

A Architectures

We used a similar architecture for all datasets, with 5 convolution layers followed by a dense layer projecting to a mean embedding.

Our architecture resembles that of Ghosh et al. (2019) but with an additional layer, ELU instead of ReLU nonlinearities, and larger latent dimensions. We list Conv (convolutional) and ConvT (transposed convolution) layers with their number of filters, kernel size, and stride.

MNIST	CIFAR-10	CelebA
Encoder		
$x \in \mathcal{R}^{28 \times 28}$	$x \in \mathcal{R}^{32 \times 32}$	$x \in \mathcal{R}^{64 \times 64}$
$\rightarrow \text{Conv}_{64,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{128,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{128,5,1} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{Conv}_{128,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{256,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{256,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{Conv}_{256,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{512,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{512,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{Conv}_{512,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{1024,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{1024,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{Conv}_{512,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{1024,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{Conv}_{1024,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{Flatten} \rightarrow \text{FC}_{32}$	$\rightarrow \text{Flatten} \rightarrow \text{FC}_{128}$	$\rightarrow \text{Flatten} \rightarrow \text{FC}_{128}$
Decoder		
$z \in \mathcal{R}^{32} \rightarrow \text{FC}_{7 \times 7 \times 256}$	$z \in \mathcal{R}^{128} \rightarrow \text{FC}_{8 \times 8 \times 512}$	$z \in \mathcal{R}^{128} \rightarrow \text{FC}_{16 \times 16 \times 512}$
$\rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{ConvT}_{512,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{1024,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{1024,5,1} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{ConvT}_{256,4,1} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{512,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{512,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{ConvT}_{128,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{256,4,2} \rightarrow \text{BN} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{256,5,2} \rightarrow \text{BN} \rightarrow \text{ELU}$
$\rightarrow \text{ConvT}_{64,4,2} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{128,4,2} \rightarrow \text{ELU}$	$\rightarrow \text{ConvT}_{128,5,2} \rightarrow \text{ELU}$
$\rightarrow \text{Conv}_{1,4,1} \rightarrow \text{Sigmoid}$	$\rightarrow \text{Conv}_{3,1,1} \rightarrow \text{Sigmoid}$	$\rightarrow \text{Conv}_{3,5,1} \rightarrow \text{Sigmoid}$

B Additional Samples

We visualize additional reconstructed test examples and samples from a post-fit GMM model with 10 mixture components on the latents.



Figure 4: **CelebA** test reconstructions from **InjFlow model**: *Top*: original test image, *Bottom*: reconstructed image.



Figure 5: **CelebA** random samples from **InjFlow model** using the post-fit Gaussian mixture distribution on the latent space.



Figure 6: **CelebA** test reconstructions from **Autoencoder**: *Top*: original test image, *Bottom*: reconstructed image.



Figure 7: **CelebA** random samples from **Autoencoder** using the post-fit Gaussian mixture distribution on the latent space.



Figure 8: **CelebA** test reconstructions from **VAE**: *Top*: original test image, *Bottom*: reconstructed image.



Figure 9: **CelebA** random samples from **VAE** using the post-fit Gaussian mixture distribution on the latent space.

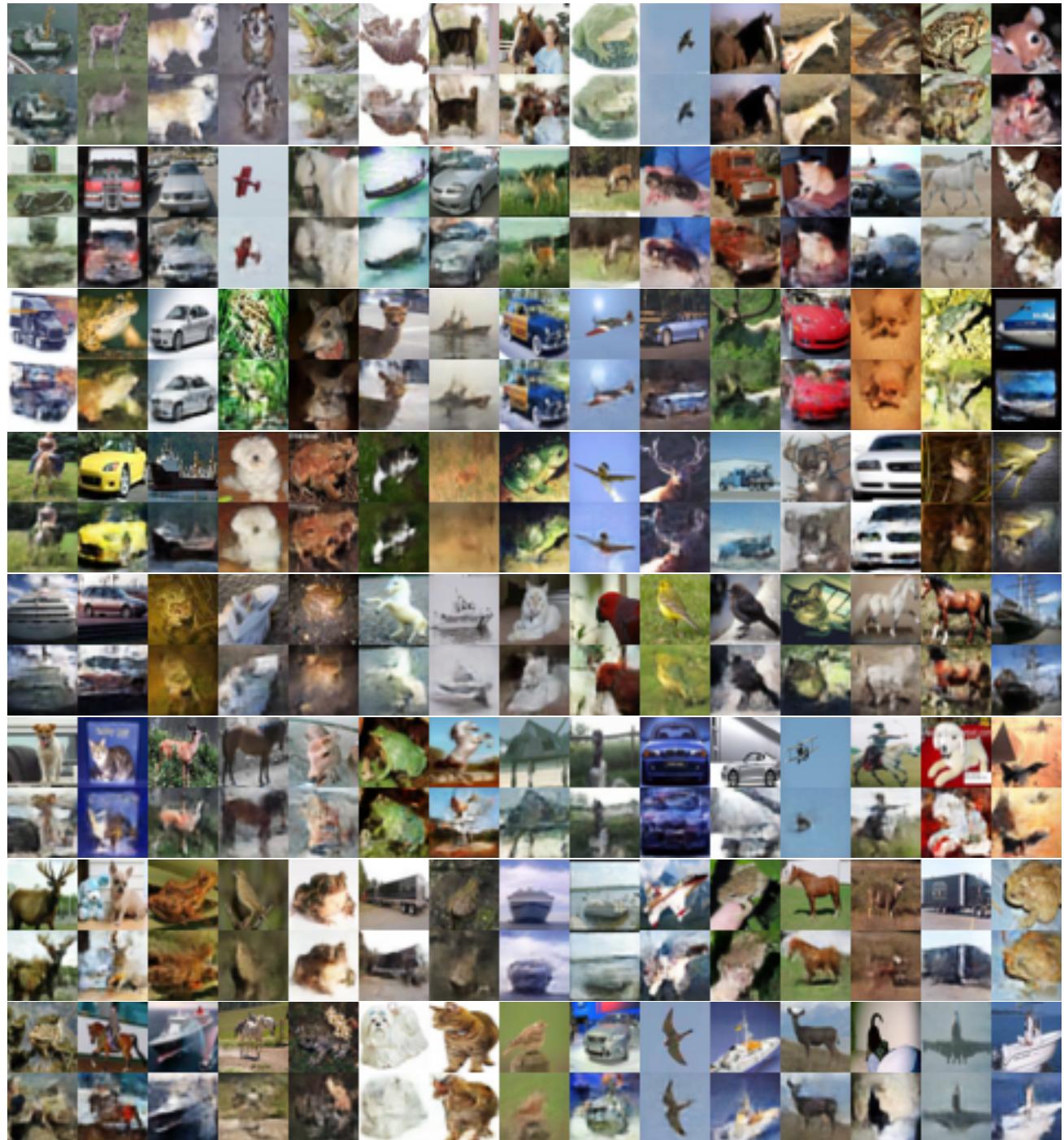


Figure 10: **CIFAR-10** test reconstructions from **InjFlow** model: *Top*: original test image, *Bottom*: reconstructed image.

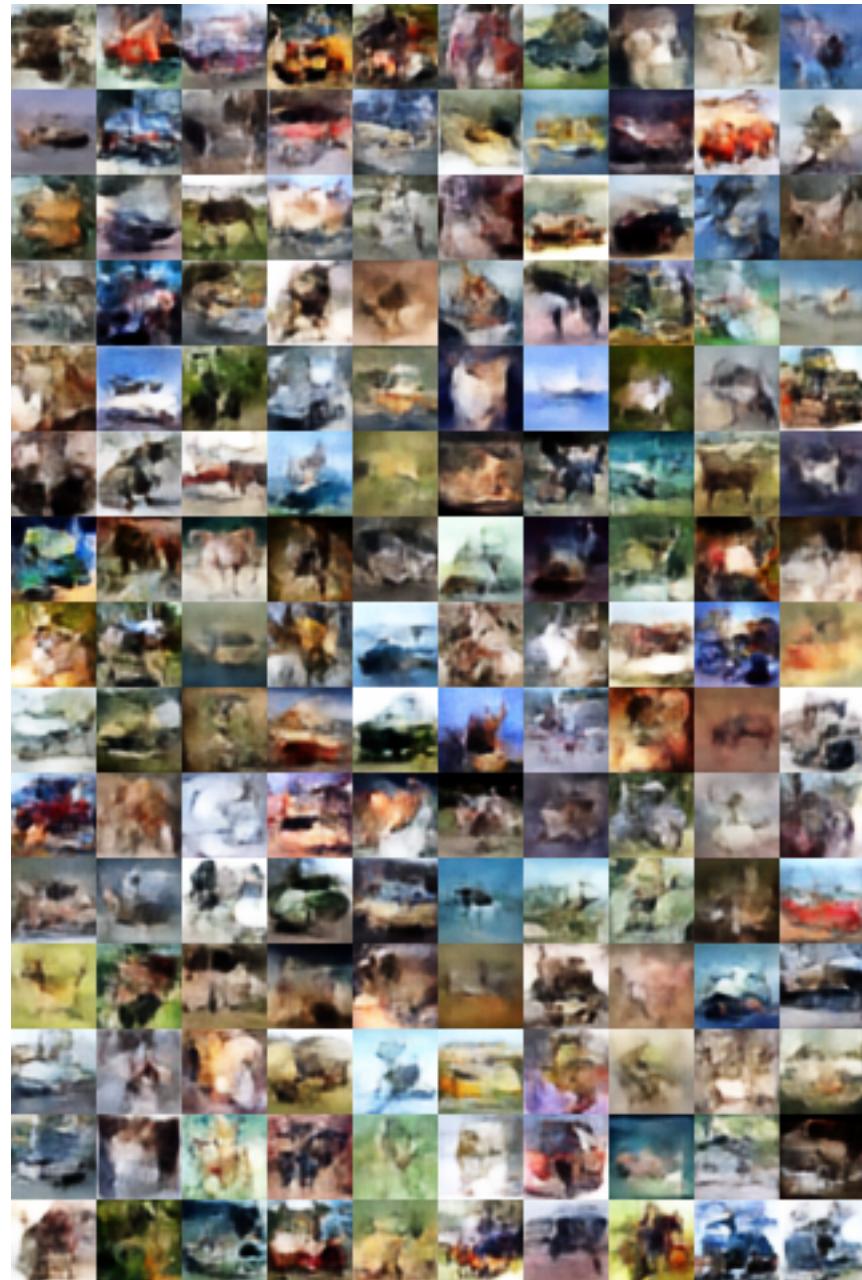


Figure 11: **CIFAR10** random samples from **InjFlow model** using the post-fit Gaussian mixture distribution on the latent space.

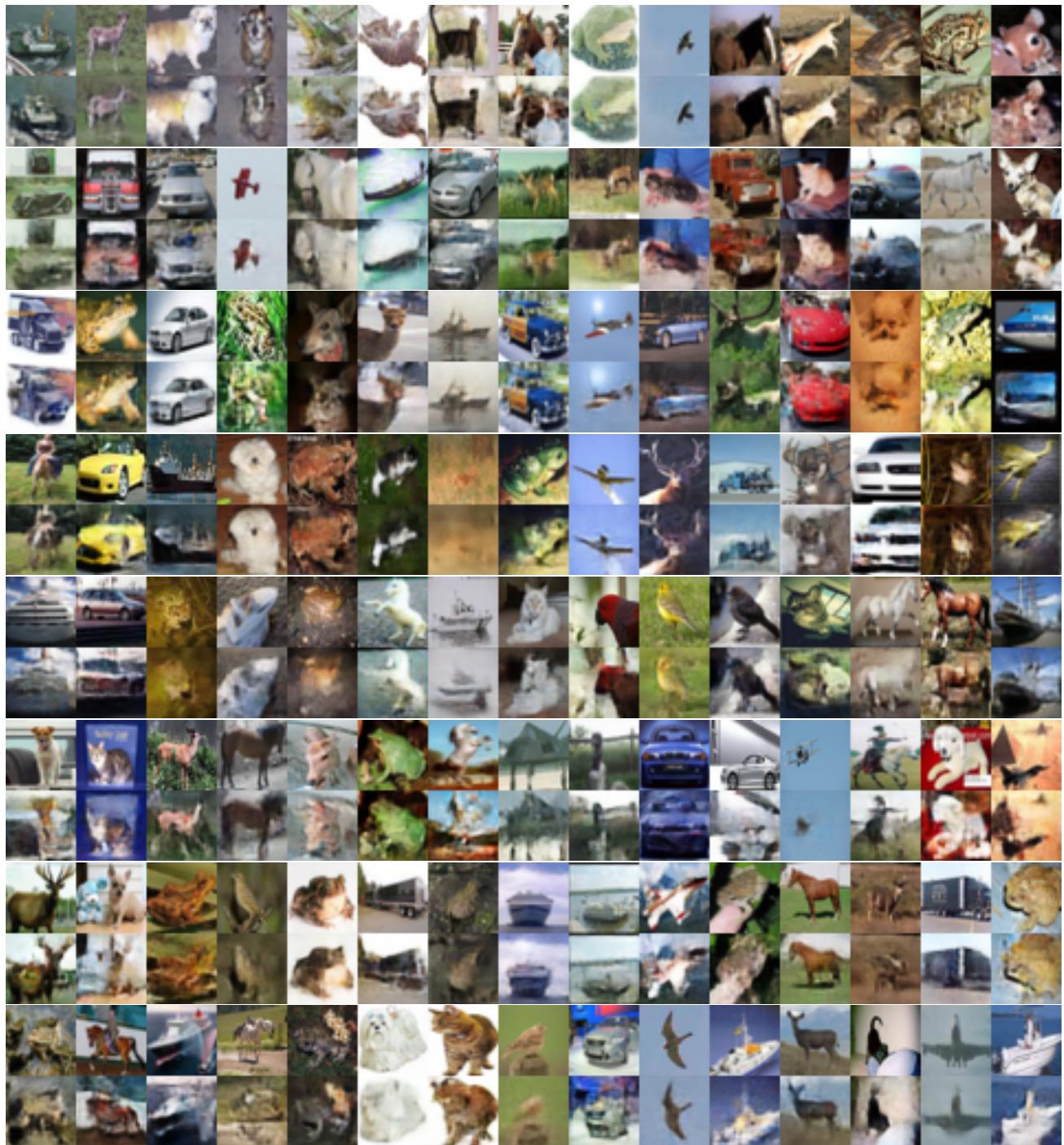


Figure 12: **CIFAR-10** test reconstructions from **Autoencoder**: *Top*: original test image, *Bottom*: reconstructed image.

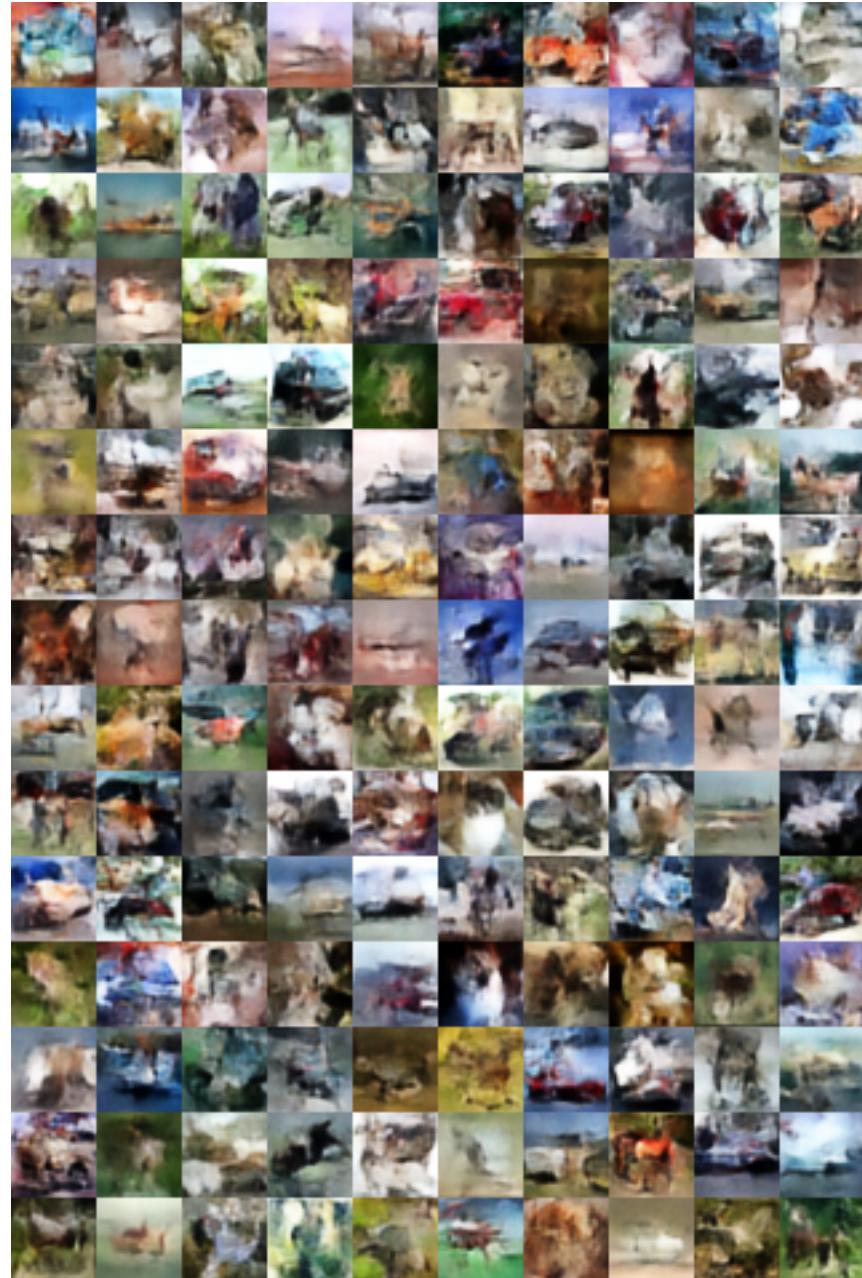


Figure 13: **CIFAR10** random samples from **Autoencoder** using the post-fit Gaussian mixture distribution on the latent space.



Figure 14: **CIFAR-10** test reconstructions from **VAE**: *Top*: original test image, *Bottom*: reconstructed image.

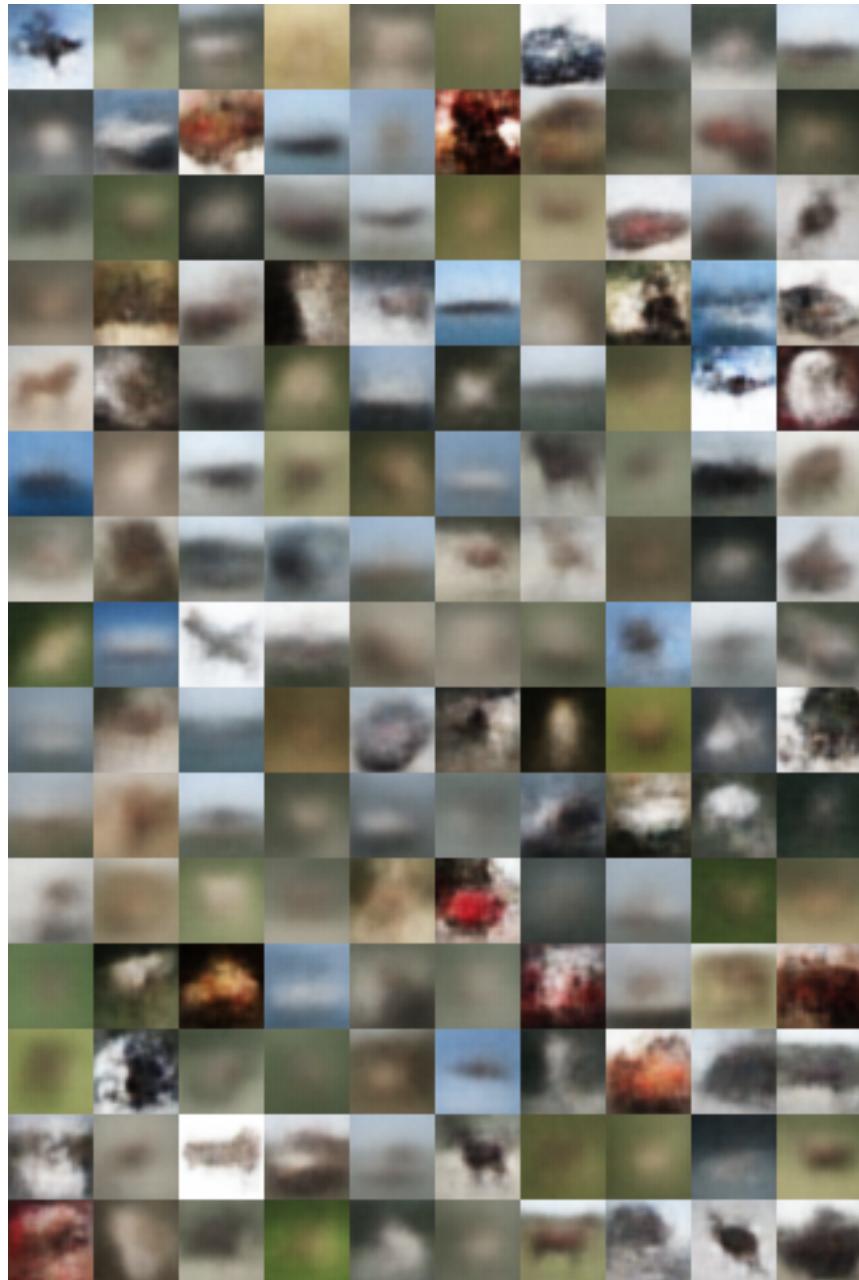


Figure 15: **CIFAR10** random samples from **VAE** using the post-fit Gaussian mixture distribution on the latent space.



Figure 16: MNIST test reconstructions from **InjFlow model**: *Top*: original test image, *Bottom*: reconstructed image.

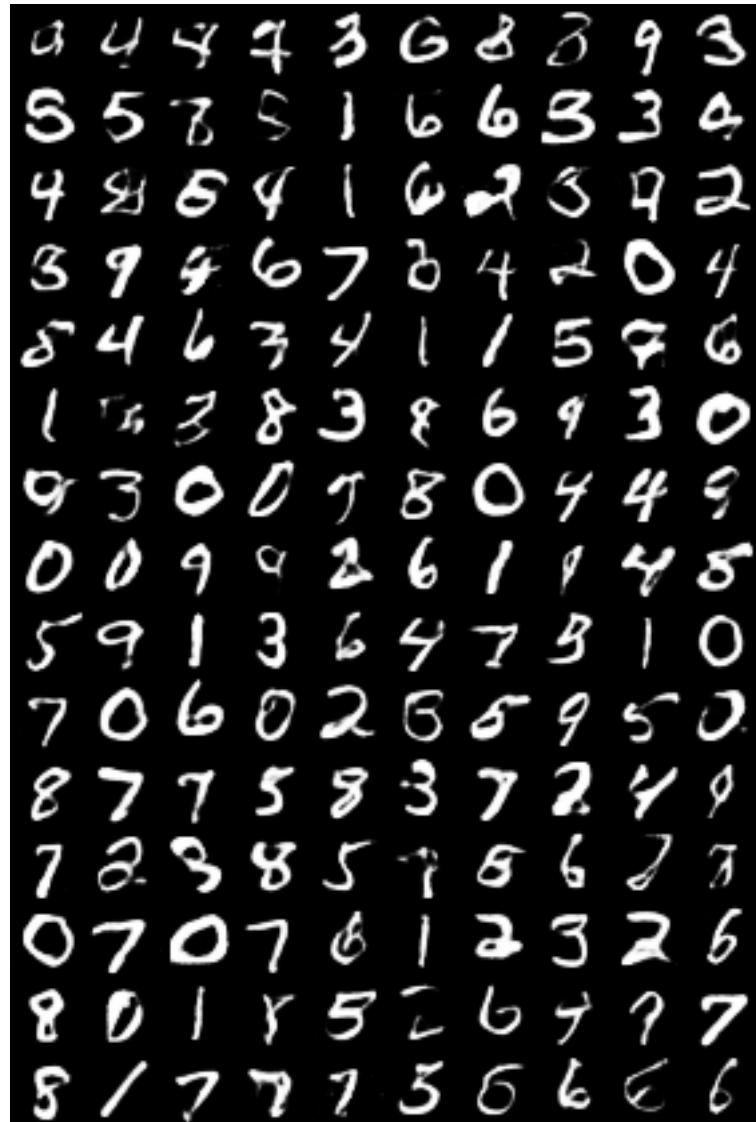


Figure 17: MNIST random samples from **InjFlow model** using the post-fit Gaussian mixture distribution on the latent space.



Figure 18: MNIST test reconstructions from **Autoencoder**: *Top*: original test image, *Bottom*: reconstructed image.



Figure 19: MNIST random samples from **Autoencoder** using the post-fit Gaussian mixture distribution on the latent space.

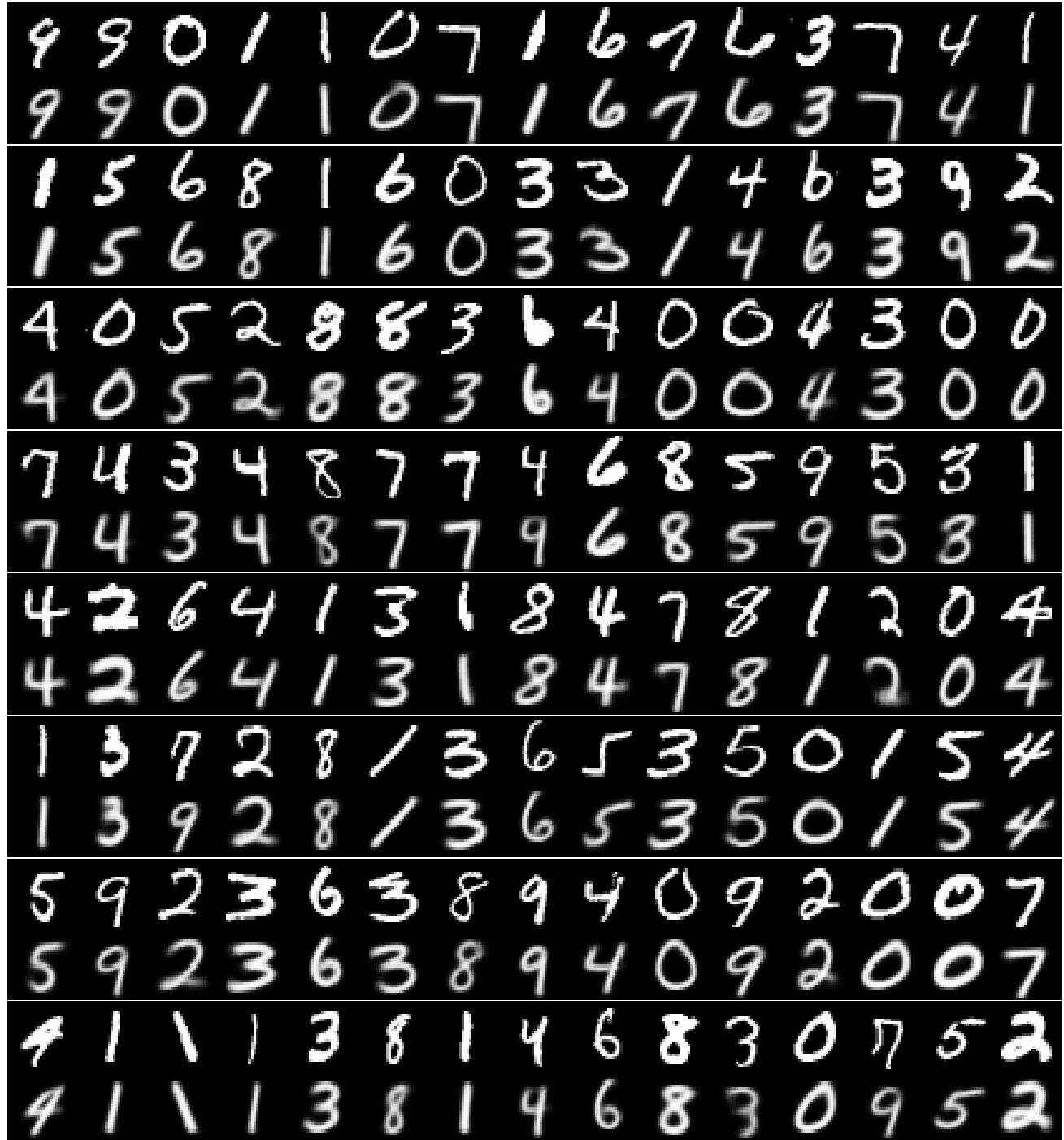
Figure 20: MNIST test reconstructions from VAE: *Top*: original test image, *Bottom*: reconstructed image.



Figure 21: MNIST random samples from VAE using the post-fit Gaussian mixture distribution on the latent space.