The Andrew and Erna Viterbi Faculty of ELECTRICAL & COMPUTER ENGINEERING

ICML International Conference On Machine Learning

$(RL)^2$

TECHNION ISRAEL INSTITUTE OF TECHNOLOGY

Unsupervised Image Representation Learning with Deep Latent Particles

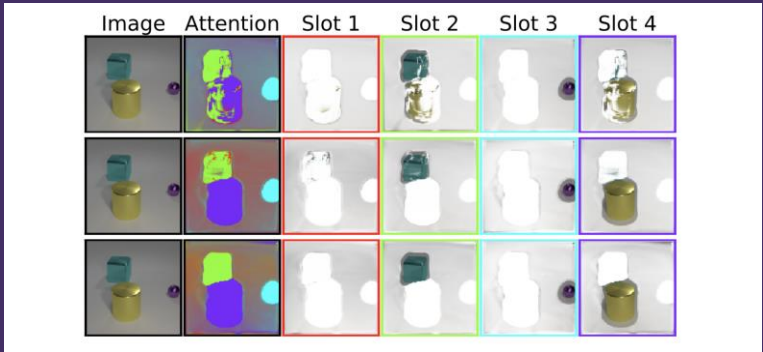
ICML 2022

taldatech.github.io/deep-latent-particles-web

Tal Daniel ◉ Aviv Tamar

Representation of Images with Physical Objects

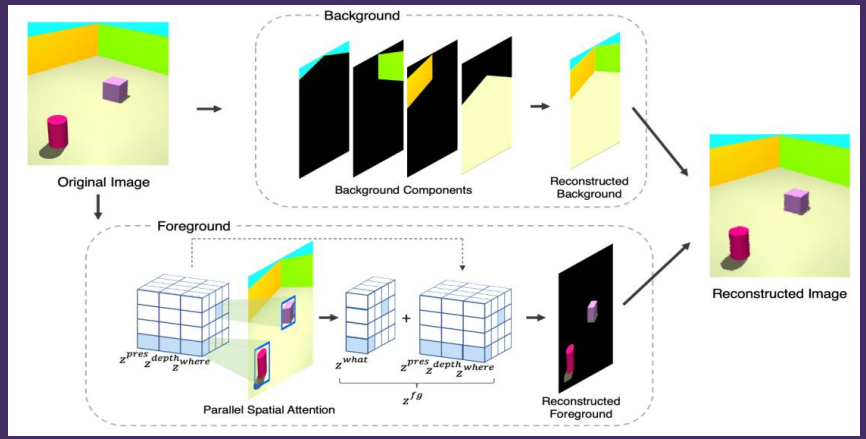
- Slot-based latent variable models
- Pros: generative, probabilistic interpretation
- Cons: complexity grows with number of objects, hard to train and interpret



Locatello et al.,“Object-Centric Learning with Slot Attention”, NeurIPS 2020
https://generallyintelligent.ai/open-source/2021-03-09-slot-attention/

Representation of Images with Physical Objects

- Patch-based object-centric latent variable models
- Pros: generative, probabilistic interpretation, non-sequential
- Cons: limited to moderate number of objects, complex filtering process

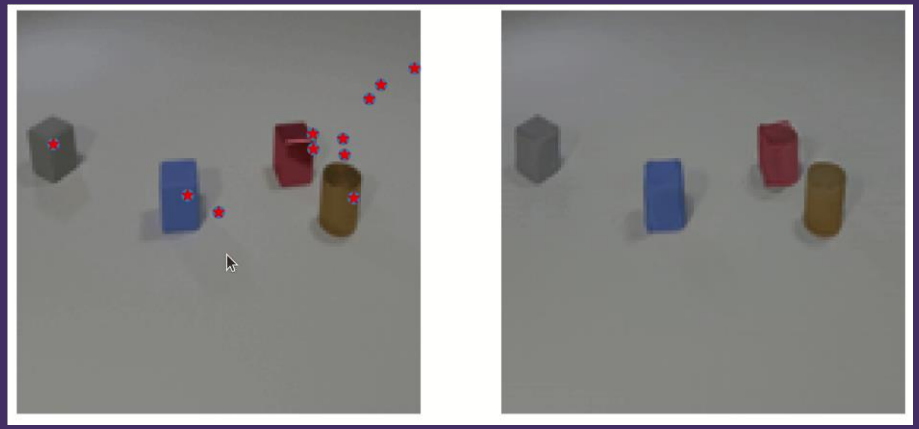


Lin et al.,“SPACE: Unsupervised Object-Oriented Scene Representation via Spatial Attention and Decomposition”, ICLR 2020

Representation of Images with Physical Objects

- Keypoints (descriptors/landmarks)
- Pros: simple, can work with a lot of objects.
- Cons: usually deterministic, limited generative capacity

Jakab et al.,“Unsupervised Learning of Object Landmarks through Conditional Image Generation”, NeurIPS 2018



Deep Latent Particles (DLP)

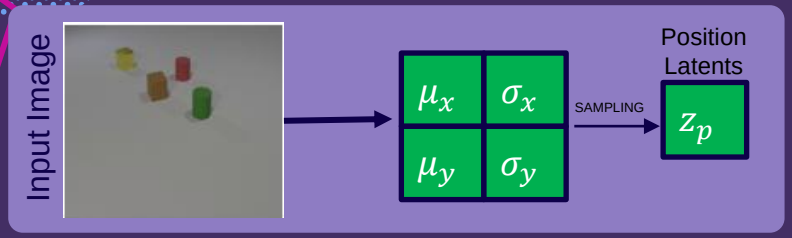
Particle: Keypoint + Features

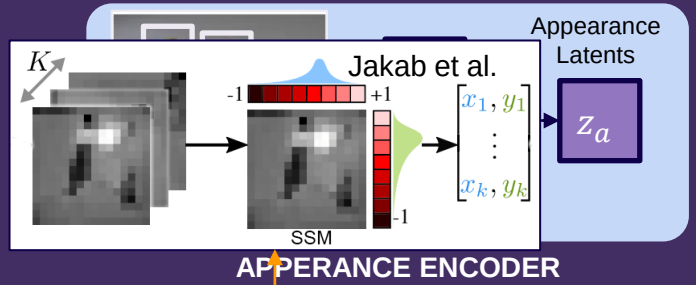
Keypoints are the latent space of a Variational Autoencoder (VAE)

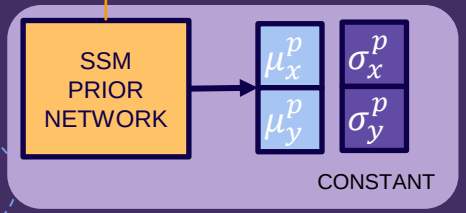
Particle positions prior based on spatial-softmax (SSM)

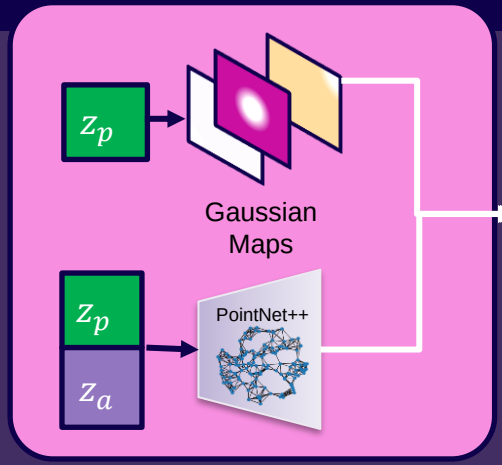
Chamfer-KL: novel modification of the KL term in the ELBO

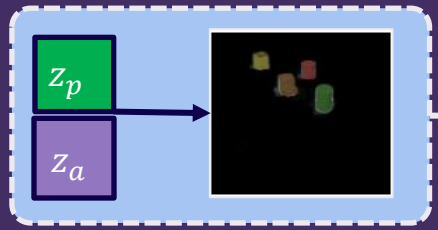
How Does DLP Work?



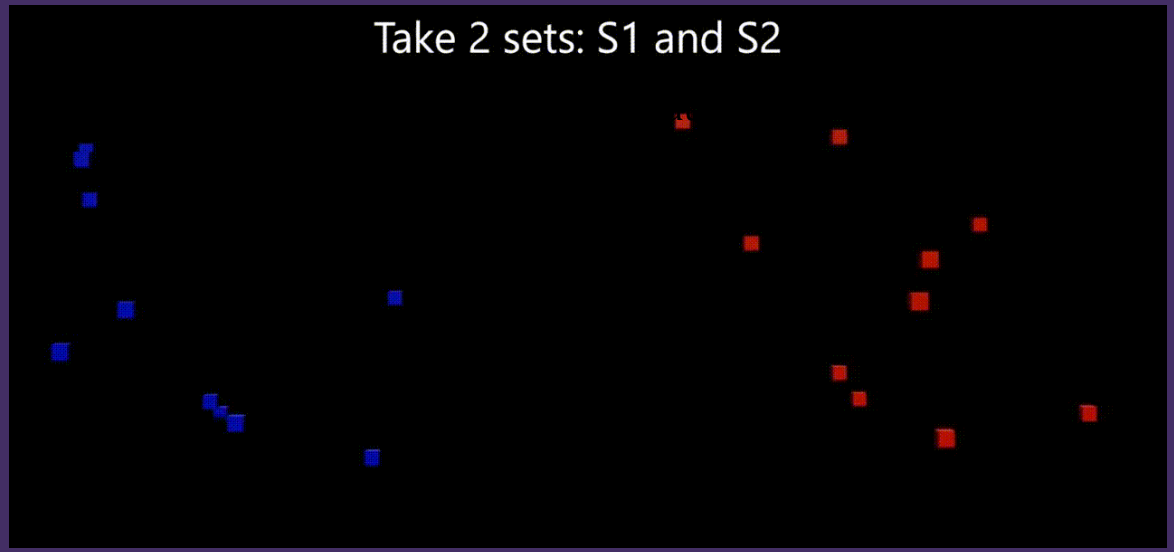








$$d_{CH-KL}(S_1, S_2) = \sum_{x \in S_1} \min_{y \in S_2} KL(x \| y) + \sum_{y \in S_2} \min_{x \in S_1} KL(x \| y)$$



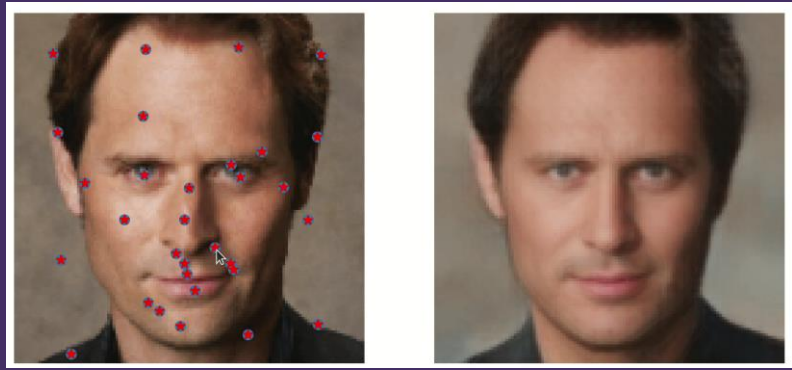
Animation by Luke Hawkes - A visual representation of the Chamfer distance function

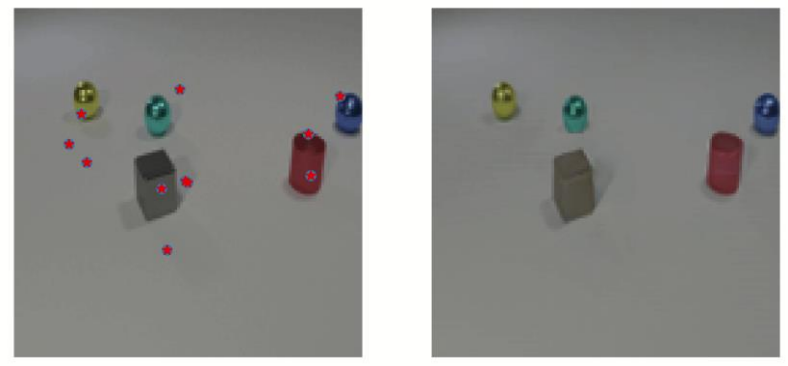
Unsupervised Keypoint Discovery

- State-of-the-art performance on the MAFL dataset
- The learned particle uncertainty is informative

Method	K (number of unsupervised KP)	Error on MAFL (lower is better)
Zhang (Zhang et al., 2018)	30	3.16
KeyNet (Jakab et al., 2018)	30	2.58
	50	2.54
Ours	25	2.87
	30	**2.56**
	50	**2.43**
Ours+ (with variance features)	25	**2.52**
	30	2.49
	50	2.42

Particle-based Image Manipulation





Particle-based Video Prediciton

- Predict the temporal change in particles with GNNs

Thanks for watching!

https://taldatech.github.io/deep-latent-particles-web

Presentation template by SlidesCarnival