SCOPS: Self-Supervised Co-Part Segmentation

Wei-Chih Hung[1], Varun Jampani[2], Sifei Liu[2], Pavlo Molchanov[2]
Ming-Hsuan Yang[1] and Jan Kautz[2]
[1] UC MERCED [2] NVIDIA

Code: https://varunjampani.github.io/scops/

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

1 Object Part Segmentation

Advantages

- Natural dense representation for non-rigid objects.
- Landmarks are sometimes ill-defined around part boundaries.

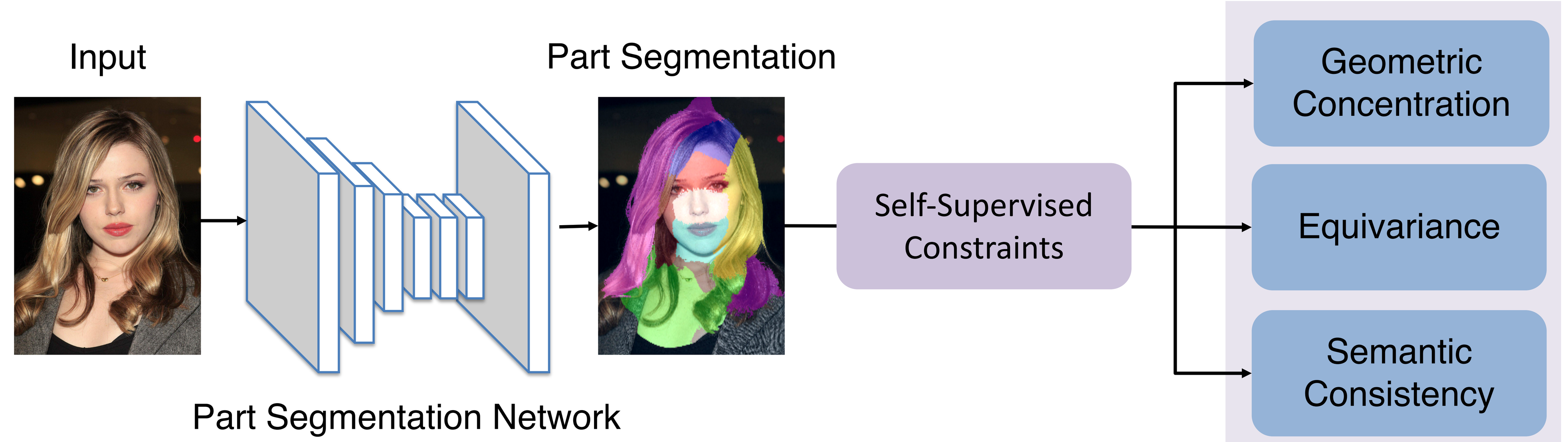
Supervised Methods

- Defining semantic part class is hard.
- Collecting annotation is also hard. [PASCAL-Part]

Proposed Method – Self-Supervised Co-Part Segmentation

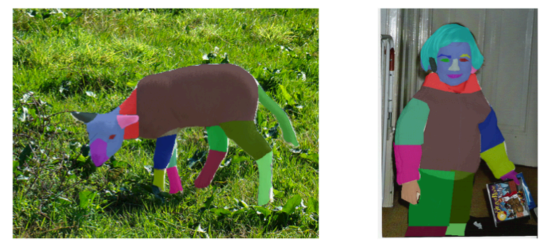
- Given an image collection of an object category, the objective is to discover consistent part segments across images.

A end-to-end trainable method that learns object part segmentation without pixel-wise annotation.



2 Desired Properties as Loss Functions

- Geometric concentration.
- Robustness to variations.
- Semantic consistency.
- Objects as union of parts.



[PASCAL-Part Dataset]

3 Concentration and Equivariance Loss

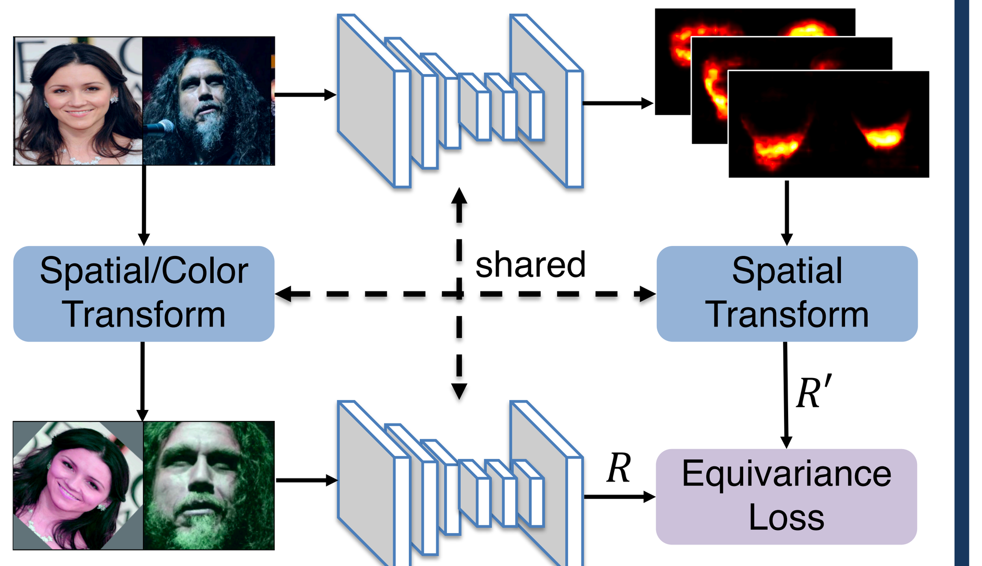
Concentration Loss

- Penalize part responses in areas far from part centric:

$$\mathcal{L}_{con} = \sum_k \sum_{u,v} ||\underbrace{\langle u, v\rangle}_{\text{Pixel Pos.}} - \underbrace{\langle c_u^k, c_v^k\rangle}_{\text{Part Centric}}||^2 \cdot \underbrace{\mathbf{R}(k,u,v)}_{\text{Part Response}}/z_k$$

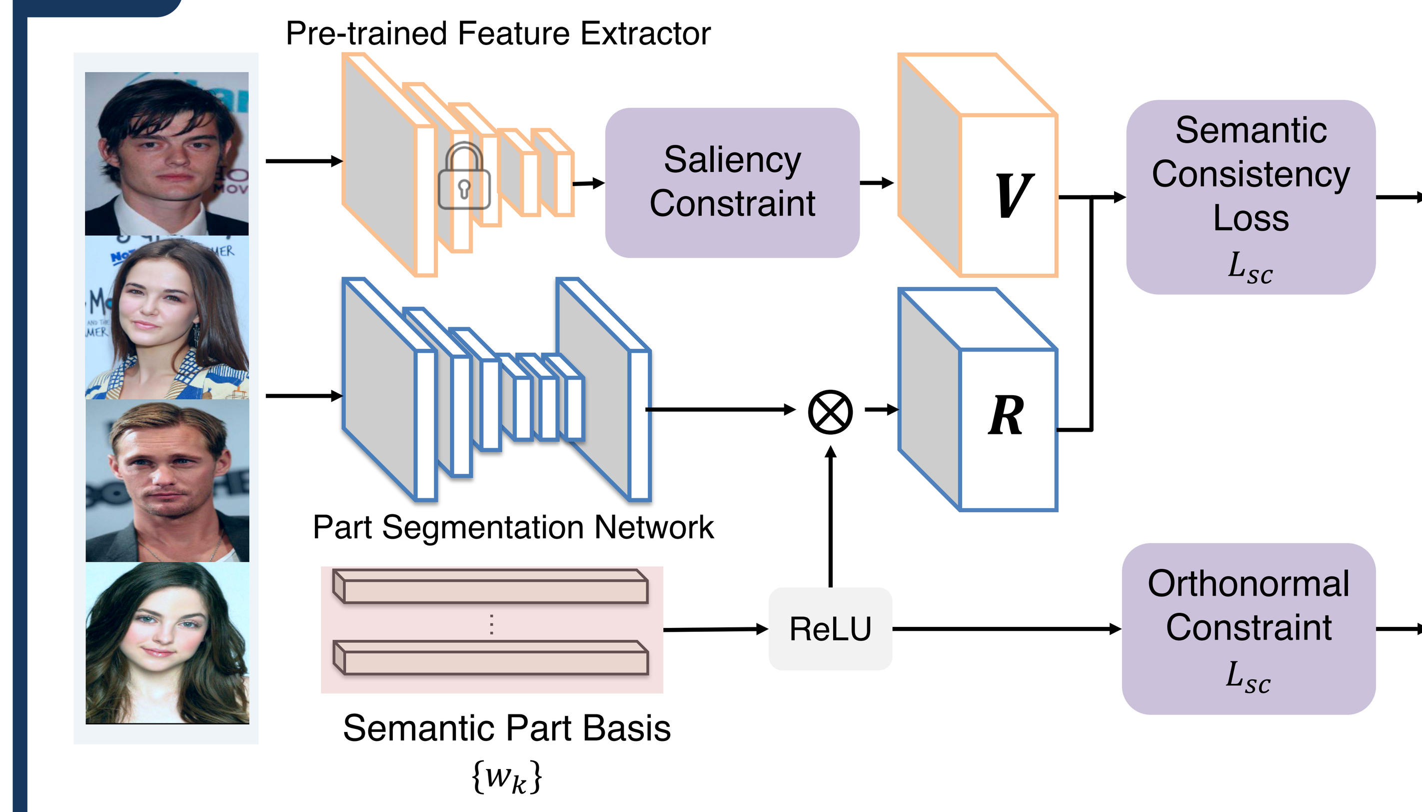
Equivariance Loss

- Segmentation should be consistent w.r.t. perturbation.
- Proposed in unsupervised landmark estimation, we extend it to segmentation.



$$\mathcal{L}_{eqv} = \underbrace{\lambda_{eqv}^s D_{KL}(\mathbf{R}'||T_s(\mathbf{R}))}_{\text{Segmentation Equivariance}} + \lambda_{eqv}^c \sum_k \underbrace{||\langle c_u^{k'}, c_v^{k'}\rangle - T_s(\langle c_u^k, c_v^k\rangle)||^2}_{\text{Part Centric Equivariance}}$$

4 Semantic Consistency Loss



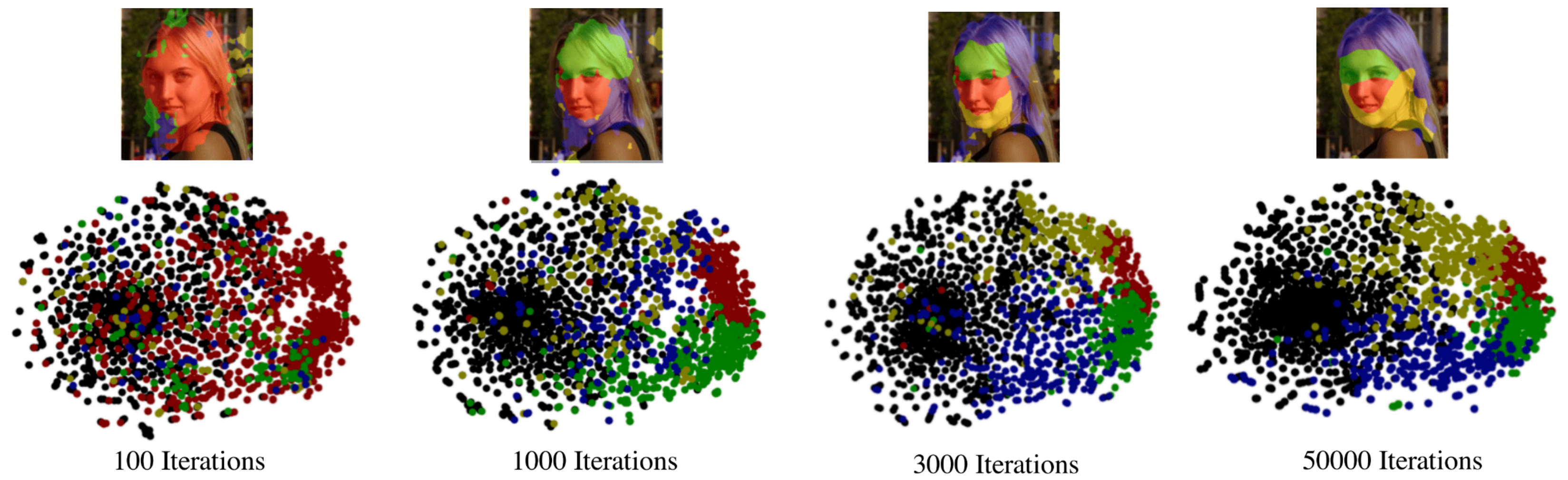
Densely reconstruct reference features with a set of trainable part basis $\{w_k\}$.

$$\mathcal{L}_{sc} = \sum_{u,v} ||\underbrace{\mathbf{V}(u,v)}_{\text{Reference Feature}} - \sum_k \underbrace{\mathbf{R}(k,u,v)}_{\text{Part Response to } \{w_k\}}\mathbf{w}_k||^2$$

Push semantic part basis $\{w_k\}$ to be an orthonormal basis.

$$\mathcal{L}_{on} = ||\underbrace{\hat{\mathbf{W}}\hat{\mathbf{W}}^T}_{\text{Normalized Basis Matrix}} - \underbrace{\mathbb{I}_K}_{\text{Identity Matrix}}||_F^2$$

Simultaneously learn part basis and pixel clusters



5 Results Visualization

6 Experimental Results

Landmarks on CelebA (Faces)

Method	Error (%)
ULD (K=8)	40.82
DFF (K=8)	31.30
SCOPS (K=4)	21.76
SCOPS (K=8)	15.01

Landmarks on CUB (Birds)

Method	CUB-001	CUB-002	CUB-003
ULD	30.12	29.36	28.19
DFF	22.42	21.62	21.98
SCOPS	18.50	18.82	21.07

Foreground Segmentation on PASCAL (IOU)

class	horse	cow	sheep	aero	bus	car	motor
DFF	49.51	56.39	51.03	48.38	58.63	56.48	54.80
DFF+CRF	50.96	57.64	52.29	50.87	58.64	57.56	55.86
SCOPS	55.76	60.79	56.95	69.02	73.82	65.18	58.53
SCOPS+CRF	**57.92**	**62.70**	**58.17**	**80.54**	**75.32**	**66.14**	**59.15**

References

ULD-1: Thewlis, J. et al. Unsupervised learning of object landmarks by factorized spatial embeddings. In *ICCV*, 2017.

ULD-2: Zhang, Y. et al. Unsupervised discovery of object landmarks as structural representations. In *CVPR*, 2018.

DFF: Collins E. et al. Deep feature factorization for concept discovery. In *ECCV*, 2018.