An Identifiability Perspective on Representation Learning

Yash Sharma

github.com/bethgelab/slow_disentanglement
arxiv.org/abs/2007.10930

github.com/brendel-group/cl-ica
arxiv.org/abs/2102.08850

Outline

1. Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding
    a. Identifiability w/ assumptions derived from natural video

Outline

1. Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding
    a. Identifiability w/ assumptions derived from natural video
2. Contrastive Learning Inverts the Data Generating Process
    a. Identifiability & InfoNCE

Outline

1. Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding
    a. Identifiability w/ assumptions derived from natural video
2. Contrastive Learning Inverts the Data Generating Process
    a. Identifiability & InfoNCE
3. Self-Supervised Learning with Data Augmentations Provably Isolates Content from Style
    a. Identifiability when augmentations leave factors invariant

Towards Nonlinear Disentanglement in Natural Data with Temporal Sparse Coding

Overview

1. Problem Statement
    a. What is disentanglement?

Overview

1. Problem Statement
    a. What is disentanglement?

Overview

1. Problem Statement
    a. What is disentanglement?
    b. What do we need to solve disentanglement?

Overview

1. Problem Statement
   a. What is disentanglement?
   b. What do we need to solve disentanglement?
   c. Can we find what we need in natural video?

Overview

1. Problem Statement
    a. What is disentanglement?
    b. What do we need to solve disentanglement?
    c. Can we find what we need in natural video?
2. Theoretical Contributions
    a. A prior based on natural statistics provably enables disentanglement

Overview

1. Problem Statement
    a. What is disentanglement?
    b. What do we need to solve disentanglement?
    c. Can we find what we need in natural video?
2. Theoretical Contributions
    a. A prior based on natural statistics provably enables disentanglement
3. Empirical Contributions
    a. Qualitative and quantitative results on existing and contributed datasets demonstrate outperformance in aggregate

What is Disentanglement?

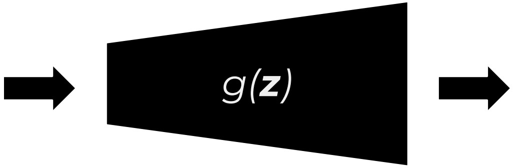
State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position

What is Disentanglement?

State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position

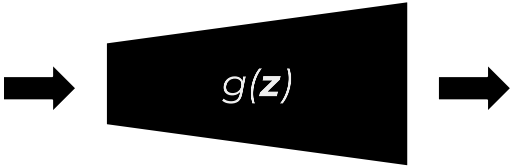


data generating process

What is Disentanglement?

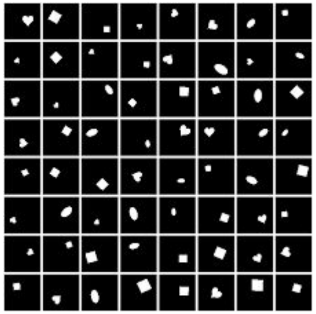
State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position



data generating process

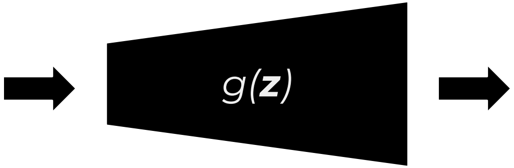
Observations, **x**:



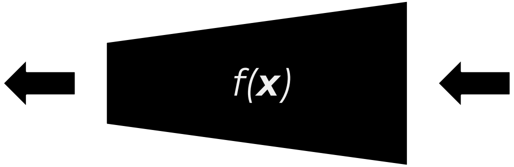
What is Disentanglement?

State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position

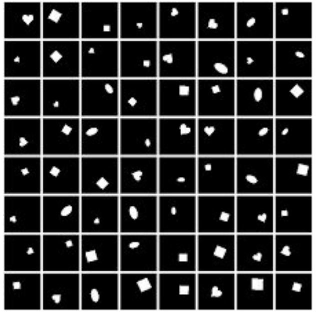


data generating process



representation learning

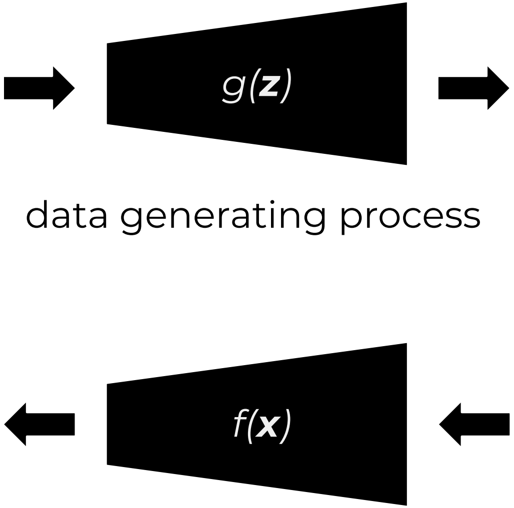
Observations, **x**:



What is Disentanglement?

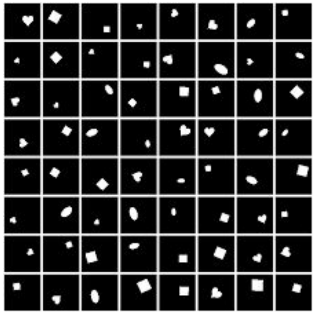
State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position



representation learning

Observations, **x**:

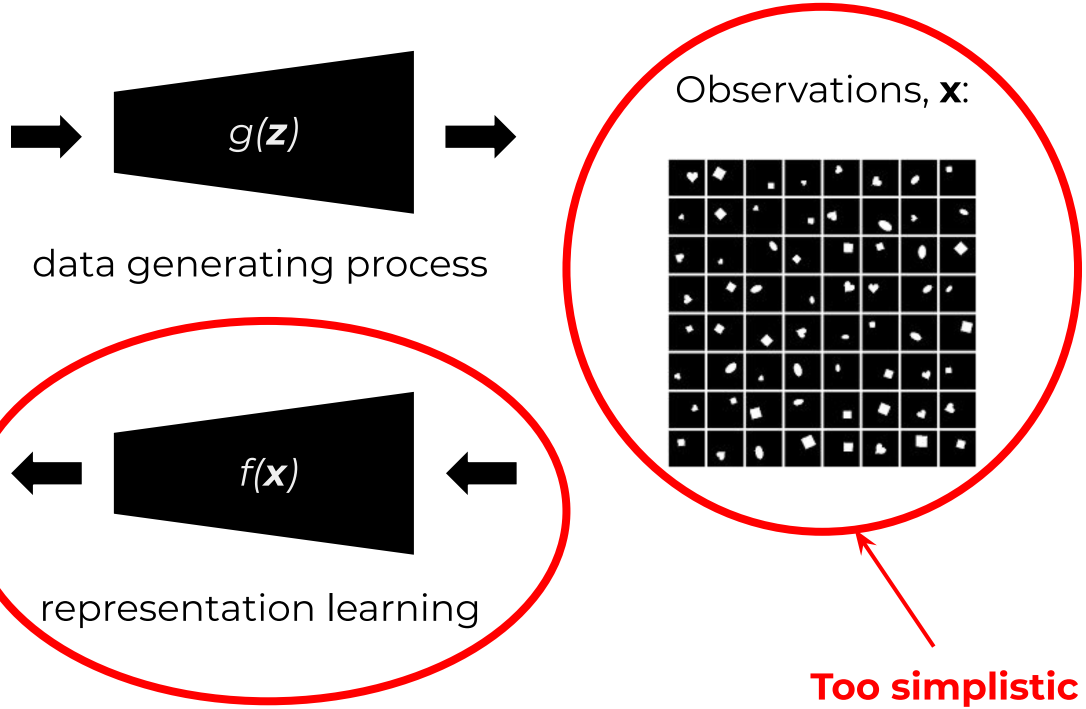


Too simplistic

What is Disentanglement?

State of the world, **z**:

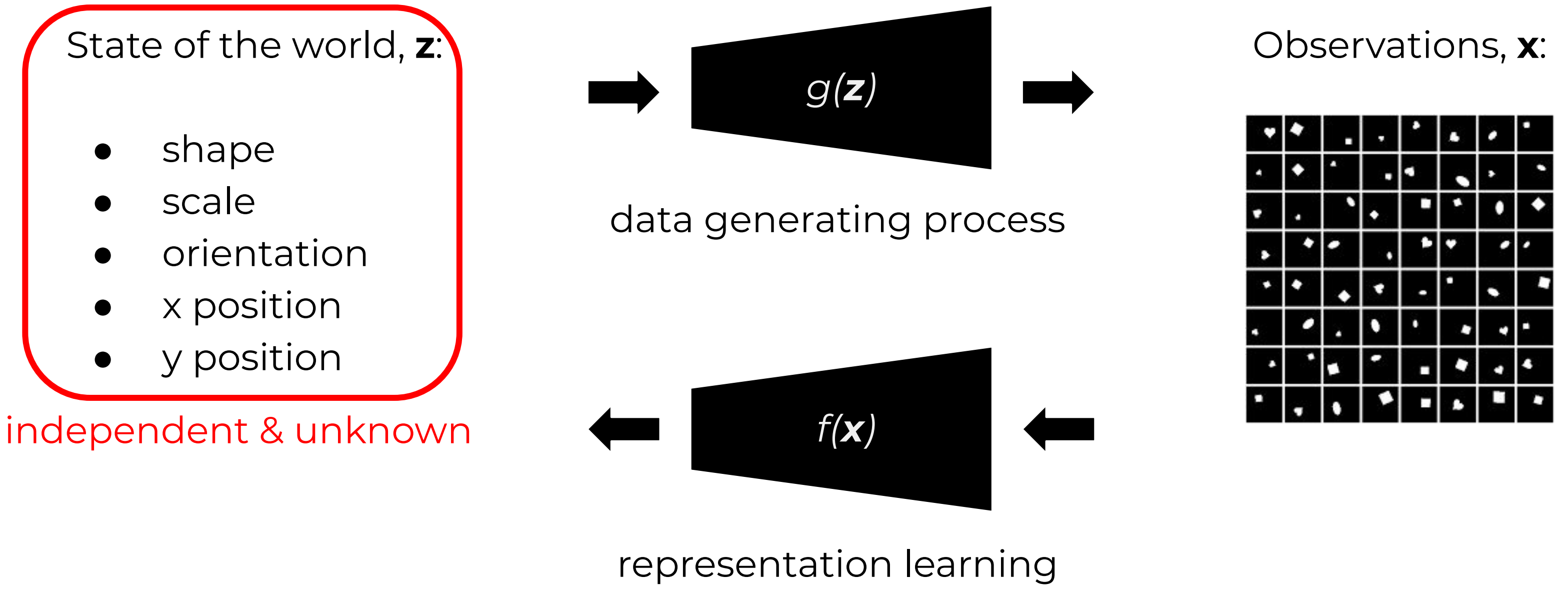
- shape
- scale
- orientation
- x position
- y position



prev solutions do not account for observed natural scene statistics

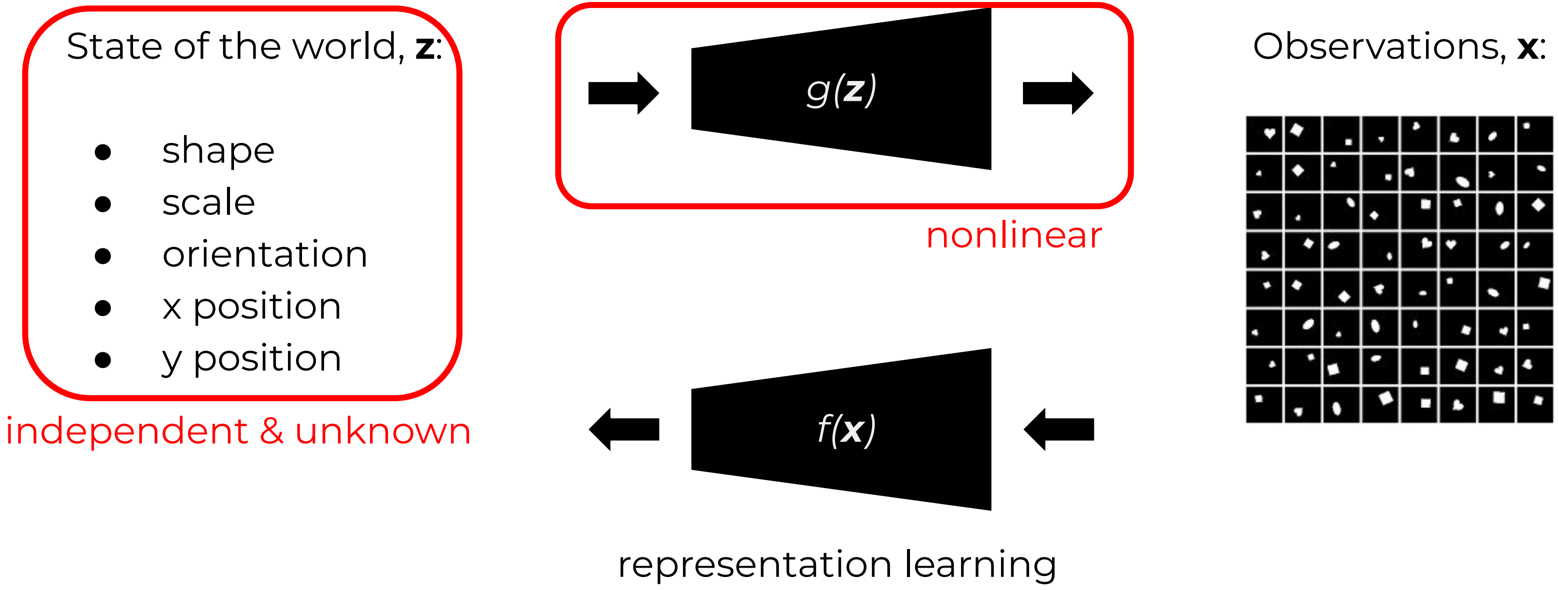
Too simplistic

Non-identifiability



Hyvärinen & Pajunen (1999) *Nonlinear independent component analysis: Existence and uniqueness results*

Non-identifiability



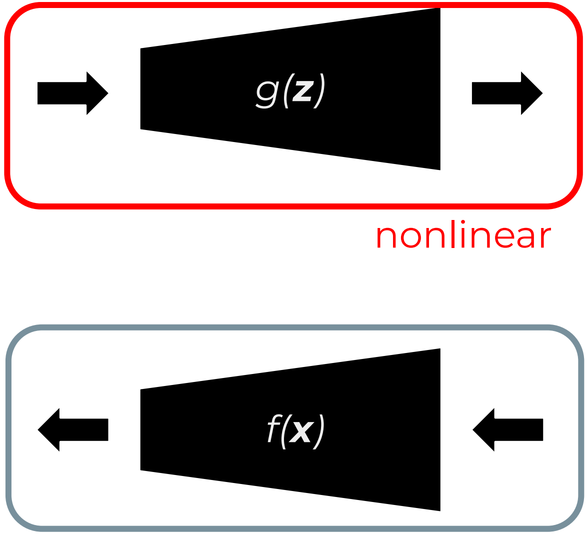
Hyvärinen & Pajunen (1999) *Nonlinear independent component analysis: Existence and uniqueness results*

Non-identifiability

State of the world, **z**:

- shape
- scale
- orientation
- x position
- y position

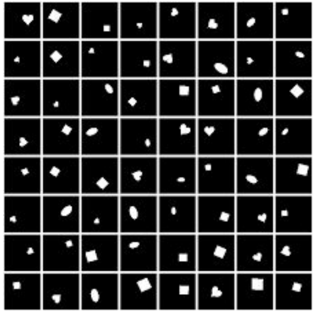
independent & unknown



nonlinear

Non-identifiable
🥺

Observations, **x**:

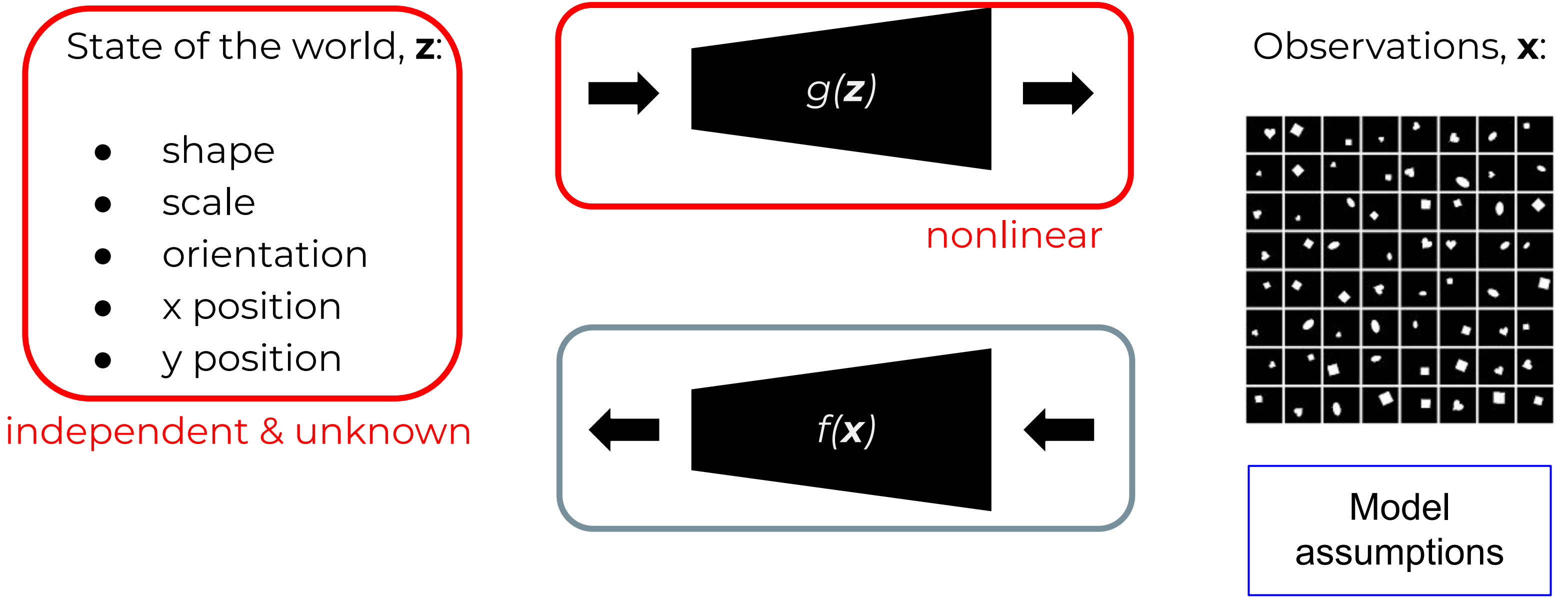


independent identically distributed (i.i.d.)

Locatello et al. (2018) *Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations*

Hyvärinen & Pajunen (1999) *Nonlinear independent component analysis: Existence and uniqueness results*

Nonlinear Disentanglement



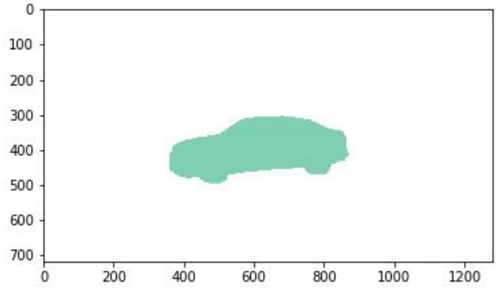
Locatello et al. (2020) *Weakly-Supervised Disentanglement Without Compromises*

Hyvärinen & Morioka (2017) *Nonlinear ICA of Temporally Dependent Stationary Sources*

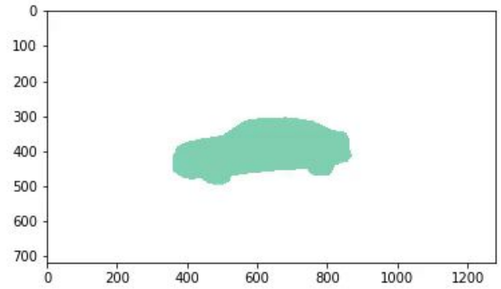
Hyvärinen & Morioka (2016) *Unsupervised Feature Extraction by Time-Contrastive Learning and Nonlinear ICA*

Natural Video

Natural Video

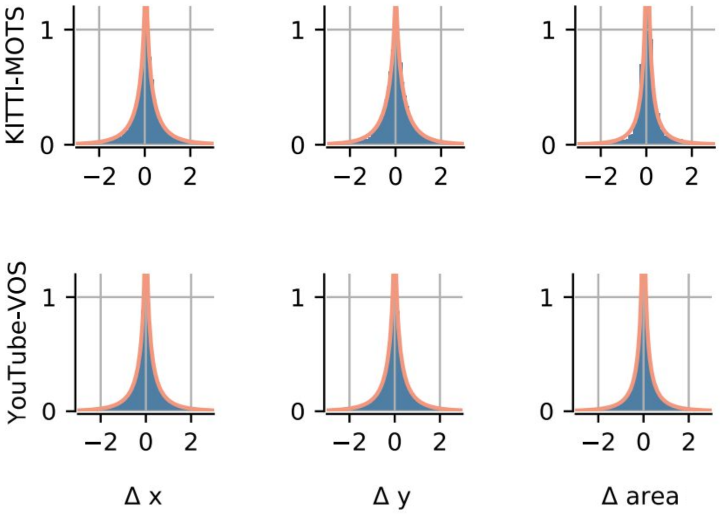


Natural Video



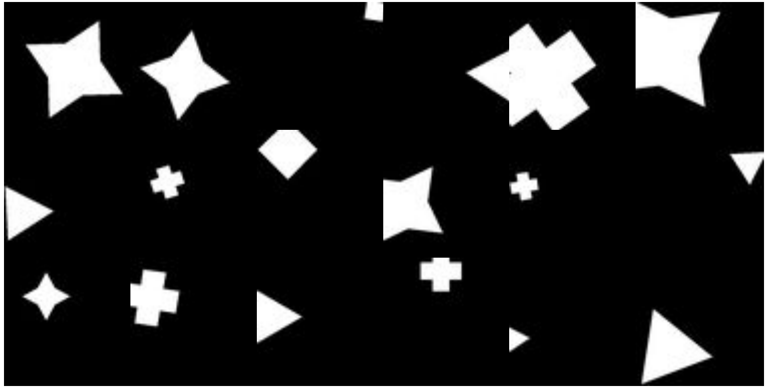
- scale
- x position
- y position

Natural Data Analysis



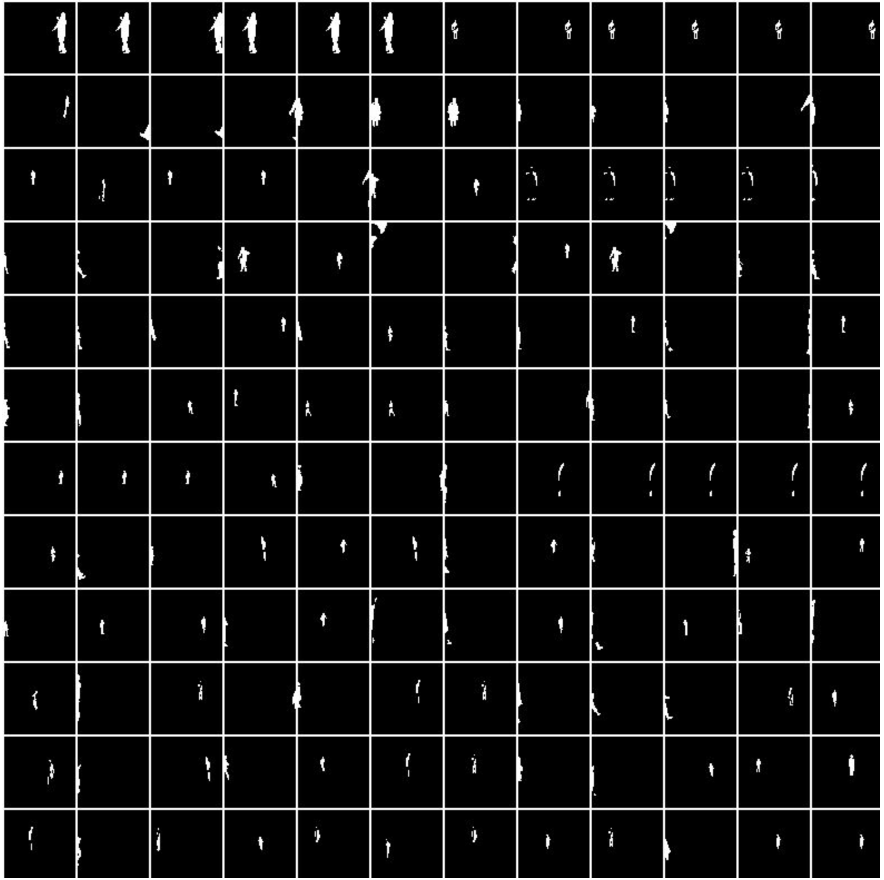
Natural Sprites

- Images generated online using renderer
- Simple, well-controlled objects
- Transitions sampled from YouTube-VOS



KITTI Masks

- Pedestrian masks extracted directly from autonomous recorded videos
- Realistic objects & transitions



The world is not *i.i.d.*

The world is not *i.i.d.*

State of the world, **z**:

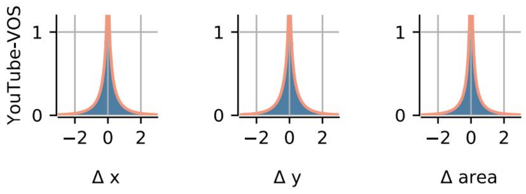
{shape, scale, orientation,
x position, y position}

The world is not *i.i.d.*

State of the world, **z**:

{shape, scale, orientation,
x position, y position}

And dynamics:

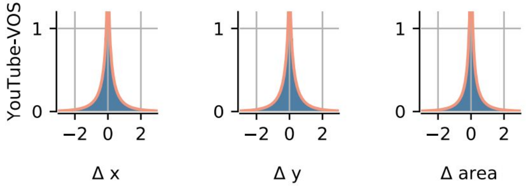


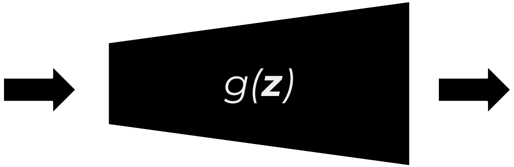
The world is not *i.i.d.*

State of the world, **z**:

{shape, scale, orientation, x position, y position}

And dynamics:



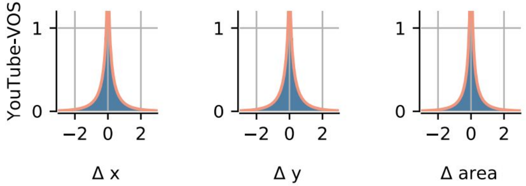


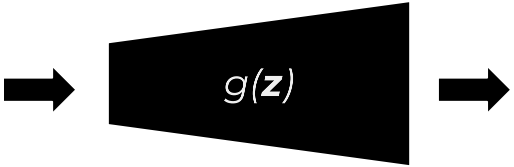
The world is not *i.i.d.*

State of the world, **z**:

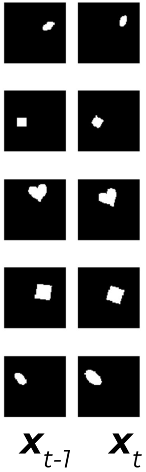
{shape, scale, orientation, x position, y position}

And dynamics:





Observations:

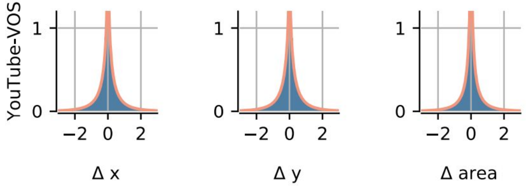


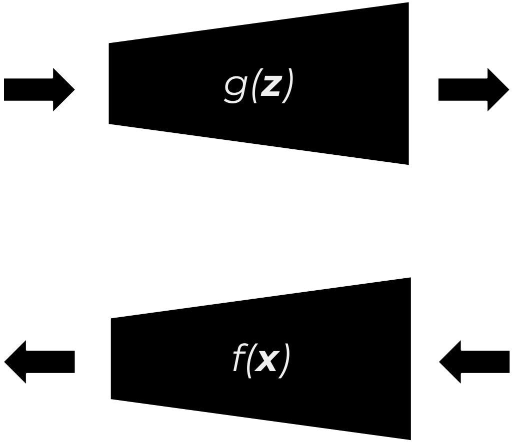
The world is not *i.i.d.*

State of the world, **z**:

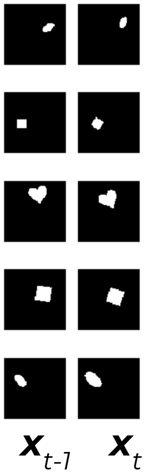
{shape, scale, orientation, x position, y position}

And dynamics:





Observations:

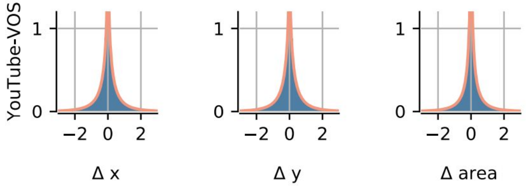


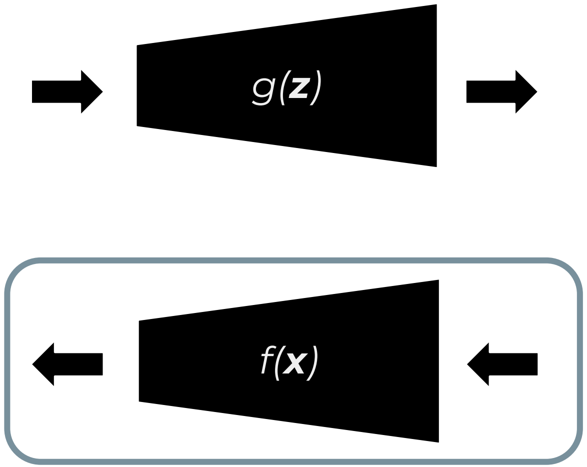
The world is not *i.i.d.*

State of the world, **z**:

{shape, scale, orientation, x position, y position}

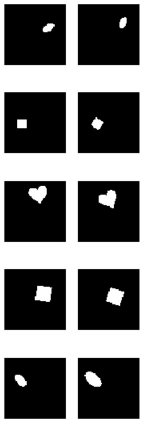
And dynamics:





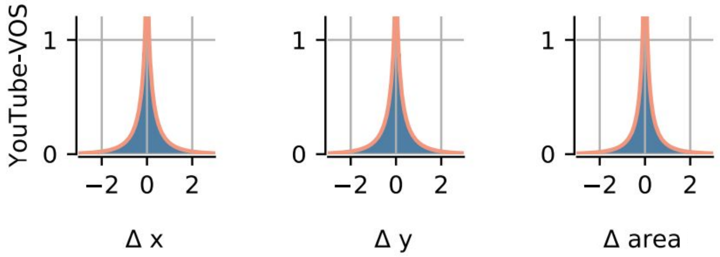
identifiable 🤩

Observations:

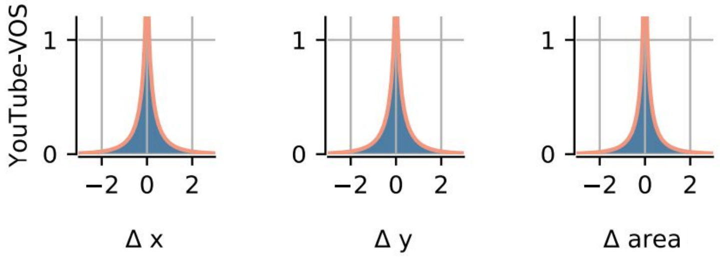


$\boldsymbol{x}_{t-1}$ $\boldsymbol{x}_t$

Identifiability Proof Intuition

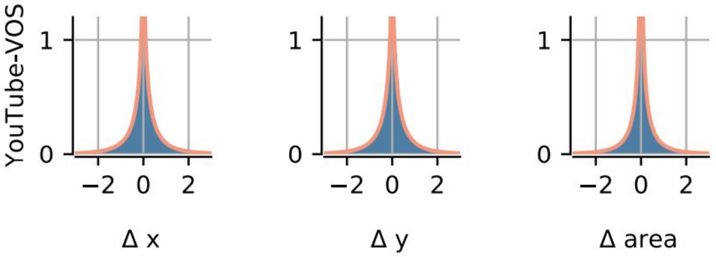


Identifiability Proof Intuition

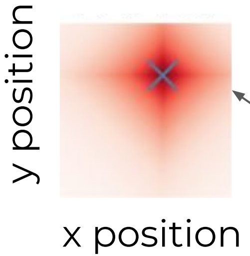


Prior:
objects in nature change sparsely

Identifiability Proof Intuition

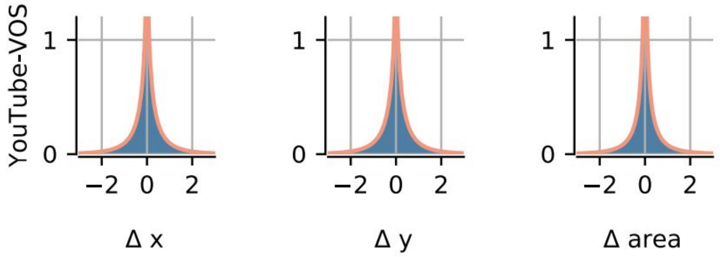


Prior:
objects in nature change sparsely

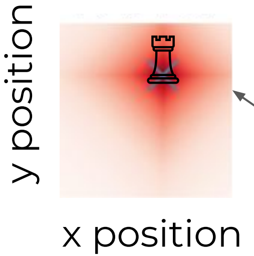


Probability of next position given previous position

Identifiability Proof Intuition

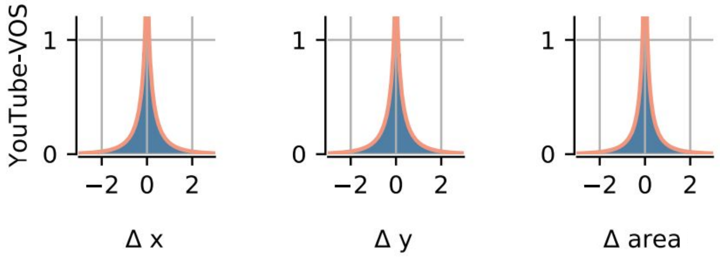


Prior:
objects in nature change sparsely

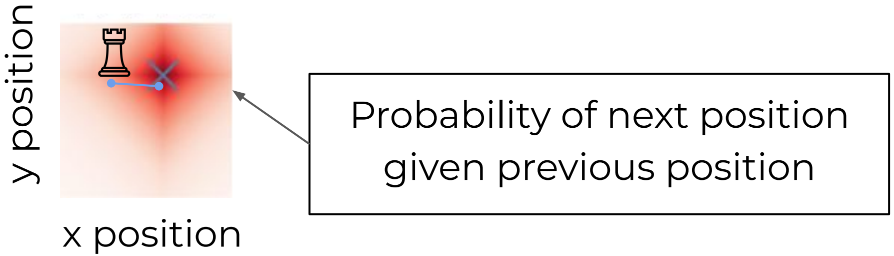


Probability of next position given previous position

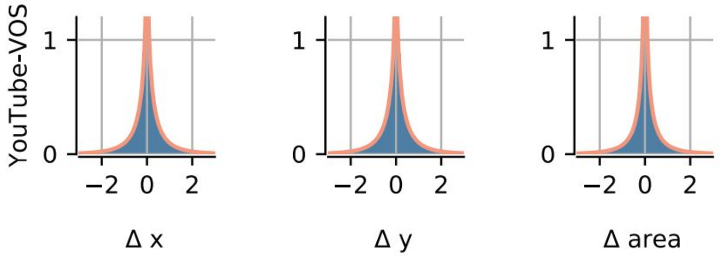
Identifiability Proof Intuition



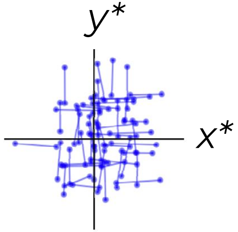
Prior:
objects in nature change sparsely



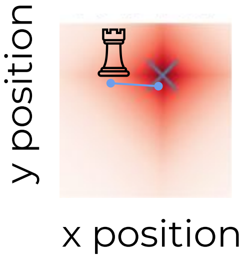
Identifiability Proof Intuition



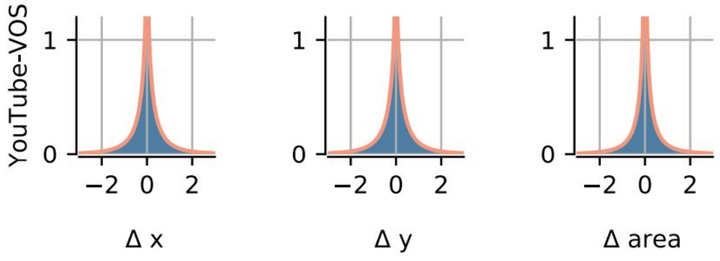
True Model



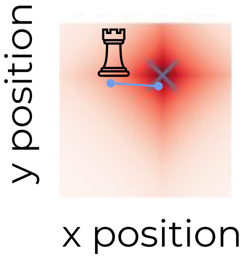
Prior:
objects in nature change sparsely



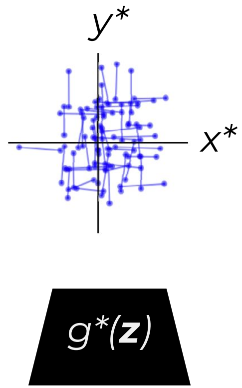
Identifiability Proof Intuition



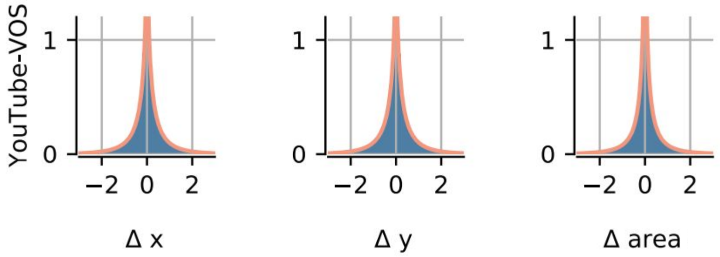
Prior:
objects in nature change sparsely



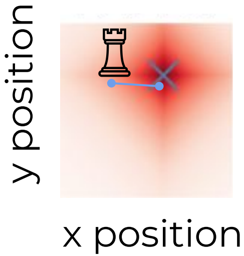
True Model



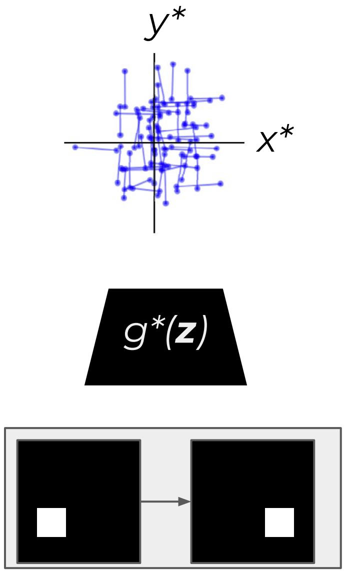
Identifiability Proof Intuition



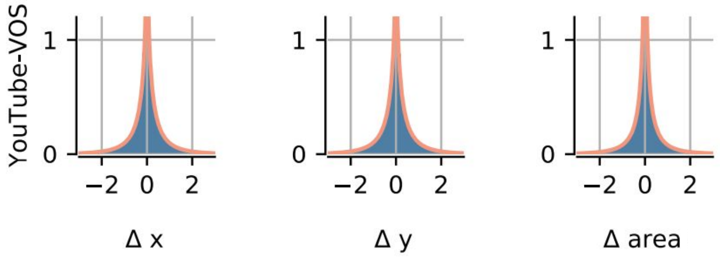
Prior:
objects in nature change sparsely



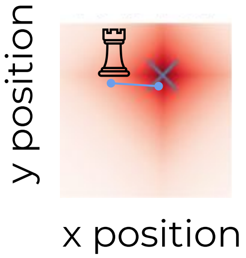
True Model

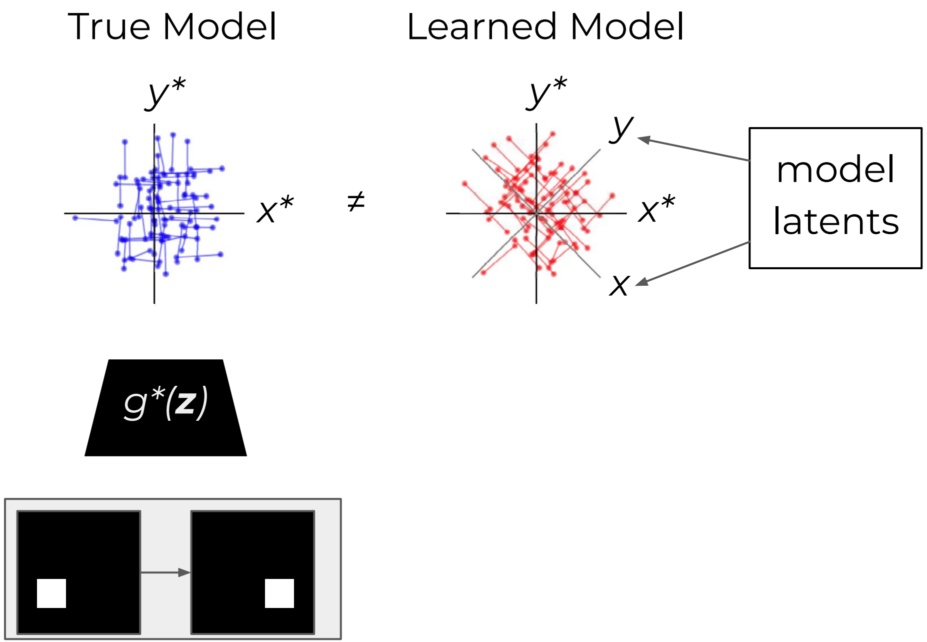


Identifiability Proof Intuition

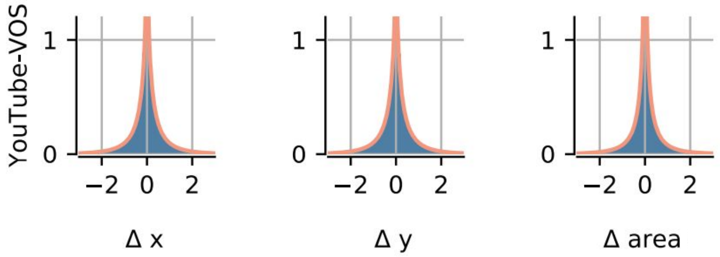


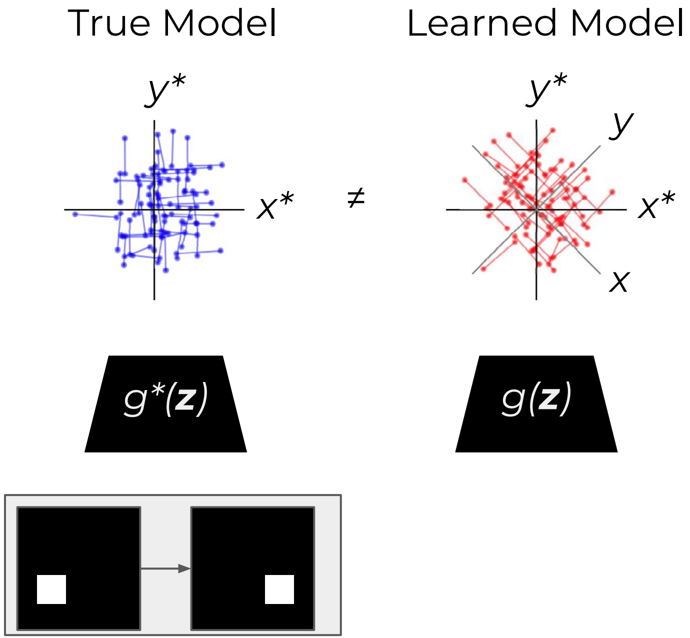
Prior:
objects in nature change sparsely



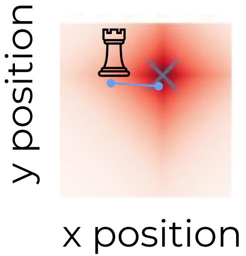


Identifiability Proof Intuition

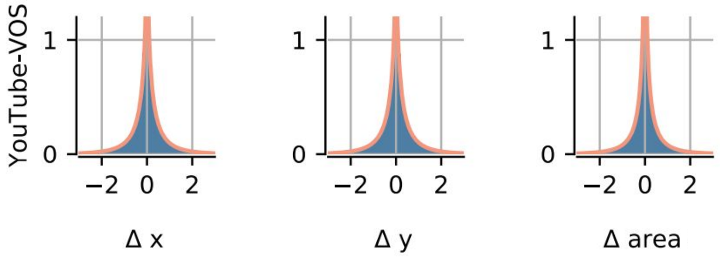




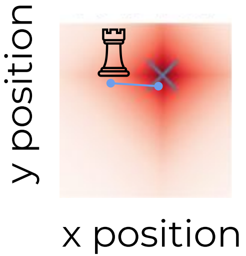
Prior:
objects in nature change sparsely

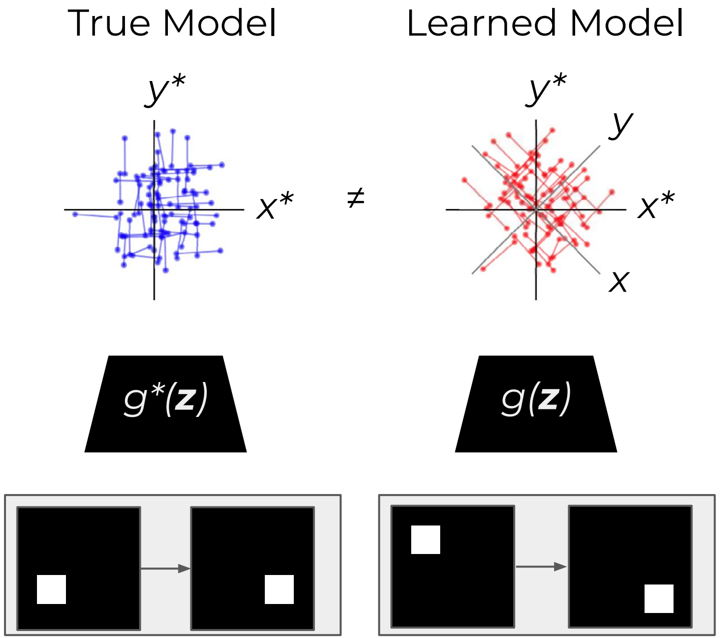


Identifiability Proof Intuition

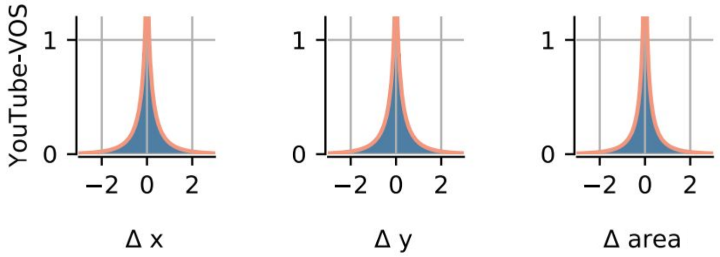


Prior:
objects in nature change sparsely

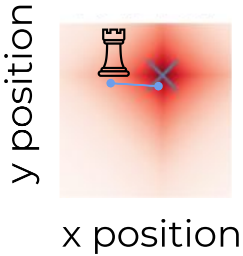


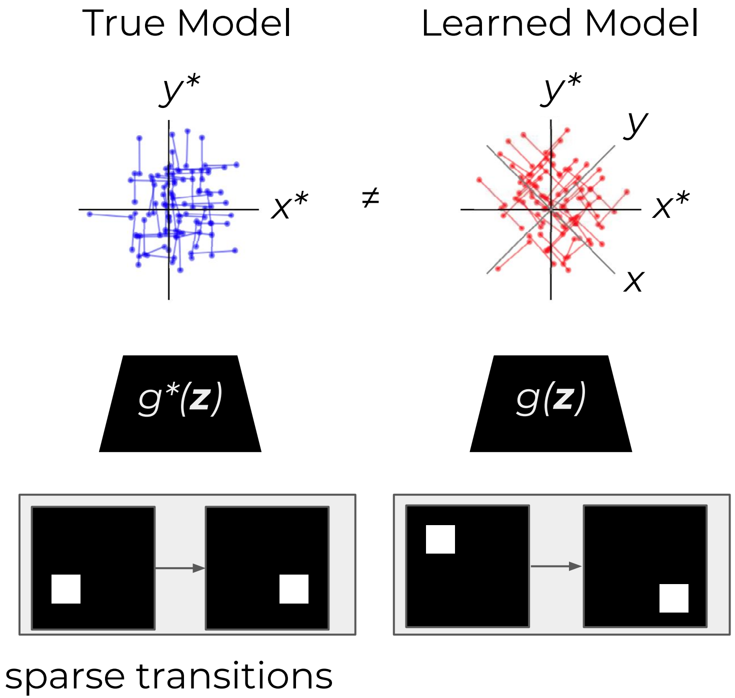


Identifiability Proof Intuition

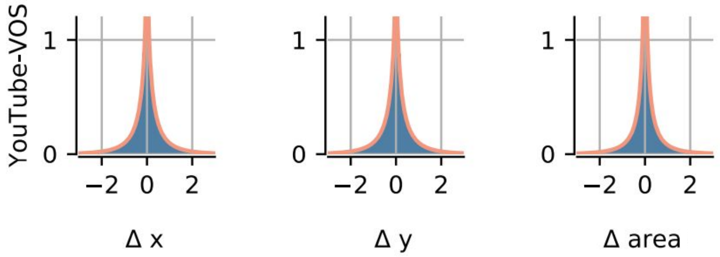


Prior:
objects in nature change sparsely

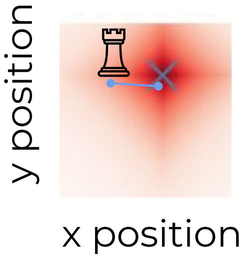


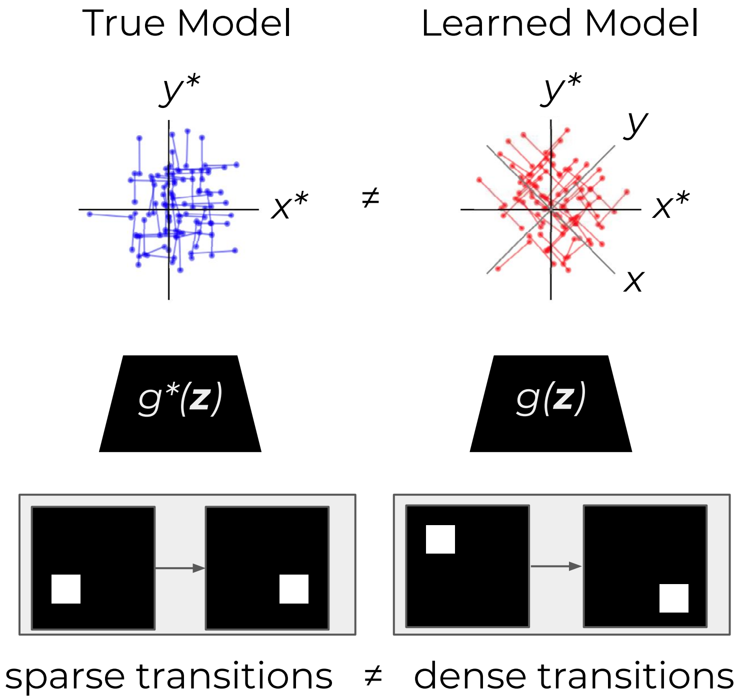


Identifiability Proof Intuition

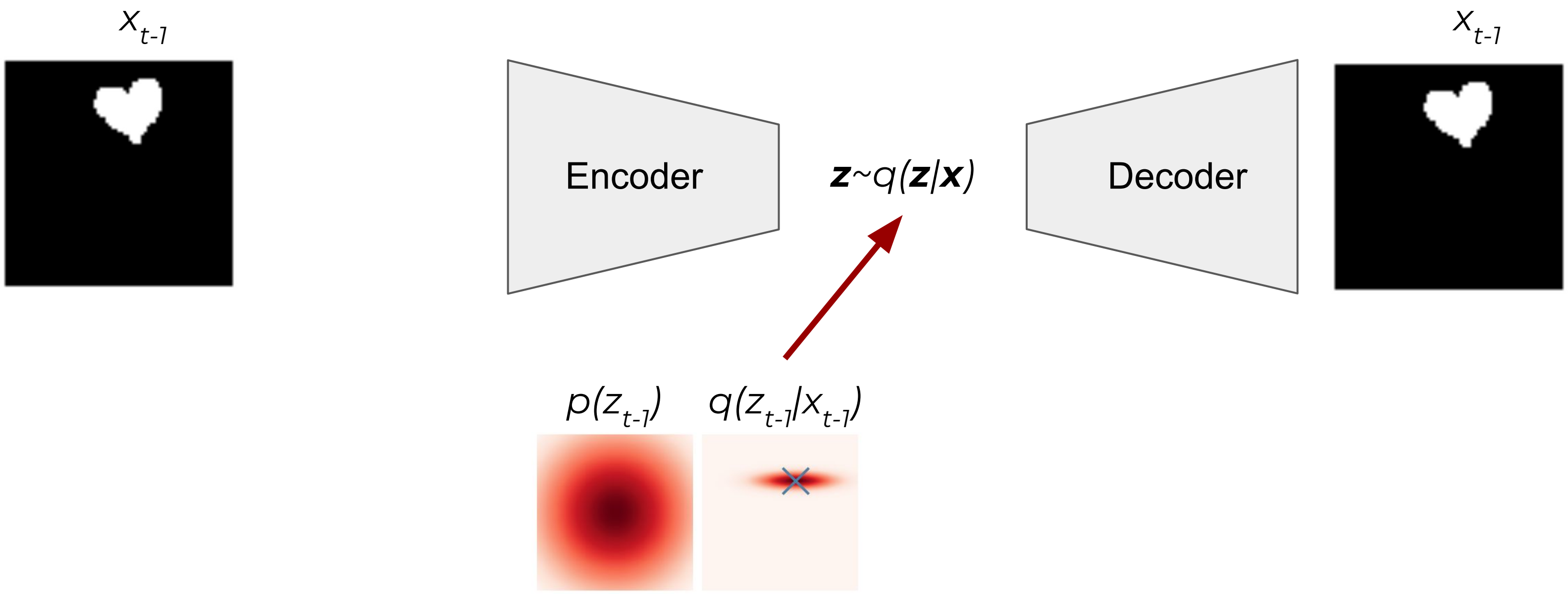


Prior:
objects in nature change sparsely

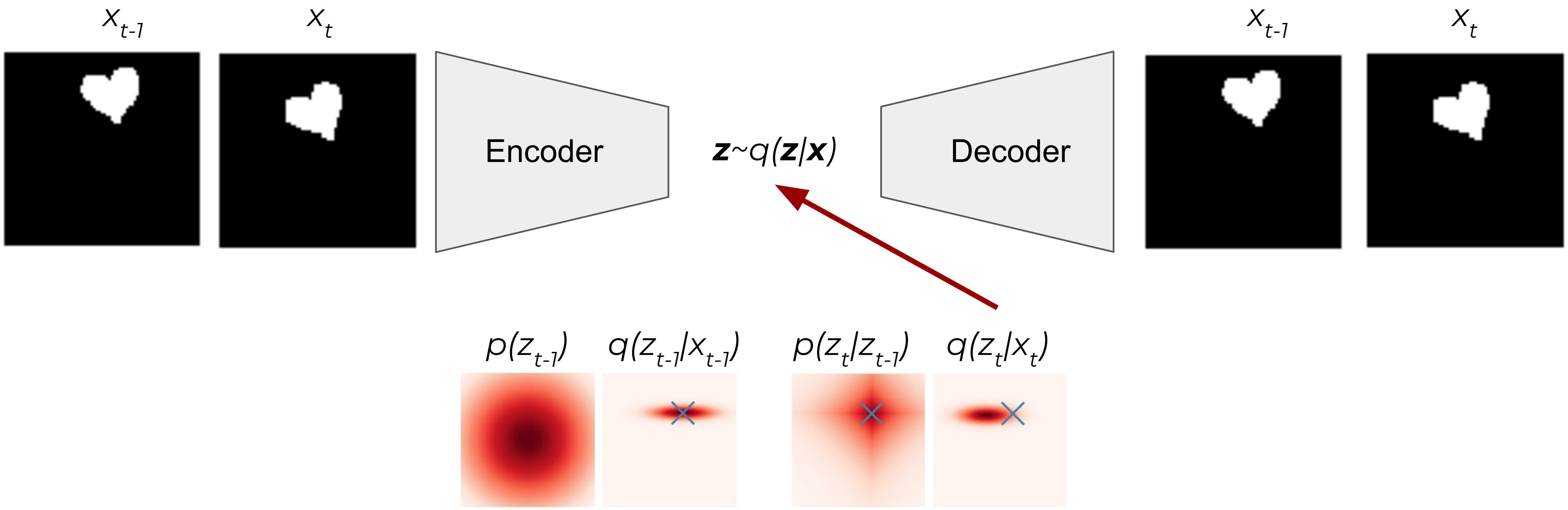




Slow Variational Autoencoder at time *t-1*



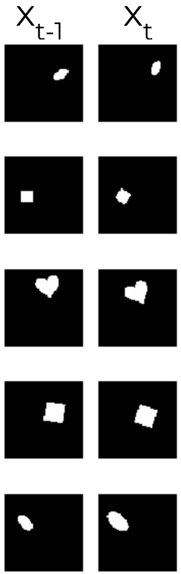
Slow Variational Autoencoder at time *t-1*



Quantitative Results - dSprites

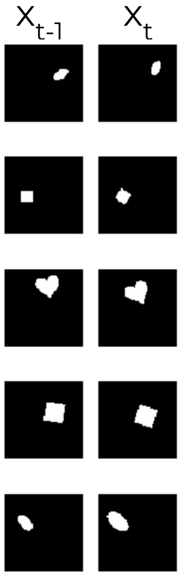
Quantitative Results - dSprites

Data



Quantitative Results - dSprites

Data



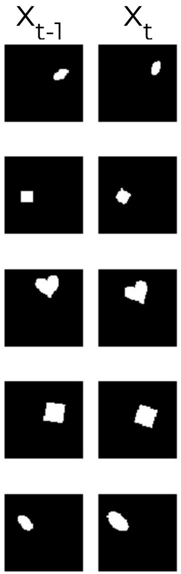
Ada-GVAE [Locatello et al. 19]

PCL [Hyvärinen et al. 17]

SlowVAE (ours)

Quantitative Results - dSprites

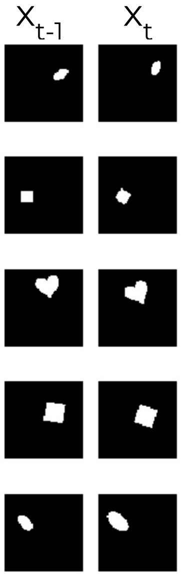
Data



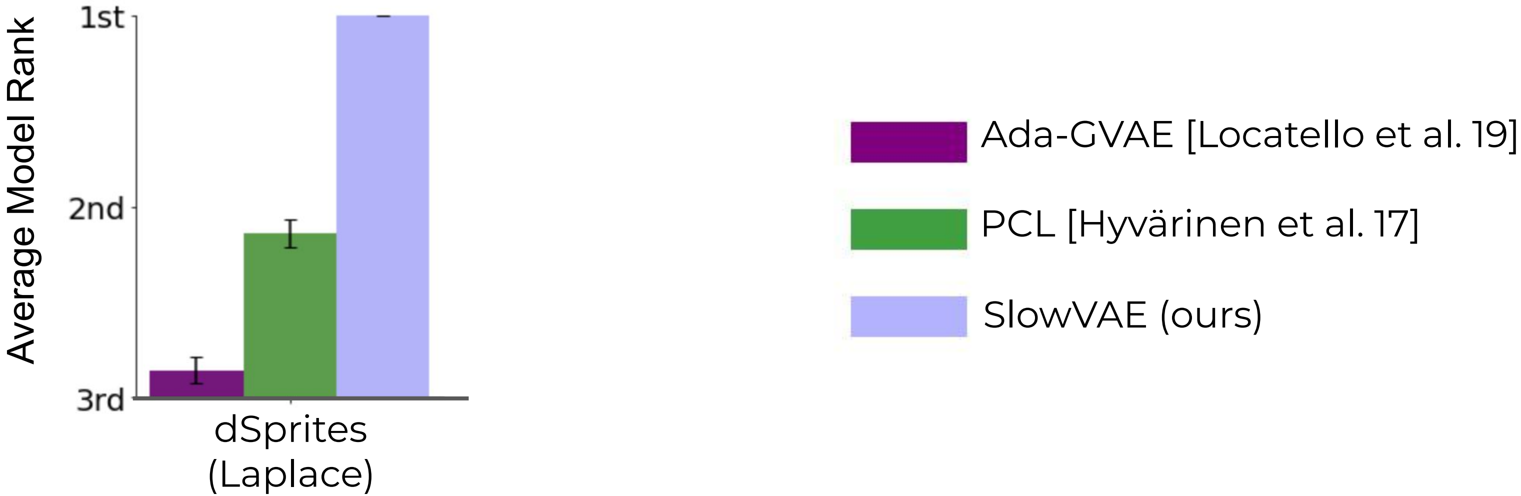
Disentanglement Performance

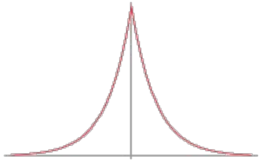
Quantitative Results - dSprites

Data



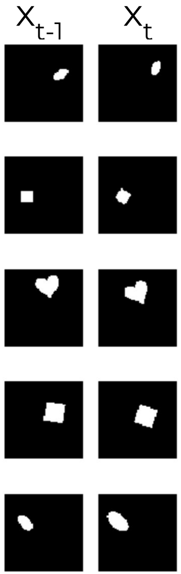
Disentanglement Performance



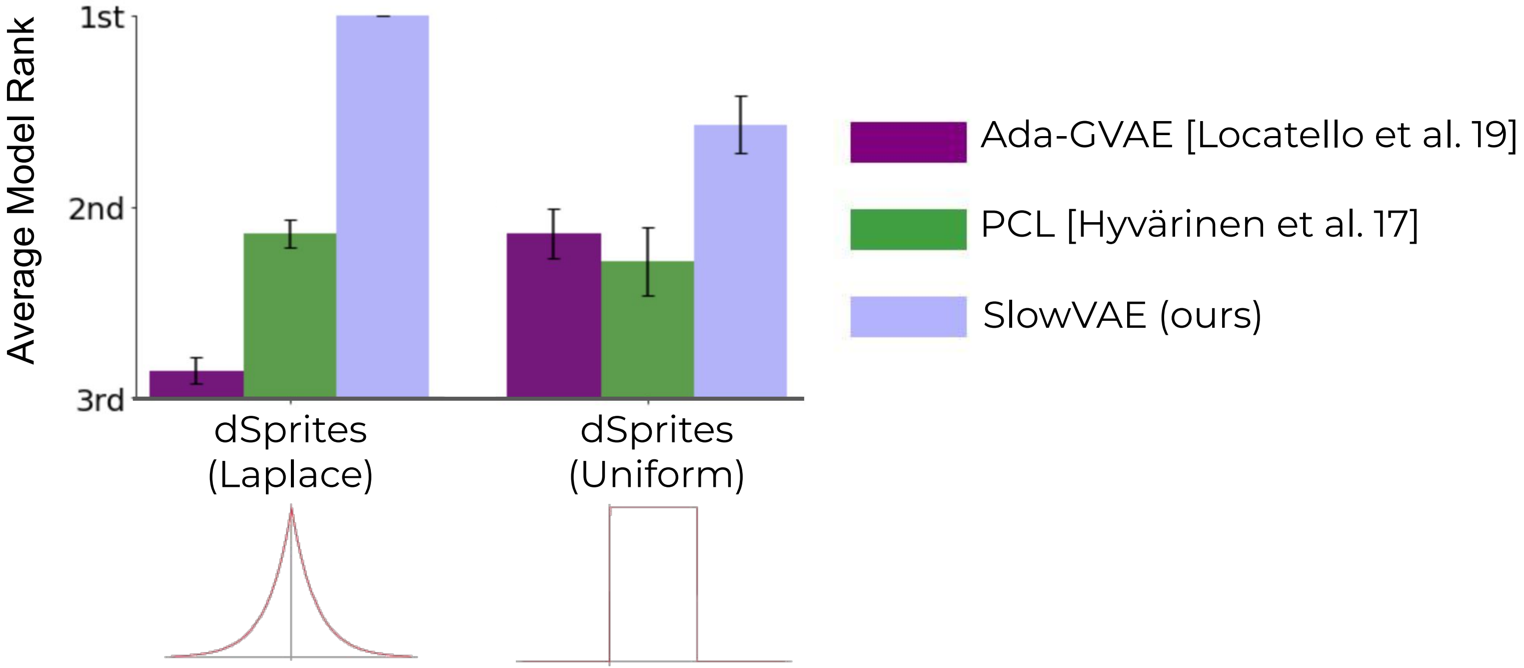


Quantitative Results - dSprites

Data

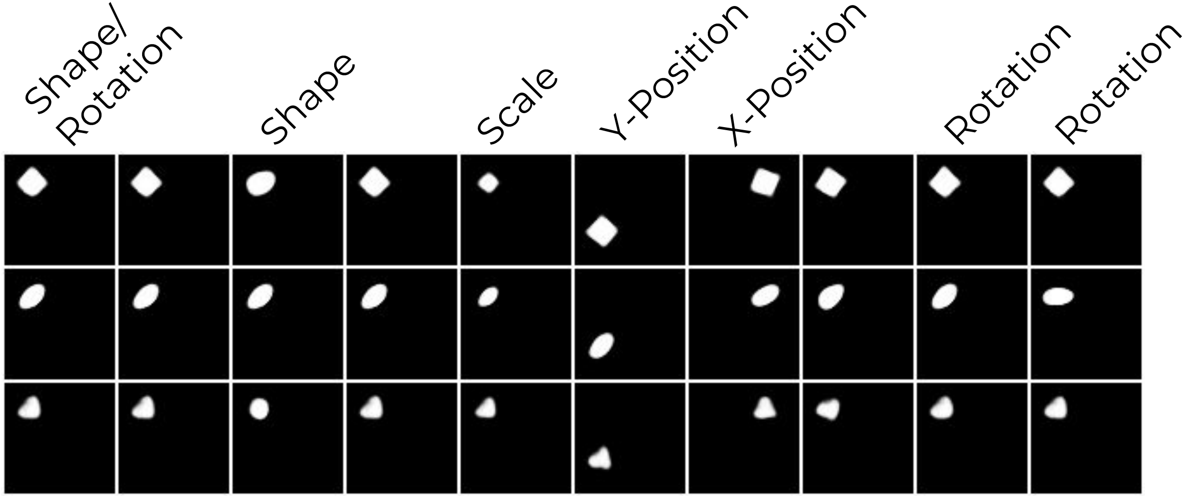


Disentanglement Performance

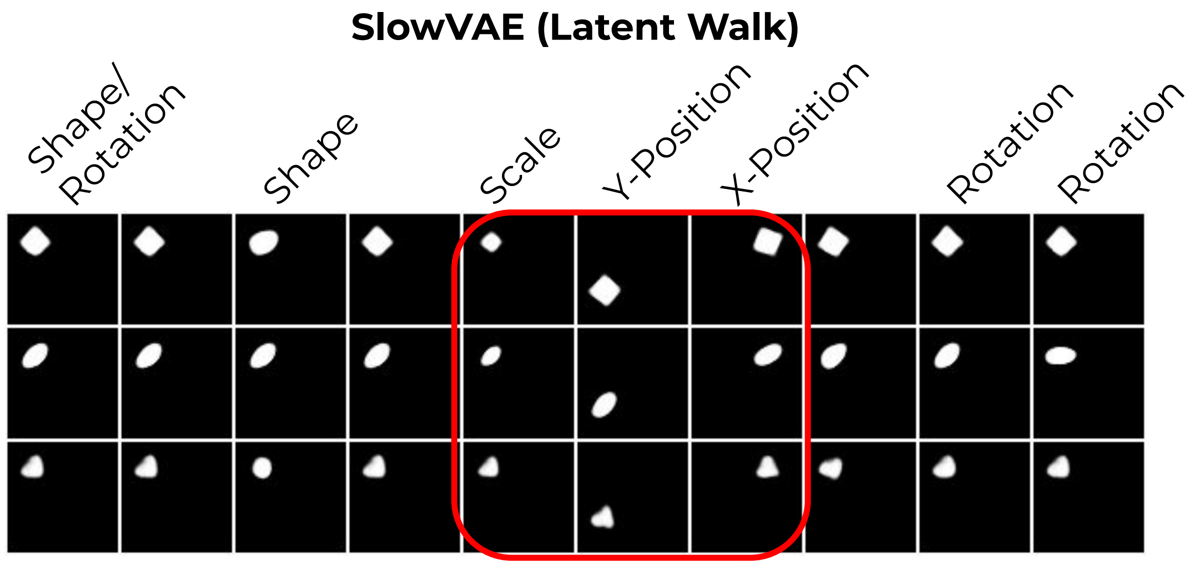


Qualitative Results - dSprites

SlowVAE (Latent Walk)

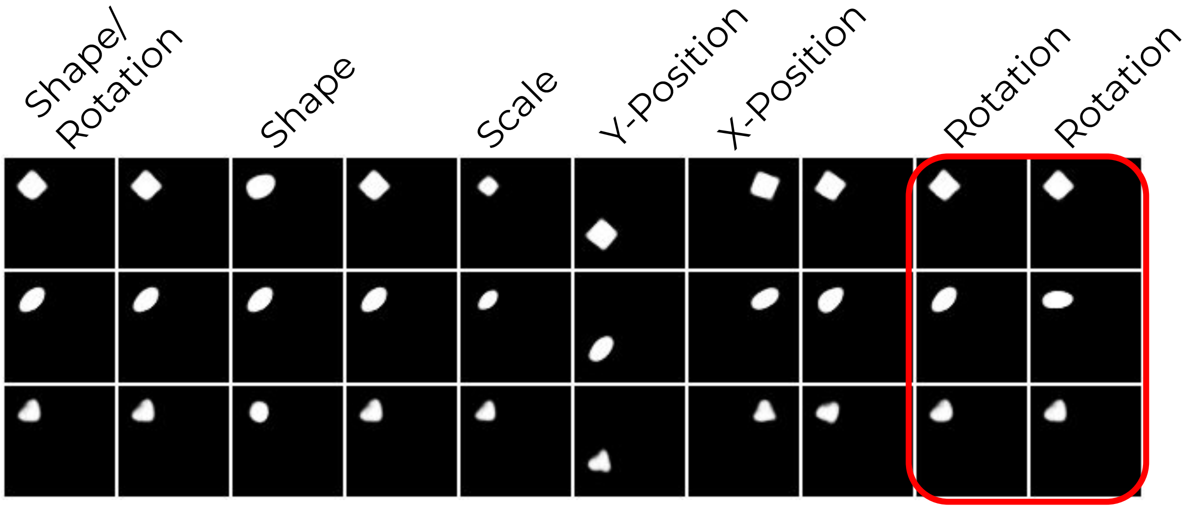


Qualitative Results - dSprites

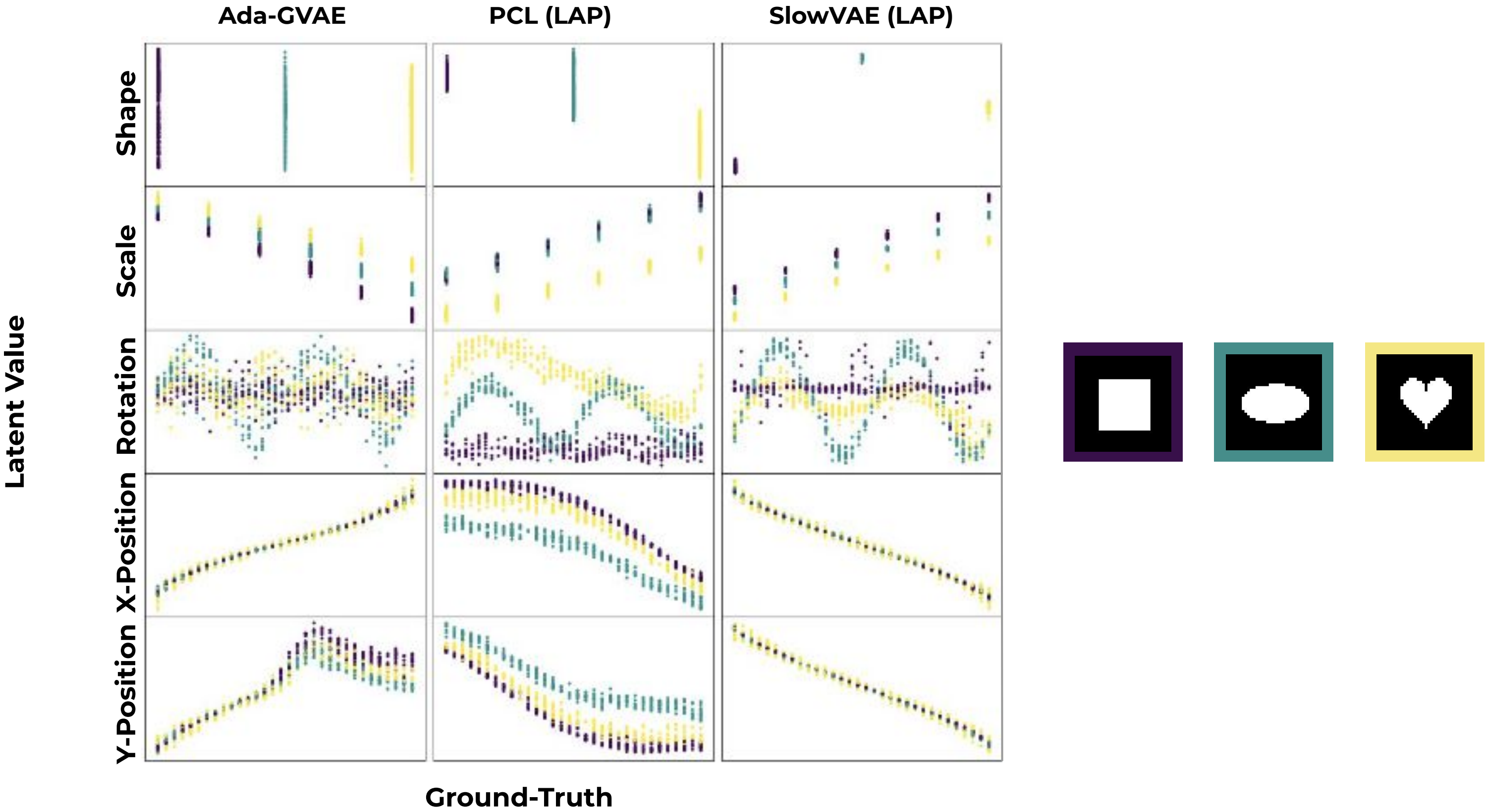


Qualitative Results - dSprites

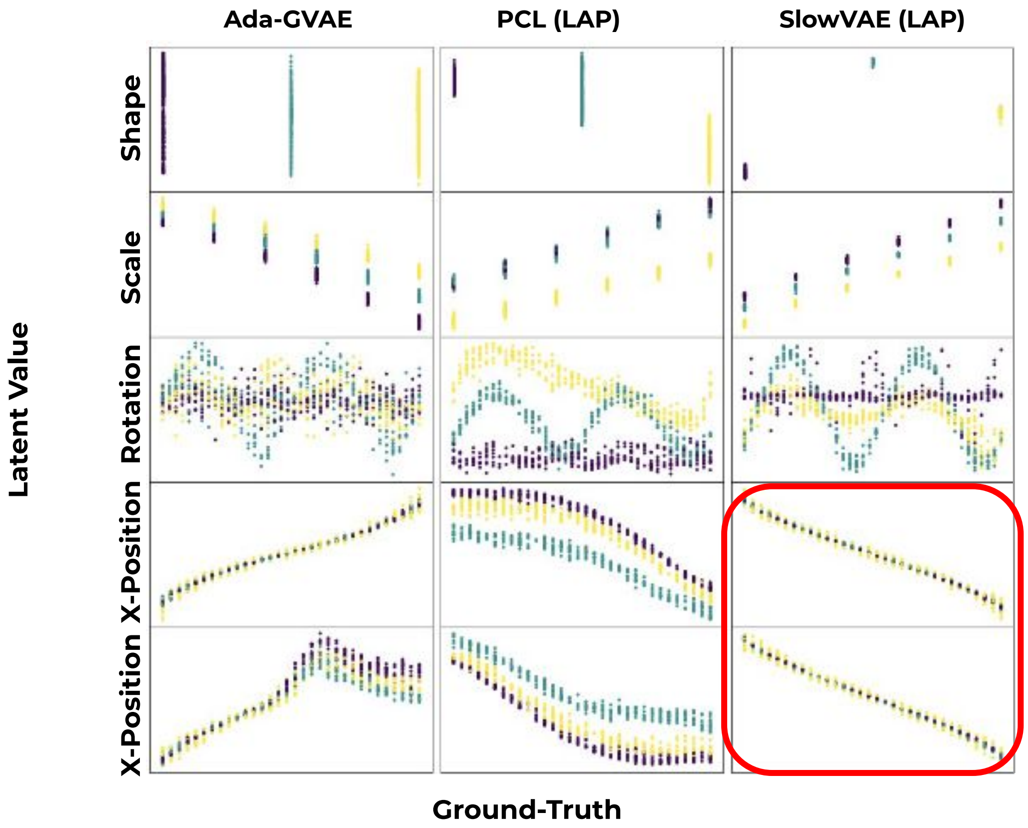
SlowVAE (Latent Walk)



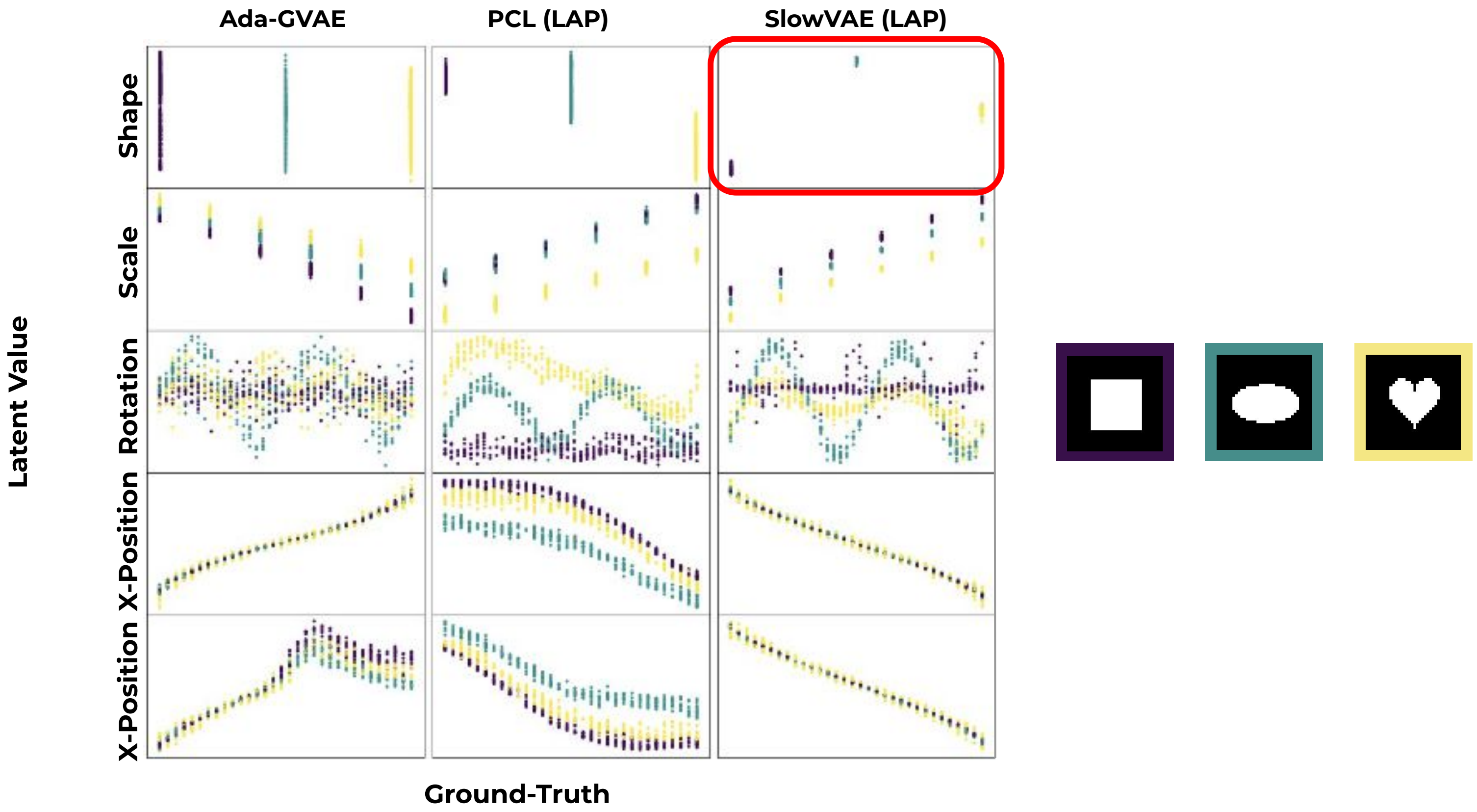
Failure Cases - dSprites



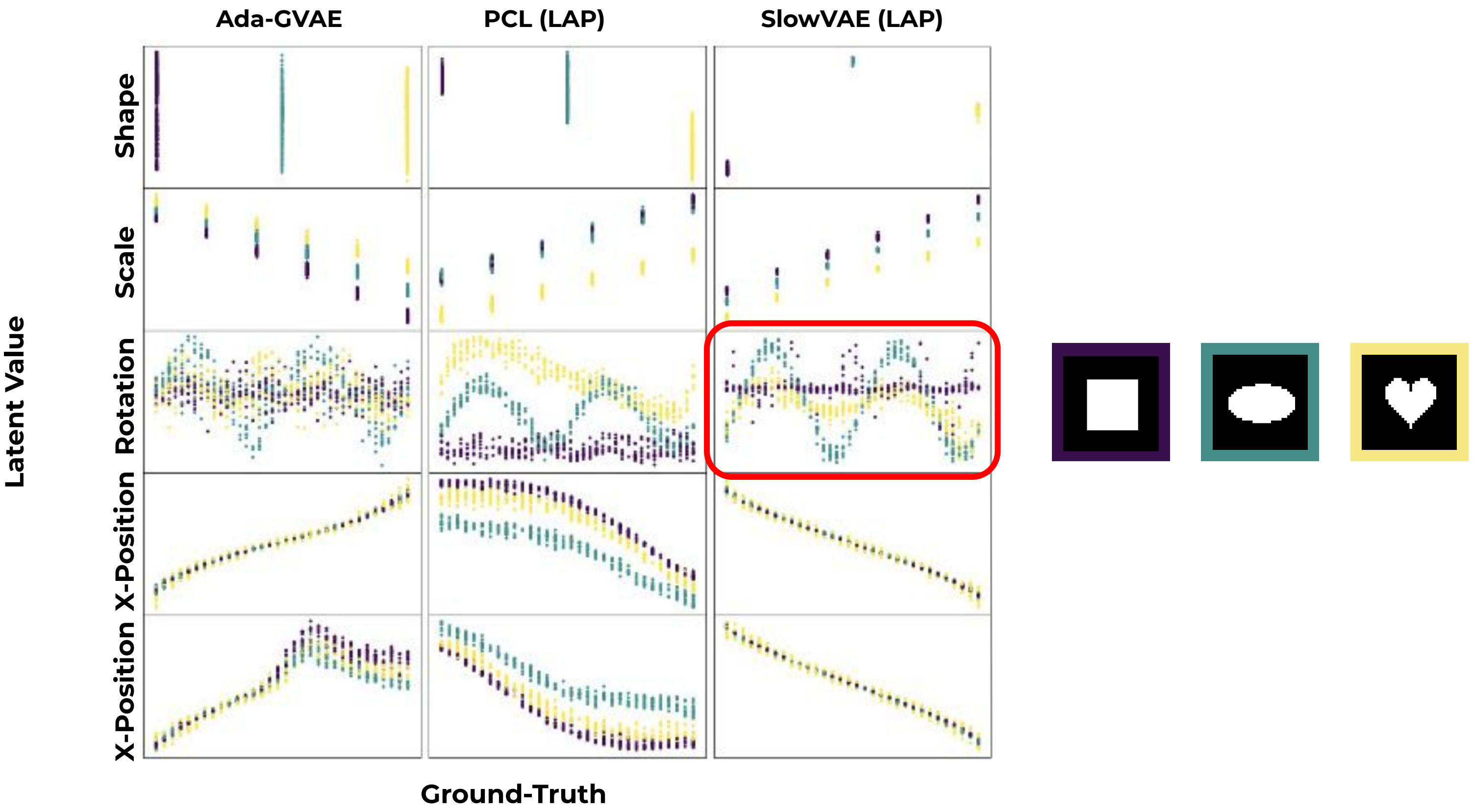
Failure Cases - dSprites



Failure Cases - dSprites

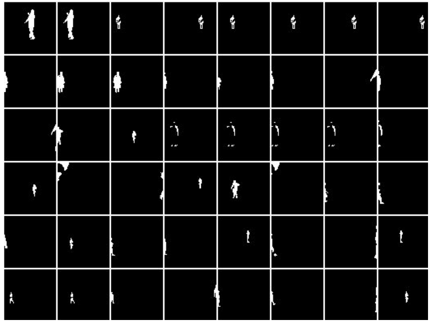


Failure Cases - dSprites



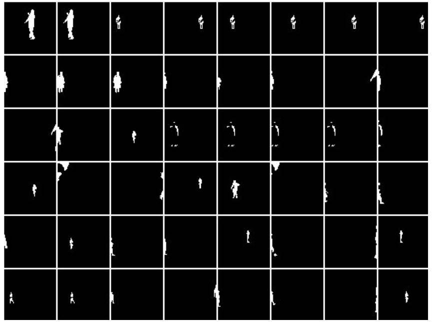
Results on Natural Data

KITTI-Masks

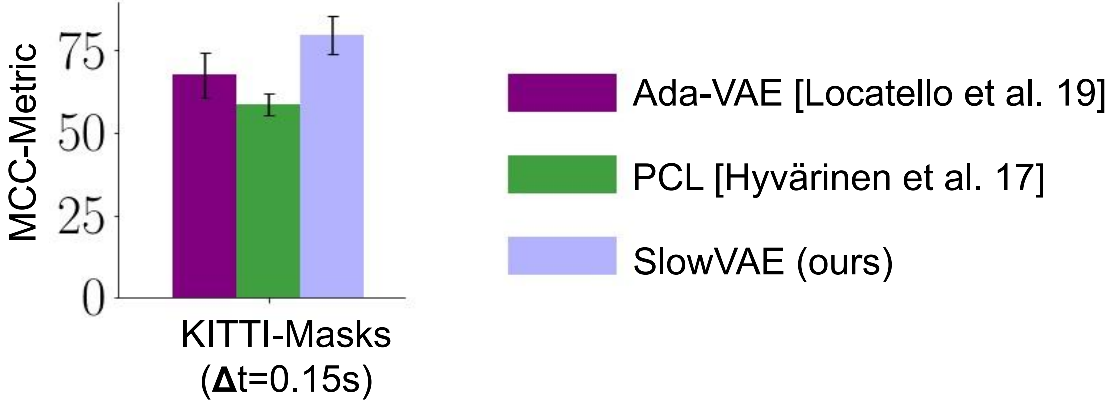


Results on Natural Data

KITTI-Masks

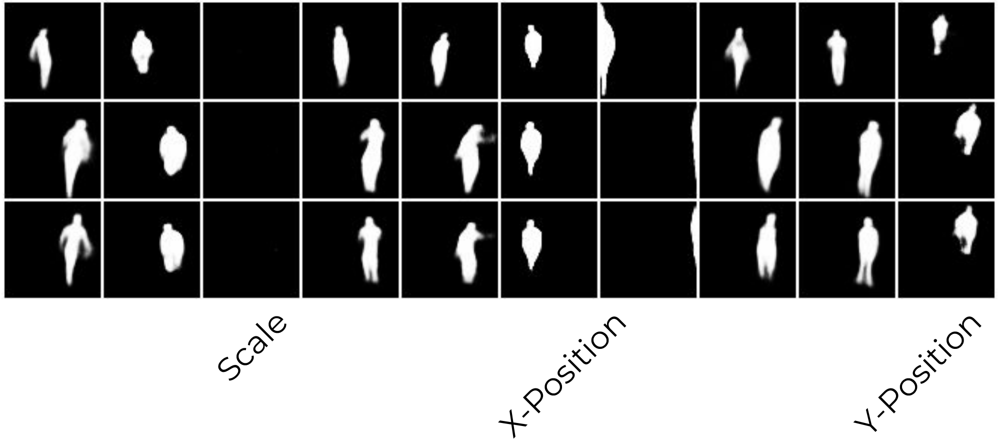


Disentanglement Performance



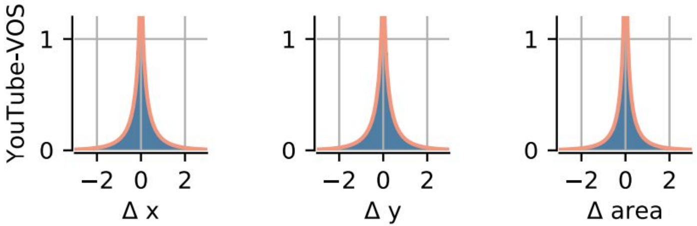
Results - KITTI-Masks

SlowVAE Latent Walk 🕺



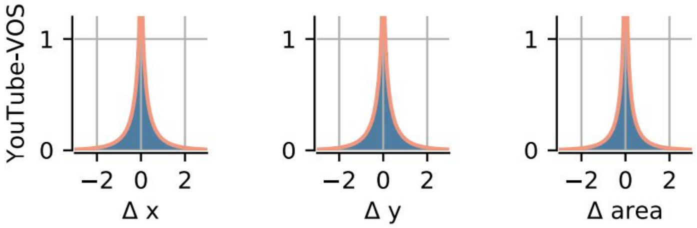
Paper Contributions

Objects in natural scenes have **sparse** marginal transition statistics

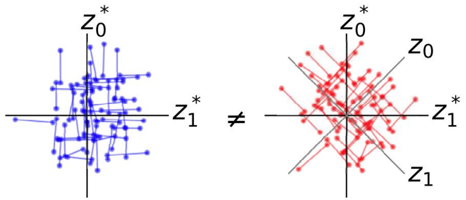


Paper Contributions

Objects in natural scenes have **sparse** marginal transition statistics

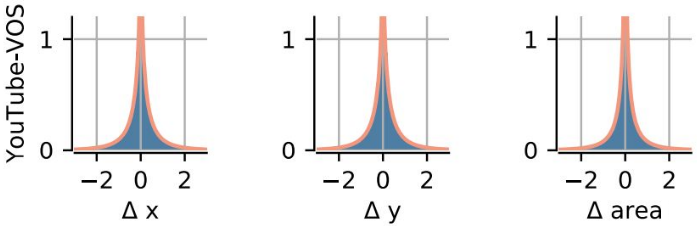


Intuitive proof for **identifiability** in Nonlinear ICA & Disentanglement

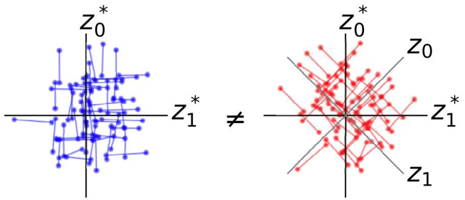


Paper Contributions

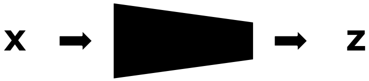
Objects in natural scenes have **sparse** marginal transition statistics



Intuitive proof for **identifiability** in Nonlinear ICA & Disentanglement

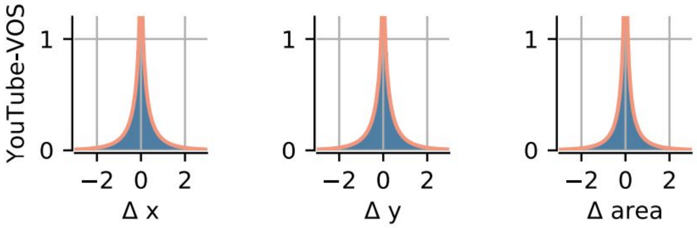


Empirical results using a **Flow** and a **VAE** based implementation of the theoretical model

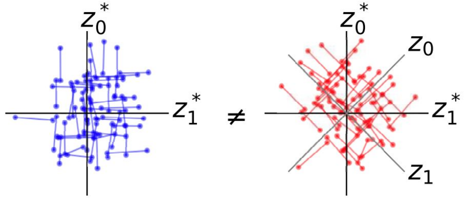


Paper Contributions

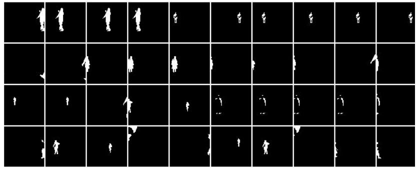
Objects in natural scenes have **sparse** marginal transition statistics

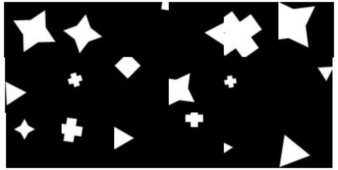


Intuitive proof for **identifiability** in Nonlinear ICA & Disentanglement

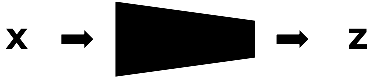


Two challenging new **datasets** to push disentanglement towards **natural** video





Empirical results using a **Flow** and a **VAE** based implementation of the theoretical model



Contrastive Learning Inverts the Data Generating Process

Overview

1. Theoretical Connection between InfoNCE & Nonlinear ICA

Overview

1. Theoretical Connection between InfoNCE & Nonlinear ICA
2. Empirical Test on robustness to mismatch (in assumptions)

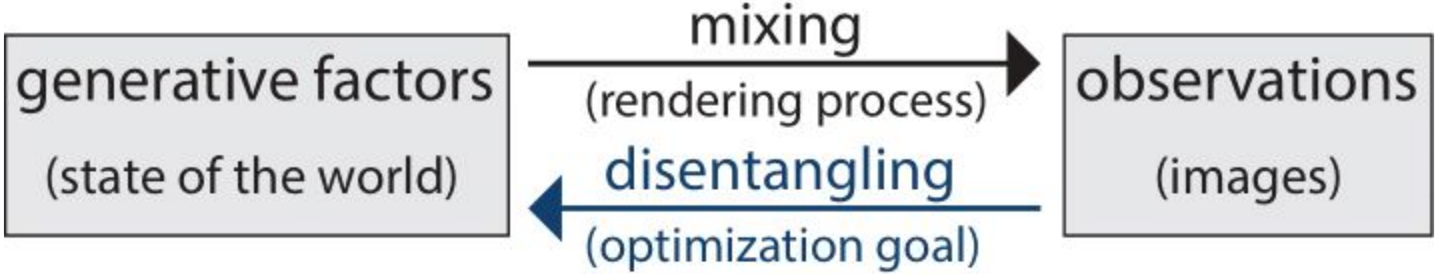
Overview

1. Theoretical Connection between InfoNCE & Nonlinear ICA
2. Empirical Test on robustness to mismatch (in assumptions)
3. Identifiability on 3DIdent

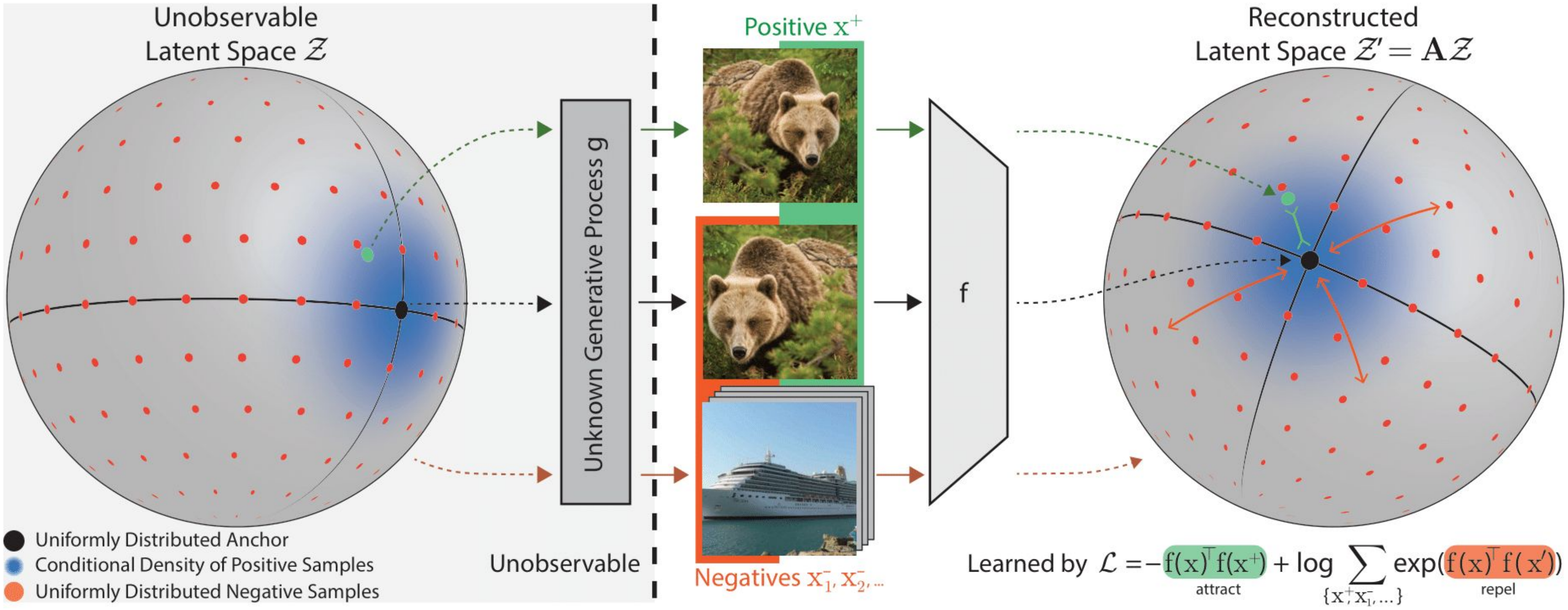
Overview

1. Theoretical Connection between InfoNCE & Nonlinear ICA
2. Empirical Test on robustness to mismatch (in assumptions)
3. Identifiability on 3DIdent
    a. complex, high-resolution images

Nonlinear ICA



Theoretical Framework



Theoretical Framework

Theorem 2. *Let $\mathcal{Z} = \mathbb{S}^{N-1}$, the ground-truth marginal be uniform, and the conditional a vMF distribution (cf. Eq. 2). Let the mixing function g be differentiable and injective. If the assumed form of q_{h}, as defined above, matches that of p, and if f is differentiable and minimizes the CL loss (1), then for fixed $\tau > 0$ and $M \to \infty$, $h = f \circ g$ is linear, i.e., f recovers the latent sources up to orthogonal linear transformations.*

$$\mathcal{L}_{\mathsf{contr}}(f;\tau,M) := \underset{\substack{(\mathbf{x},\tilde{\mathbf{x}})\sim p_{\mathsf{pos}} \\ \{\mathbf{x}_i^-\}_{i=1}^M \overset{\text{i.i.d.}}{\sim} p_{\mathsf{data}}}}{\mathbb{E}} \left[-\log \frac{e^{f(\mathbf{x})^\top f(\tilde{\mathbf{x}})/\tau}}{e^{f(\mathbf{x})^\top f(\tilde{\mathbf{x}})/\tau} + \sum_{i=1}^{M} e^{f(\mathbf{x}_i^-)^\top f(\tilde{\mathbf{x}})/\tau}} \right]. \tag{1}$$

$$p(\mathbf{z}) = |\mathcal{Z}|^{-1}, \qquad p(\mathbf{z}|\tilde{\mathbf{z}}) = C_p^{-1} e^{\kappa \mathbf{z}^\top \tilde{\mathbf{z}}}$$
$$\text{with} \quad C_p := \int e^{\kappa \mathbf{z}^\top \tilde{\mathbf{z}}} \, d\tilde{\mathbf{z}} = \text{const.}, \quad \mathbf{x} = g(\mathbf{z}). \tag{2}$$

$$q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z}) = C_h(\tilde{\mathbf{z}})^{-1} e^{h(\tilde{\mathbf{z}})^\top h(\mathbf{z})/\tau}$$
$$\textit{with} \quad C_h(\mathbf{z}) := \int e^{h(\tilde{\mathbf{z}})^\top h(\mathbf{z})/\tau} \, d\mathbf{z},$$

Theoretical Framework

Theorem 1 ($\mathcal{L}_{\mathsf{contr}}$ converges to the cross-entropy between latent distributions)**.** *If the ground-truth marginal distribution p is uniform, then for fixed $\tau > 0$, as the number of negative samples $M \to \infty$, the (normalized) contrastive loss converges to*

$$\lim_{M \to \infty} \mathcal{L}_{\mathsf{contr}}(f; \tau, M) - \log M + \log |\mathcal{Z}| = \mathop{\mathbb{E}}_{\mathbf{z} \sim p(\mathbf{z})} \left[H(p(\cdot|\mathbf{z}), q_h(\cdot|\mathbf{z})) \right] \tag{14}$$

where H is the cross-entropy between the ground-truth conditional distribution p over positive pairs and a conditional distribution q_{h} parameterized by the model f, and $C_h(\mathbf{z}) \in \mathbb{R}^+$ is the partition function of q_{h} (see Appendix A.1.1):

$$q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z}) = C_h(\mathbf{z})^{-1} e^{h(\tilde{\mathbf{z}})^\top h(\mathbf{z})/\tau}$$
$$\textit{with} \quad C_h(\mathbf{z}) := \int e^{h(\tilde{\mathbf{z}})^\top h(\mathbf{z})/\tau} \, \mathrm{d}\tilde{\mathbf{z}}. \tag{15}$$

Theoretical Framework

Proposition 1 (Minimizers of the cross-entropy maintain the dot product)**.** *Let* $\mathcal{Z} = \mathbb{S}^{N-1}$, $\tau > 0$ *and consider the ground-truth conditional distribution of the form* $p(\tilde{\mathbf{z}}|\mathbf{z}) = C_p^{-1}\exp(\kappa\tilde{\mathbf{z}}^\top\mathbf{z})$. *Let* h *map onto a hypersphere with radius* $\sqrt{\tau\kappa}$.[3] *Consider the conditional distribution* q_h *parameterized by the model, as defined above in Theorem 1, where the hypothesis class for* h *is assumed to be sufficiently flexible such that* $p(\tilde{\mathbf{z}}|\mathbf{z})$ *and* $q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z})$ *can match. If* h *is a minimizer of the cross-entropy* $\mathbb{E}_{p(\tilde{\mathbf{z}}|\mathbf{z})}[-\log q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z})]$, *then* $p(\tilde{\mathbf{z}}|\mathbf{z}) = q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z})$ *and* $\forall \mathbf{z}, \tilde{\mathbf{z}} : \kappa\mathbf{z}^\top\tilde{\mathbf{z}} = h(\mathbf{z})^\top h(\tilde{\mathbf{z}})$.

Proposition 2 (Extension of the Mazur-Ulam theorem to hyperspheres and the dot product)**.** *Let* $\mathcal{Z} = \mathbb{S}^{N-1}$. *If* $h : \mathcal{Z} \to \mathcal{Z}$ *maintains the dot product up to a constant factor, i.e.,* $\forall \mathbf{z}, \tilde{\mathbf{z}} \in \mathcal{Z} : \kappa\mathbf{z}^\top\tilde{\mathbf{z}} = h(\mathbf{z})^\top h(\tilde{\mathbf{z}})$, *then* h *is an orthogonal linear transformation.*

Theoretical Framework

Theorem 5. *Let $\mathcal{Z}$ be a convex body in $\mathbb{R}^N$, $h = f \circ g : \mathcal{Z} \to \mathcal{Z}$, and δ be a metric. Further, let the ground-truth marginal distribution be uniform and the conditional distribution be as (5). Let the mixing function g be differentiable and injective. If the assumed form of q_{h} matches that of p, i.e.,*

$$q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z}) = C_q^{-1}(\mathbf{z}) e^{-\delta(h(\tilde{\mathbf{z}}), h(\mathbf{z}))/\tau}$$
$$\textit{with} \quad C_q(\mathbf{z}) := \int e^{-\delta(h(\tilde{\mathbf{z}}), h(\mathbf{z}))/\tau} \, \mathrm{d}\tilde{\mathbf{z}}, \tag{7}$$

and if f is differentiable and minimizes the $\mathcal{L}_{\delta\text{-contr}}$ objective in (6) for $M \to \infty$, we find that $h = f \circ g$ is invertible and affine, i.e., we recover the latent sources up to affine transformations.

$$p(\mathbf{z}) = |\mathcal{Z}|^{-1}, \qquad p(\mathbf{z}|\tilde{\mathbf{z}}) = C_p^{-1} e^{-\delta(\mathbf{z}, \tilde{\mathbf{z}})}$$
$$\text{with} \quad C_p(\mathbf{z}) := \int e^{-\delta(\mathbf{z}, \tilde{\mathbf{z}})} \, \mathrm{d}\tilde{\mathbf{z}}, \quad \mathbf{x} = g(\mathbf{z}), \tag{5}$$

$$\mathcal{L}_{\delta\text{-contr}}(f; \tau, M) \quad := \tag{6}$$
$$\underset{\substack{(\mathbf{x}, \tilde{\mathbf{x}}) \sim p_{\mathrm{pos}} \\ \{\mathbf{x}_i^-\}_{i=1}^M \overset{\text{i.i.d.}}{\sim} p_{\mathrm{data}}}}{\mathbb{E}} \left[-\log \frac{e^{-\delta(f(\mathbf{x}), f(\tilde{\mathbf{x}}))/\tau}}{e^{-\delta(f(\mathbf{x}), f(\tilde{\mathbf{x}}))/\tau} + \sum_{i=1}^{M} e^{-\delta(f(\mathbf{x}_i^-), f(\tilde{\mathbf{x}}))/\tau}} \right].$$

Theoretical Framework

Theorem 3. *Let δ be a semi-metric and $\tau, \lambda > 0$ and let the ground-truth marginal distribution p be uniform. Consider a ground-truth conditional distribution $p(\tilde{\mathbf{z}}|\mathbf{z}) = C_p^{-1}(\mathbf{z}) \exp(-\lambda\delta(\tilde{\mathbf{z}}, \mathbf{z}))$ and the model conditional distribution*

$$q_{\mathrm{h}}(\tilde{\mathbf{z}}|\mathbf{z}) = C_h^{-1}(\mathbf{z}) e^{-\delta(h(\tilde{\mathbf{z}}), h(\mathbf{z}))/\tau}$$
$$\text{with} \quad C_h(\mathbf{z}) := \int_{\mathcal{Z}} e^{-\delta(h(\tilde{\mathbf{z}}), h(\mathbf{z}))/\tau} \mathrm{d}\tilde{\mathbf{z}}. \tag{61}$$

Then the cross-entropy between p and q_{h} is given by

$$\lim_{M\to\infty} \mathcal{L}_{\delta\text{-contr}}(f; \tau, M) - \log M + \log |\mathcal{Z}| = \underset{\mathbf{z}\sim p(\mathbf{z})}{\mathbb{E}} \left[H(p(\cdot|\mathbf{z}), q_{\mathrm{h}}(\cdot|\mathbf{z})\right], \tag{62}$$

which can be implemented by sampling data from the accessible distributions.

Theorem 4. *Let $\mathcal{Z} = \mathcal{Z}'$ be a convex body in $\mathbb{R}^N$. Let the mixing function g be differentiable and invertible. If the assumed form of q_{h} as defined in (4) matches that of p, and if f is differentiable and minimizes the cross-entropy between p and q_{h}, then we find that $h = f \circ g$ is affine, i.e., we recover the latent sources up to affine transformations.*

Theoretical Framework

Theorem 6. *Let $\mathcal{Z}$ be a convex body in $\mathbb{R}^N$, $h : \mathcal{Z} \to \mathcal{Z}$, and δ be an L^α metric for $\alpha \geq 1, \alpha \neq 2$. Further, let the ground-truth marginal distribution be uniform and the conditional distribution be as (5), and let the mixing function g be differentiable and invertible. If the assumed form of $q_{\mathrm{h}}(\cdot|\mathbf{z})$ matches that of $p(\cdot|\mathbf{z})$, i.e., both use the same metric δ up to a constant scaling factor, and if f is differentiable and minimizes the $\mathcal{L}_{\delta\text{-contr}}$ objective in (6) for $M \to \infty$, we find that $h = f \circ g$ is a composition of input independent permutations, sign flips and rescaling.*

Theorem D. *Suppose $1 \leq \alpha \leq \infty$ and $\alpha \neq 2$. An $n \times n$ matrix $\mathbf{A}$ is an isometry of L^α-norm if and only if $\mathbf{A}$ is a generalized permutation matrix, i.e., $\forall \mathbf{z} : (\mathbf{A}\mathbf{z})_{\mathbf{i}} = \alpha_{\mathbf{i}} \mathbf{z}_{\sigma(\mathbf{i})}$, with $\alpha_i = \pm 2$ and σ being a permutation.*

Proof. See Li & So (1994). Note that this can also be concluded from the Banach-Lamperti Theorem (Lamperti et al., 1958). □

$$p(\mathbf{z}) = |\mathcal{Z}|^{-1}, \qquad p(\mathbf{z}|\tilde{\mathbf{z}}) = C_p^{-1} e^{-\delta(\mathbf{z},\tilde{\mathbf{z}})}$$

$$\text{with} \quad C_p(\mathbf{z}) := \int e^{-\delta(\mathbf{z},\tilde{\mathbf{z}})} \, \mathrm{d}\tilde{\mathbf{z}}, \quad \mathbf{x} = g(\mathbf{z}), \tag{5}$$

$$\mathcal{L}_{\delta\text{-contr}}(f; \tau, M) := \tag{6}$$

$$\mathop{\mathbb{E}}_{\substack{(\mathbf{x},\tilde{\mathbf{x}}) \sim p_{\text{pos}} \\ \{\mathbf{x}_i^-\}_{i=1}^M \overset{\text{i.i.d.}}{\sim} p_{\text{data}}}} \left[-\log \frac{e^{-\delta(f(\mathbf{x}),f(\tilde{\mathbf{x}}))/\tau}}{e^{-\delta(f(\mathbf{x}),f(\tilde{\mathbf{x}}))/\tau} + \sum_{i=1}^{M} e^{-\delta(f(\mathbf{x}_i^-),f(\tilde{\mathbf{x}}))/\tau}} \right].$$

Different Assumptions, Different Losses

$\bigcirc$ + vMF $\longrightarrow \exp(f(x)^\top f(x'))$

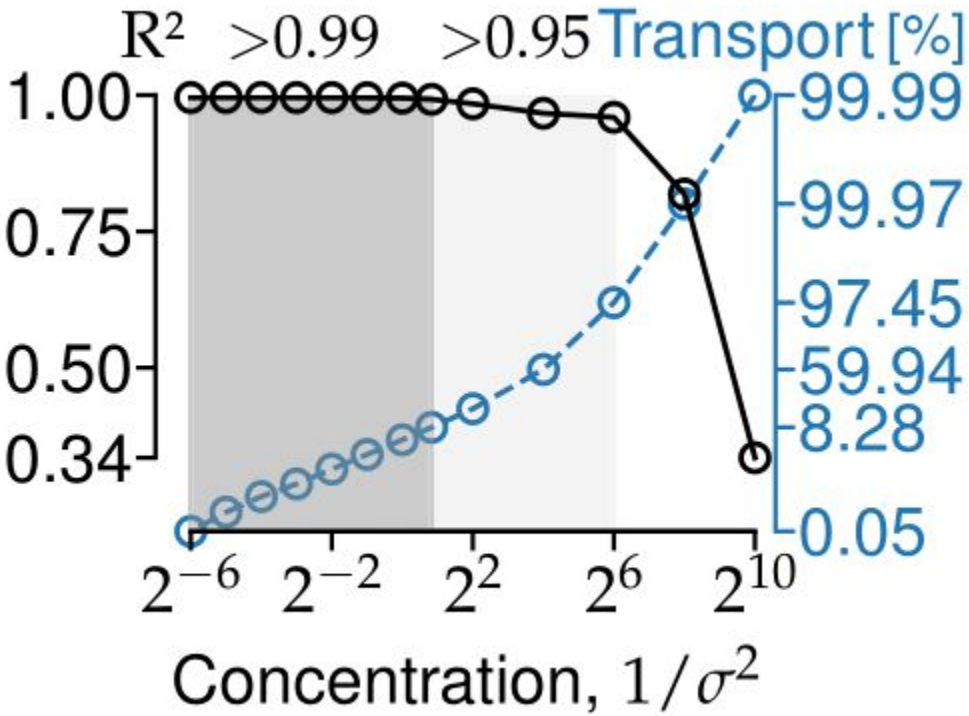
$\square$ + Normal $\longrightarrow \exp(-\|f(x) - f(x')\|_2)$

$\square$ + Laplace $\longrightarrow \exp(-\|f(x) - f(x')\|_1)$

Empirical Results

Generative process g			Model f			R^2 Score [%]		
Space	$p(\cdot)$	$p(\cdot\|\cdot)$	Space	$q_{\mathrm{h}}(\cdot\|\cdot)$	M.	Identity	Supervised	Unsupervised
Sphere	Uniform	vMF(κ=1)	Sphere	vMF(κ=1)	✓	66.98 ± 2.79	99.71 ± 0.05	99.42 ± 0.05
Sphere	Uniform	vMF(κ=10)	Sphere	vMF(κ=1)	✗	—〃—	—〃—	99.86 ± 0.01
Sphere	Uniform	Laplace(λ=0.05)	Sphere	vMF(κ=1)	✗	—〃—	—〃—	99.91 ± 0.01
Sphere	Uniform	Normal(σ=0.05)	Sphere	vMF(κ=1)	✗	—〃—	—〃—	99.86 ± 0.00
Box	Uniform	Normal(σ=0.05)	Unbounded	Normal	✗	67.93 ± 7.40	99.78 ± 0.06	99.60 ± 0.02
Box	Uniform	Laplace(λ=0.05)	Unbounded	Normal	✗	—〃—	—〃—	99.64 ± 0.02
Box	Uniform	Laplace(λ=0.05)	Unbounded	GenNorm(β=3)	✗	—〃—	—〃—	99.70 ± 0.02
Box	Uniform	Normal(σ=0.05)	Unbounded	GenNorm(β=3)	✗	—〃—	—〃—	99.69 ± 0.02
Sphere	Normal(σ=1)	Laplace(λ=0.05)	Sphere	vMF(κ=1)	✗	63.37 ± 2.41	99.70 ± 0.07	99.02 ± 0.01
Sphere	Normal(σ=1)	Normal(σ=0.05)	Sphere	vMF(κ=1)	✗	—〃—	—〃—	99.02 ± 0.02
Unbounded	Laplace(λ=1)	Normal(σ=1)	Unbounded	Normal	✗	62.49 ± 1.65	99.65 ± 0.04	98.13 ± 0.14
Unbounded	Normal(σ=1)	Normal(σ=1)	Unbounded	Normal	✗	63.57 ± 2.30	99.61 ± 0.17	98.76 ± 0.03

Empirical Results



Empirical Results

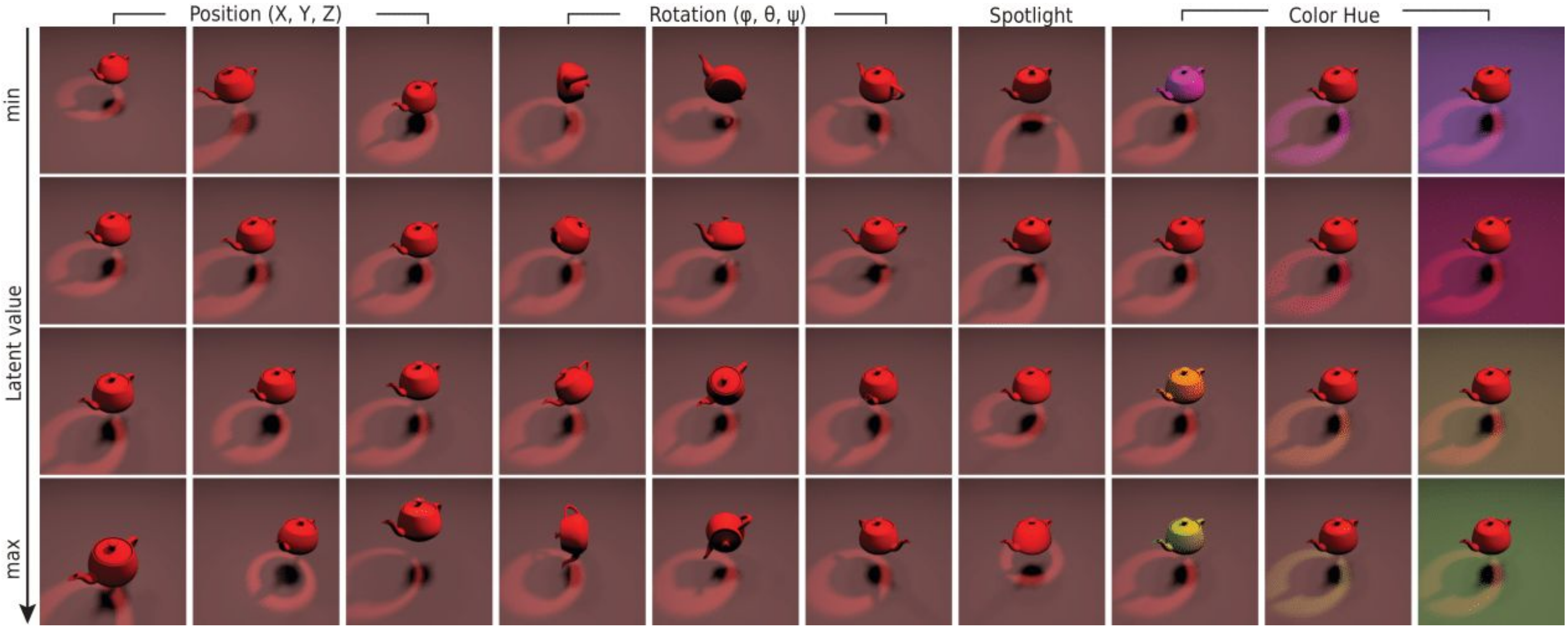
Generative process g			Model f			MCC Score [%]		
Space	$p(\cdot)$	$p(\cdot\|\cdot)$	Space	$q_{\mathrm{h}}(\cdot\|\cdot)$	M.	Identity	Supervised	Unsupervised
Box	Uniform	Laplace(λ=0.05)	Box	Laplace	✓	46.55 ± 1.34	99.93 ± 0.03	98.62 ± 0.05
Box	Uniform	GenNorm(β=3; λ=0.05)	Box	GenNorm(β=3)	✓	—"—	—"—	99.90 ± 0.06
Box	Uniform	Normal(σ=0.05)	Box	Normal	✗	—"—	—"—	99.77 ± 0.01
Box	Uniform	Laplace(λ=0.05)	Box	Normal	✗	—"—	—"—	99.76 ± 0.02
Box	Uniform	GenNorm(β=3; λ=0.05)	Box	Laplace	✗	—"—	—"—	98.80 ± 0.02
Box	Uniform	Laplace(λ=0.05)	Unbounded	Laplace	✗	—"—	99.97 ± 0.03	98.57 ± 0.02
Box	Uniform	GenNorm(β=3; λ=0.05)	Unbounded	GenNorm(β=3)	✗	—"—	—"—	99.85 ± 0.01
Box	Uniform	Normal(σ=0.05)	Unbounded	Normal	✗	—"—	—"—	58.26 ± 3.00
Box	Uniform	Laplace(λ=0.05)	Unbounded	Normal	✗	—"—	—"—	59.67 ± 2.33
Box	Uniform	Normal(σ=0.05)	Unbounded	GenNorm(β=3)	✗	—"—	—"—	43.80 ± 2.15

KITTI Masks

Table 3. **KITTI Masks**. Mean $\pm$ standard deviation over 10 random seeds. $\overline{\Delta t}$ indicates the average temporal distance of frames used.

	Model	Model Space	MCC [%]
$\overline{\Delta t} = 0.05s$	SlowVAE	Unbounded	66.1 ± 4.5
	Laplace	Unbounded	77.1 ± 1.0
	Laplace	Box	74.1 ± 4.4
	Normal	Unbounded	58.3 ± 5.4
	Normal	Box	59.9 ± 5.5
$\overline{\Delta t} = 0.15s$	SlowVAE	Unbounded	79.6 ± 5.8
	Laplace	Unbounded	79.4 ± 1.9
	Laplace	Box	80.9 ± 3.8
	Normal	Unbounded	60.2 ± 8.7
	Normal	Box	68.4 ± 6.7

3DIdent


Position (X, Y, Z)
Rotation (φ, θ, ψ)
Spotlight
Color Hue
min
Latent value
max

3DIdent

Dataset	Model f			Identity [%]	Unsupervised [%]	
$p(\cdot\|\cdot)$	Space	$q_{\mathrm{h}}(\cdot\|\cdot)$	M.	R^2	R^2	MCC
Normal	Box	Normal	✓	5.25 ± 1.20	96.73 ± 0.10	98.31 ± 0.04
Normal	Unbounded	Normal	✗	—″—	96.43 ± 0.03	54.94 ± 0.02
Laplace	Box	Normal	✗	—″—	96.87 ± 0.08	98.38 ± 0.03
Normal	Sphere	vMF	✗	—″—	65.74 ± 0.01	42.44 ± 3.27

Ongoing Work

Ongoing Work

1. Extend framework to object-centric methods
    a. MONet, IODINE, Slot Attention etc.

Ongoing Work

1. Extend framework to object-centric methods
    a. MONet, IODINE, Slot Attention etc.
2. Extend framework to data augmentations
    a. Content & Style Disambiguation
    b. Invariant factors == delta conditional

Self-supervised learning with data augmentations provably isolates content from style

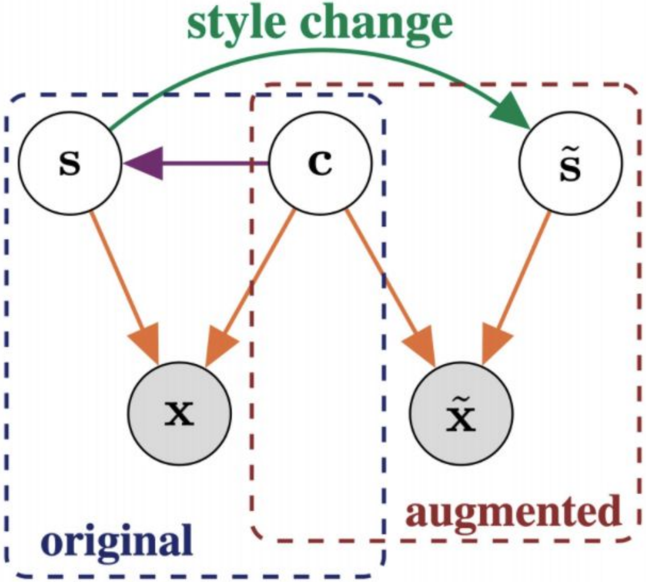
with **Julius von Kügelgen*, Yash Sharma*, Luigi Gresele***, Wieland Brendel, Michel Besserve, Francesco Locatello

Formalise generation $x = f(z)$ and augmentation $\tilde{x} = f(\tilde{z})$ processes as latent variable model with a content-style partition $z = (c, s)$:

- *invariant content* c*:* always shared between pairs $(x, \tilde{x})$ of views;
- *varying style* s*:* may change across pairs $(x, \tilde{x})$ of views.

Allow causal dependence of style on content (*Causal3DIdent* dataset):

augmented view $\tilde{x}$ = counterfactual under soft style intervention on x.



Theory: Can identify* invariant content partition in generative and discriminative learning with entropy maximisation (e.g., SimCLR).

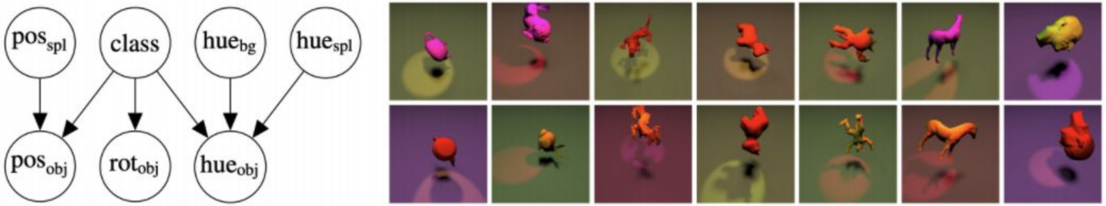


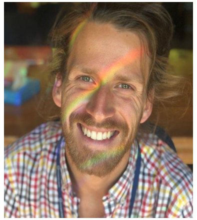
Figure 2: *(Left)* Causal graph for the *Causal3DIdent* dataset. *(Right)* Two samples from each object class.

*up to invertible transformation

Ongoing Work

1. Extend framework to object-centric methods
    a. MONet, IODINE, Slot Attention etc.
2. Extend framework to data augmentations
    a. Content & Style Disambiguation
    b. Invariant factors == delta conditional
3. Extend framework for causal discovery
    a. Robustness in downstream tasks?

Thank you for your attention!


David Klindt


Lukas Schott


Roland Zimmermann


Steffen Schneider


Ivan Ustyuzhaninov


Wieland Brendel


Matthias Bethge


Dylan Paiton

Funding: imprs-is TÜBINGEN AI CENTER DFG

If you are interested in this research, feel free to reach out!

Mail: ysharma1126@gmail.com
Twitter: @yash_j_sharma