Universal Weakly Supervised Segmentation by Pixel-to-Segment Contrastive Learning

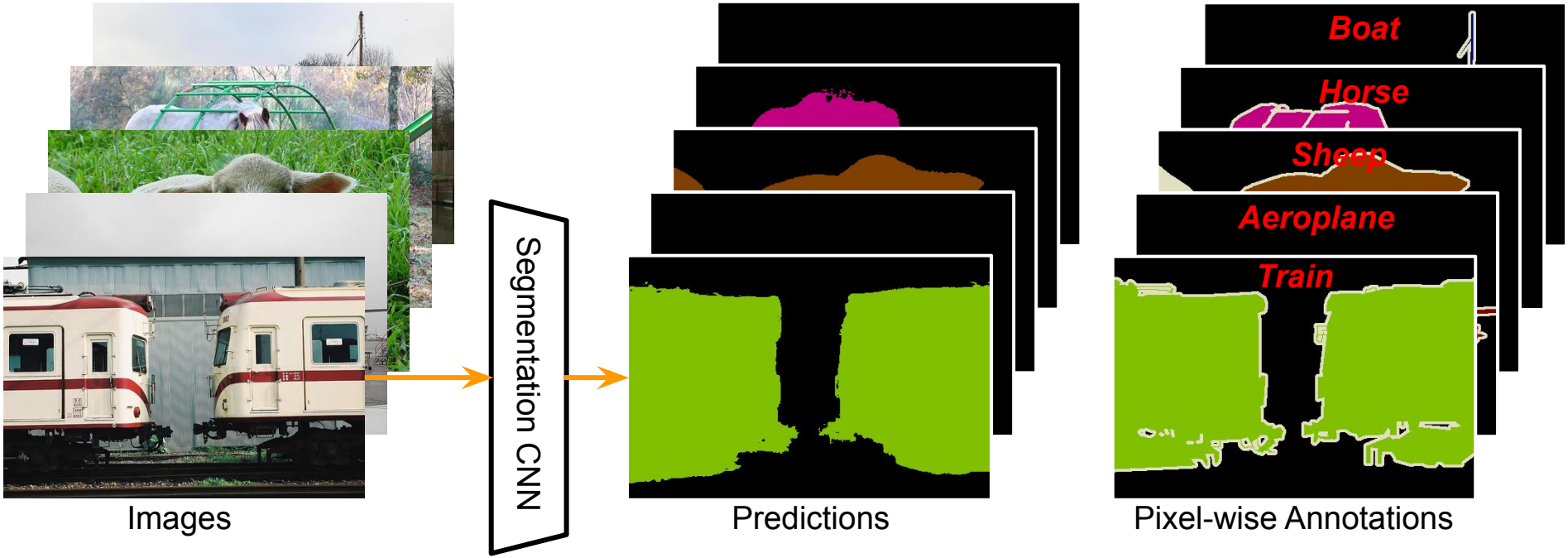
Tsung-Wei Ke

Jyh-Jing Hwang

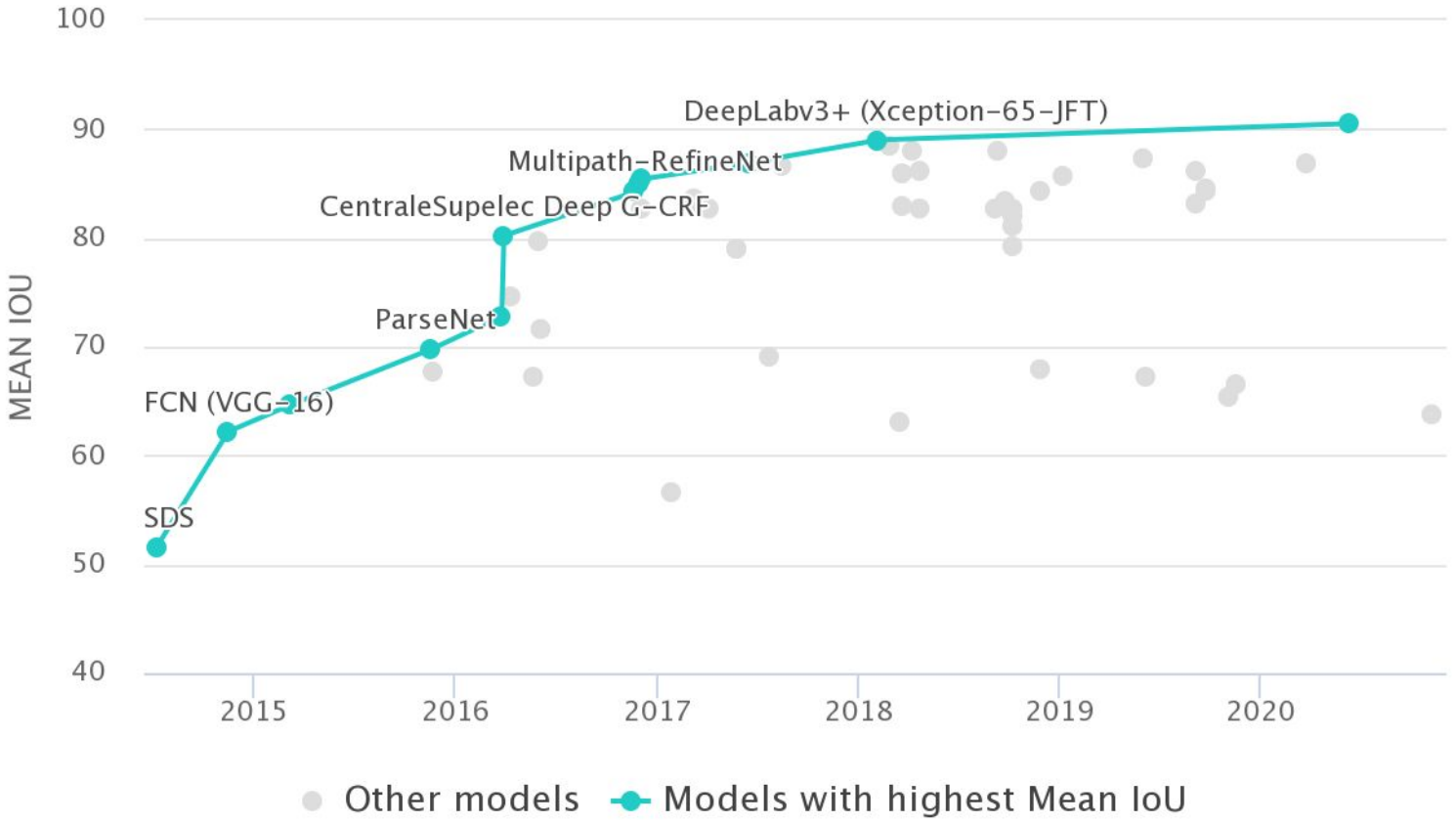
Stella X. Yu

Berkeley
UNIVERSITY OF CALIFORNIA

Semantic Segmentation: Classify Pixels into Semantic Categories



State-of-the-art Methods Require Pixel-wise Annotations



https://paperswithcode.com/sota/semantic-segmentation-on-pascal-voc-2012

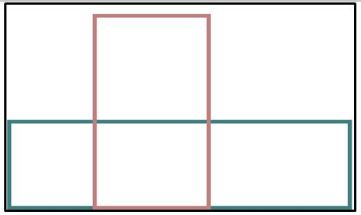
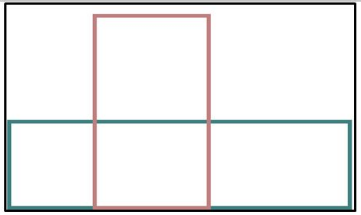
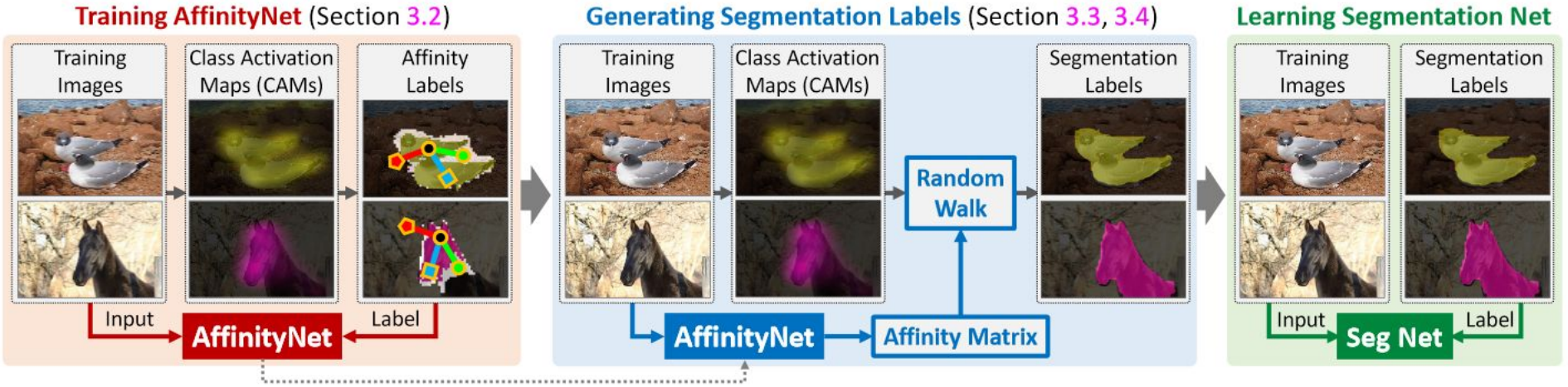
Image	Image Tags	Boxes
	Person Motorbike	
Supervision	Coarse	
Current Methods	Class Activation Map	

Image	Image Tags	Boxes
	Person Motorbike	
Supervision	Coarse	
Current Methods	Class Activation Map	



Learning Pixel-level Semantic Affinity with Image-level Supervision for Weakly Supervised Semantic Segmentation. Jiwoon Ahn and Suha Kwak. CVPR 2018.

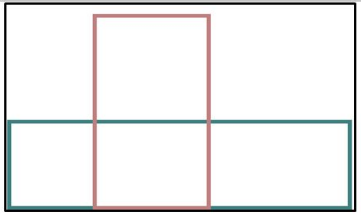
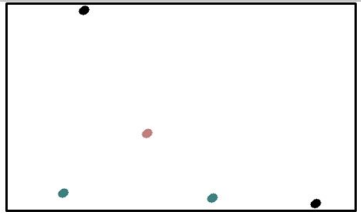
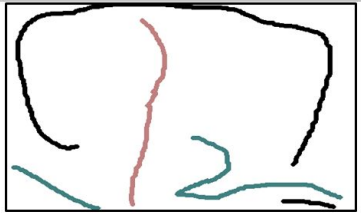
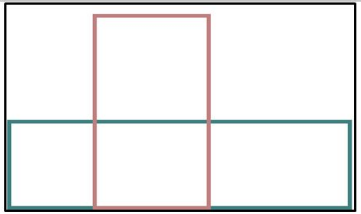
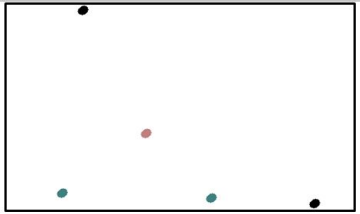
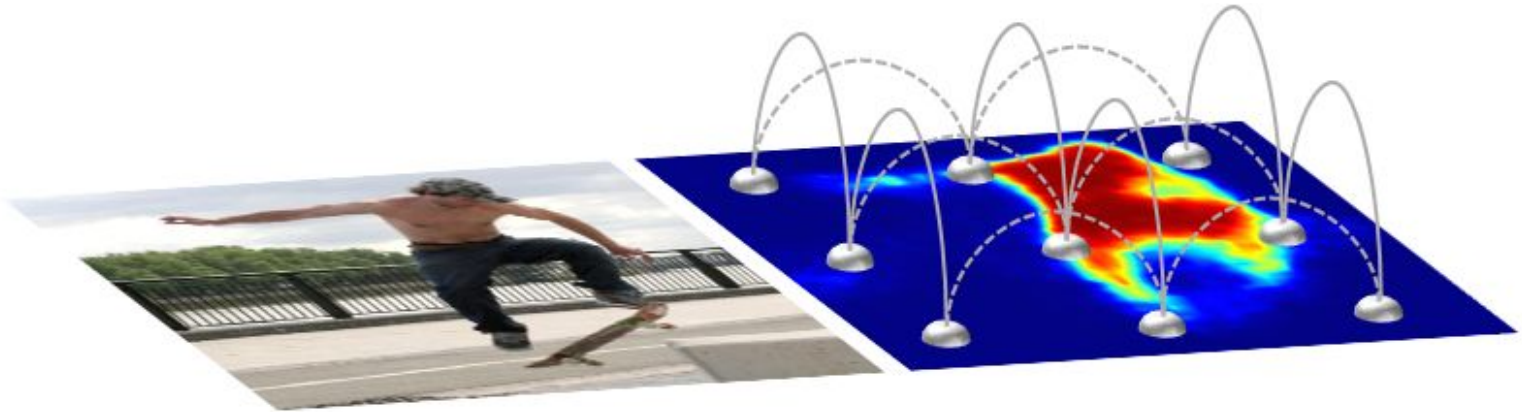
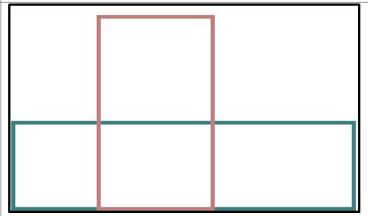
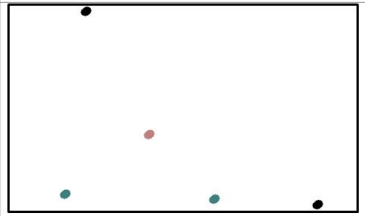
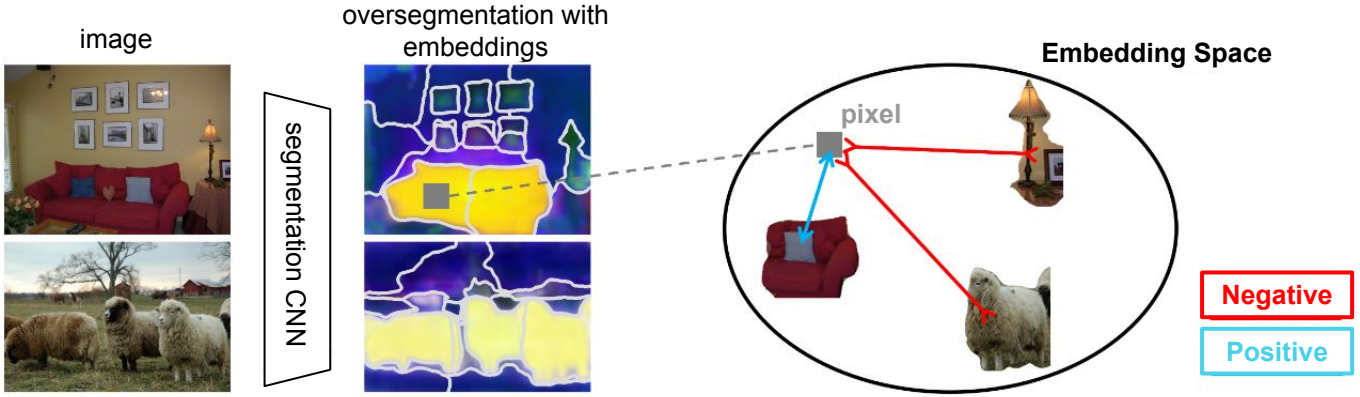
Image	Image Tags	Boxes	Points	Scribbles
	Person Motorbike			
Supervision	Coarse		Sparse	
Current Methods	Class Activation Map		Conditional Random Fields	

Image	Image Tags	Boxes	Points	Scribbles
	Person Motorbike			
Supervision	Coarse		Sparse	
Current Methods	Class Activation Map		Conditional Random Fields	

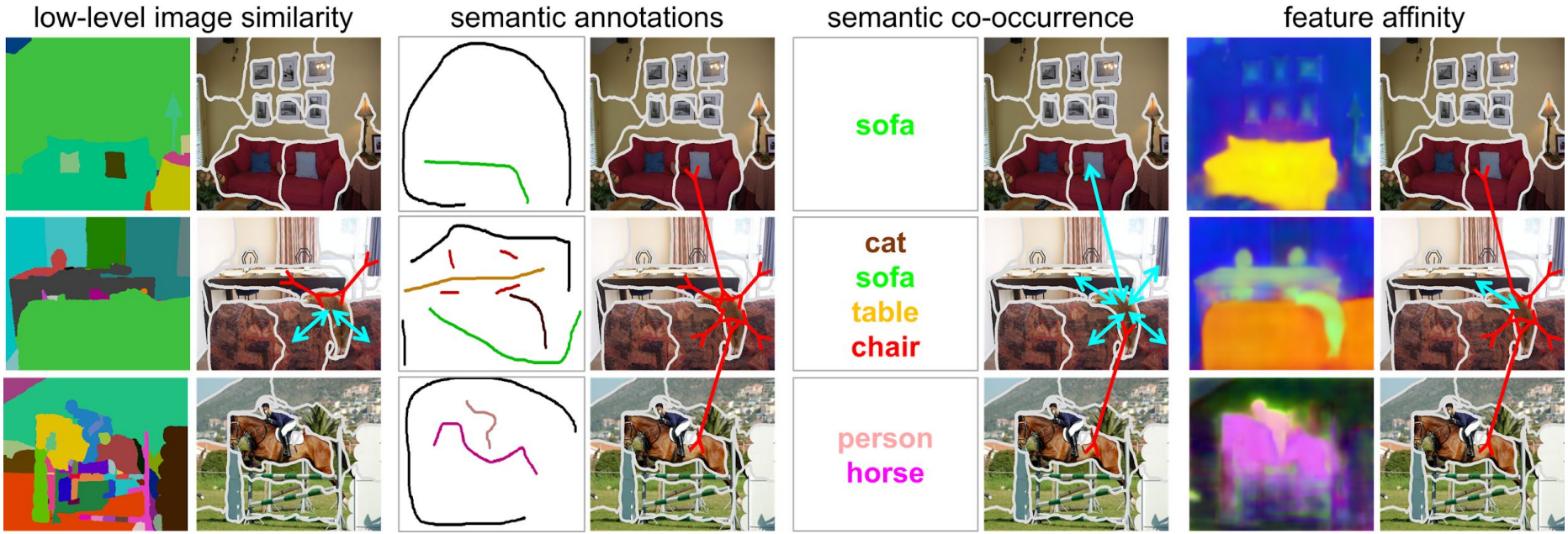


Normalized Cut Loss for Weakly-supervised CNN Segmentation. Tang et al. CVPR 2018.

Image	Image Tags	Boxes	Points	Scribbles
	Person Motorbike			
Supervision	Coarse		Sparse	
Current Methods	Class Activation Map		Conditional Random Fields	
Our Method	**single pixel-to-segment contrastive learning loss formulation**			



Our SPML: Contrasts Pixels with Segments on 4 Types of Relationships



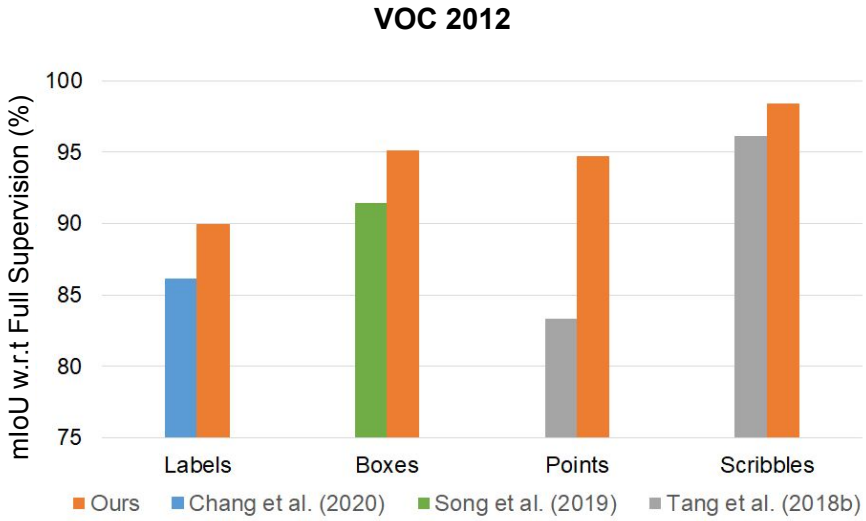
$$L(i) = \lambda_I L_{\text{SegSort}^+}(i, \mathcal{V}^+, \mathcal{V}^-) + \lambda_C L_{\text{SegSort}^+}(i, \mathcal{C}^+, \mathcal{C}^-) + \lambda_O L_{\text{SegSort}^+}(i, \mathcal{O}^+, \mathcal{O}^-) + \lambda_A L_{\text{SegSort}^+}(i, \hat{\mathcal{C}}^+, \hat{\mathcal{C}}^-)$$

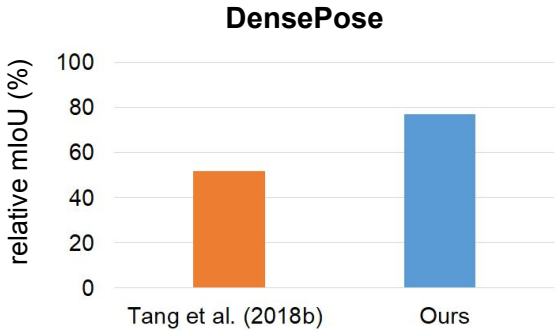
Contrastive loss for pixel $\boldsymbol{i}$ with positive segments $\boldsymbol{C}^+$, negative segments $\boldsymbol{C}^-$:

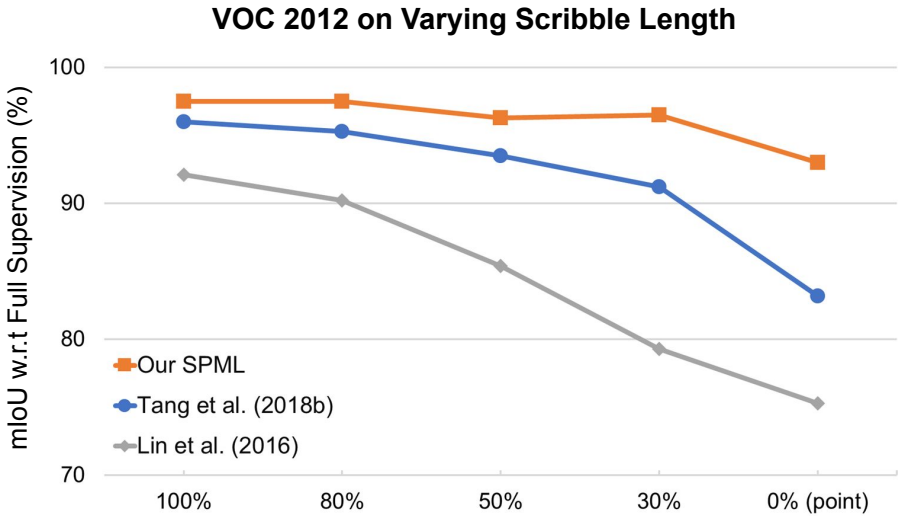
$$L^i_{\text{SegSort}} = -\log \frac{\exp(\kappa \boldsymbol{p}_s^\top \boldsymbol{e}_i)}{\sum_{l \in \Omega} \exp(\kappa \boldsymbol{p}_l^\top \boldsymbol{e}_i)}.$$

SegSort: Segmentation by Discriminative Sorting of Segments. Hwang et al. ICCV 2019.

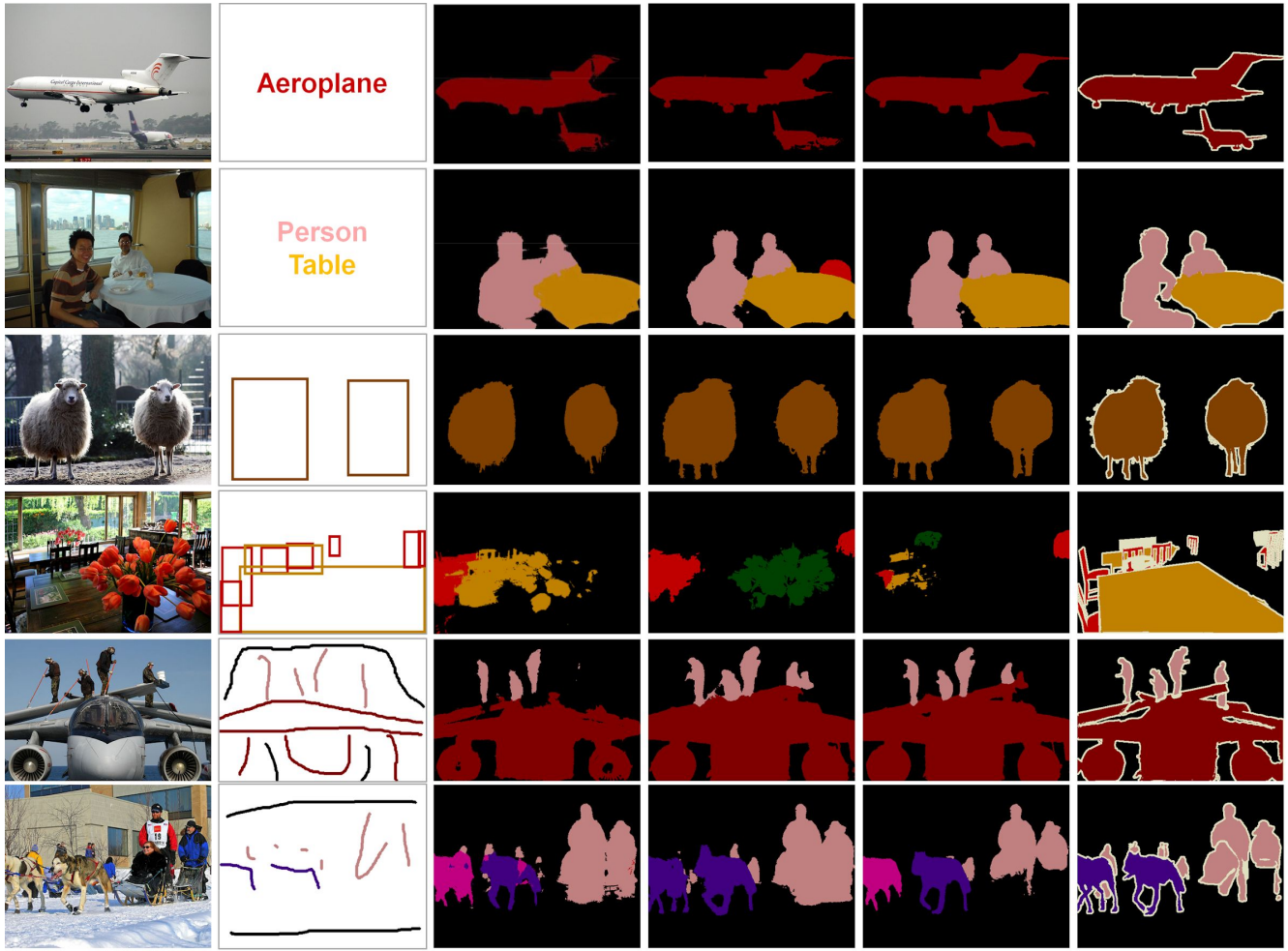
Beats All Weak Supervision SOTA's on Pascal VOC & DensePose

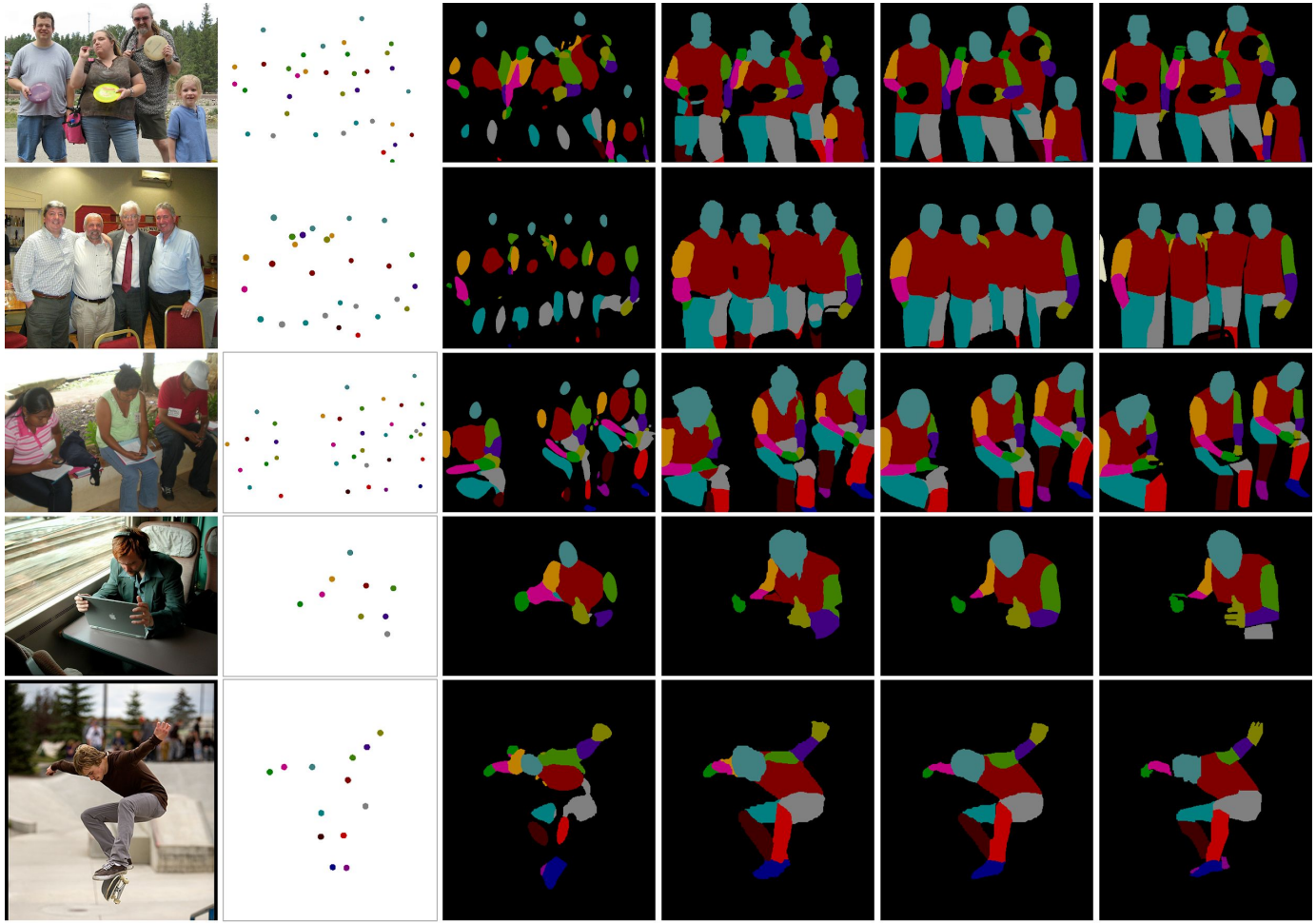






Pascal: Varying sparsity of scribbles and point annotation


Aeroplane
Person
Table
(a) Image
(b) Annotation
(c) Baseline
(d) Our SPML
(e) Full Supervision
(f) Ground Truth



(a) Image (b) Annotation (c) Baseline (d) Our SPML (e) Full Supervision (f) Ground Truth

Context-Aware Segment Retrieval via Learned Pixel-wise Feature

Code available at https://github.com/twke18/SPML