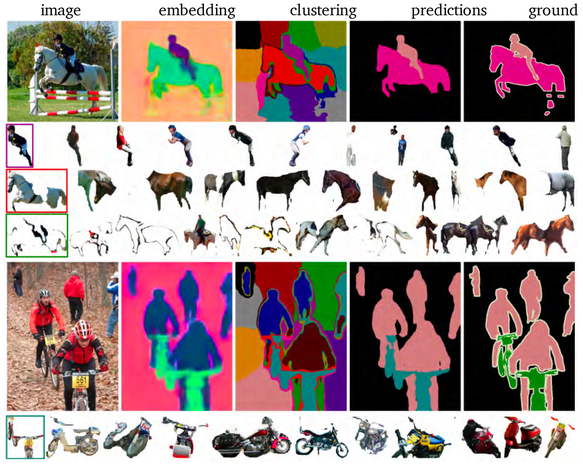
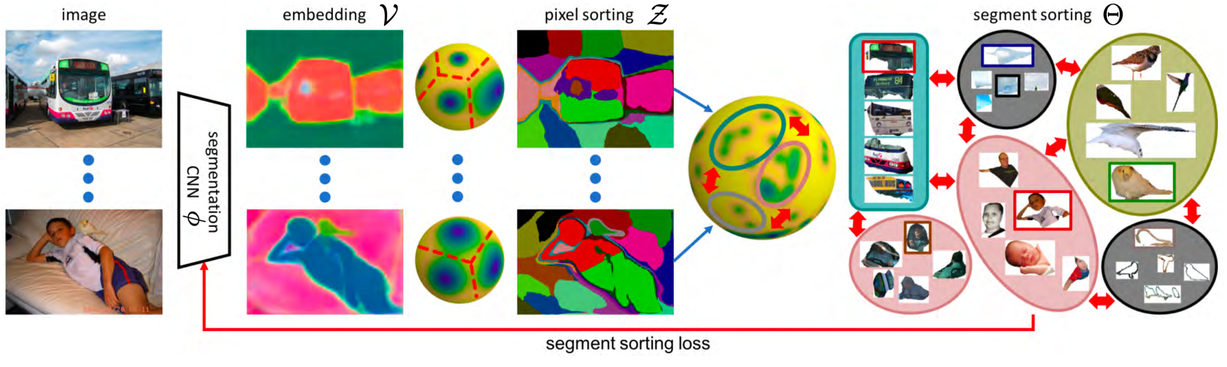
SegSort: Segmentation by Discriminative Sorting of Segments

Jyh-Jing Hwang, Stella X. Yu, Jianbo Shi,
Maxwell D. Collins, Tien-Ju Yang, Xiao Zhang, Liang-Chieh Chen

Supervised Segmentation

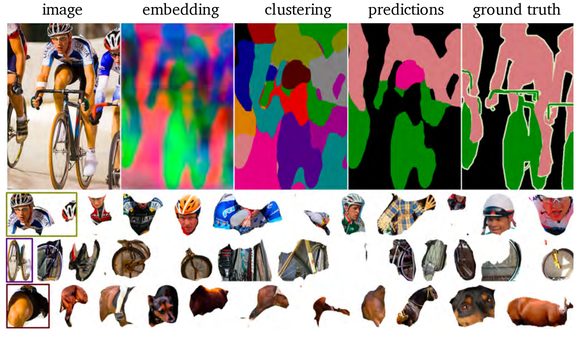


SegSort: Unified Metric Learning Framework for Segmentation

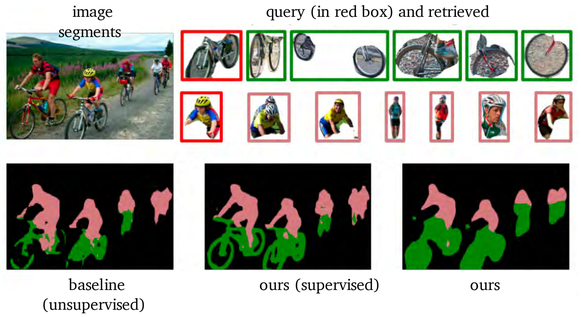


- Basic idea: Assuming independent normal distributions for individual segments, we seek a maximum likelihood estimation of the feature mapping $\mathcal{V}$, so that the feature induced partitioning $\mathcal{Z}$ in the image and Θ clustering across images provide maximum discrimination among segments.
- von-Mises Fisher distribution: $f(\boldsymbol{v} \mid \boldsymbol{\mu}, \kappa) = C_d(\kappa) \exp(\kappa \boldsymbol{\mu}^\top \boldsymbol{v})$
- Single MLE objective: $\min_{\phi, \mathcal{Z}, \Theta} -\log P(\mathcal{V}, \mathcal{Z} \mid \Theta) = \min_{\phi, \mathcal{Z}, \Theta} -\sum_i \log f_{z_i}(\boldsymbol{v}_i \mid \theta_{z_i})$
- Optimize in a two-stage EM: $[\ \mathcal{Z} \circlearrowright \Theta\ ] \circlearrowright \mathcal{V}$

Unsupervised Segmentation



Contributions



1. First deep end-to-end non-parametric segmentation
2. First deep unsupervised semantic segmentation
3. Interpretability from retrieved segments

Pixel Sorting: Partition Each Image

- EM for K-mean clustering based on $\mathcal{V}$ for each image
- Update $\mathcal{Z}$: $p(z_i = s | \boldsymbol{v}_i, \Theta) = \frac{f_s(\boldsymbol{v}_i \mid \Theta)}{\sum_{l=1}^k f_l(\boldsymbol{v}_l | \Theta)}$
- Update Θ: $\hat{\mu}_c = \frac{\sum_i \boldsymbol{v}_i p(z_i = c \mid \boldsymbol{v}_i, \Theta)}{||\sum_i \boldsymbol{v}_i p(z_i = c \mid \boldsymbol{v}_i, \Theta)||} = \frac{\sum_{i \in \mathcal{R}_c} \boldsymbol{v}_i}{||\sum_{i \in \mathcal{R}_c} \boldsymbol{v}_i||}$
- Alternative for unsupervised semantic segmentation:

Segment Sorting: Organize All Segments

- Unsupervised vMF Loss:
  - Maximize discrimination among segments.

$$L^i_{\text{vMF}} = -\log p_\phi(c \mid \boldsymbol{v}_i, \Theta) = -\log \frac{\exp(\kappa \boldsymbol{\mu}_c^\top \boldsymbol{v}_i)}{\sum_{l=1}^k \exp(\kappa \boldsymbol{\mu}_l^\top \boldsymbol{v}_i)}$$

- Supervised vMF-NCA Loss:
  - Maximize discrimination between *different-class* segments.

$$L^i_{\text{vMF-N}} = -\log \sum_{s \in C_i^+} p'_\phi(z_i = s \mid \boldsymbol{v}_i, \Theta) = -\log \frac{\sum_{s \in C_i^+} \exp(\kappa \boldsymbol{\mu}_s^\top \boldsymbol{v}_i)}{\sum_{l \neq c} \exp(\kappa \boldsymbol{\mu}_l^\top \boldsymbol{v}_i)}$$

- Memory bank caching segment prototypes across batches.