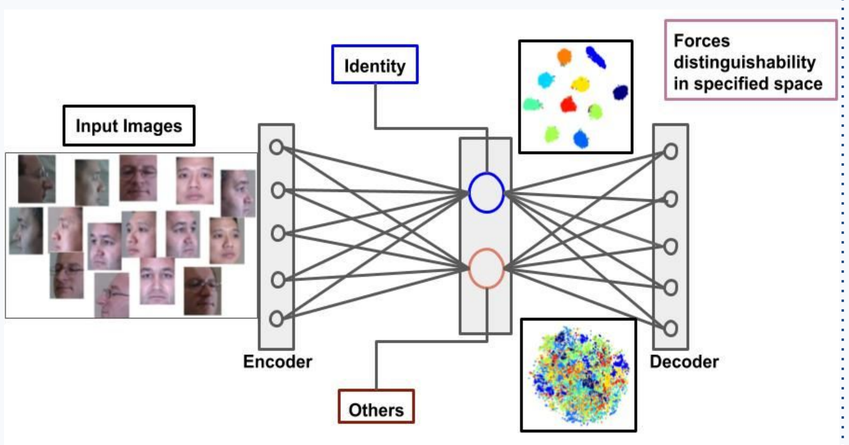
PrOSe: **Pr**oduct of **O**rthogonal **S**phere**s** Parameterization for Disentangled Representation Learning

Ankita Shukla*, Sarthak Bhagat*, Shagun Uppal*, Saket Anand*, Pavan Turaga^

*IIIT- Delhi, ^Arizona State University

Motivation

- **Disentangled Representations** encode different attributes in different dimensions

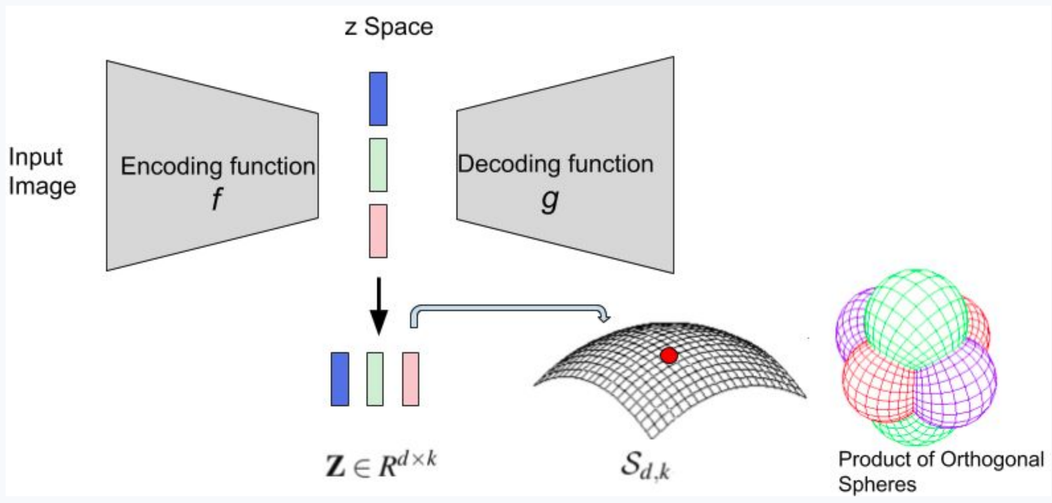


- **Physical Factors of Image Formation**
  - Lighting Variables
  - Pose Variables
  - Deformation Variables

Is hyper - spherical a good model for factors beyond ...above?

- Semantic factors can be approximated to a good degree from low-level physical factors

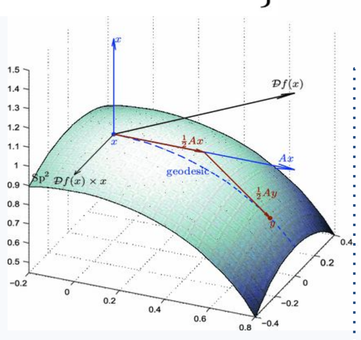
Approach Overview



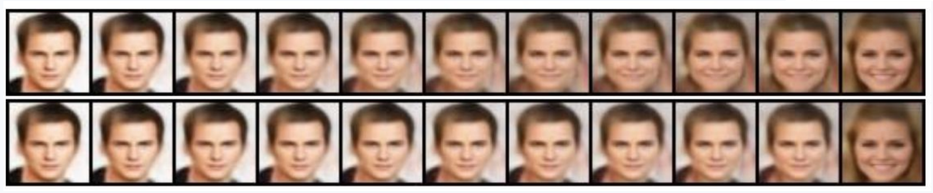
- Stiefel Manifold

$$\mathcal{S}_{d,k} = \{\mathbf{U} \in \mathbb{R}^{d \times k} : \mathbf{U}^\top \mathbf{U} = \mathbf{I}\}$$

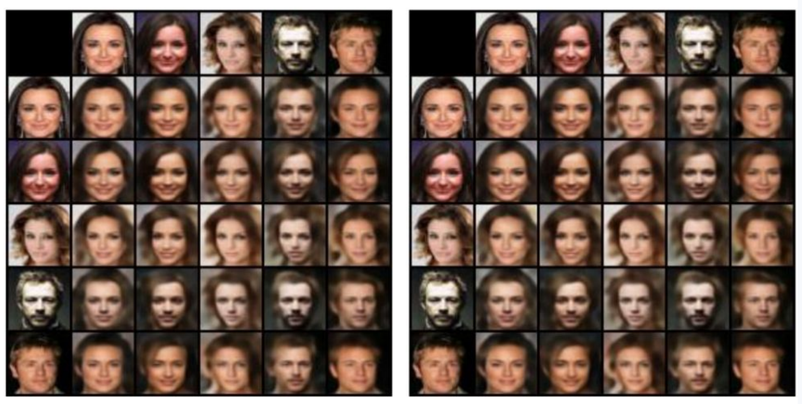
Cayley transformation for preserving the orthogonality



- Interpolation along identity partition



- Image synthesis/attribute transfer



Formulation and Optimization

$$\min_{\theta} \quad \mathcal{L}_{disentangle} + \mathcal{L}_{orth}$$

$$\mathcal{L}_{orth} = \left\| \mathbf{Z}^\top \mathbf{Z} - \mathbf{I} \right\|_F^2$$

$$\mathbf{Z} \in \mathbb{R}^{d \times k} = [\mathbf{z}^1 \mathbf{z}^2 \ldots, \mathbf{z}^k]$$

Optimize the Stiefel Manifold Constraint

$$\mathbf{Z}_{new} = \left(\mathbf{I} + \frac{\tau}{2}\mathbf{A}\right)^{-1} \left(\mathbf{I} - \frac{\tau}{2}\mathbf{A}\right)\mathbf{Z},$$

$$\mathbf{A} = \mathbf{J}\mathbf{Z}^\top - \mathbf{Z}\mathbf{J}^\top$$

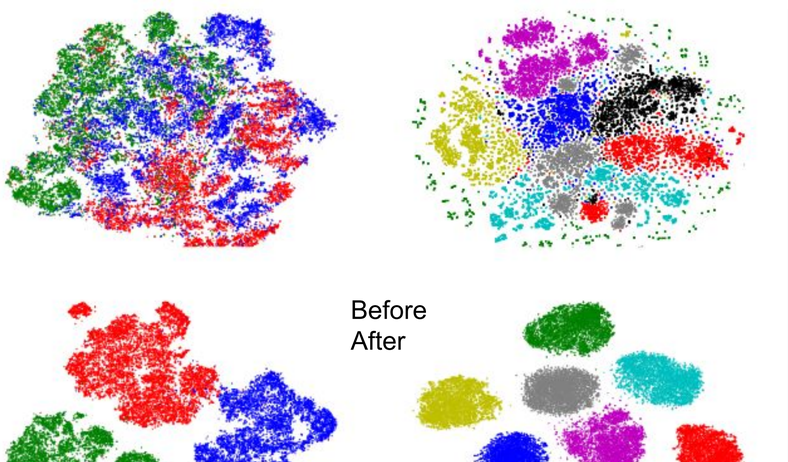
Results

Method	Body	Skin	Vest	Hair	Arm	Leg	Avg.
MIX	66.5	77.2	90.0	56.2	63.1	89.4	73.7
+ **PrOSe**	70.1	75.5	88.5	63.2	72.1	94.4	**77.3**
β-VAE	66.8	78.4	89.7	57.2	63.3	90.1	74.0
+ **PrOSe**	70.2	77.3	88.1	65.3	73.0	93.7	**77.7**
Factor-VAE	66.7	78.6	90.3	57.3	63.1	90.1	74.4
+ **PrOSe**	69.8	77.1	89.0	63.8	72.7	93.7	**77.7**

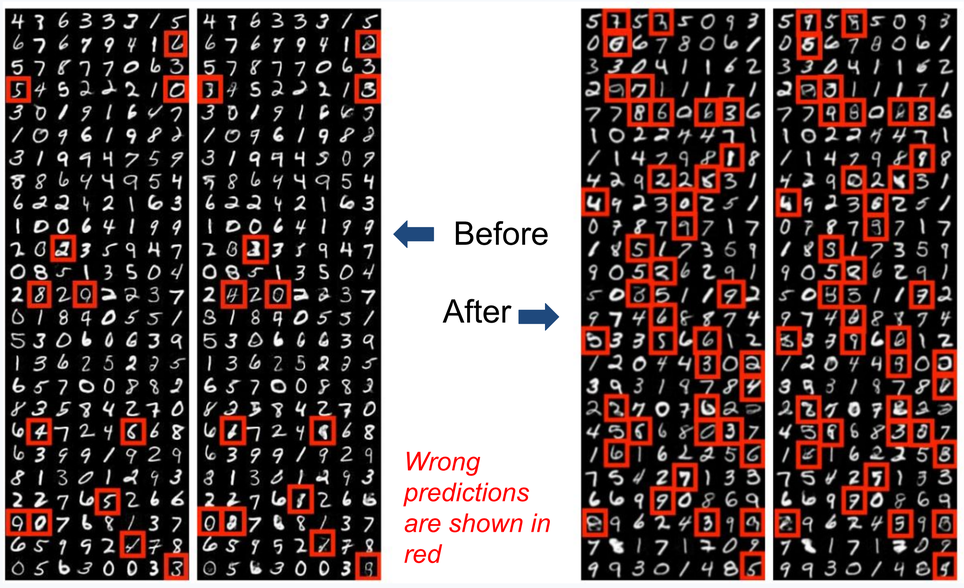
mAP values for different attributes with and without ProSe

Attribute	MIX	+**PrOSe**	β-VAE	+**PrOSe**	Factor-VAE	+ **PrOSe**
Eyebrows	79.4	79.5	78.8	79.4	79.4	79.8
Attractive	72.6	80.4	73.6	74.6	74.4	76.2
Bangs	91.7	90.2	92.6	92.8	93.2	94.2
Black Hair	71.9	75.6	72.4	78.2	76.5	78.1
Blonde Hair	87.2	92.0	88.4	90.2	88.2	88.4
Makeup	76.5	78.0	76.2	77.8	77.1	77.5
Male	86.2	83.1	84.6	86.4	83.8	83.2
Mouth	72.0	80.6	73.8	74.1	74.4	75.0
No Beard	86.3	89.6	86.0	86.3	85.2	85.8
Wavy Hair	65.7	71.9	66.2	67.0	66.8	67.0
Hat	95.2	94.8	93.5	95.4	93.0	93.8
Lipstick	79.8	80.5	82.4	83.2	83.3	87.5
Average	80.3	**83.0**	80.7	**82.1**	81.3	**82.2**

Effect on Latent Space



- How well have we disentangled



References

- Higgins, Irina, et al. "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework." *ICLR* 2.5 (2017): 6.
- Kim, Hyunjik, and Andriy Mnih. "Disentangling by Factorising." *International Conference on Machine Learning*. 2018.
- Hu, Qiyang, et al. "Disentangling factors of variation by mixing them." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018