Flow Factorized Representation Learning

Yue Song[1,2], Andy Keller[2], Nicu Sebe[1], and Max Welling[2]
[1]DISI, University of Trento, Trento, Italy
[2]AMLab, University of Amsterdam, the Netherlands

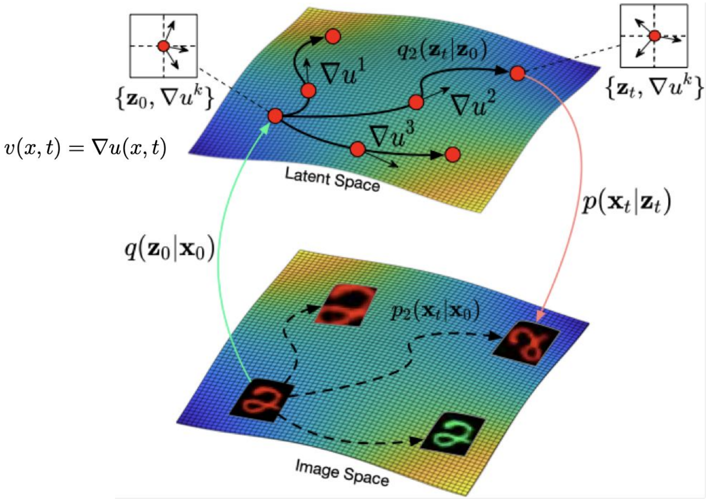
UNIVERSITÀ DI TRENTO

UNIVERSITEIT VAN AMSTERDAM

Flow Factorized VAE

Novel definitions of ***generalized equivariance*** and ***disentanglement***.

$$\frac{d\,\rho(x,t)}{dt} = -\nabla \cdot (v(x,t)\rho(x,t))$$



Generalized Equivariance:

$$p_k(\boldsymbol{x}_t|\boldsymbol{x}_0) = \int_{\boldsymbol{z}_0,\boldsymbol{z}_t} q(\boldsymbol{z}_0|\boldsymbol{x}_0)q_k(\boldsymbol{z}_t|\boldsymbol{z}_0)p(\boldsymbol{x}_t|\boldsymbol{z}_t)$$

Disentanglement:

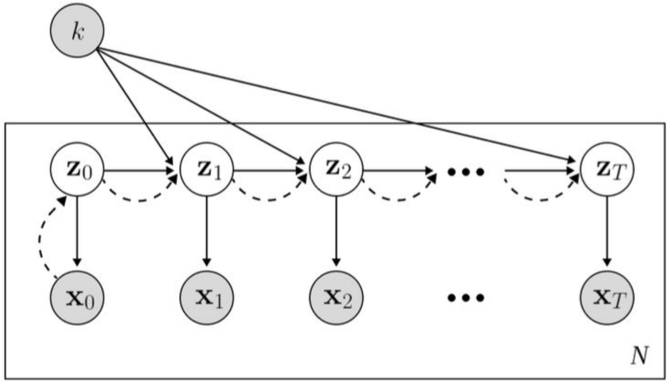
Distinct tangent bundles following OT

Fluid-Dynamic Optimal Transport:

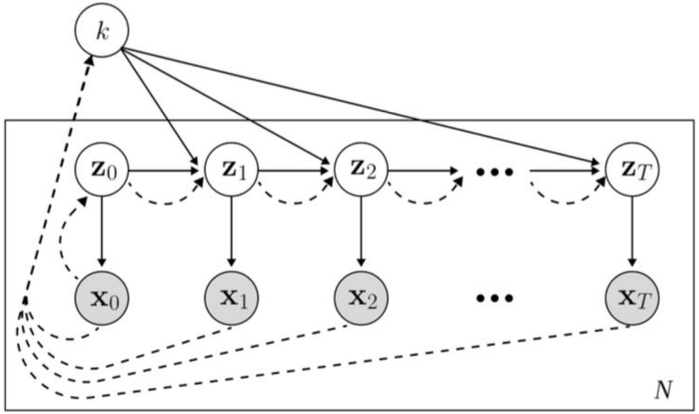
Hamilton-Jacobi Eq. : $\frac{\partial}{\partial t}u^k(\boldsymbol{z},t) + \frac{1}{2}||\nabla_{\boldsymbol{z}} u^k(\boldsymbol{z},t)||^2 = f(\boldsymbol{z},t)$

The Generative Model

Supervised



Weakly-supervised



The joint distribution is factorized as follows:

$$p(\bar{\boldsymbol{x}}, \bar{\boldsymbol{z}}, k) = p(k)p(\boldsymbol{z}_0)p(\boldsymbol{x}_0|\boldsymbol{z}_0)\prod_{t=1}^{T} p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1}, k)p(\boldsymbol{x}_t|\boldsymbol{z}_t).$$

Prior&Posterior Time Evolution

For both the prior and posterior, since the induced velocity field advects the probability density, we have the normalizing-flow-like conditional update:

$$p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k) = p(\boldsymbol{z}_{t-1})\left|\frac{df(\boldsymbol{z}_{t-1},k)}{d\boldsymbol{z}_{t-1}}\right|^{-1} \qquad q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k) = q(\boldsymbol{z}_{t-1})\left|\frac{dg(\boldsymbol{z}_{t-1},k)}{d\boldsymbol{z}_{t-1}}\right|^{-1}$$

where the function f&g are defined as:

$$\boldsymbol{z}_t = f(\boldsymbol{z}_{t-1},k) = \boldsymbol{z}_{t-1} + \nabla_{\boldsymbol{z}}\psi^k(\boldsymbol{z}_{t-1}) \qquad \boldsymbol{z}_t = g(\boldsymbol{z}_{t-1},k) = \boldsymbol{z}_{t-1} + \nabla_{\boldsymbol{z}}u^k$$

Prior. Since we have no prior knowledge of the sequence, as a minimally informative prior for random trajectories, we use diffusion equation for the prior and simply take:

$$\psi^k = -D_k \log p(\boldsymbol{z}_t) \qquad \partial_t p(\boldsymbol{z}_t) = -\nabla\cdot\Big(p(\boldsymbol{z}_t)\nabla\psi\Big) = D_k\nabla^2 p(\boldsymbol{z}_t)$$

Posterior. We paramterize the potentials as $u^k(\boldsymbol{z},t) = \texttt{MLP}([\boldsymbol{z};t])$. The posterior evolves as:

$$\log q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k) = \log q(\boldsymbol{z}_{t-1}) - \log|1 + \nabla^2_{\boldsymbol{z}}u^k|$$

Evidence Lower Bound

Inference with observed k (supervised). When k is observed, we factorize the posterior as:

$$q(\bar{\boldsymbol{z}}|\bar{\boldsymbol{x}},k) = q(\boldsymbol{z}_0|\boldsymbol{x}_0)\prod_{t=1}^{T} q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k)$$

We derive the following upper bound as:

$$\begin{aligned}\log p(\bar{\boldsymbol{x}}|k) \geq \sum_{t=0}^{T} \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}}|k)}\left[\log p(\boldsymbol{x}_t|\boldsymbol{z}_t,k)\right] - \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}}|k)}\left[\mathrm{D}_{\mathrm{KL}}\left[q_\theta(\boldsymbol{z}_0|\boldsymbol{x}_0)||p(\boldsymbol{z}_0)\right]\right] \\ - \sum_{t=1}^{T} \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}}|k)}\left[\mathrm{D}_{\mathrm{KL}}\left[q_\theta(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k)||p(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k)\right]\right]\end{aligned}$$

Inference with latent k (weakly supervised). We treat k as a latent variable and define the approximate posterior as:

$$q(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}}) = q(k|\bar{\boldsymbol{x}})q(\boldsymbol{z}_0|\boldsymbol{x}_0)\prod_{t=1}^{T} q(\boldsymbol{z}_t|\boldsymbol{z}_{t-1},k)$$

The new ELBO is derived as:

$$\begin{aligned}\log p(\bar{\boldsymbol{x}}) &= \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\left[\log \frac{p(\bar{\boldsymbol{x}},\bar{\boldsymbol{z}},k)}{q(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\frac{q(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}{p(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\right] \\ &\geq \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\left[\log \frac{p(\bar{\boldsymbol{x}}|\bar{\boldsymbol{z}},k)p(\bar{\boldsymbol{z}}|k)}{q(\bar{\boldsymbol{z}}|\bar{\boldsymbol{x}},k)}\frac{p(k)}{q(k|\bar{\boldsymbol{x}})}\right] \\ &= \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\left[\log p(\bar{\boldsymbol{x}}|\bar{\boldsymbol{z}},k)\right] + \mathbb{E}_{q_\theta(\bar{\boldsymbol{z}},k|\bar{\boldsymbol{x}})}\left[\log \frac{p(\bar{\boldsymbol{z}}|k)}{q(\bar{\boldsymbol{z}}|\bar{\boldsymbol{x}},k)}\right] + \mathbb{E}_{q_\gamma(k|\bar{\boldsymbol{x}})}\left[\log \frac{p(k)}{q(k|\bar{\boldsymbol{x}})}\right]\end{aligned}$$

Quantitative Evaluation

Our approach achieves better equivariance error and improved likelihood than previous baselines.

Methods	Supervision?	Equivariance Error (↓)			Log-likelihood (↑)
		Scaling	Rotation	Coloring	
VAE [47]	No (✗)	1275.31±1.89	1310.72±2.19	1368.92±2.33	-2206.17±1.83
β-VAE [35]	No (✗)	741.58±4.57	751.32±5.22	808.16±5.03	-2224.67±2.35
FactorVAE [46]	No (✗)	659.71±4.89	632.44±5.76	662.18±5.26	-2209.33±2.47
SlowVAE [49]	Weak (✓)	461.59±5.37	447.46±5.46	398.12±4.83	-2197.68±2.39
TVAE [45]	Yes (✓)	505.19±2.77	493.28±3.37	451.25±2.76	-2181.13±1.87
PoFlow [79]	Yes (✓)	234.78±2.91	231.42±2.98	240.57±2.58	-2145.03±2.01
Ours	Yes (✓)	**185.42±2.35**	**153.54±3.10**	**158.57±2.95**	**-2112.45±1.57**
Ours	Weak (✓)	193.84±2.47	157.16±3.24	165.19±2.78	-2119.94±1.76

Table 1: Equivariance error $\mathcal{E}_k$ and log-likelihood $\log p(\boldsymbol{x}_t)$ on MNIST [54].

Methods	Supervision?	Equivariance Error (↓)				Log-likelihood (↑)
		Floor Hue	Wall Hue	Object Hue	Scale	
VAE [47]	No (✗)	6924.63±8.92	7746.37±8.77	4383.54±9.26	2609.59±7.41	-11784.69±4.87
β-VAE [35]	No (✗)	2243.95±12.48	2279.23±13.97	2188.73±12.61	2037.94±11.72	-11924.83±5.64
FactorVAE [46]	No (✗)	1985.75±13.26	1876.41±11.93	1902.83±12.27	1657.32±11.05	-11802.17±5.69
SlowVAE [49]	Weak (✓)	1247.36±12.49	1314.86±11.41	1102.28±12.17	1058.74±10.96	-11674.89±5.74
TVAE [45]	Yes (✓)	1225.47±9.82	1246.32±9.54	1261.79±9.86	1142.01±9.37	-11475.48±5.18
PoFlow [79]	Yes (✓)	885.46±10.37	916.71±10.49	912.48±9.86	924.39±10.05	-11335.84±4.95
Ours	Yes (✓)	**613.29±8.93**	**653.45±9.48**	**605.79±8.63**	**599.71±9.34**	**-11215.42±5.71**
Ours	Weak (✓)	690.84±9.57	717.74±10.65	681.59±9.02	653.58±9.57	-11279.61±5.89

Table 2: Equivariance error $\mathcal{E}_k$ and log-likelihood $\log p(\boldsymbol{x}_t)$ on Shapes3D [10].

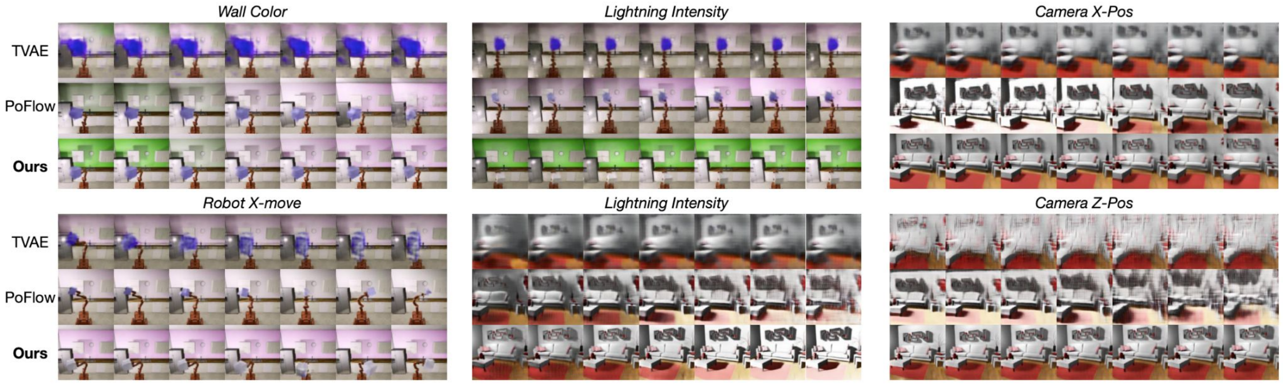
Methods	Lighting Intensity	Lighting X-dir	Lighting Y-dir	Lighting Z-dir	Camera X-pos	Camera Y-pos	Camera Y-pos
TVAE [45]	11477.81	12568.32	11807.34	11829.33	11539.69	11736.78	11951.45
PoFlow [79]	8312.97	7956.18	8519.39	8871.62	8116.82	8534.91	8994.63
Ours	**5798.42**	**6145.09**	**6334.87**	**6782.84**	**6312.95**	**6513.68**	**6614.27**

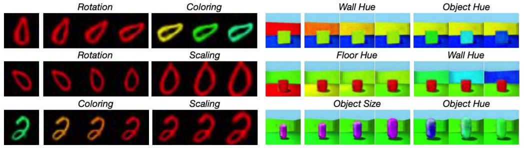
Table 3: Equivariance error (↓) on Falcol3D [61].

Methods	Robot X-move	Robot Y-move	Camera Height	Object Scale	Lighting Intensity	Lighting Y-dir	Object Color	Wall Color
TVAE [45]	8441.65	8348.23	8495.31	8251.34	8291.70	8741.07	8456.78	8512.09
PoFlow [79]	6572.19	6489.35	6319.82	6188.59	6517.40	6712.06	7056.98	6343.76
Ours	**3659.72**	**3993.33**	**4170.27**	**4359.78**	**4225.34**	**4019.84**	**5514.97**	**3876.01**

Table 4: Equivariance error (↓) on Isaac3D [61].

Qualitative Evaluation





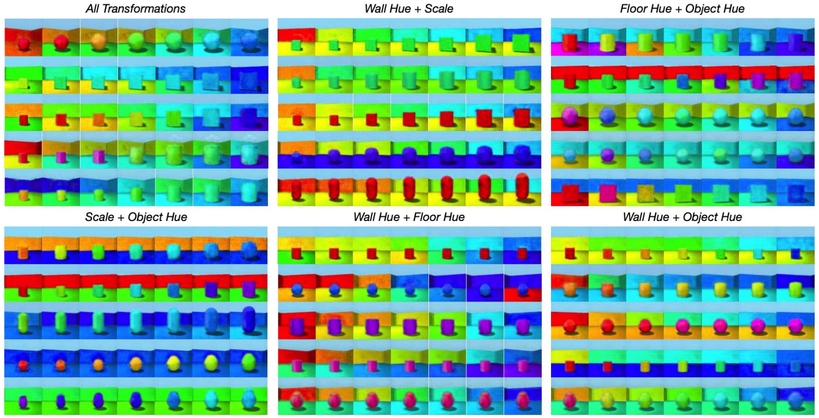
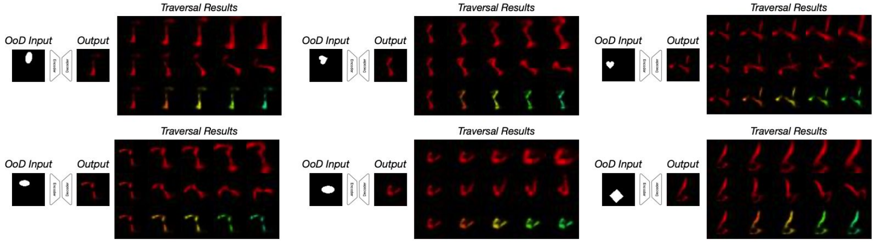


Figure 6: Examples of combining different transformations simultaneously during the latent evolution.



Thank you!