Representation Learning with Deep Generative Models

Petra Poklukar
50% Seminar

Outline

- Representation learning with deep generative models
- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Compact Task Embeddings
- Theory
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

past

present
&
future

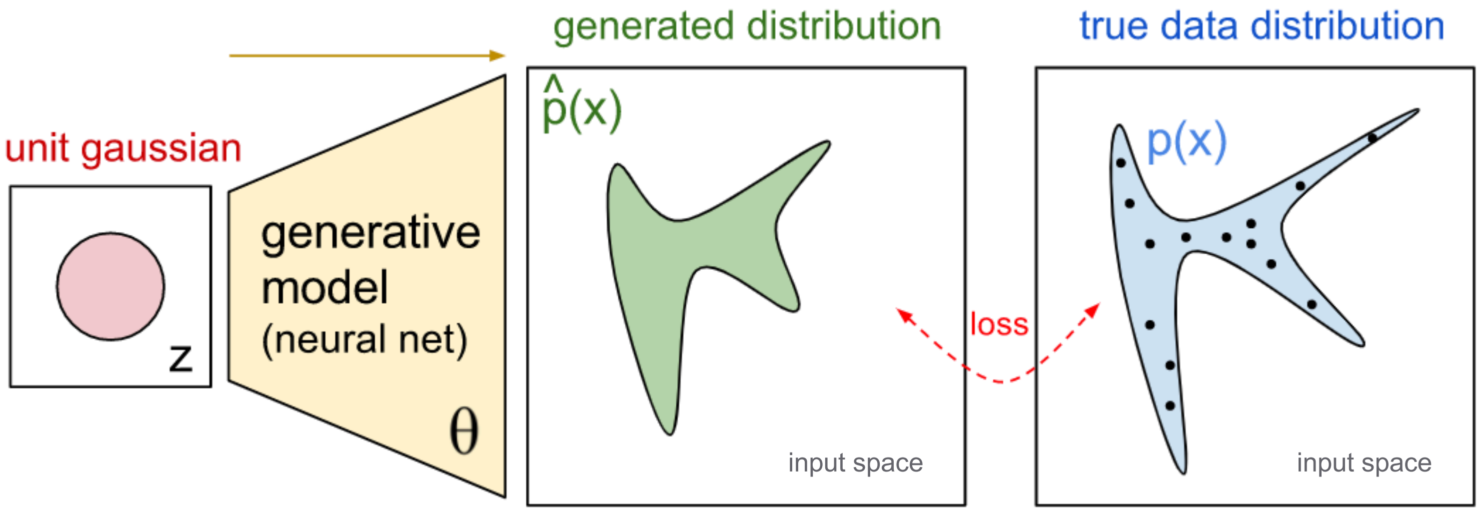
Outline

- **Representation learning with deep generative models**
- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Compact Task Embeddings
- Theory
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

past

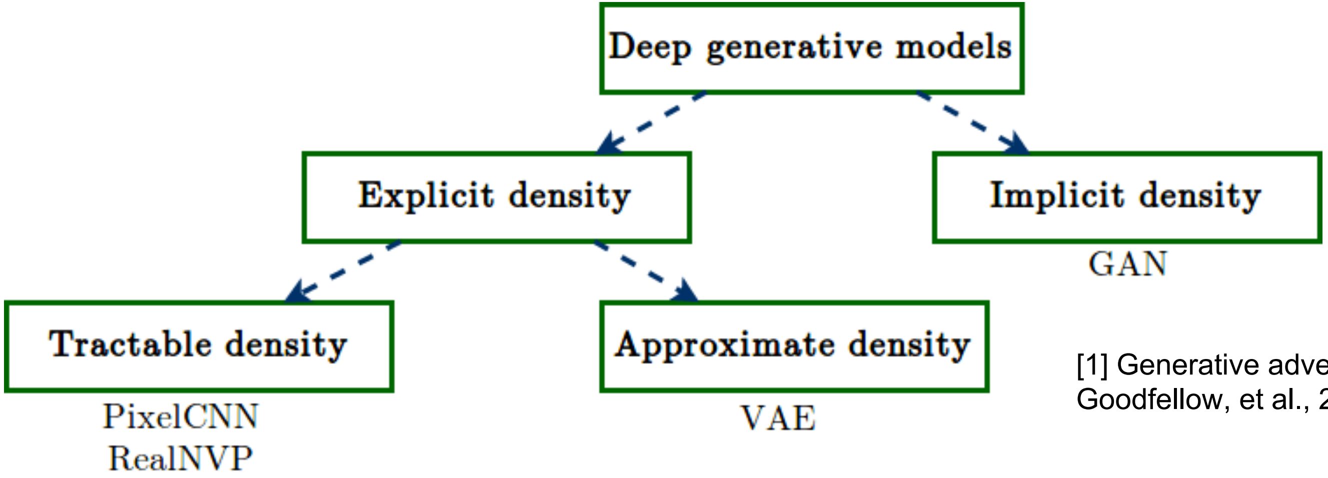
present
&
future

Deep Generative Models (DGM): definition



https://openai.com/blog/generative-models/

DGMs: taxonomy



[1] Generative adversarial nets, Goodfellow, et al., 2014.

[2] Auto-encoding variational bayes, Kingma, et al, 2013.

[3[Stochastic backpropagation and variational inference in deep latent gaussian models, Rezende, et al, 2014.

[4] Conditional image generation with pixelcnn decoders, Van den Oord, et al., 2016.

[5] Density estimation using realNVP, Dinh, et al, 2017

DGMs: the big picture

Density estimation

Quality of $p_{model}(x)$

Flows

Sampling

Sample quality

GANs

Representation learning

Latent space structure

VAEs

[6] Adaptive Density Estimation for Generative Models, Lucas et al. NeurIPS 2019

[7] Glow: Generative flow with invertible 1x1 convolutions, Kingma et al. NeurIPS 2018.

[8] Manifold-Valued Image Generation with Wasserstein Generative Adversarial Nets, Huang et al. *AAAI* 2019.

[9] Neural discrete representation learning, van den Oord et al. NIPS 2017

[10] Improved precision and recall metric for assessing generative models, Kynkäänniemi et al. NeurIPS 2019.

[11] Continuous Hierarchical Representations with Poincaré Variational Auto-Encoders, Emile, et al. NeurIPS 2019.

[12] Residual flows for invertible generative modelling, Chen et al. NeurIPS 2019

[13] Explicitly disentangling image content from translation and rotation with spatial-VAE, Bepler et al. NeurIPS 2019.

Representation learning: why?

“The performance of machine learning methods is heavily dependent on the choice of data representation (or features) on which they are applied.” [14]

- eliminate manual feature engineering
- extract useful non-linear information

[14] Bengio, Yoshua, et al. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.

What is a good representation?

"The one that makes the subsequent learning tasks easier."

- Low dimensional
- Captures similarities
- View invariant
- Disentangled
- Reflects input manifold

Evaluation depends on our goal

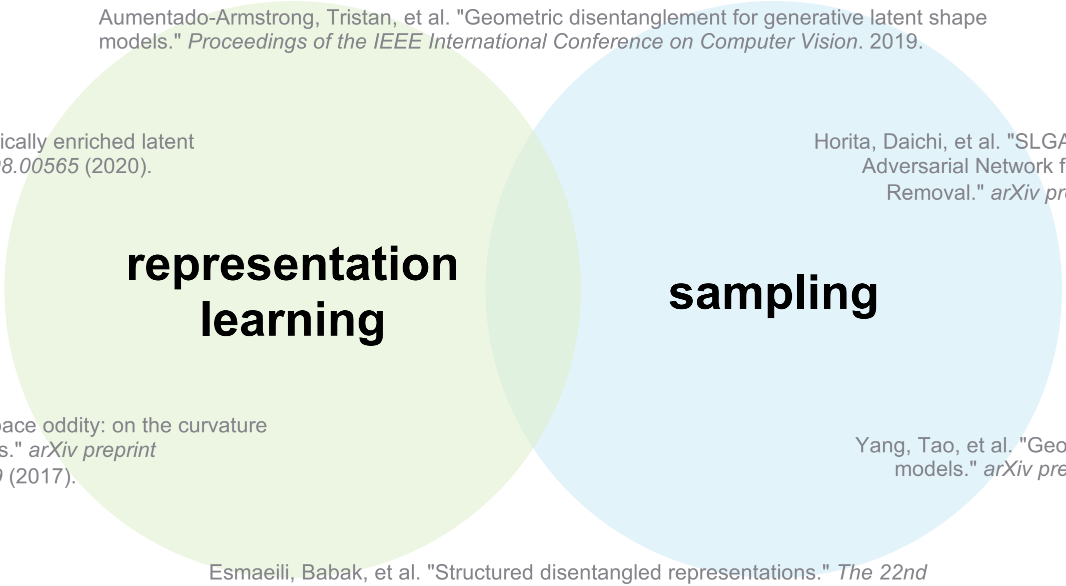
Representation learning with DGMs

“if generated data **samples** are **realistic,** then the underlying **structure** of the explanatory factors must be **captured**”

Aumentado-Armstrong, Tristan, et al. "Geometric disentanglement for generative latent shape models." *Proceedings of the IEEE International Conference on Computer Vision*. 2019.

Arvanitidis, Georgios, et al. ”Geometrically enriched latent spaces." *arXiv preprint arXiv:2008.00565* (2020).

Horita, Daichi, et al. "SLGAN: Style-and Latent-guided Generative Adversarial Network for Desirable Makeup Transfer and Removal." *arXiv preprint arXiv:2009.07557* (2020).



Arvanitidis, Georgios, et al. "Latent space oddity: on the curvature of deep generative models." *arXiv preprint arXiv:1710.11379* (2017).

Yang, Tao, et al. "Geodesic clustering in deep generative models." *arXiv preprint arXiv:1809.04747* (2018).

Esmaeili, Babak, et al. "Structured disentangled representations." *The 22nd International Conference on Artificial Intelligence and Statistics*. PMLR, 2019.

My work so far: statistics

conference publications: 1

workshops: 4

journal submissions: 2

collaborators: 12+

submissions in pipeline: 3

past

present & near future

Personal website: https://people.kth.se/~poklukar/

My work so far: statistics

conference publications: 1

workshops: 4

journal submissions: 2

collaborators: 12+

submissions in pipeline: 3

past

present & near future

Personal website: https://people.kth.se/~poklukar/

Outline

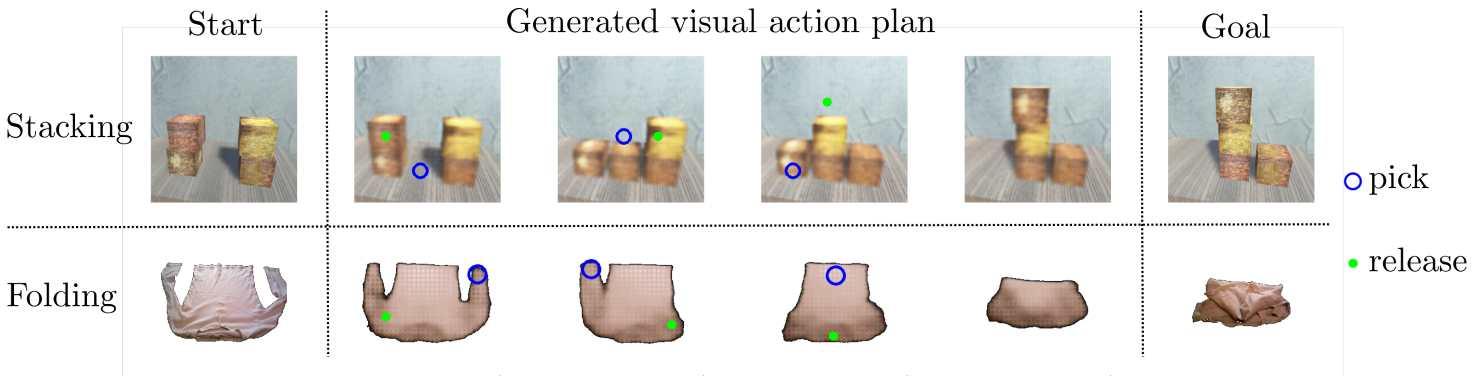
- Representation learning with deep generative models
- **Applications**
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Compact Task Embeddings
- Theory
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

past

present
&
future

Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation

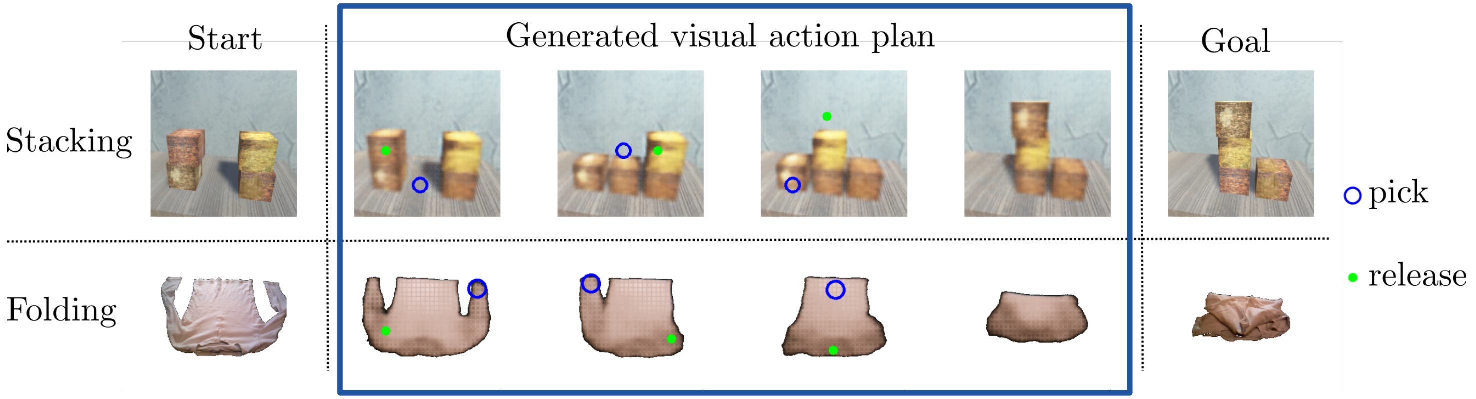
Martina Lippi*, **Petra Poklukar***, Michael C. Welle*, Anastasiia Varava, Hang Yin, Alessandro Marino and Danica Kragic



- published at IROS 2020
- journal extension to be submitted by the end of November

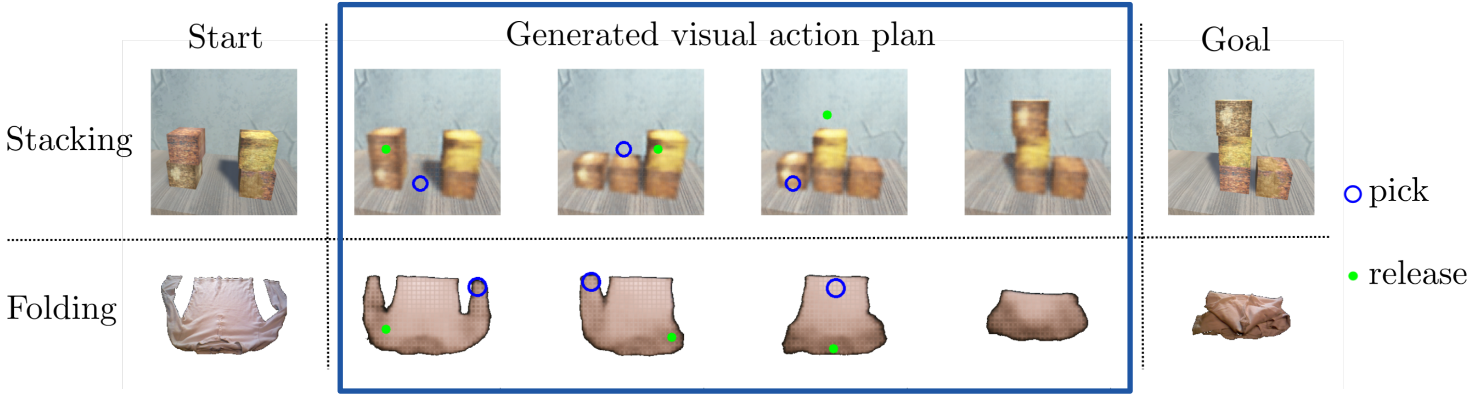
Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation

Martina Lippi*, **Petra Poklukar***, Michael C. Welle*, Anastasiia Varava, Hang Yin, Alessandro Marino and Danica Kragic



using low-dimensional
representations learned by a VAE

What is a good representation for visual action planning?

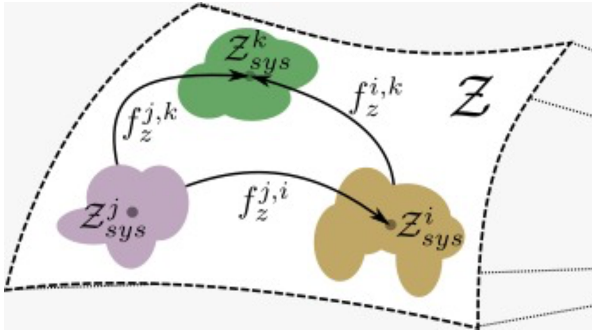


- low-dimensional
- extracts features representing each state
- cluster the extracted features

Latent Space Roadmap (LSR): challenges

1. How to extract the states from observations fulfilling our expectations?

representation learning



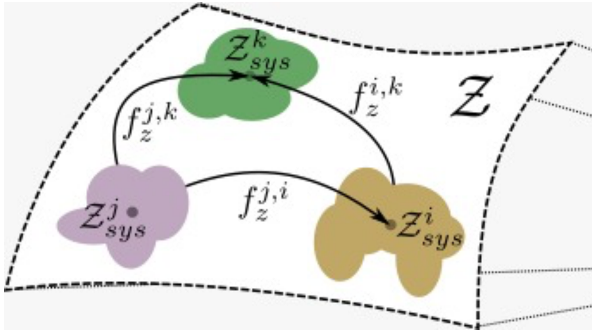
Latent Space Roadmap (LSR): challenges

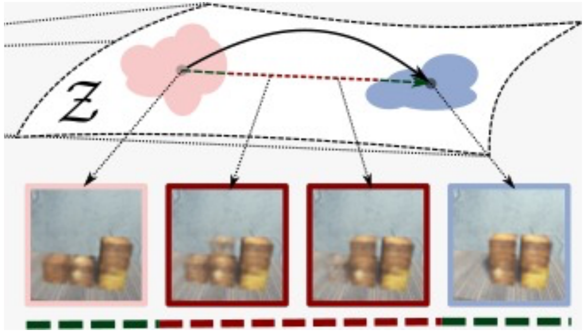
1. How to extract the states from observations fulfilling our expectations?

representation learning

informed sampling

2. How to navigate in the latent space to generate paths containing meaningful states?





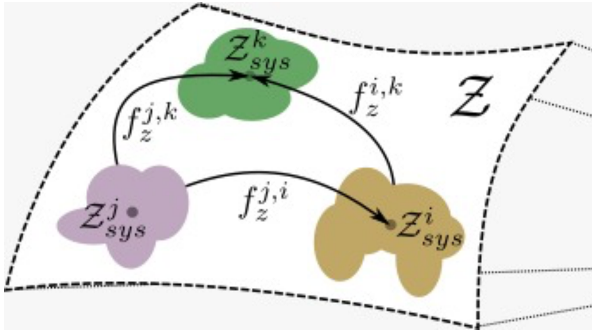
Latent Space Roadmap (LSR): contributions

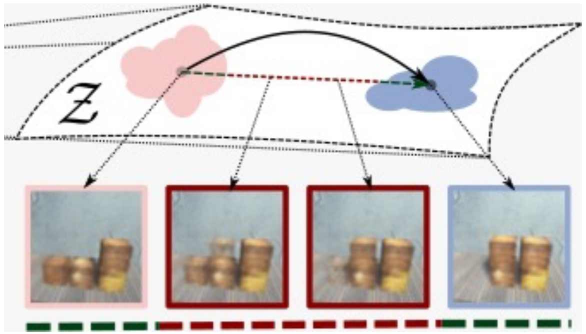
1. How to extract the states from observations fulfilling our expectations?

representation learning

informed sampling

2. How to navigate in the latent space to generate paths containing meaningful states?





- contrastive loss added to a VAE

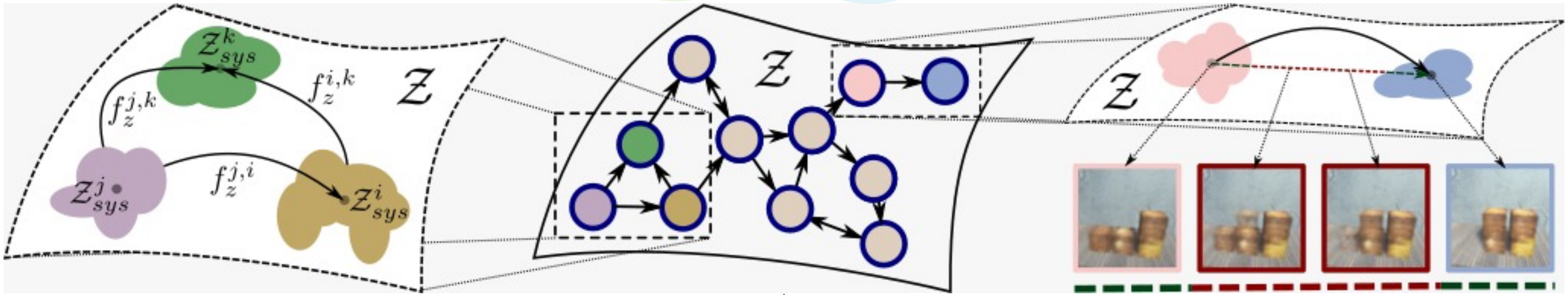
Latent Space Roadmap (LSR): contributions

1. How to extract the states from observations fulfilling our expectations?

representation learning

informed sampling

2. How to navigate in the latent space to generate paths containing meaningful states?



- contrastive loss added to a VAE

- LSR

How did we succeed?

squeeze the dataset

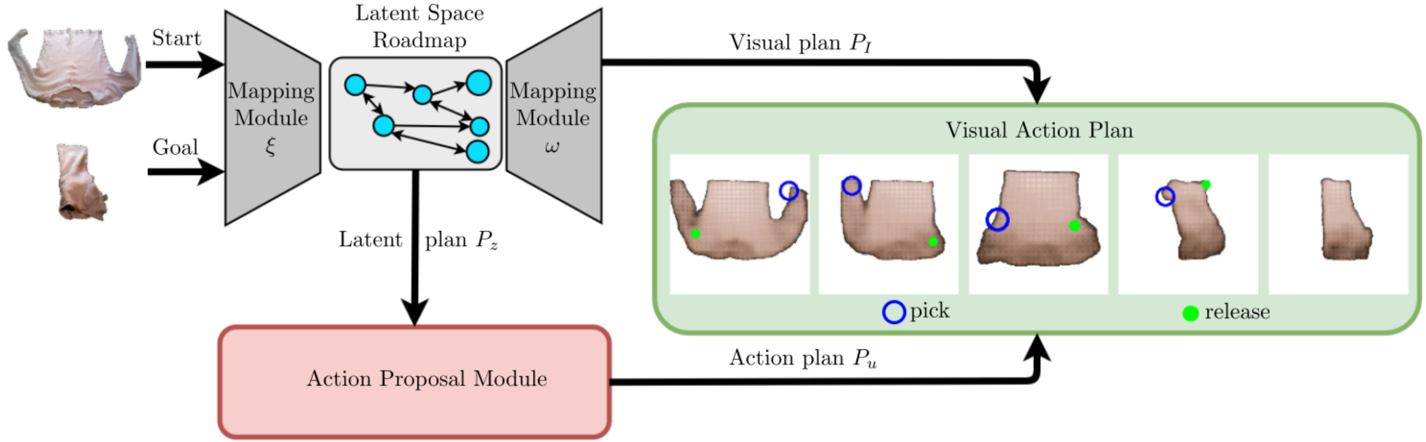
a)

b)

$\rho = (1, (\circ\,\text{pick}, \bullet\,\text{place}))$

$\rho = (0, ())$

LSR: overview

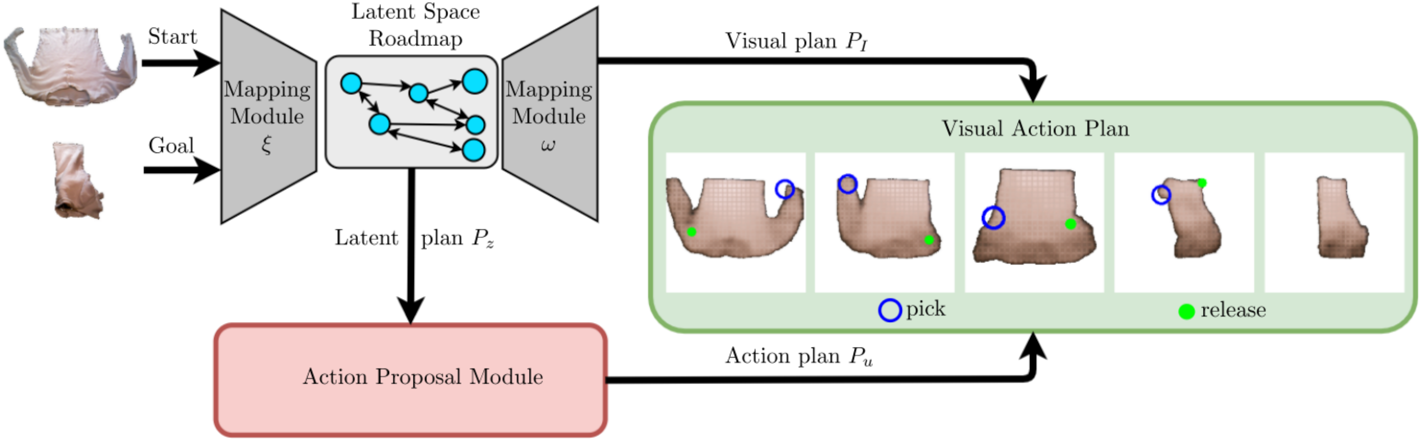


Contributions:

- augmented the VAE loss for better representation learning
- defined the LSR for informed sampling
- used representations to learn actions
- evaluated the method in simulation and real world

Lippi, Martina, et al. "Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation." *arXiv preprint arXiv:2003.08974* (2020).

LSR: overview



Contributions:
- augmented the VAE loss for better representation learning
- defined the LSR for informed sampling
- used representations to learn actions
- evaluated the method in simulation and real world

Extra in the journal:
- removed a bunch of hyperparameters
- extensive ablation study on all components

Lippi, Martina, et al. "Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation." *arXiv preprint arXiv:2003.08974* (2020).

Outline

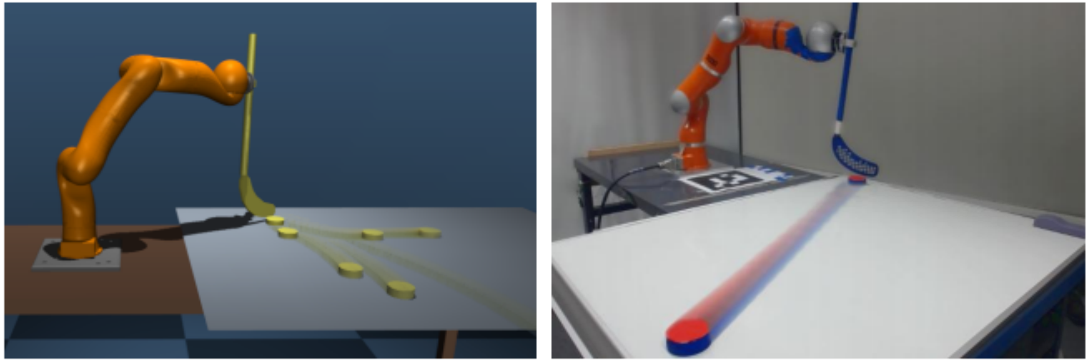
- Representation learning with deep generative models
- **Applications**
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Compact Task Embeddings
- Theory
  - A framework for evaluating disentangled representations
  - Manifold Approximation using Two Discrete Datasets
- What is missing?

past

present
&
future

Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models

Ali Ghadirzadeh*, **Petra Poklukar***, Ville Kyrki, Danica Kragic and Mårten Björkman



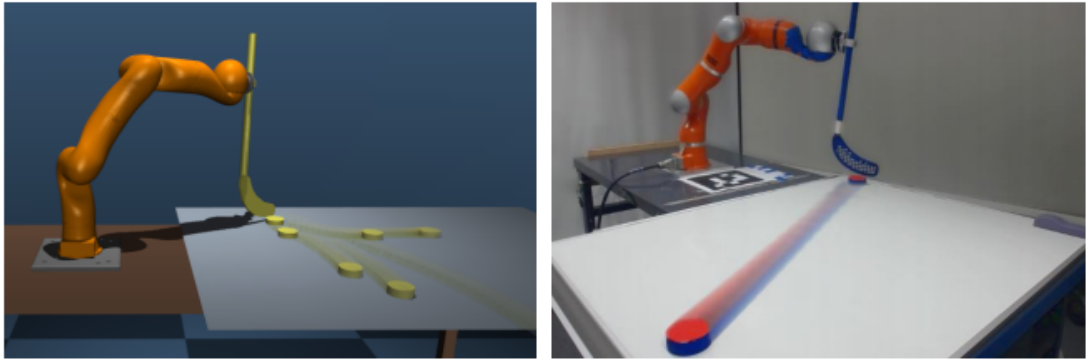
Arndt, Karol, et al. "Meta reinforcement learning for sim-to-real domain adaptation." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

- Submitted to the Journal of Machine Learning Research

Ghadirzadeh, Ali, et al. "Data-efficient visuomotor policy training using reinforcement learning and generative models." *arXiv preprint arXiv:2007.13134* (2020).

Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models

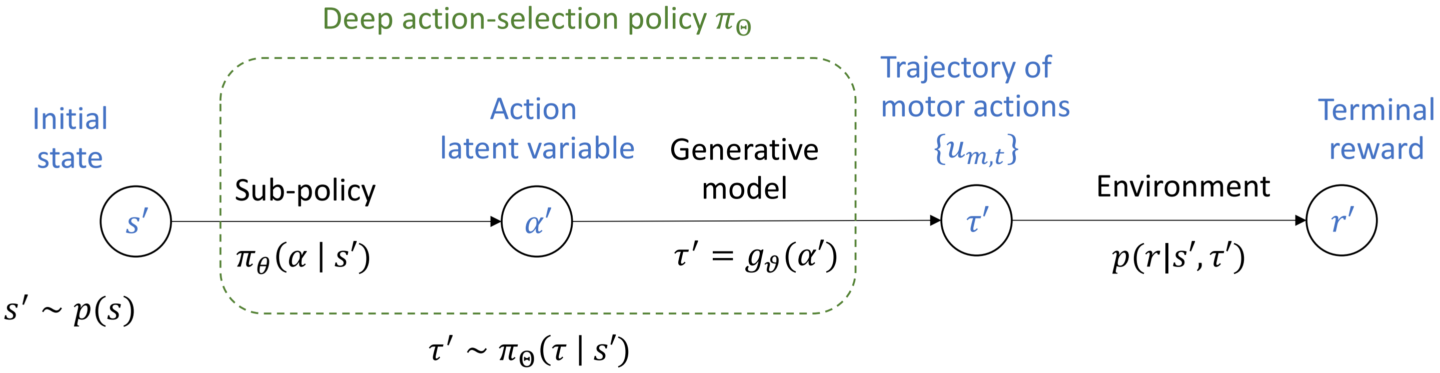
Ali Ghadirzadeh*, **Petra Poklukar***, Ville Kyrki, Danica Kragic and Mårten Björkman



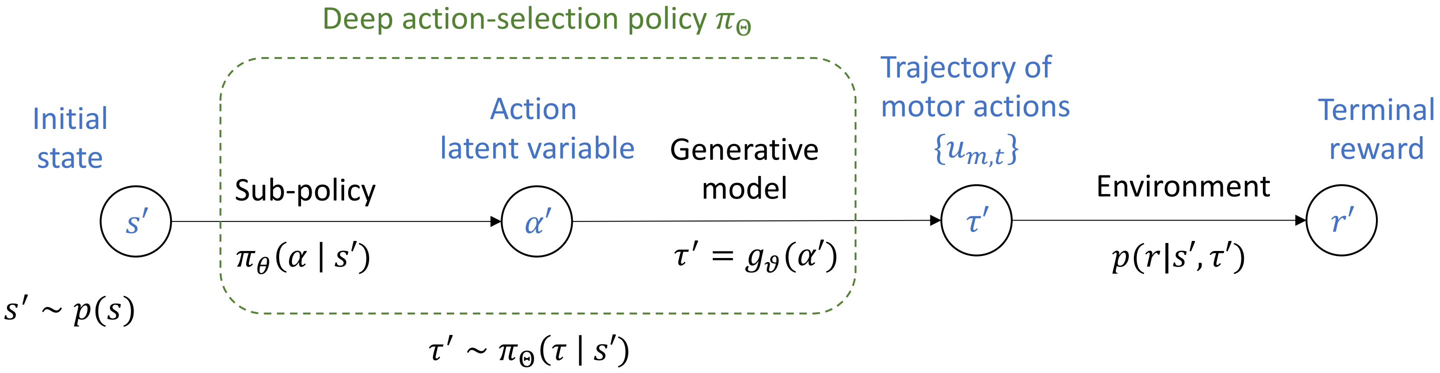
Arndt, Karol, et al. "Meta reinforcement learning for sim-to-real domain adaptation." *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020.

use DGMs to reduce the complexity of the problem

How to integrate DGMs into RL?

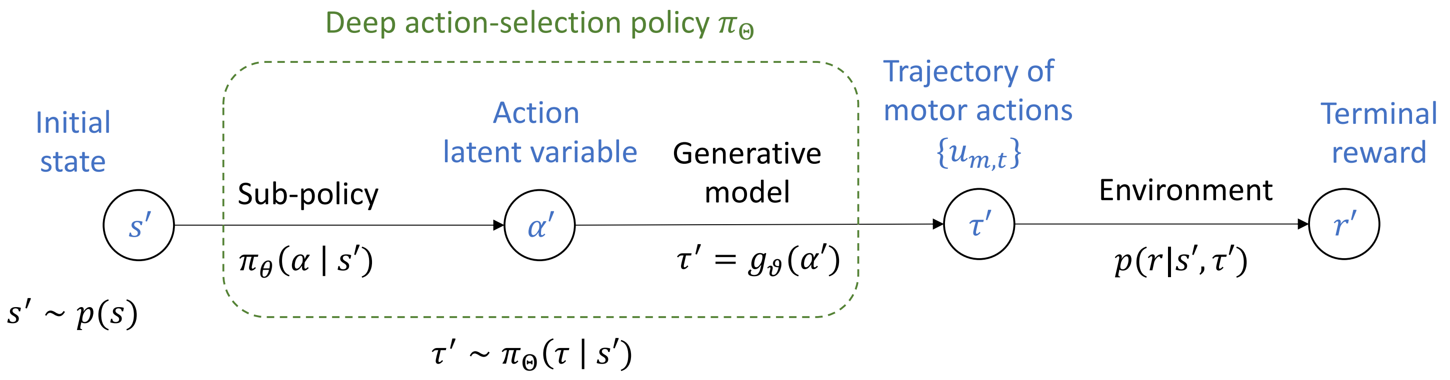


What is a good representation for data-efficient policy training?



- low dimensional
- captures similarities ?
- disentangled ?
- reflects input manifold ?

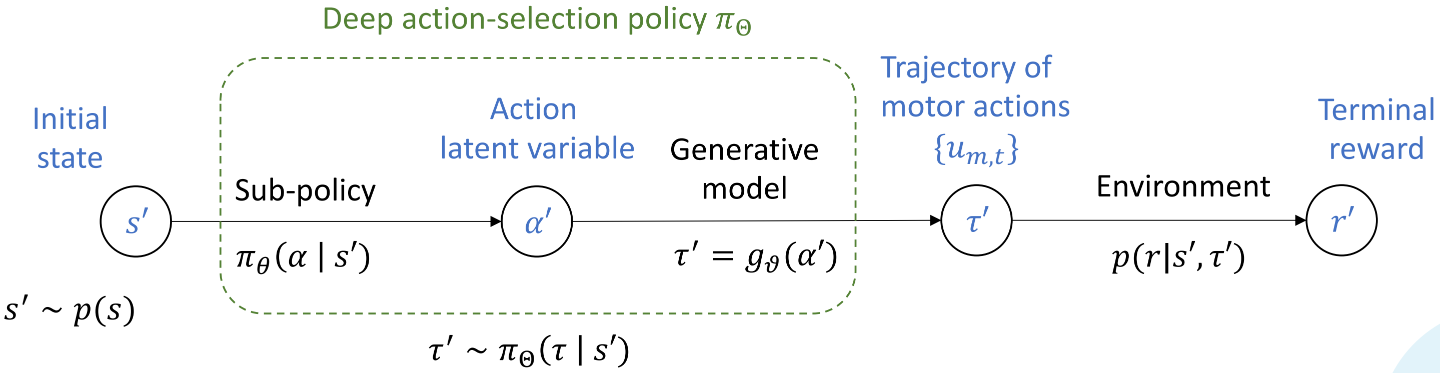
What is a good representation for data-efficient policy training?



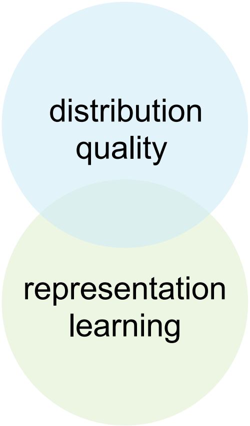
- low dimensional
- captures similarities ?
- disentangled ?
- reflects input manifold ?

Let's try to relate the policy performance with characteristics of DGMs

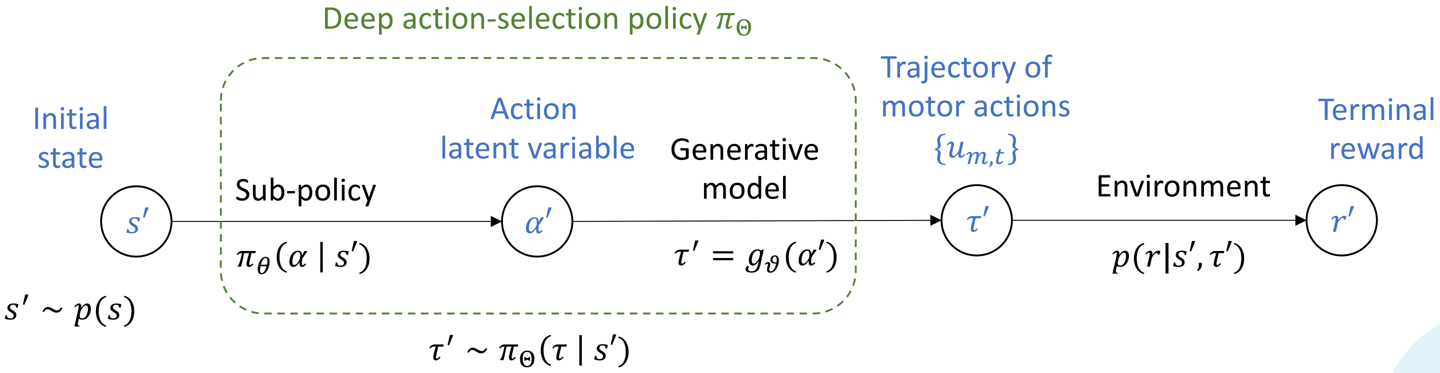
Good representation for data-efficient policy training: hypothesis



1. Enable generating trajectories that:
   - resemble training motion trajectories
   - are distinct and valid
2. Are disentangled
3. Are “locally simple”



Good representation for data-efficient policy training: evaluation

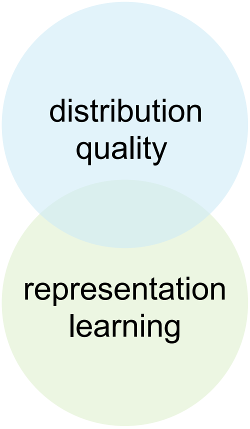


1. Enable generating trajectories that:
   - resemble training motion trajectories
   - are distinct and valid
2. Are disentangled
3. Are "locally simple"

precision and recall [15]

disentangling precision and recall

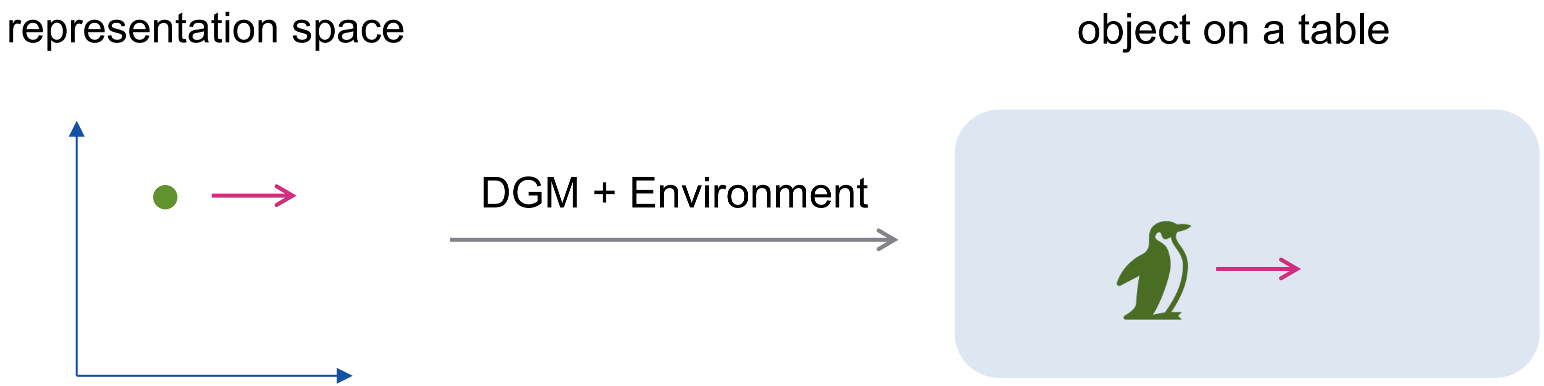
latent local linearity



[15] Kynkäänniemi, Tuomas, et al. "Improved precision and recall metric for assessing generative models." *Advances in Neural Information Processing Systems*. 2019.

What is a disentangled representation?

≈ **Definition [disentanglement]:** one independent factor of variation or an underlying generative factor, present in the data, is associated with exactly one latent dimension. [14]

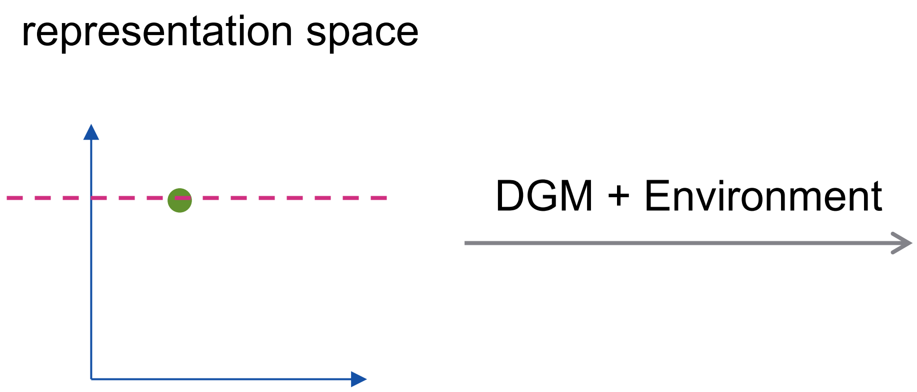


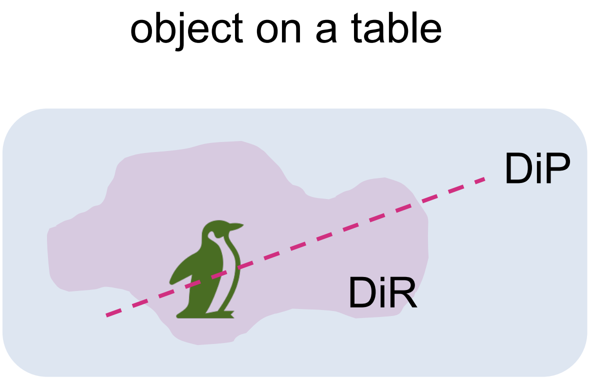
[14] Bengio, Yoshua, et al. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.

Disentangling precision and recall: idea

- fixing one latent dimension yields a limited set of end states
- compare it to the training dataset to estimate how limited

Disentangling precision and recall: intuition





- **disentangling precision (DiP)**: quantifies the effect of limiting representations on the end states.
- **disentangling recall (DiR)**: measures how many different aspects of the end states are captured in the latent space

Outline

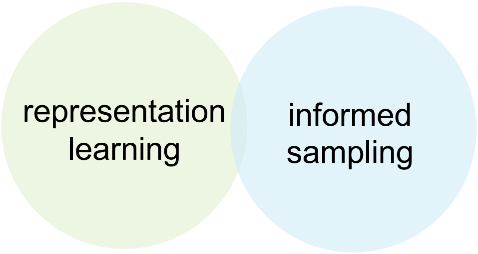
- Representation learning with deep generative models
- **Applications**
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Latent Task Embeddings
- Theory
  - A framework for evaluating disentangled representations
  - Manifold Approximation using Two Discrete Datasets
- What is missing?

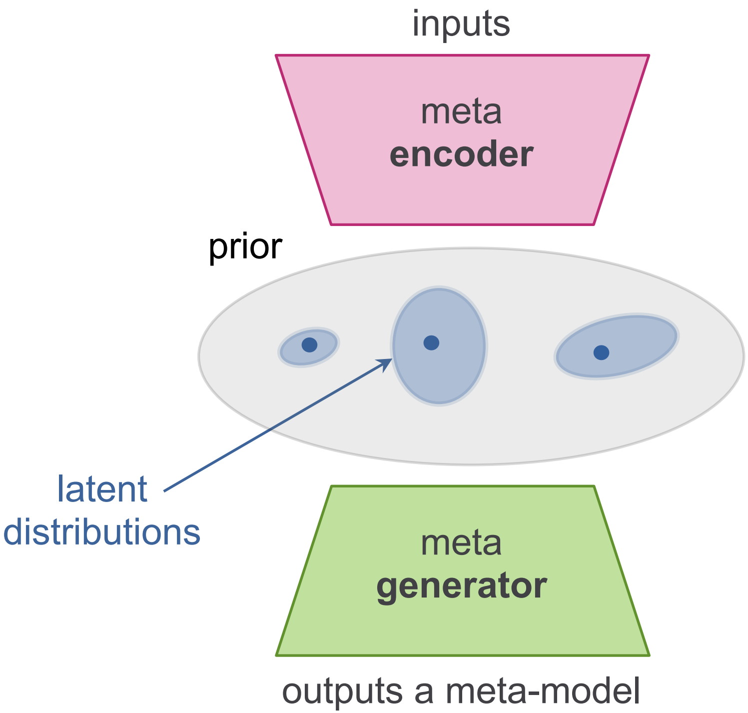
present
&
future

Variational Model-Agnostic Meta-Learning with Latent Task Embeddings

Petra Poklukar*, Ali Ghadirzadeh*, Xi Chen*,
Chelsea Finn, Mårten Björkman and Danica Kragic

”***fast adaptation*** to a new meta-task using only a ***few datapoints*** and training ***iterations***“





Outline

- Representation learning with deep generative models
- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Latent Task Embeddings
- **Theory**
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

present
&
future

Challenges with learning disentangled representations

$\approx$ **Definition [disentanglement]:** one independent factor of variation or an underlying generative factor, present in the data, is associated with exactly one latent dimension. [14]

- No unified definition of neither *generative factor* nor *disentangled representation*

[14] Bengio, Yoshua, et al. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.

Challenges with learning disentangled representations

≈ **Definition [disentanglement]:** one independent factor of variation or an underlying generative factor, present in the data, is associated with exactly one latent dimension. [14]

- No unified definition of neither *generative factor* nor *disentangled representation*
- Current disentanglement metrics [BetaVAE, FactorVAE, MIG, DCI, …] have many issues:
  - Rely on ground truth labels for generative factors
  - Tuned for the specific model
  - Not consistent [16]

[14] Bengio, Yoshua, et al. "Representation learning: A review and new perspectives." *IEEE transactions on pattern analysis and machine intelligence* 35.8 (2013): 1798-1828.

[16] Locatello, Francesco, et al. "Challenging common assumptions in the unsupervised learning of disentangled representations." *international conference on machine learning*. 2019.

A framework for evaluating disentangled representations

[with Michael]

Aim is to build a controlled environment for evaluation:

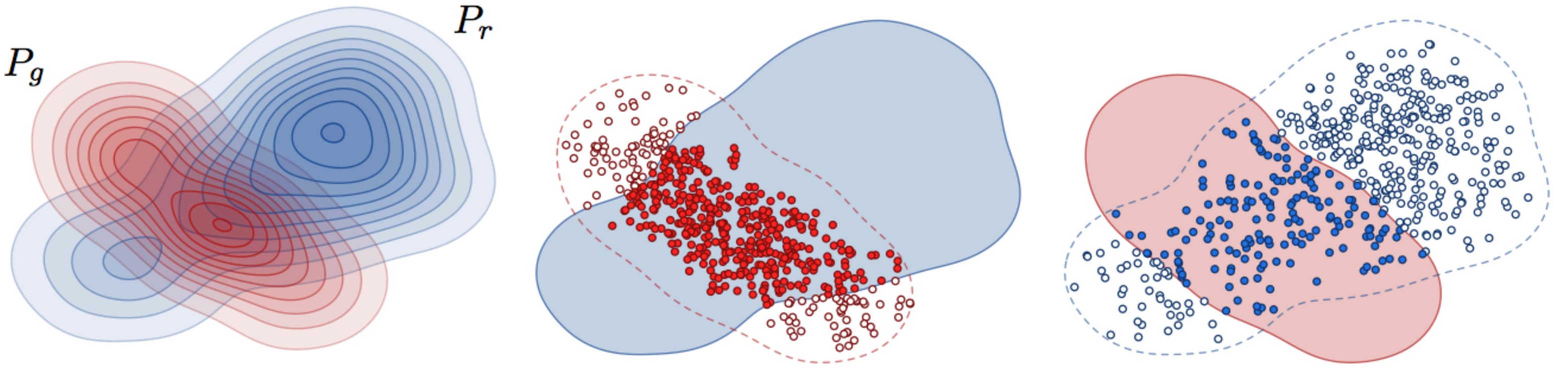
- Model independent
- Enables to “set the definitions”

Outline

- Representation learning with deep generative models
- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Latent Task Embeddings
- **Theory**
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

present
&
future

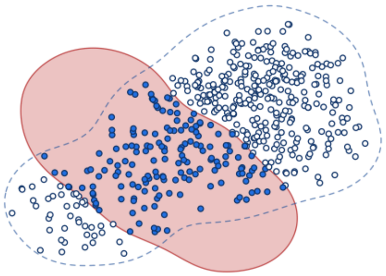
Precision and recall for assessing generative models

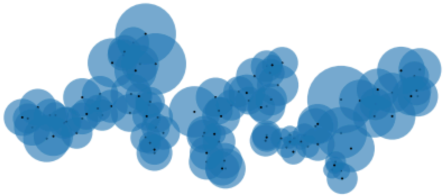


(a) Example distributions (b) Precision (c) Recall

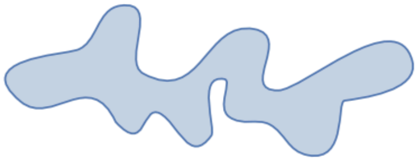
[15] Kynkäänniemi, Tuomas, et al. "Improved precision and recall metric for assessing generative models." *Advances in Neural Information Processing Systems*. 2019.

Precision and recall: manifold estimation

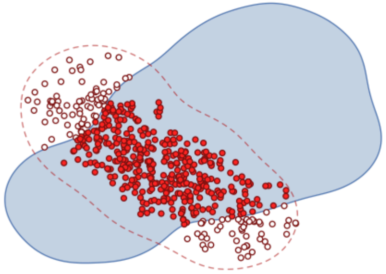




(b) Approx. manifold

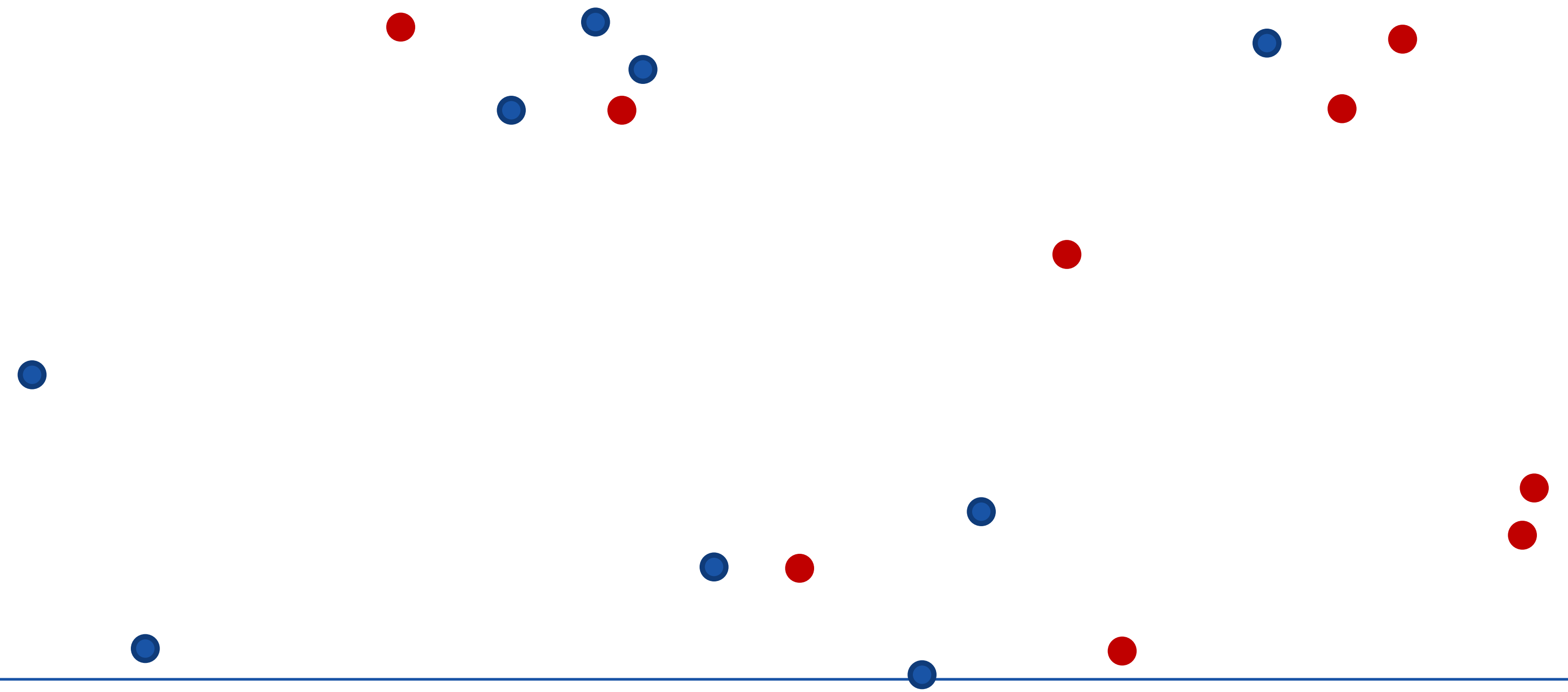


(a) True manifold

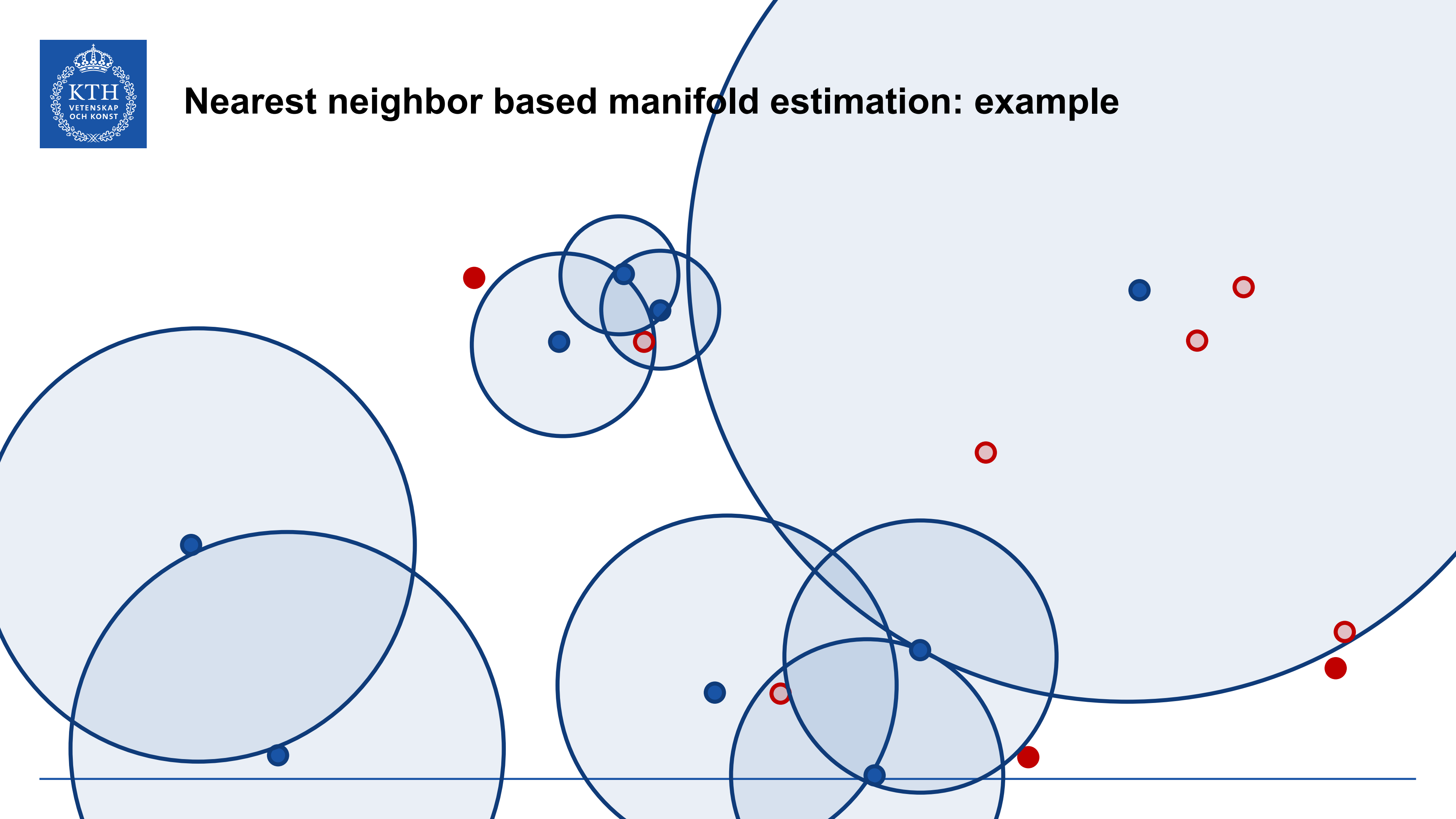


[15] Kynkäänniemi, Tuomas, et al. "Improved precision and recall metric for assessing generative models." *Advances in Neural Information Processing Systems*. 2019.

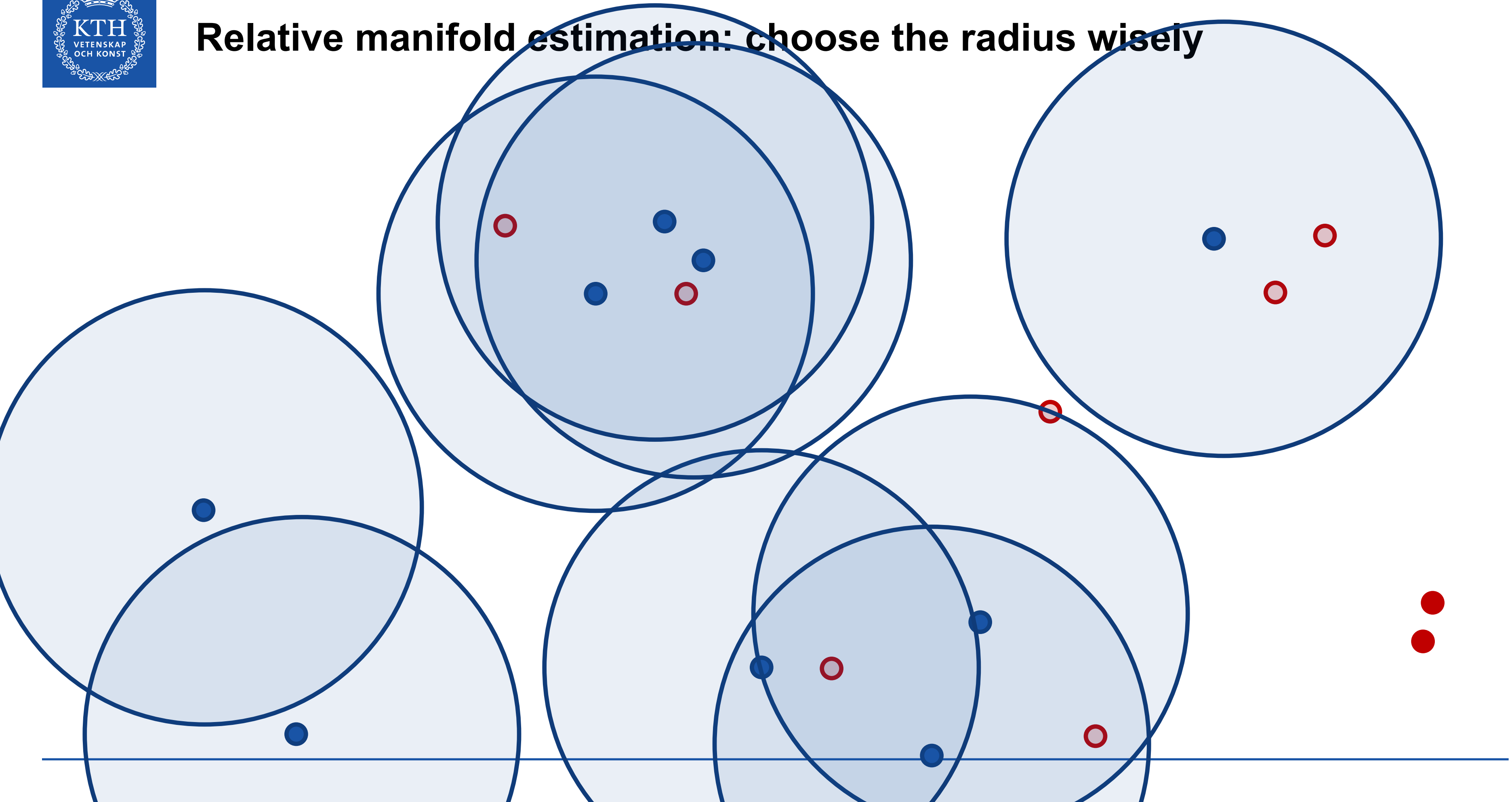
Nearest neighbor based manifold estimation: example



Nearest neighbor based manifold estimation: example



Relative manifold estimation: choose the radius wisely



Outline

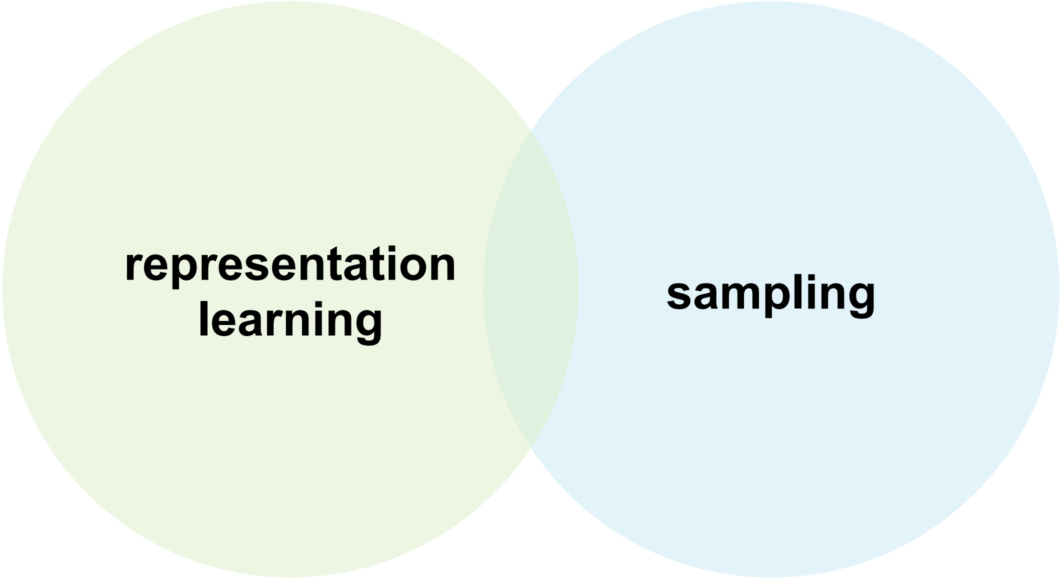
- Representation learning with deep generative models
- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Latent Task Embeddings
- Theory
  - A framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- **What is missing?**

present
&
future

Representation learning with DGMs: missing parts

1. Coverage of the representation space

2. Hierarchical *probabilistic* view of *similarities* among representations



PhD roadmap: Representation Learning with DGMs

- Applications
  - Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
  - Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
  - Variational Model-Agnostic Meta-Learning with Latent Task Embeddings
- Theory
  - A Framework for Evaluating Disentangled Representations
  - Relative Manifold Approximation using Two Discrete Datasets
- What is missing?

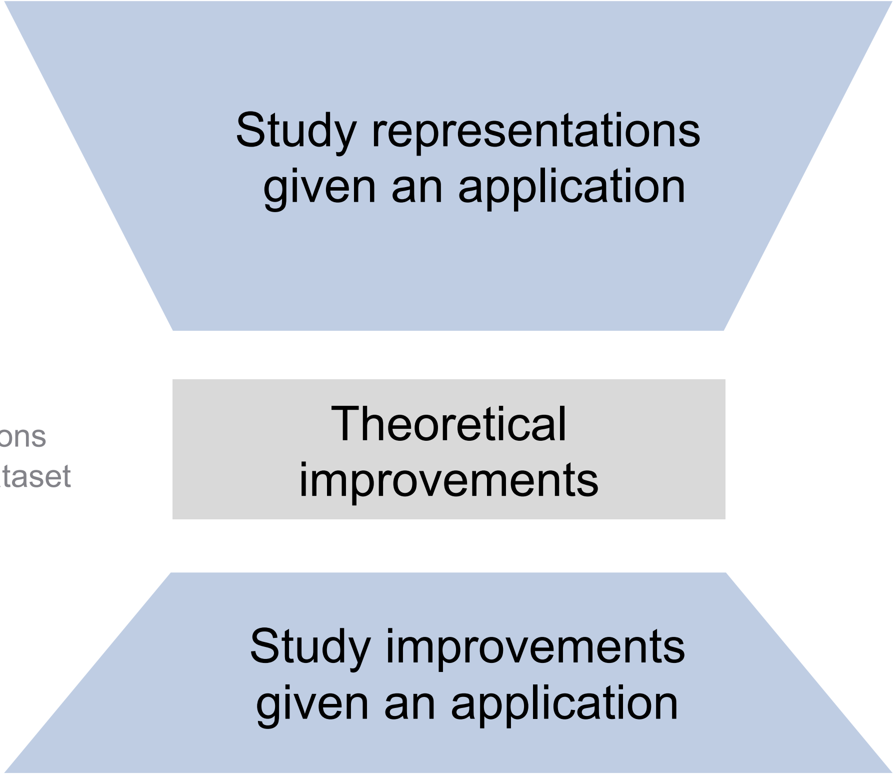
PhD roadmap: Representation Learning with DGMs

Applications

- Latent Space Roadmap for Visual Action Planning of Deformable and Rigid Object Manipulation
- Data-Efficient Visuomotor Policy Training using Reinforcement Learning and Generative Models
- Variational Model-Agnostic Meta-Learning with Latent Task Embeddings

Theory

- A Framework for Evaluating Disentangled Representations
- Relative Manifold Approximation using Two Discrete Dataset
- and more…



Representation Learning with Deep Generative Models

Petra Poklukar; https://people.kth.se/~poklukar/
50% Seminar