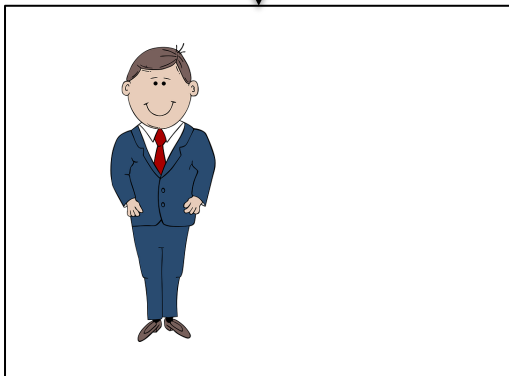
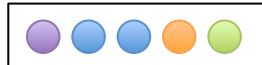
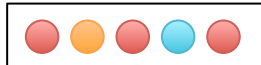
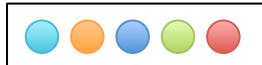


Observed image

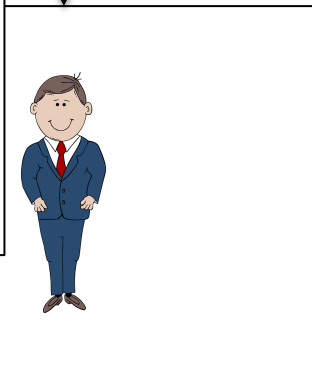




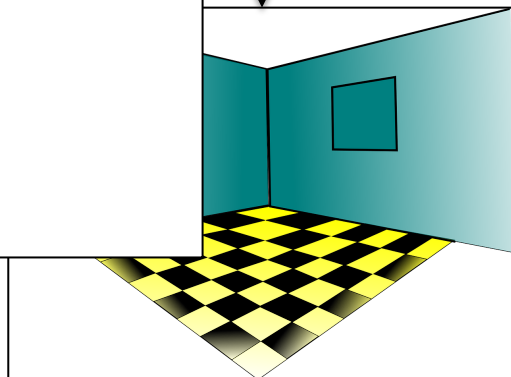




Layer 1



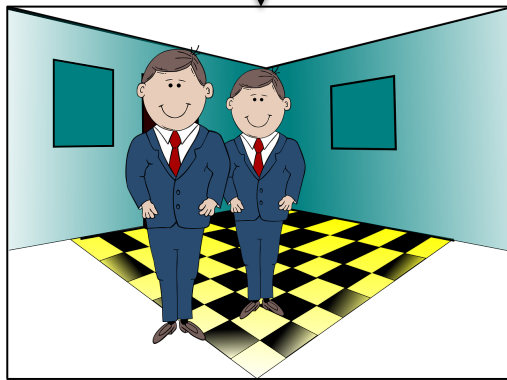
Layer 2



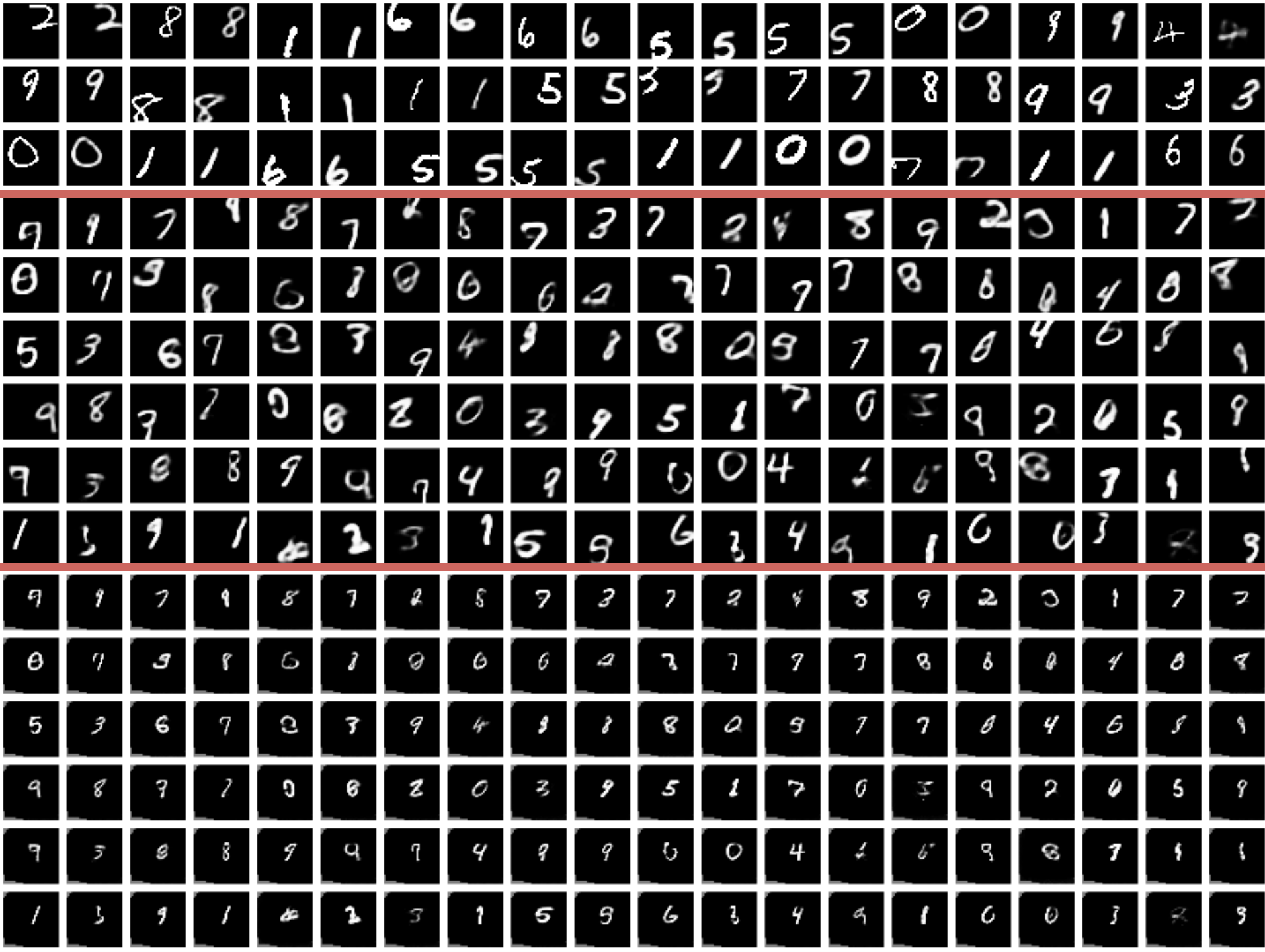
Layer 3



LayerComposite







**VAE  
samples**



**ST-VAE  
samples**



**ST-VAE samples  
in canonical pose**





**VAE  
samples**

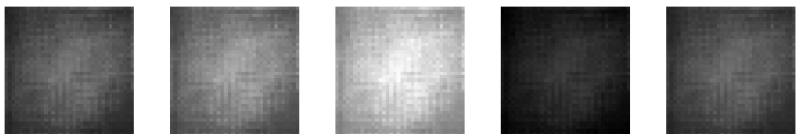
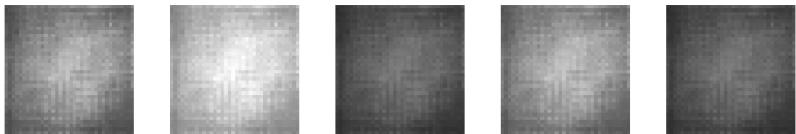


**ST-VAE  
samples**

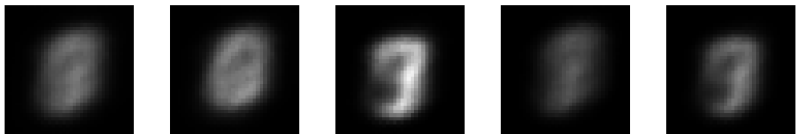
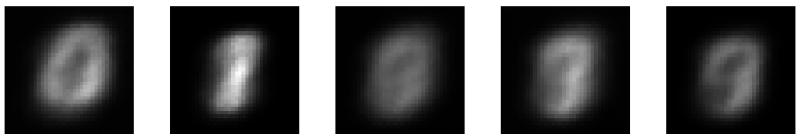


**ST-VAE samples  
in canonical pose**

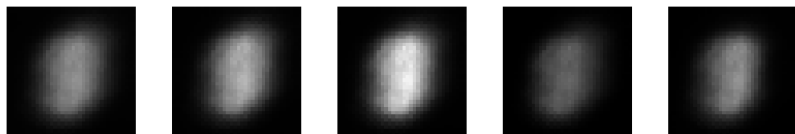
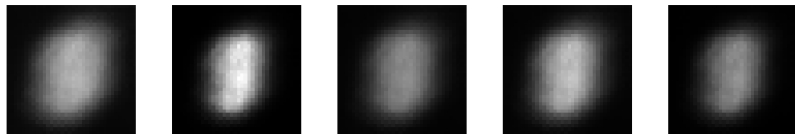




**t=1**



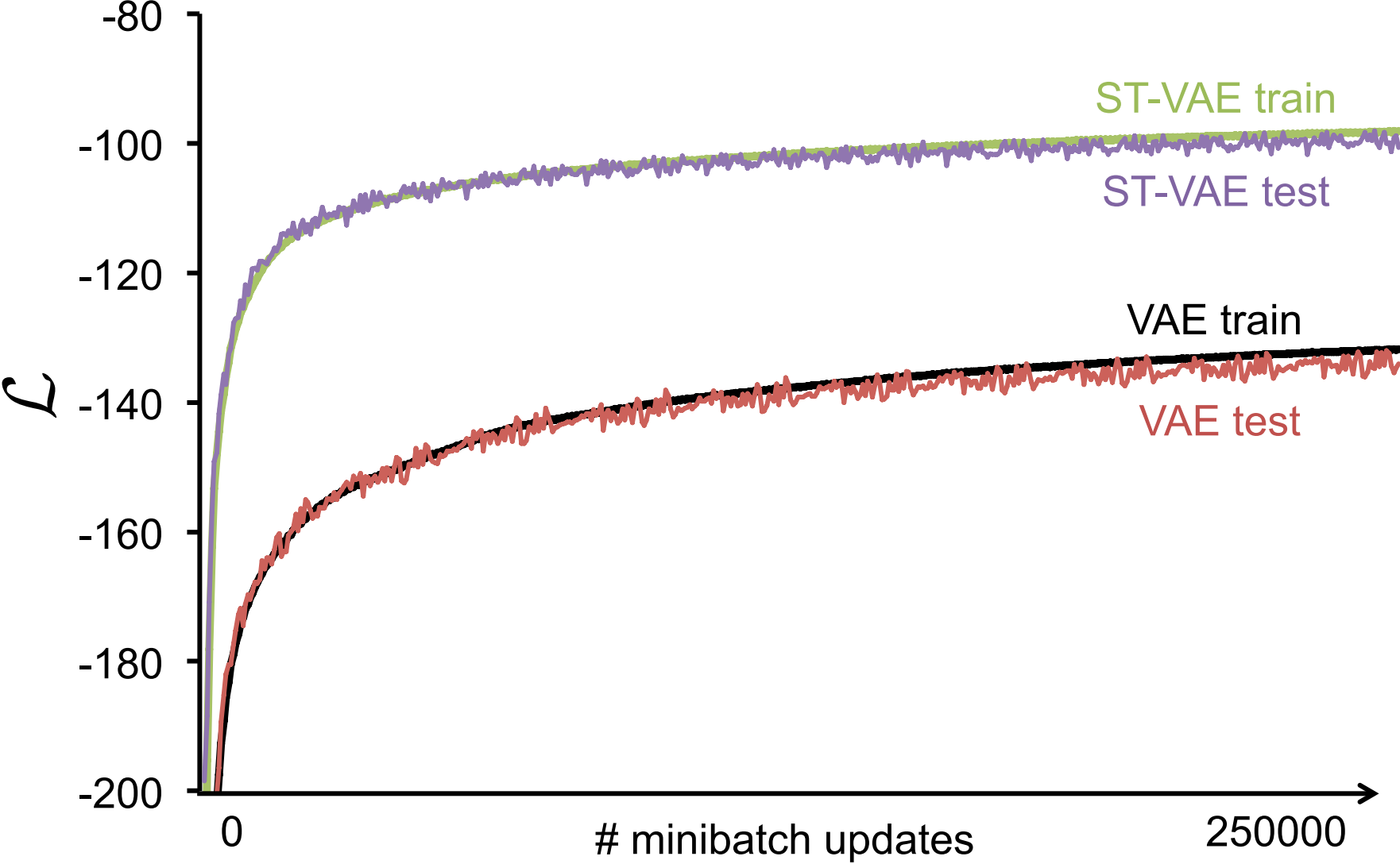
**t=20**



**t=10**



**t=200**



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

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

|   |   |    |    |   |   |    |    |   |   |    |    |   |   |    |    |   |   |    |    |
|---|---|----|----|---|---|----|----|---|---|----|----|---|---|----|----|---|---|----|----|
| 3 | 3 | 33 | 33 | 4 | 4 | 41 | 41 | 8 | 1 | 23 | 23 | 3 | 9 | 8  | 8  | 9 | 3 | 39 | 39 |
| 3 | 7 | 73 | 73 | 9 | 3 | 37 | 37 | 1 | 5 | 57 | 4  | 8 | 6 | 6  | 6  | 8 | 1 | 6  | 6  |
| 7 | 4 | 41 | 41 | 0 | 9 | 00 | 00 | 2 | 9 | 92 | 8  | 1 | 2 | 21 | 21 | 3 | 2 | 3  | 37 |

Reconstructed  
ForegroundReconstructed  
BackgroundReconstructed  
Composite

Input Image

4

9

94

94

1

7

71

71

2

2

22

22

4

4

41

41

9

3

3

3