Appendix

In the case of AE, reconstruction loss is a mean squared error (MSE) between an input x and output x' of the network: $\mathcal{L}(x, x') = \|x - x'\|^2$. The network can be trained using standard machine learning techniques for training such as backpropagation. We utilise a regularized version of a standard AE which is known as *Sparse Autoencoder*. While L1 regularization is applied to the weight matrix of the final dense layer of the encoder which produced the latent vectors to make it sparse, L2 regularization is utilized for output of this layer to prevent its growth and overfitting [25].

Another version of considered autoencoders is *Sliced-Wasserstein Autoencoder (SWAE)*. This is a generative model with a simple implementation and which does not require adversarial training [20]. SWAE objective consists of a Wasserstein distance W_c between the distribution of input p_X and a decoder $p_{X'}$, and is regularized with the sliced-Wasserstein distance SW_c between the distribution of encoded training samples p_Z and, in our experiments, a uniform distribution in the embedding space q_Z:

$$arg\min_{\phi,\theta} \quad W_c(p_X, p_{X'}) + \lambda SW_c(p_Z, q_Z), \tag{1}$$

where ϕ and θ are the parameters of probabilistic encoder and decoder respectively.

Finally, we consider *VAE* and *β-VAE* in our autoencoder study for feature extraction. In the case of VAE [19], variational lower bound is:

$$\mathcal{L}(\theta, \phi; x) = -D_{KL}(q_\phi(z|x)||p_\theta(z)) + \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)], \tag{2}$$

assuming that the prior $p_\theta(z)$ is a unit Gaussian distribution $\mathcal{N}(0, I)$ and the approximate posterior $q_\phi(z|x)$ is a Gaussian $\mathcal{N}(\mu, \sigma^2)$ with parameters μ and σ as outputs of the encoder. The lower bound $-\mathcal{L}(\theta, \phi; x)$ must be minimized w.r.t. ϕ and θ. We can notice in the right hand side the regularization term in the form of KL divergence and the reconstruction term in the form of expected likelihood.

In the case of β-VAE [13], the beta-variational loss can be defined with one Lagrangian multiplier hyperparameter β:

$$\mathcal{L}(\theta, \phi; x, z, \beta) = -\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] + \beta D_{KL}(q_\phi(z|x)||p(z)). \tag{3}$$

The smaller values of β, less than one, encourage the expression to be in a form of an autoencoder, with the value $\beta = 1$ being a standard VAE explained above, the greater values restrict the representation capacity of the latent space. We tested the values of β in a range from 0.1 to 100.

Method type	**Convolution**	**Latent dim.**	**Neighborhood hit**	**Silhouette**	**Calinski-Harabasz**	**Davies-Bouldin**
Sparse AE	2D	64	0.985±0.005	-0.006±0.047	282.7±66.7	4.2±1.8
Sparse AE	2D	128	0.978±0.005	-0.024±0.01	122.8±48.8	9.2±2.8
Sparse AE	2D	256	0.974±0.004	-0.04±0.044	90.9±70.3	15.1±12.6
SWAE	2D	2	**0.994±0.001**	0.305±0.049	1131.1±145.4	1.5±0.2
SWAE	2D	4	0.941±0.005	-0.063±0.034	27.1±13.5	15.1±9.1
SWAE	2D	8	0.974±0.004	0.04±0.035	431.3±179.3	2.5±0.7
SWAE	2D	16	0.965±0.005	-0.006±0.043	302.1±78.9	4.7±1.5
SWAE	2D	32	0.957±0.006	-0.047±0.032	150.1±78.5	5.1±2.3
SWAE	2D	64	0.956±0.005	-0.061±0.061	109.9±132.3	9.8±4.8
SWAE	2D	128	0.974±0.003	0.051±0.077	416.6±223.5	2.6±0.6
VAE	2D	128	0.993±0.001	0.131±0.068	375.8±170.8	3.5±2.6
VAE	2D	256	**0.994±0.001**	0.099±0.089	305.7±190.3	4.4±1.9
β(2)-VAE	2D	32	0.991±0.003	0.195±0.055	560.7±177.7	2.5±1.0
β(2)-VAE	2D	64	0.992±0.001	0.141±0.106	426.8±141.0	2.9±1.2
β(2)-VAE	2D	128	0.993±0.001	0.113±0.062	359.6±123.3	3.2±0.8
β(2)-VAE	2D	256	0.992±0.001	0.171±0.064	453.8±85.5	3.1±1.4
β(4)-VAE	2D	128	0.986±0.003	0.226±0.049	794.4±264.5	2.3±1.3
β(4)-VAE	2D	256	0.978±0.006	0.258±0.095	1277.7±610.9	1.4±0.3
β(6)-VAE	2D	128	0.973±0.004	0.287±0.028	1135.1±266.5	1.3±0.3
β(6)-VAE	2D	256	0.921±0.027	0.209±0.039	986.0±246.0	1.6±0.4
β(8)-VAE	2D	32	0.947±0.009	**0.317±0.036**	**1504.8±399.1**	1.3±0.6
β(8)-VAE	2D	64	0.964±0.01	0.287±0.083	1284.5±491.7	1.5±0.6
β(8)-VAE	2D	128	0.912±0.027	0.272±0.036	1278.4±154.8	1.3±0.2
β(8)-VAE	2D	256	0.867±0.043	0.247±0.03	1284.8±206.5	1.4±0.2
β(10)-VAE	2D	128	0.888±0.028	0.292±0.029	1382.1±236.7	1.4±0.3
β(10)-VAE	2D	256	0.775±0.028	0.211±0.043	1108.4±231.3	1.4±0.5
β(20)-VAE	2D	32	0.346±0.269	0.039±0.132	282.0±629.1	76.7±51.4
β(20)-VAE	2D	64	0.814±0.014	0.257±0.022	1303.7±218.9	**1.2±0.3**
β(20)-VAE	2D	128	0.742±0.037	0.225±0.041	1062.9±183.9	1.3±0.2
β(20)-VAE	2D	256	0.588±0.037	0.102±0.03	638.3±98.6	2.2±0.7
β(100)-VAE	2D	128	0.228±0.003	-0.027±0.004	2.6±2.7	64.3±42.1
β(100)-VAE	2D	256	0.227±0.003	-0.022±0.005	0.8±0.4	118.4±102.8
Baseline	–	–	0.977±0.008	-0.061±0.033	90.7±63.1	8.8±5.8

Table 1: Metrics scores of all models performing feature extraction on the MCMC ensemble.

Method type	**Convolution**	**Latent dim.**	**Neighborhood hit**	**Silhouette**	**Calinski-Harabasz**	**Davies-Bouldin**
Sparse AE	2D	64	0.677±0.006	-0.044±0.022	735.7±284.5	4.6±2.1
Sparse AE	2D	128	0.671±0.017	-0.063±0.029	642.0±193.7	3.7±0.8
Sparse AE	2D	256	0.673±0.01	-0.052±0.015	657.6±131.5	6.8±5.0
VAE	2D	128	0.633±0.001	-0.025±0.014	617.9±83.8	5.9±1.8
VAE	2D	256	0.59±0.01	-0.028±0.01	608.2±95.1	4.7±1.1
β(4)-VAE	2D	128	0.483±0.014	-0.061±0.011	549.7±19.0	7.3±3.3
AE	3D	64	0.775±0.008	-0.091±0.011	408.4±84.8	5.7±2.7
AE	3D	128	0.772±0.01	-0.129±0.016	353.7±115.2	5.7±1.0
AE	3D	256	**0.782±0.009**	-0.085±0.015	458.8±48.2	4.9±0.9
Sparse AE	3D	256	0.77±0.006	-0.111±0.02	339.1±100.7	6.3±3.3
SWAE	3D	32	0.773±0.014	-0.083±0.008	655.6±104.0	4.7±1.9
SWAE	3D	64	0.773±0.01	-0.101±0.015	583.6±85.7	5.3±1.4
SWAE	3D	128	0.762±0.018	-0.093±0.026	552.8±152.3	4.7±1.1
β(0.1)-VAE	3D	256	0.723±0.011	**-0.005±0.03**	**831.3±224.1**	**3.6±0.3**
VAE	3D	256	0.592±0.024	-0.02±0.016	797.2±58.9	9.6±2.4
β(2)-VAE	3D	256	0.514±0.011	-0.05±0.014	640.6±80.3	10.5±2.2
β(4)-VAE	3D	256	0.421±0.012	-0.08±0.017	583.9±60.8	15.7±3.7
β(10)-VAE	3D	256	0.301±0.004	-0.094±0.011	400.2±21.3	15.6±3.5
Baseline	–	–	0.641±0.011	-0.112±0.029	449.3±156.8	9.2±3.9

Table 2: Metrics scores of all models performing feature extraction on the Drop Dynamics ensemble.

Layer type	**Output Shape**	**Details**
Input	(batch size, 3, h, w, 1)	height = h, width = w
Conv3D	(batch size, 1, h/2, w/2, 64)	kernel size = (3, 3, 3), stride = (3, 2, 2)
Conv3D	(batch size, 1, h/4, w/4, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Conv3D	(batch size, 1, h/8, w/8, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Conv3D	(batch size, 1, h/16, w/16, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Flatten	(batch size, 1, (h/16) · (w/16) · 64)	reshape before dense layer
Dense	(batch size, num. of units)	first dense layer of encoder
AE: Dense	(batch size, latent dimension)	second dense layer
VAE: Dense $(\mu, \log\sigma)$	(batch size, latent dimension)	two parallel dense layers for VAE
VAE: Sample z	(batch size, latent dimension)	reparameterization trick for VAE
Dense	(batch size, 1, (h/16) · (w/16) · 64)	first dense layer of decoder
Reshape	(batch size, 1, (h/16) · (w/16) · 64)	reshape before deconvolutions
Conv3DTranspose	(batch size, 1, h/8, w/8, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Conv3DTranspose	(batch size, 1, h/4, w/4, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Conv3DTranspose	(batch size, 1, h/2, w/2, 64)	kernel size = (1, 3, 3), stride = (1, 2, 2)
Conv3DTranspose	(batch size, 1, h, w, 64)	kernel size = (3, 3, 3), stride = (3, 2, 2)
Conv3DTranspose	(batch size, 1, h, w, 1)	kernel size = (3, 3, 3), stride = (1, 1, 1)

Table 3: 3D AE/VAE architecture used on Drop Dynamics ensemble. The difference is highlighted in italics in the bottleneck.