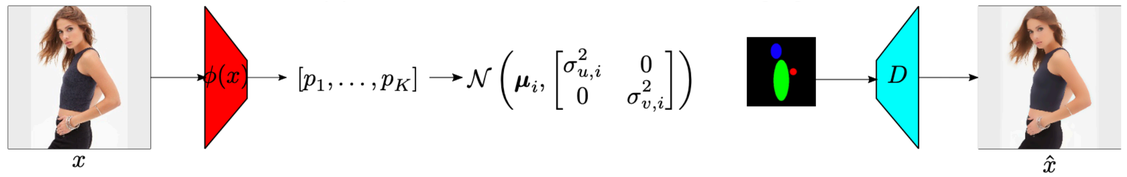
Unsupervised Part Discovery by Unsupervised Disentanglement

Sandro Braun, Patrick Esser, Björn Ommer

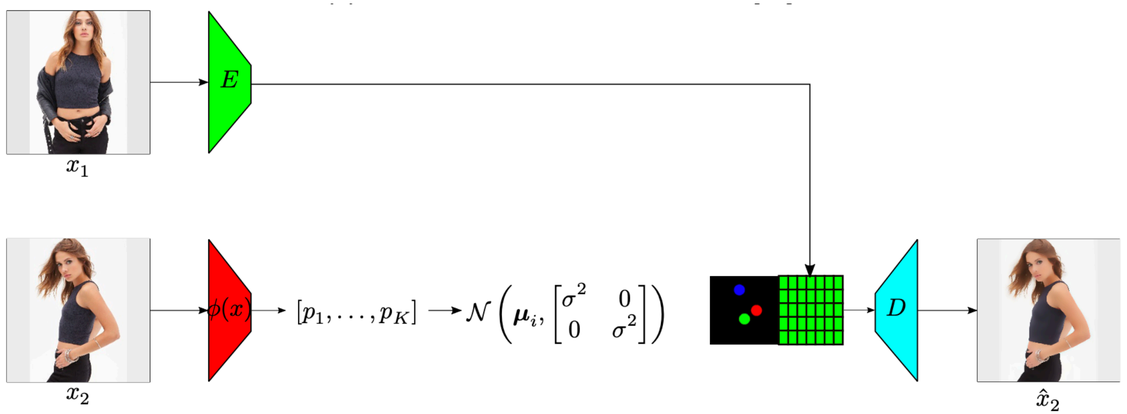
Previous Work on Unsupervised Keypoint Learning

Zhang [48]



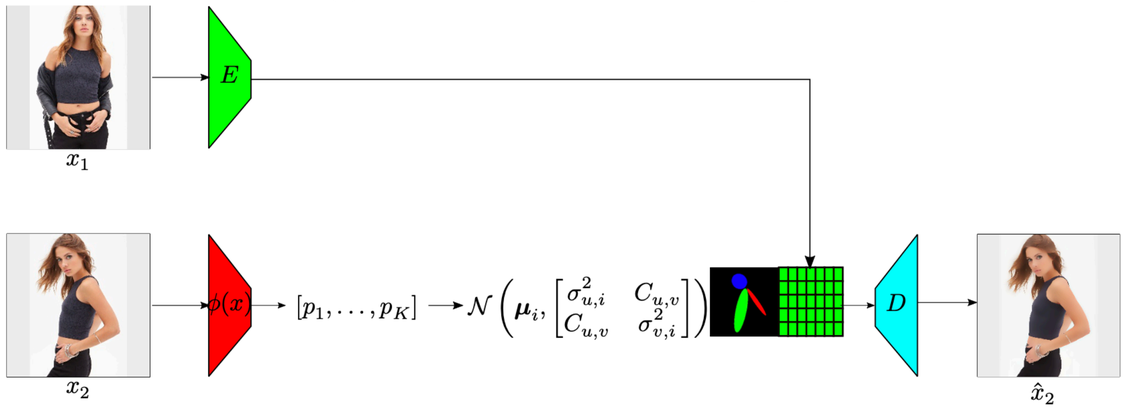
Diagonal Covariance

Jakab [14]



Isotropic Covariance

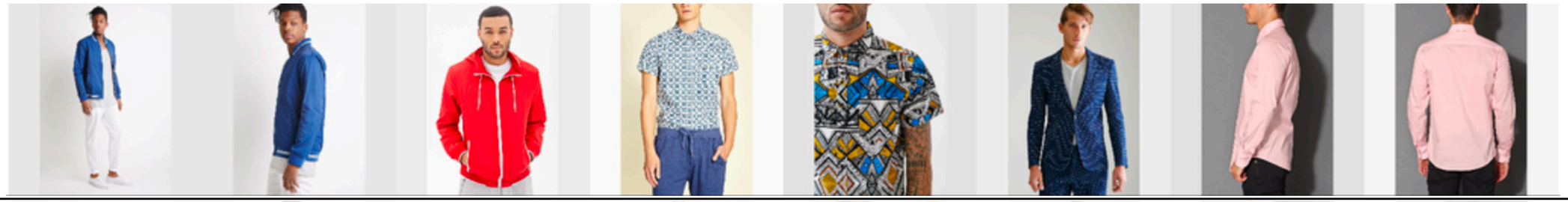
Lorenz [25]



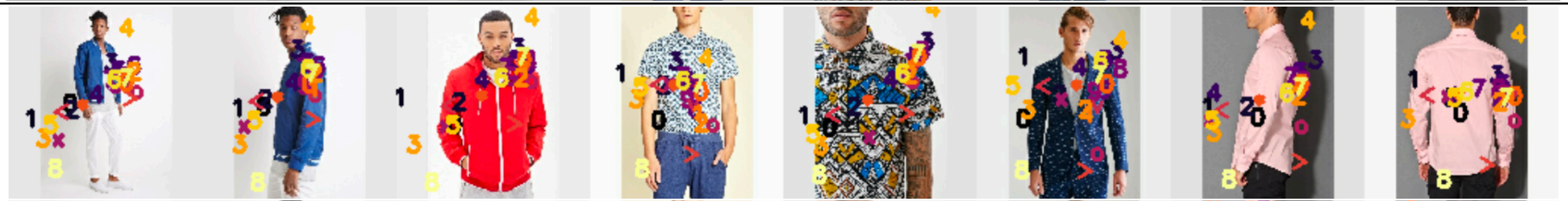
Full Covariance

… uses Gaussian parametric shape models

Input



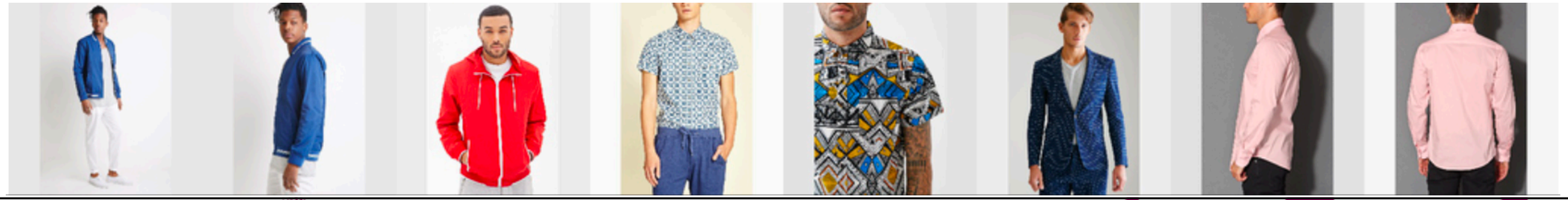
Jakab [14]



Zhang [48]

Lorenz [25]

Input



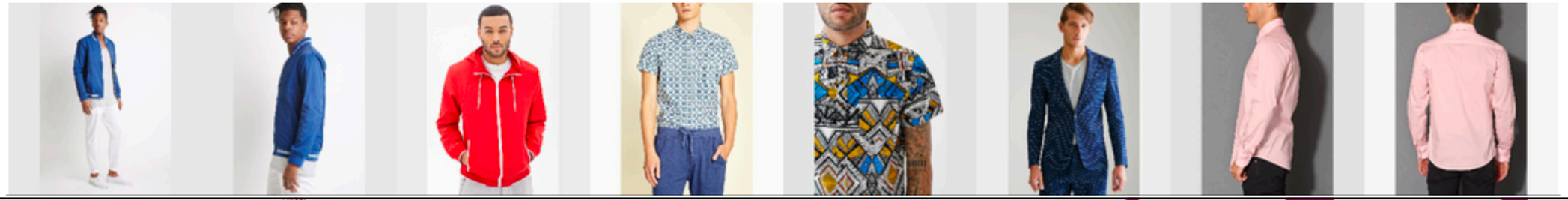
Jakab [14]
+ CRF

Zhang [48]
+ CRF

Lorenz [25]
+ CRF

GT

Input



Jakab [14]
+ CRF

Zhang [48]
+ CRF

Lorenz [25]
+ CRF

Ours

GT

„Just remove the parametric shape model“

Sounds simple

„But simple is still hard“
- Tesla Battery Day 2020, 1:13:41

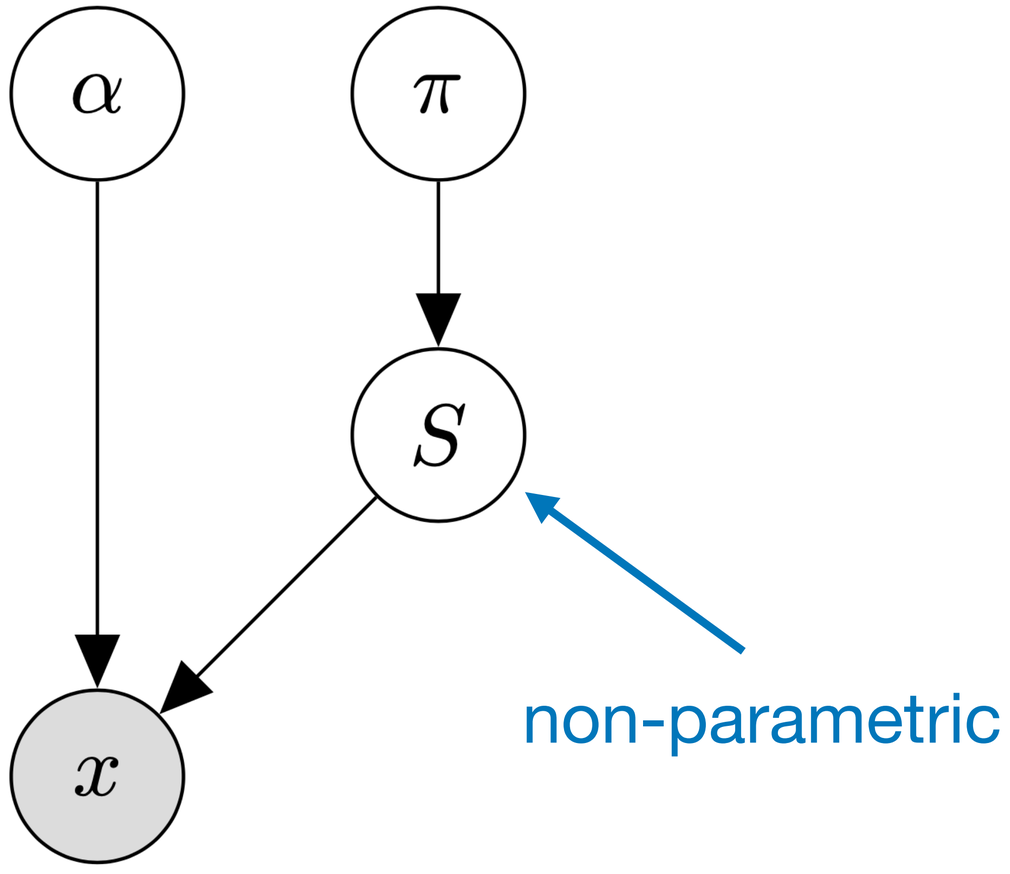
https://youtu.be/l6T9xIeZTds?t=4421

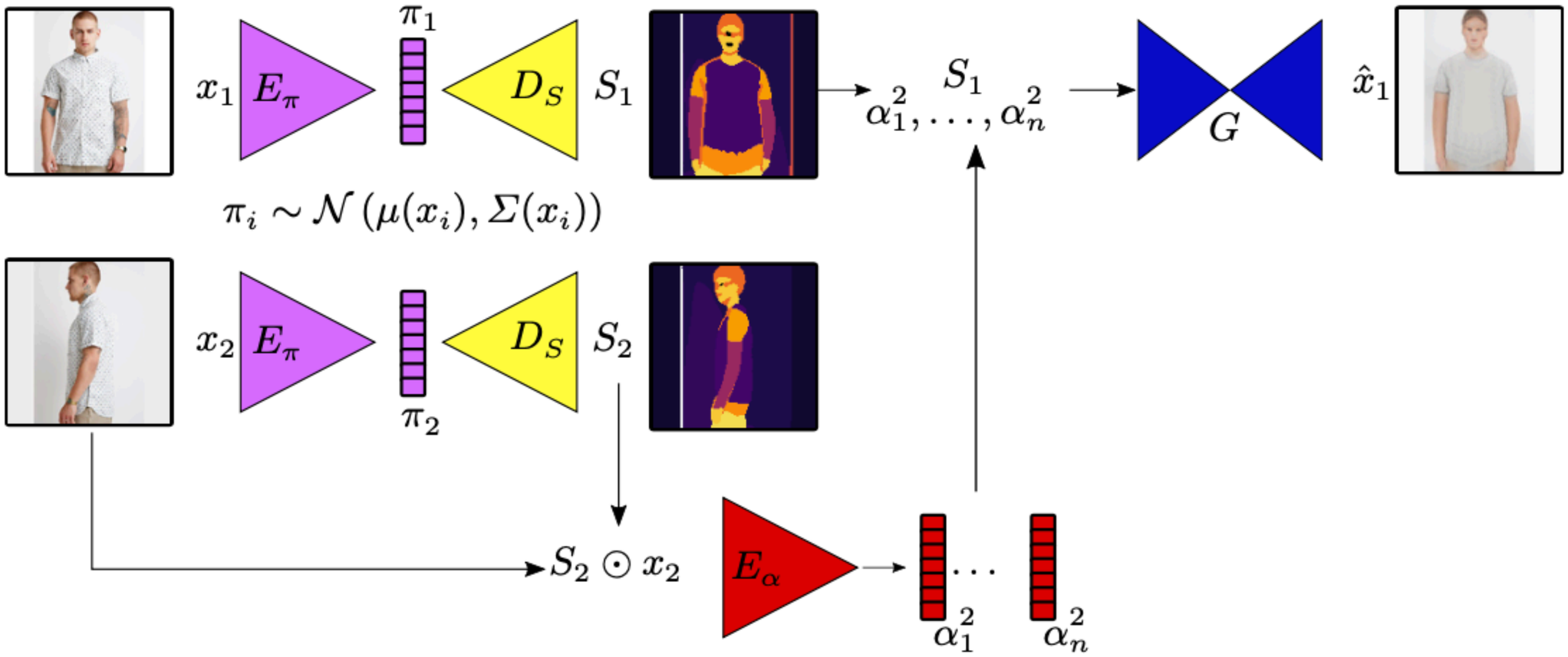
High level ideas

1. part segmentations S are independent of latent variable appearance
—> requires disentanglement!

2. part segmentations S
as latent variables with suitable priors
—> which priors?

3. unsupervised
—> generative!



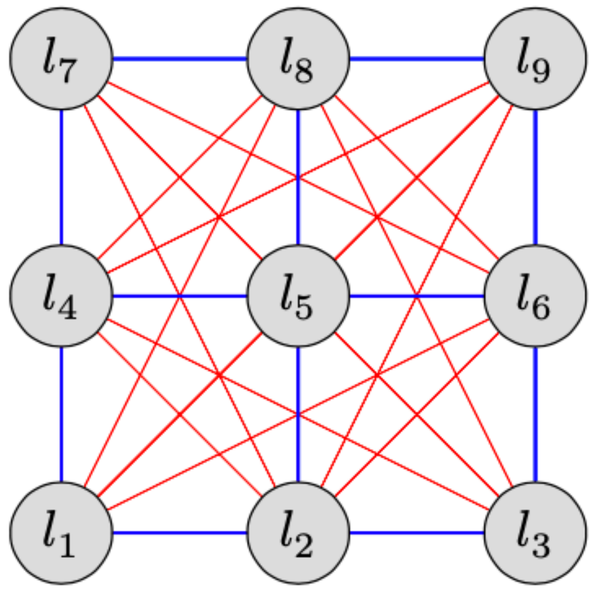


$$
\begin{aligned}
E_\pi &: \min \mathcal{L}_{rec} + \lambda_{\text{variational}} \mathrm{KL}\left(q(\pi|x)\|p(\pi)\right) + \lambda_{\text{adversarial}} I_T(\pi, \alpha) \\
E_\alpha &: \min \mathcal{L}_{rec} \\
D_S &: \min \mathcal{L}_{rec} + \lambda_{\text{GMRF}} \mathrm{KL}\left(q(l|x)\|p(l)\right) + \lambda_{H(p)} H(p)
\end{aligned}
$$

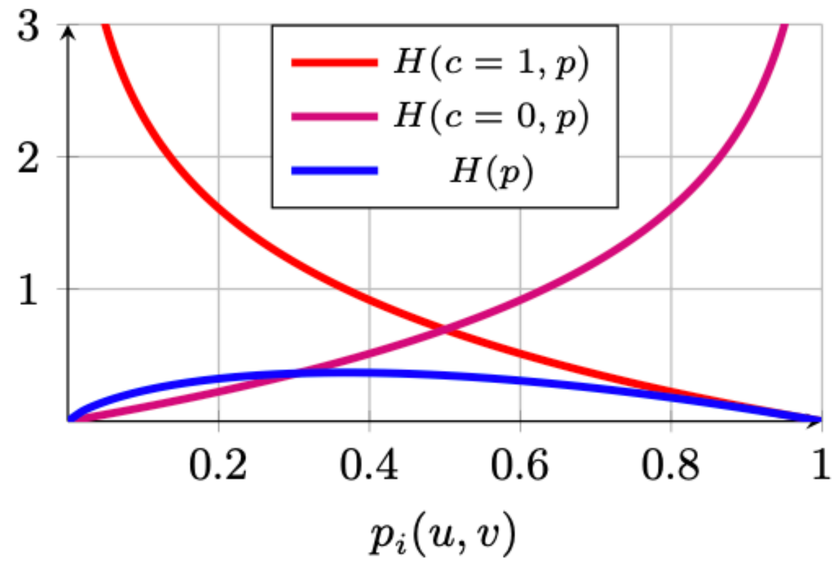
Shape and appearance disentanglement

Priors on segmentation variable S

Generative model, hence unsupervised



(a) **Gaussian Markov Random Field**. Without any prior assumptions, any image pixel is dependent on any other pixel, (**dense** connectivity). In a GMRF, we only allow adjacent pixels to interact (**sparse** connectivity).



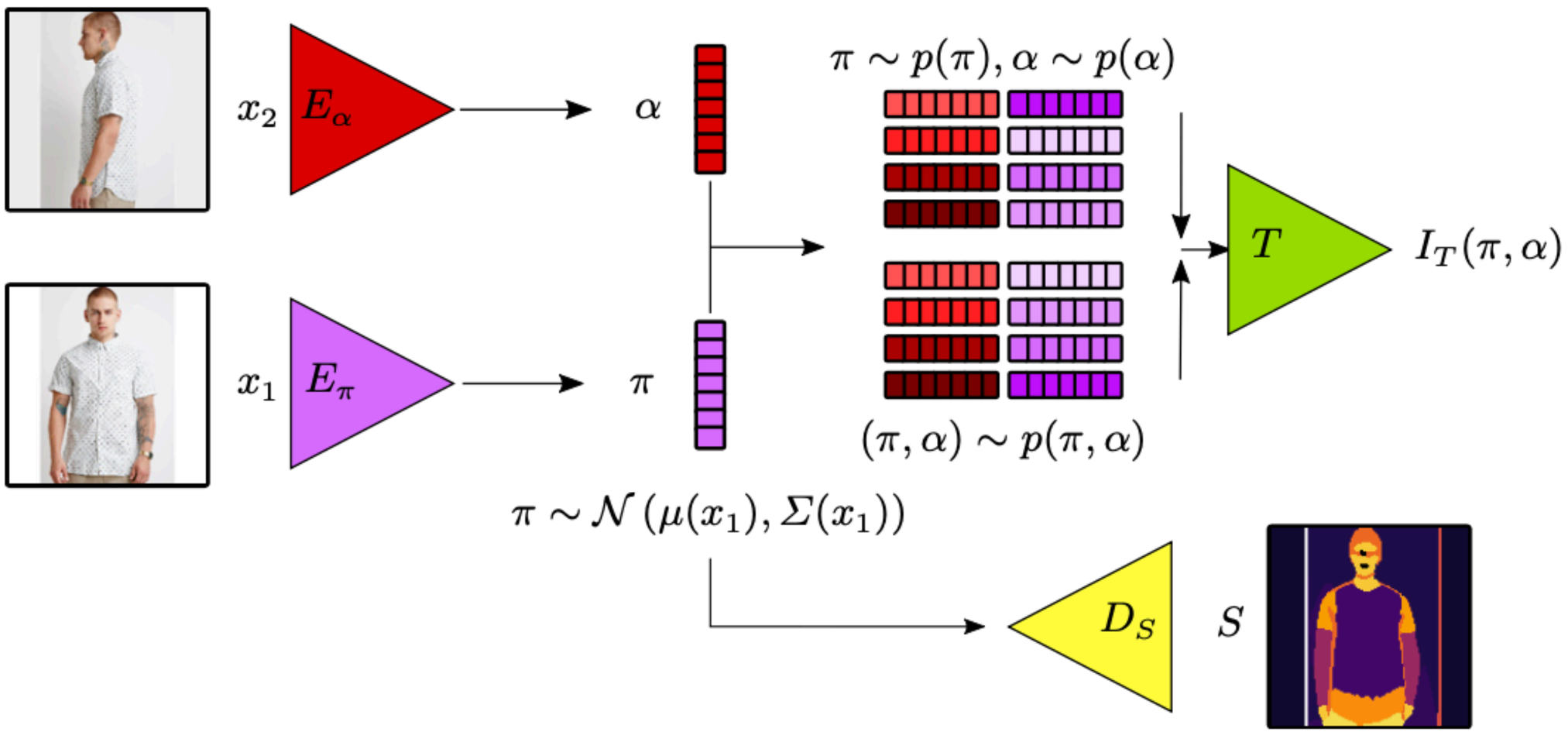
(b) **Entropy Regularization**. To keep the learned segmentation S close to a categorical distribution, we regularize the entropy of part probabilities $p_i(u,v)$.

„smooth“

„hard“, i.e. categorical

$$D_S \;:\; \min\; \mathcal{L}_{rec} + \lambda_{\text{GMRF}}\,\text{KL}\left(q(l|x)\|p(l)\right) + \lambda_{H(p)}H(p)$$

Priors on segmentation variable S

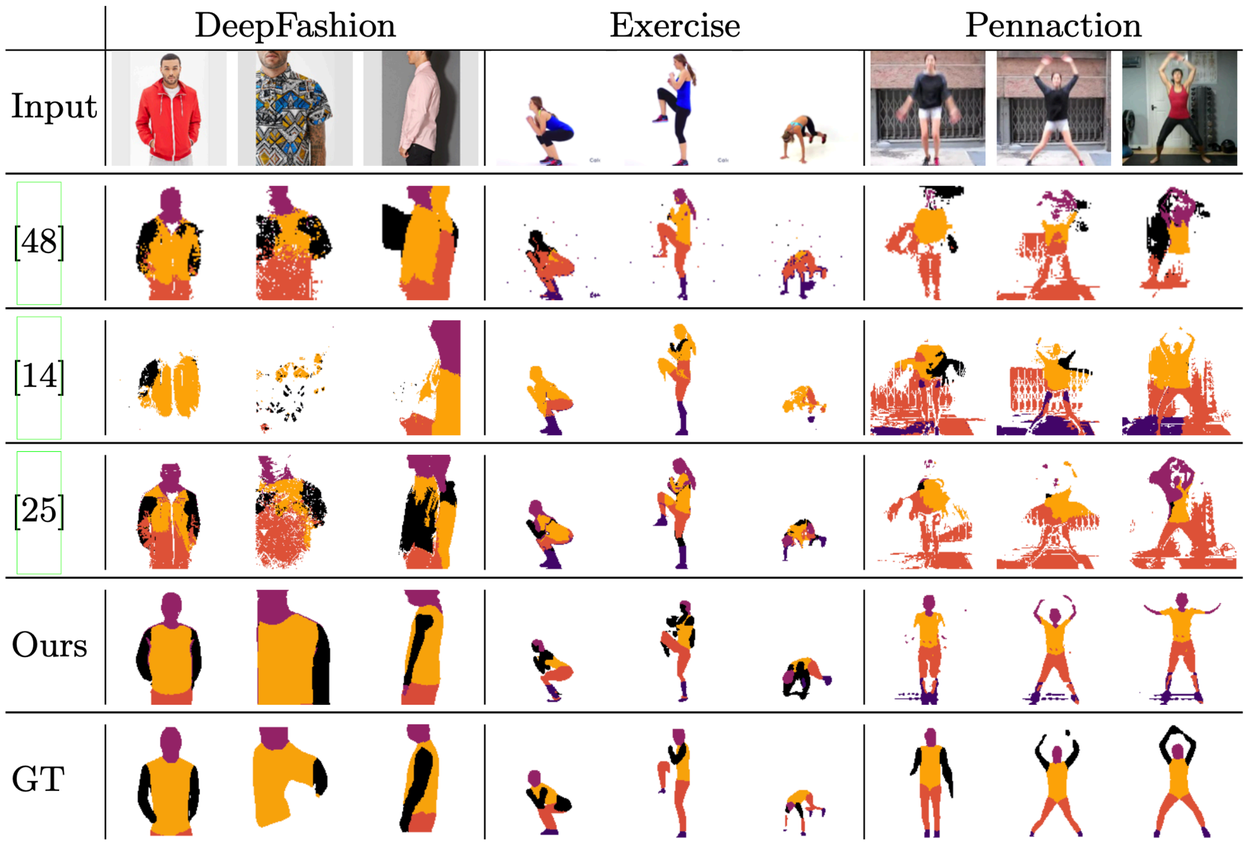


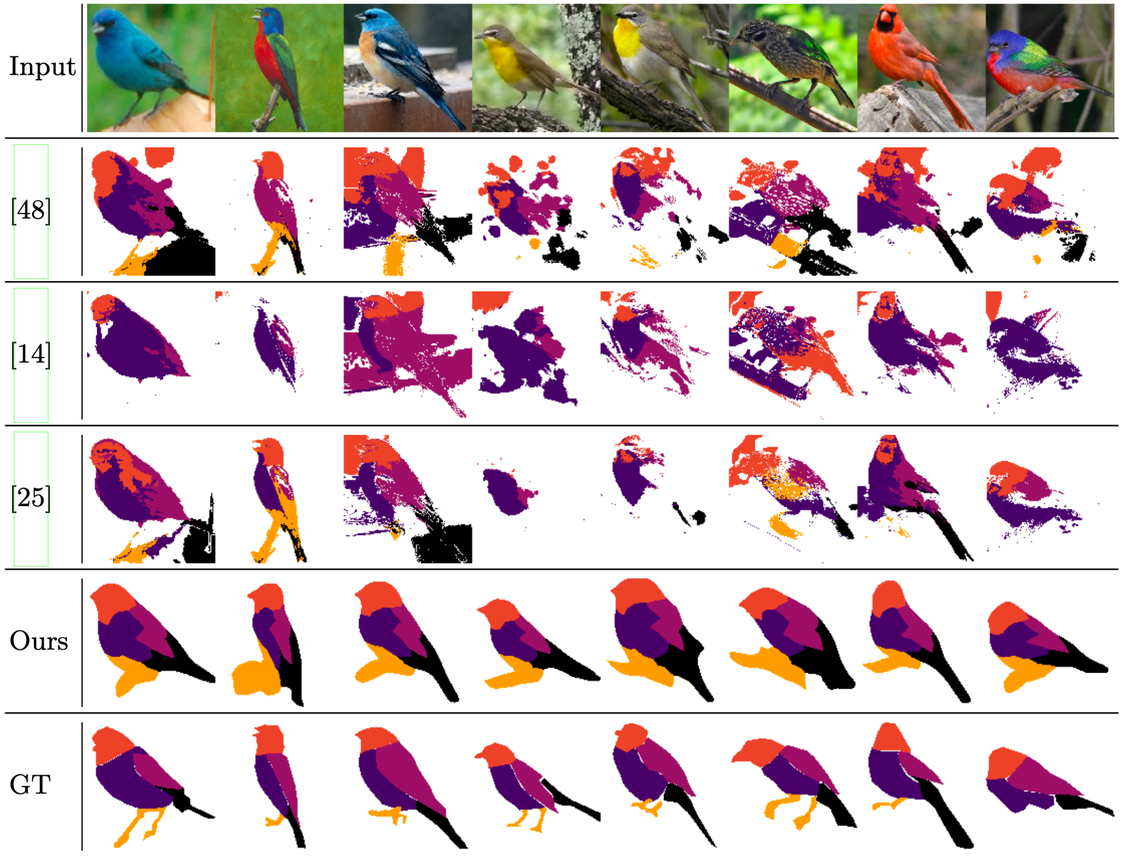
$$E_\pi \ : \ \min \ \mathcal{L}_{rec} + \lambda_{\text{variational}} \, \text{KL}\left(q(\pi|x) \| p(\pi)\right) + \lambda_{\text{adversarial}} I_T(\pi, \alpha)$$

Shape and appearance disentanglement
Esser et al., 2019 https://arxiv.org/abs/1910.10223

	disentanglement	GMRF	$\mathcal{L}_{H(p)}$			
						
1.	✗	✗	✗			
2.	variational + adversarial	✗	✗			
3.	variational + adversarial	✓	✗			
4.	only variational	✓	✓			
5.	Full Model, variational + adversarial	✓	✓			

Part Segmentation across Object Categories



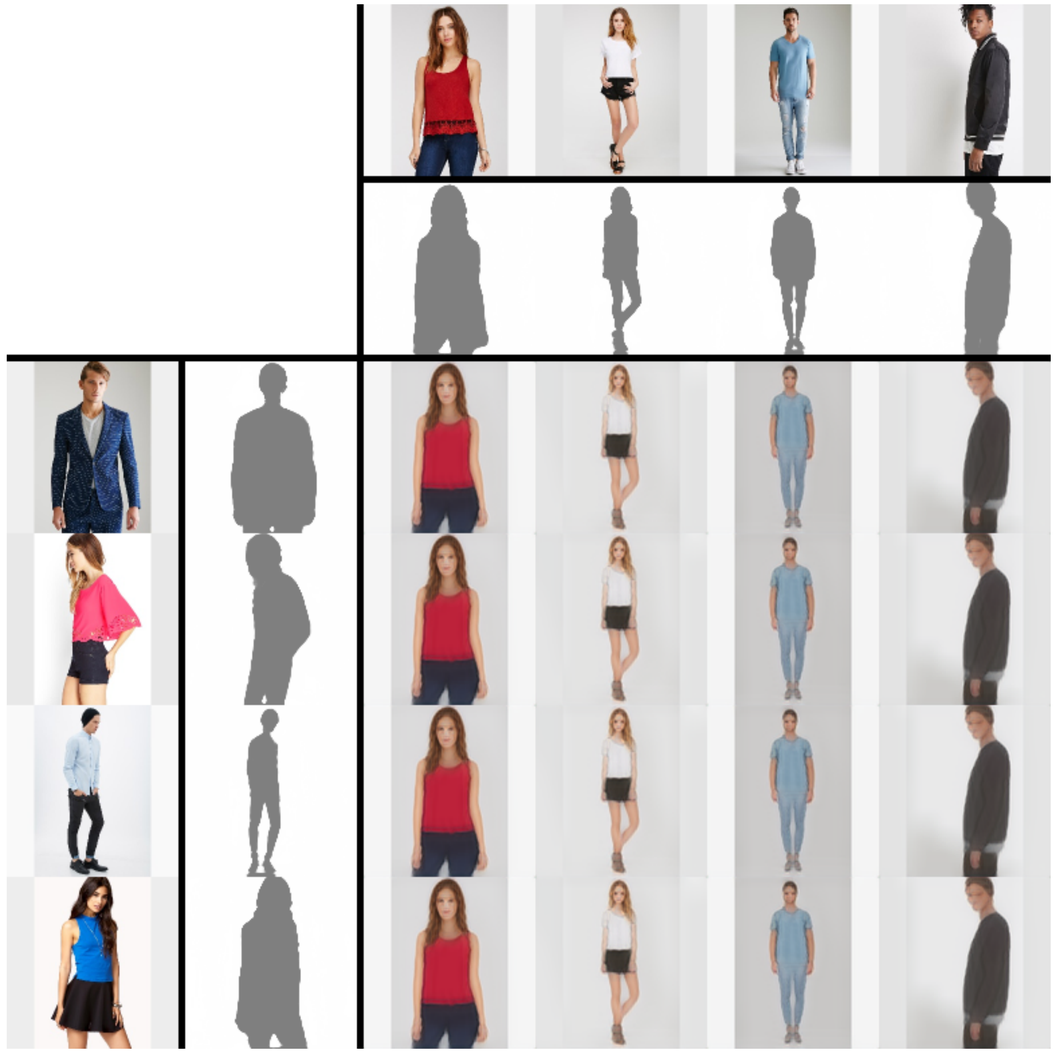


Part-based Appearance Transfer

transferring

no

appearance

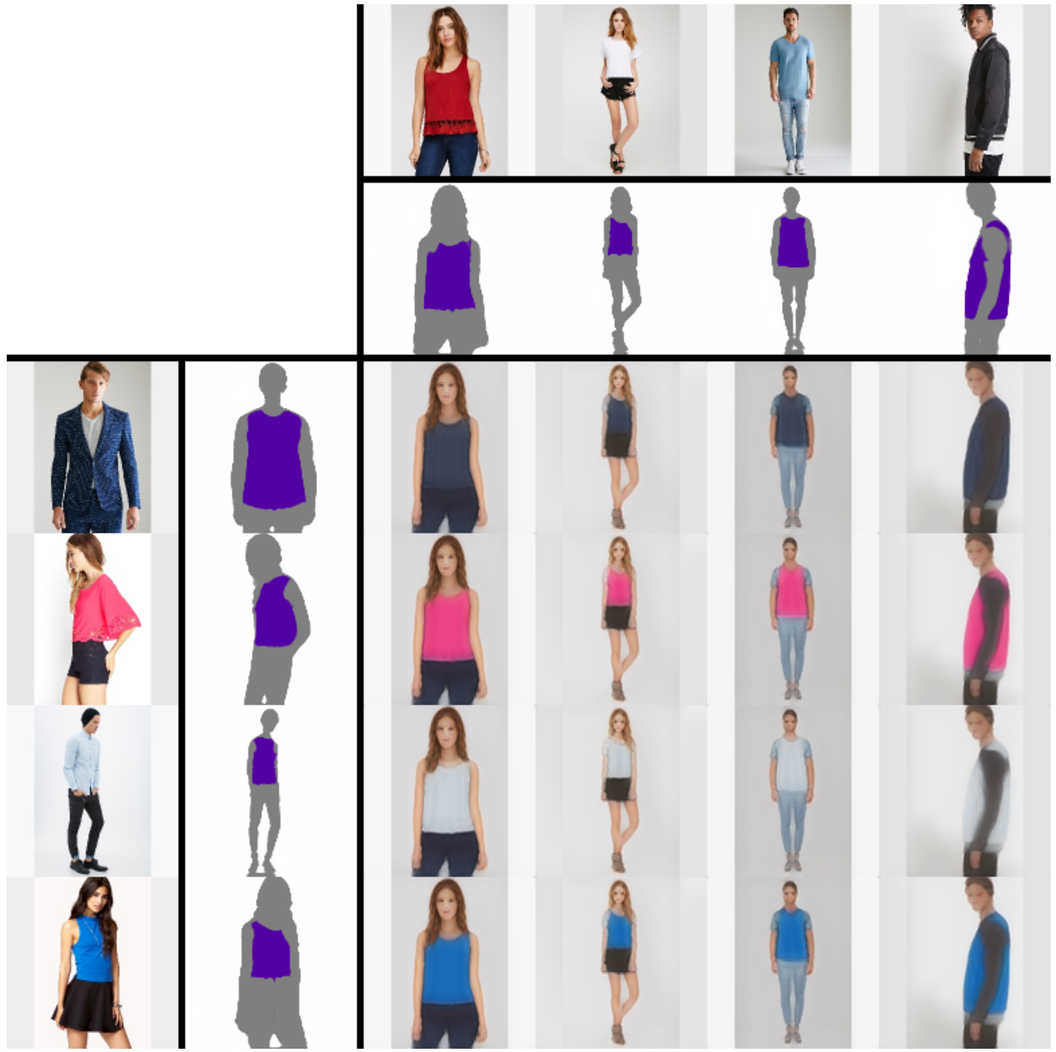


Part-based Appearance Transfer

transferring

chest

appearance

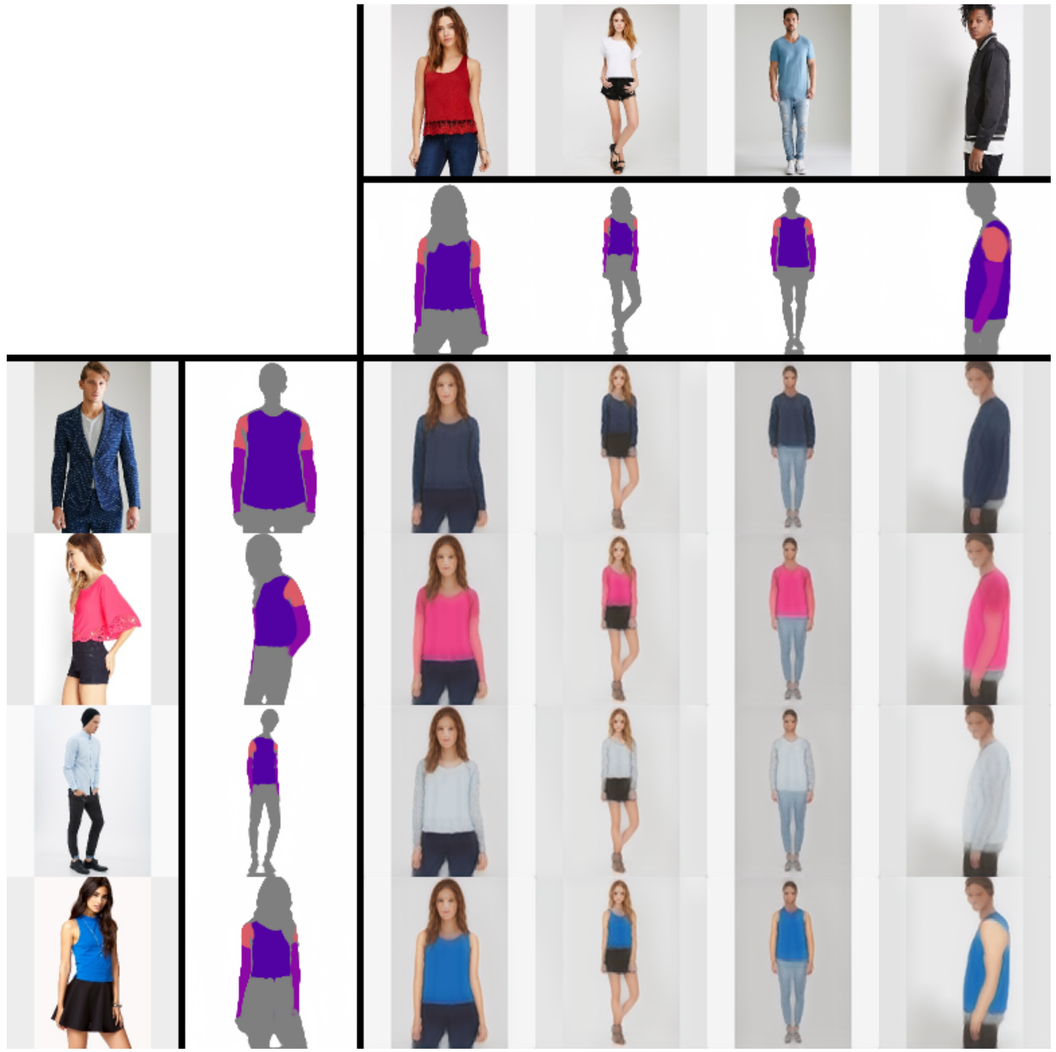


Part-based Appearance Transfer

transferring

chest and arm

appearance

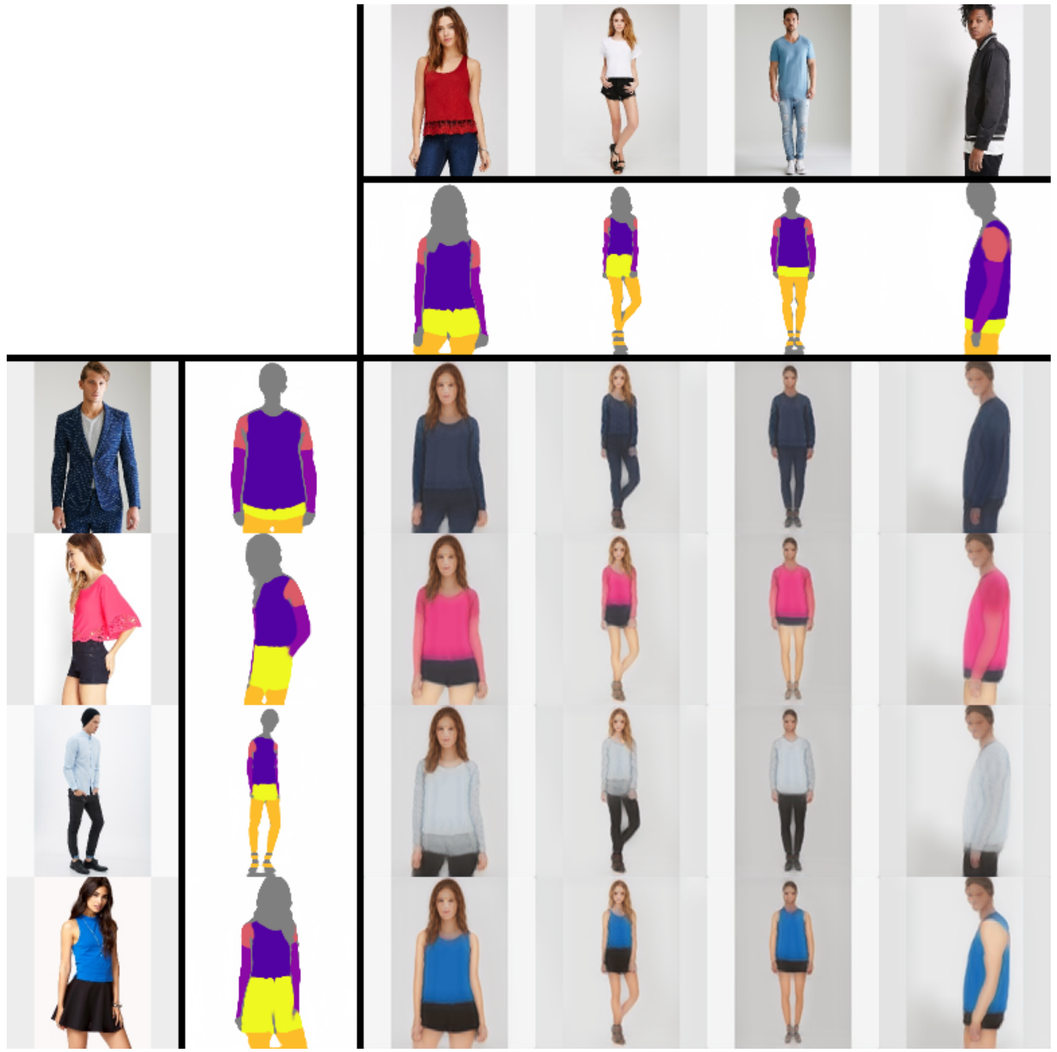


Part-based Appearance Transfer

transferring

chest, arm, hip and leg

appearance



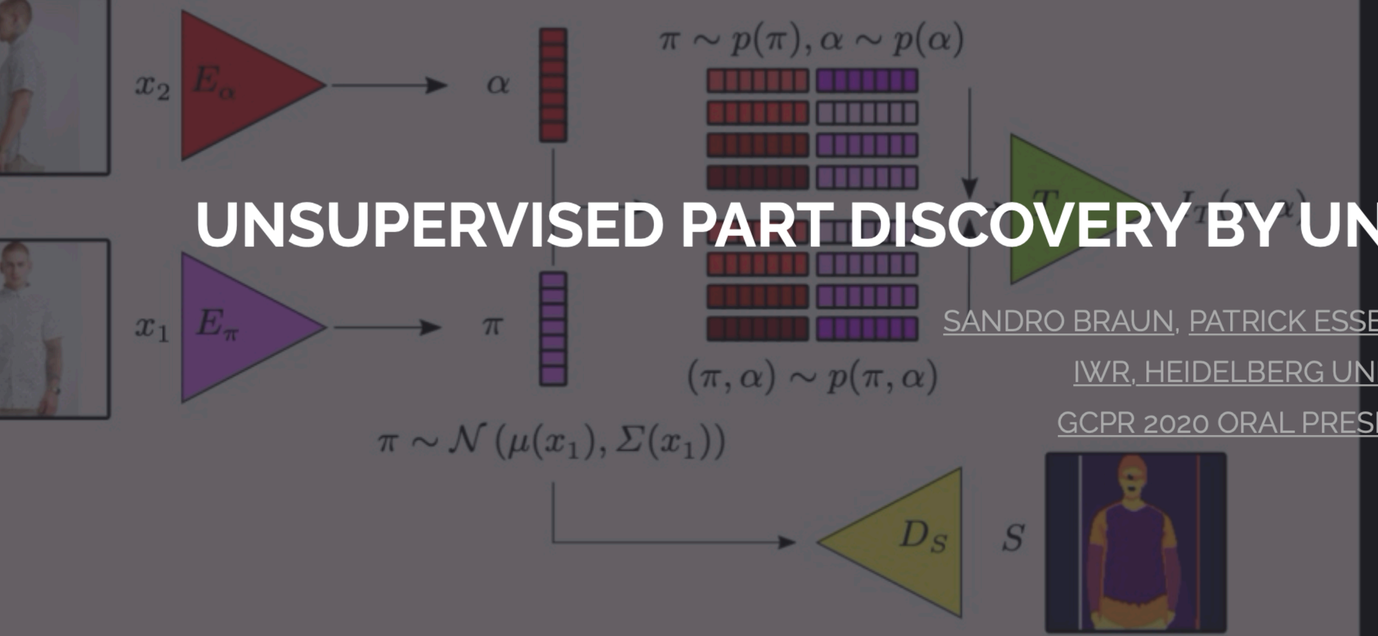
Quantitative Evaluation

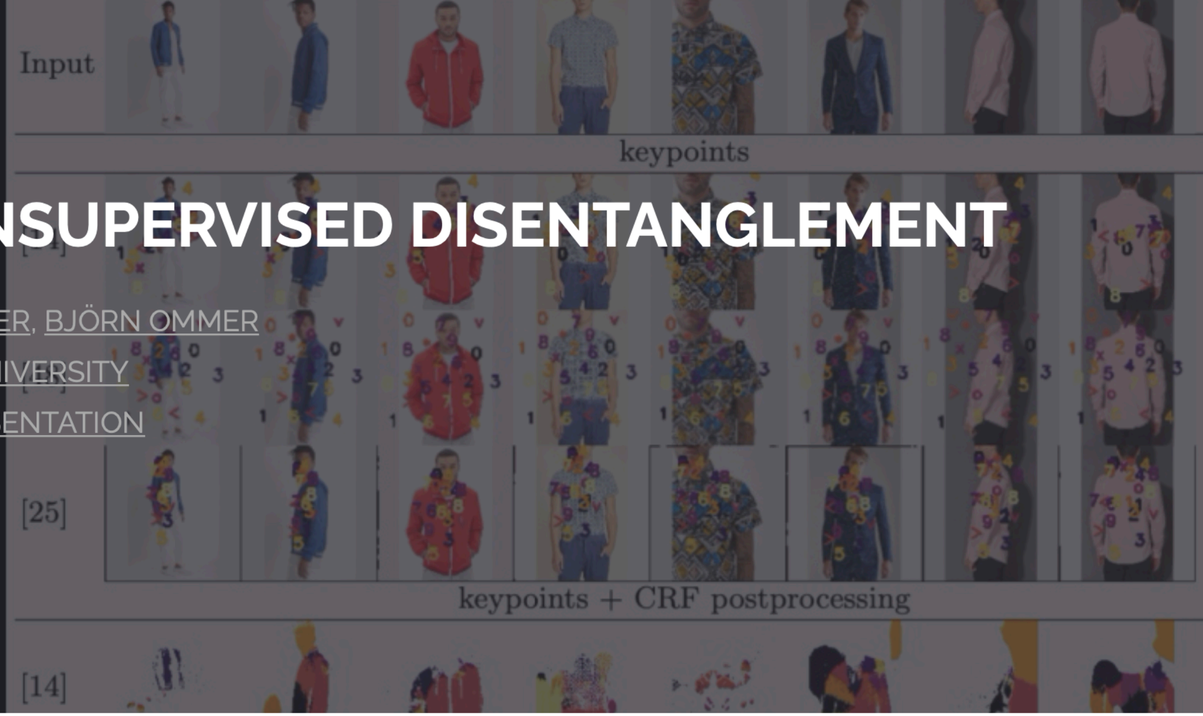
Dataset	Method	Arms	Feet	Head	Legs	Torso	Overall
DeepFashion	[48] + CRF	0.194	0.000	0.598	0.293	0.376	0.292
DeepFashion	[14] + CRF	0.052	0.000	0.118	0.108	0.244	0.104
DeepFashion	[25] + CRF	0.215	0.000	**0.606**	0.309	0.322	0.290
DeepFashion	Ours	**0.508**	0.000	0.530	**0.500**	**0.722**	**0.452**
Exercise	[48] + CRF	0.043	**0.230**	0.096	0.433	0.335	0.227
Exercise	[14] + CRF	0.101	0.190	0.000	0.469	0.357	0.223
Exercise	[25] + CRF	0.212	0.213	**0.366**	**0.445**	0.441	**0.336**
Exercise	Ours	**0.253**	0.104	0.340	0.428	**0.504**	0.326
Pennaction	[48] + CRF	0.066	0.000	**0.327**	0.379	0.442	0.243
Pennaction	[14] + CRF	0.050	**0.122**	0.000	0.316	0.455	0.189
Pennaction	[25] + CRF	0.038	0.000	0.105	0.312	0.402	0.171
Pennaction	Ours	**0.094**	0.101	0.237	**0.371**	**0.484**	**0.257**

Improved Segmentation performance in terms of IoU

α	PCK@2.5 %	PCK@5 %	PCK@10 %
VU-Net [8]	31.64	54.90	80.83
Lorenz et al. [25]	14.50	37.50	69.63
Ours	**41.56**	**65.76**	**83.12**

Improved shape consistency in terms of PCK





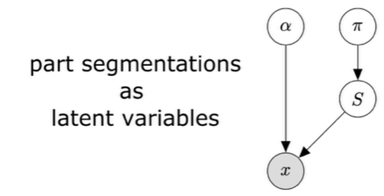
UNSUPERVISED PART DISCOVERY BY UNSUPERVISED DISENTANGLEMENT

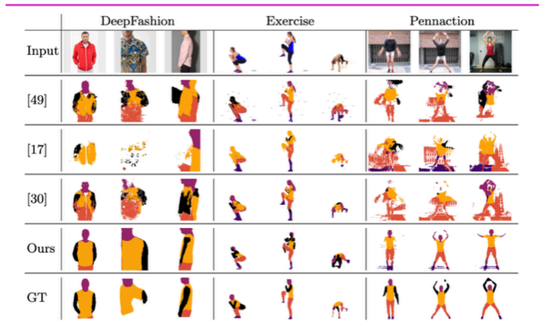
SANDRO BRAUN, PATRICK ESSER, BJÖRN OMMER

IWR, HEIDELBERG UNIVERSITY

GCPR 2020 ORAL PRESENTATION

-page: compvis.github.io/unsupervised-part-segmentation





Our part segmentation method leverages a disentangled representation for shape and appearance to discover semantic part

ABSTRACT

We address the problem of discovering part segmentations of articulated objects without supervision. In contrast to keypoints, part segmentations provide information about part localizations on the level of individual pixels. Capturing both locations and semantics, they are an attractive target for supervised learning approaches. However, large annotation costs limit the scalability of supervised algorithms to other object categories than humans. Unsupervised approaches potentially allow to use much more data at a lower cost. Most existing unsupervised approaches focus on learning abstract representations to be refined with supervision into the final representation. Our approach leverages a generative model consisting of two disentangled representations for an object's