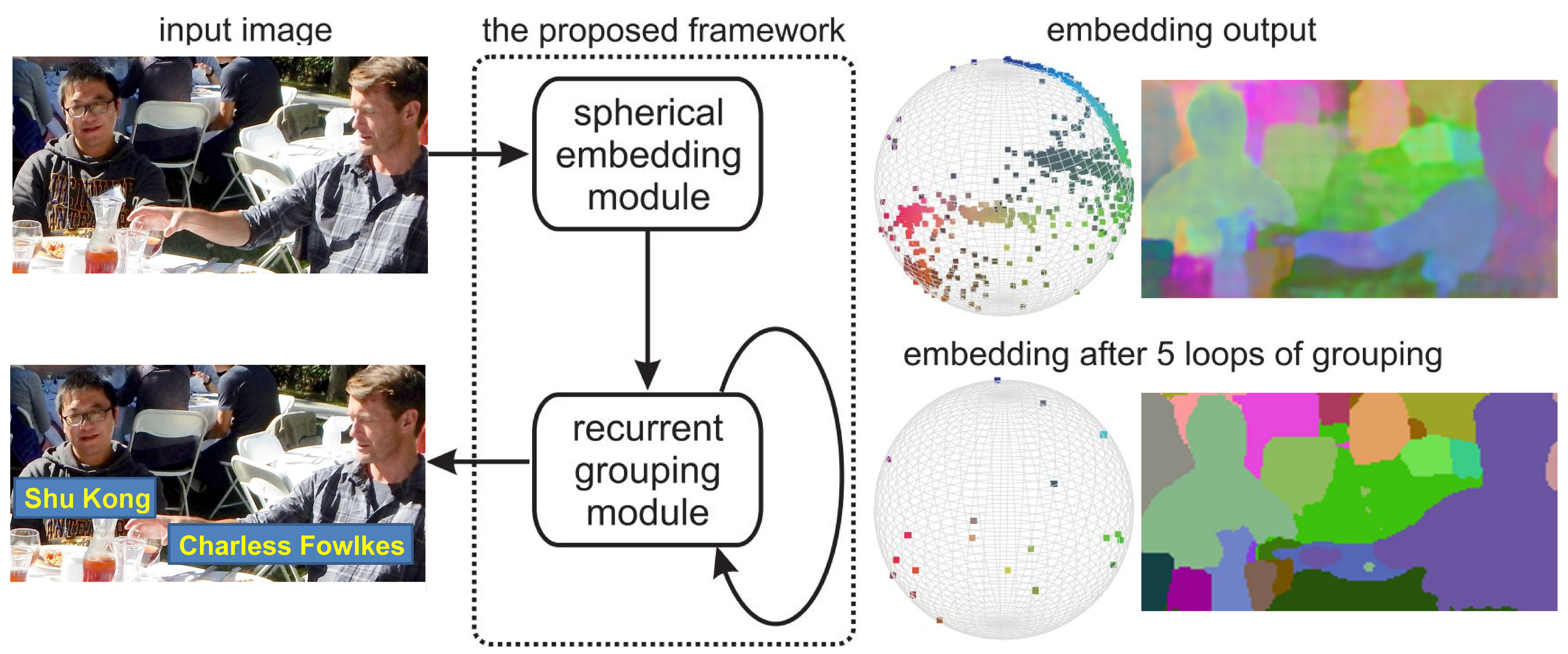
Recurrent Pixel Embedding for Instance Grouping

Shu Kong, Charless Fowlkes
Department of Computer Science, University of California, Irvine, California, USA

Highlight



A differentiable end-to-end framework for solving pixel-level grouping problems; conceptually simple and theoretically abundant. Substantial improvements over state-of-the-art instance segmentation for object proposal generation

- **Spherical Embedding Module**: regresses pixels into a spherical embedding space so that pixels from the same/different groups are more clustered/separable.
- **Recurrent Grouping Module**: group embedding vectors through a variant of mean-shift clustering implemented as a recurrent neural network, which is differentiable with convergent dynamics and probabilistic interpretability.

Learning Pixel Embeddings

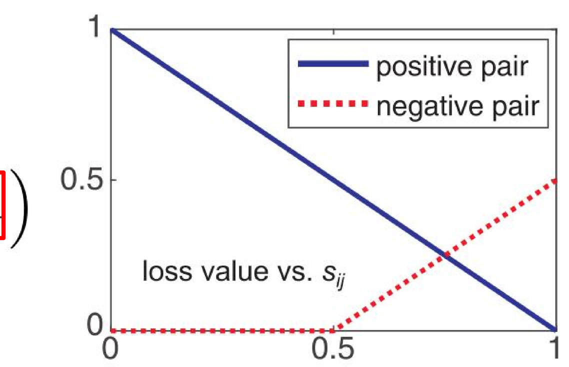
Pairwise Loss for Spherical Pixel Embeddings

$$s_{ij} = \frac{1}{2}\left(1 + \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\|_2 \|\mathbf{x}_j\|_2}\right)$$

same group → similarity near 1

different groups → similarity below a specified margin.

$$\ell = \sum_{k=1}^{M} \sum_{i,j=1}^{N_k} \frac{w_i^k w_j^k}{N_k} \Big(\underbrace{\boxed{\mathbf{1}_{\{y_i=y_j\}}(1-s_{ij})}}_{\text{positive pairs}} + \underbrace{\boxed{\mathbf{1}_{\{y_i \neq y_j\}}[s_{ij}-\alpha]_+}}_{\text{negative pairs}} \Big)$$



Instance-aware Pixel Weighting

Bounds on loss that are independent of embedding dimension

Proposition 1 *For n vectors $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$, the total intra-pixel similarity is bounded as $\sum_{i\neq j} \mathbf{x}_i^T \mathbf{x}_j \geq -\sum_{i=1}^{n} \|\mathbf{x}_i\|_2^2$. In particular, for n vectors on the hyper-sphere where $\|\mathbf{x}_i\|_2 = 1$, we have $\sum_{i\neq j} \mathbf{x}_i^T \mathbf{x}_j \geq -n$.*

$$\|\sum_{q=1}^{Q}\sum_{i\in\mathcal{I}_q} w_i \mathbf{x}_i\|^2 = \underbrace{\boxed{\sum_{q=1}^{Q}\|\sum_{i\in\mathcal{I}_q} w_i \mathbf{x}_i\|^2}}_{\text{positive pairs}} + \underbrace{\boxed{\sum_{p\neq q}\Big(\sum_{i\in\mathcal{I}_p} w_i \mathbf{x}_i\Big)^T\Big(\sum_{j\in\mathcal{I}_q} w_j \mathbf{x}_j\Big)}}_{\text{negative pairs}}$$

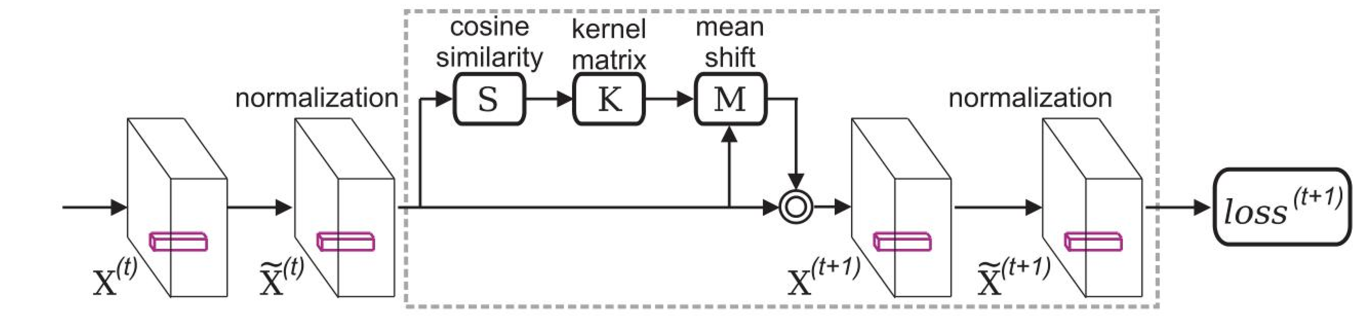
Setting $w_i = \frac{1}{|\mathcal{I}_q|}$ for each pixel i belonging to instance q yields a loss independent of instance number and size.

Margin Selection

Related to Tammes' circle packing problem -- maximizing the smallest distance among n points on a sphere: $\max_{\mathbf{x}_i \in \mathbb{R}^3} \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|_2$

A safe (trivial) strategy -- for *n* instances embedded in *n*/2 dimensions, zero loss with a α = 0.5 by placing pairs of groups antipodally along each of n/2 orthogonal axes.

Recurrent Mean-Shift Grouping



Gaussian Blurring Mean Shift ($\eta = 1$) moves each embedding vector to the weighted mean of its neighbors.

von Mises-Fisher distribution as kernel for non-parametric density estimation on the sphere.

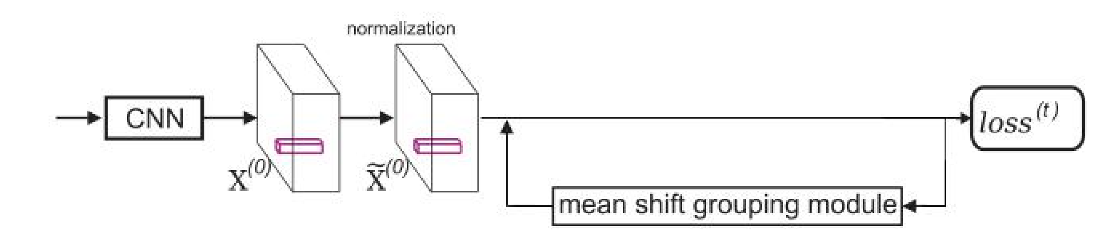
$$\mathbf{S} = \mathbf{X}^T\mathbf{X}$$

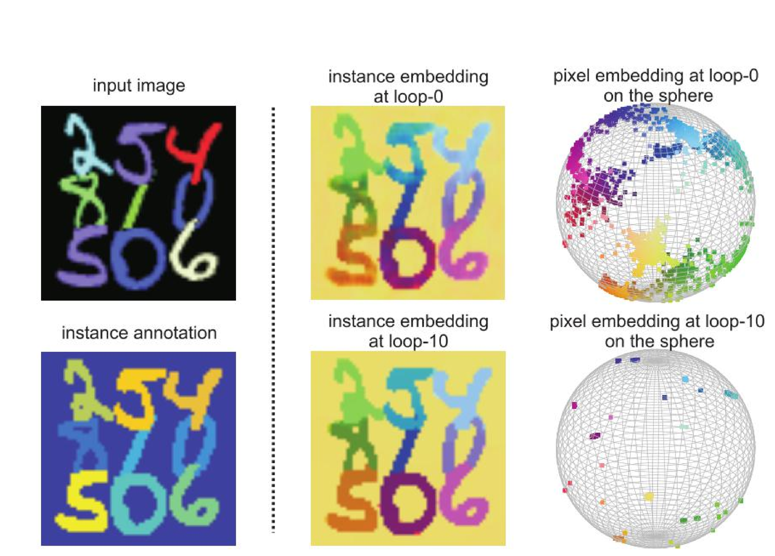
$$\mathbf{K} = \exp(\delta^2 \mathbf{S})$$

$$\mathbf{D} = \text{diag}(\mathbf{K}^T\mathbf{1})$$

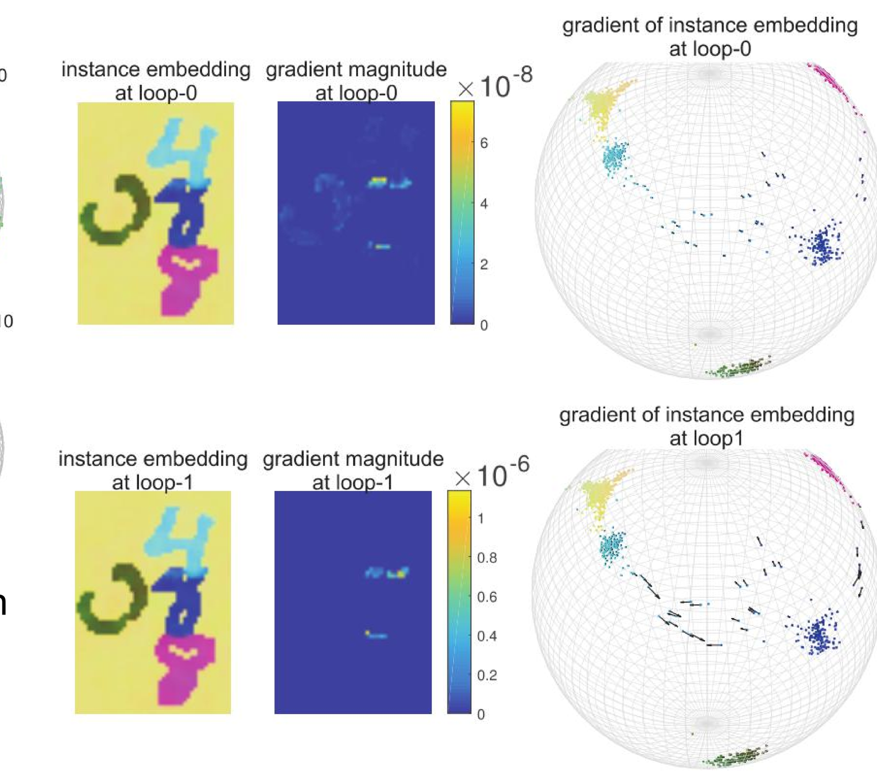
$$\mathbf{M} = \mathbf{X}\mathbf{K}\mathbf{D}^{-1} - \mathbf{X}$$

$$\mathbf{X} \leftarrow \mathbf{X} + \eta\mathbf{M}$$





The recurrent grouping module focuses on correcting embedding errors that won't be resolved during subsequent clustering.

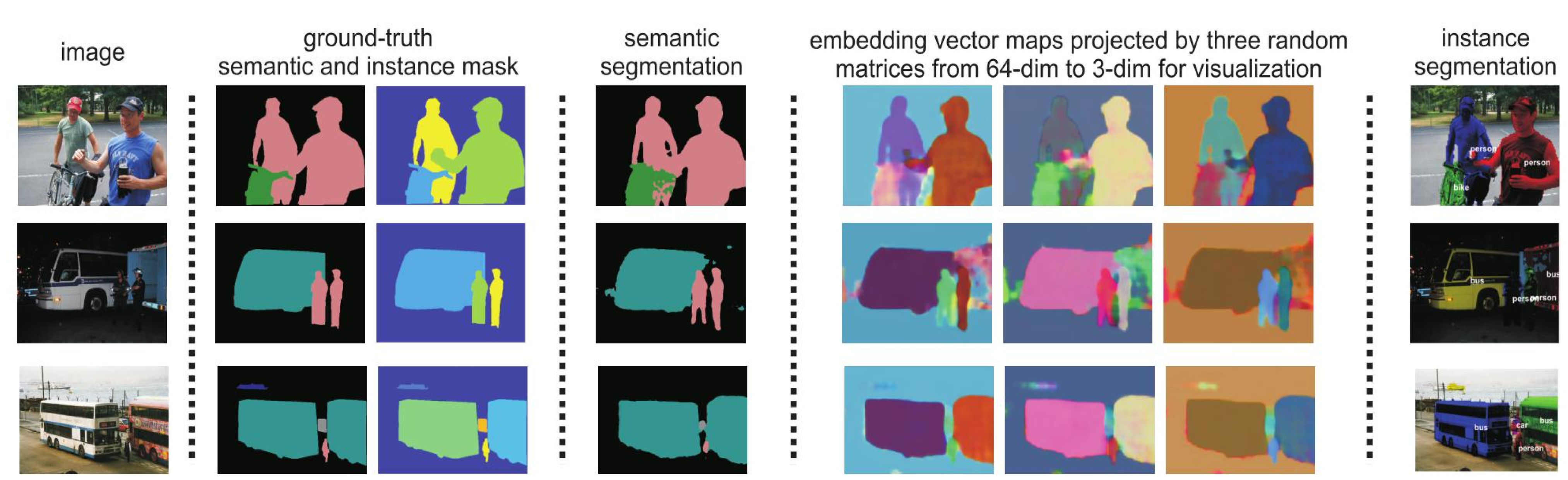


Experiments

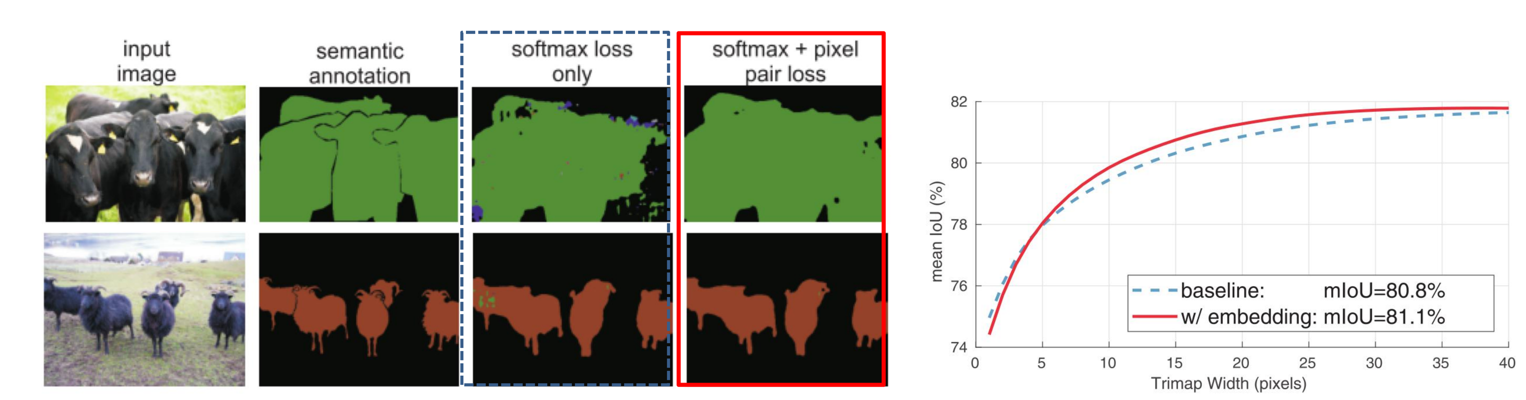
Instance Embedding

Learned instance embeddings are correlated with scale and location.

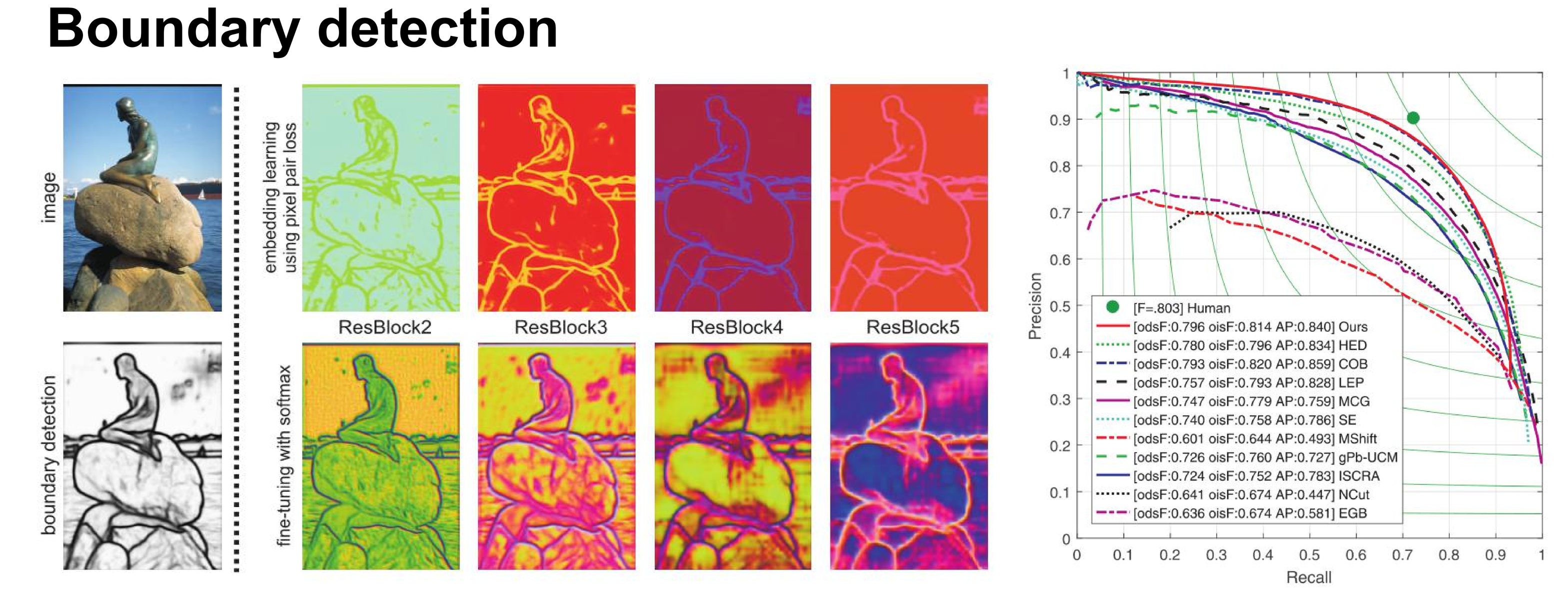
Semantic Instance Segmentation



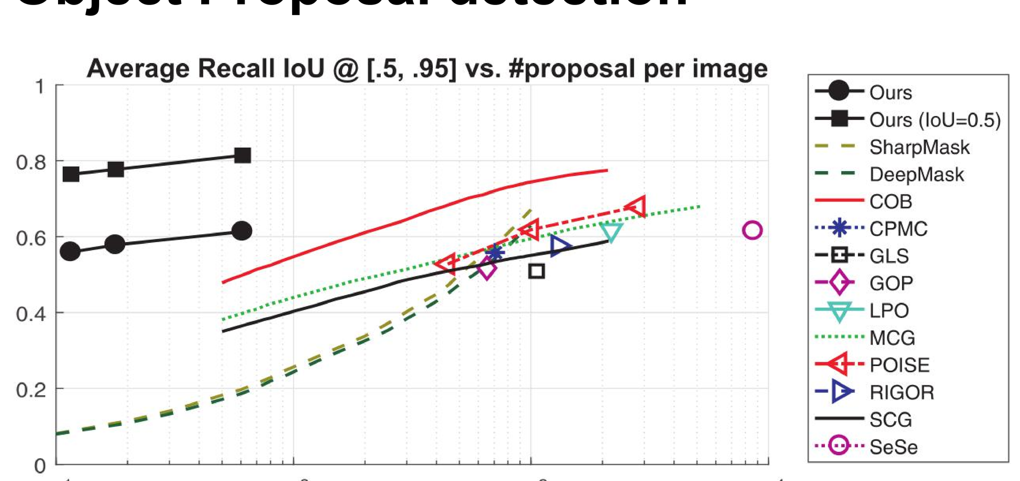
Semantic Segmentation



Boundary detection



Object Proposal detection



Object proposal detection AR@IoU=0.5 and various number of proposals per image.

#prop.	SCG	MCG	COB	inst-DML	Ours
10	-	-	-	0.558	0.769
60	0.624	0.652	0.738	0.667	0.814

Semantic Instance Detection : VOC 2012 Validation

Method	plane	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	motor	person	plant	sheep	sofa	train	tv	mean
SDS	58.8	0.5	60.1	34.4	29.5	60.6	40.0	73.6	6.5	52.4	31.7	62.0	49.1	45.6	47.9	22.6	43.5	26.9	66.2	66.1	43.8
Chen et al.	63.6	0.3	61.5	43.9	33.8	67.3	46.9	74.4	8.6	52.3	31.3	63.5	48.8	47.9	48.3	26.3	40.1	33.5	66.7	67.8	46.3
PFN	76.4	15.6	74.2	54.1	26.3	73.8	31.4	92.1	17.4	73.7	48.1	82.2	81.7	72.0	48.4	23.7	57.7	64.4	88.9	72.3	58.7
MNC	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	63.5
Li et al.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	65.7
R2-IOS	87.0	6.1	90.3	67.9	48.4	86.2	68.3	90.3	24.5	84.2	29.6	91.0	71.2	79.9	60.4	42.4	67.4	61.7	94.3	82.1	66.7
Assoc. Embed.	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	35.1
inst-DML	69.7	1.2	78.2	53.8	42.2	80.1	57.4	88.8	16.0	73.2	57.9	88.4	78.9	80.0	68.0	28.0	61.5	61.3	87.5	70.4	62.1
Ours	85.9	10.0	74.3	54.6	43.7	81.3	64.1	86.1	17.5	77.5	57.0	89.2	77.8	83.7	67.9	31.2	62.5	63.3	88.6	74.2	64.5

Demo, code and model can be found here: https://www.ics.uci.edu/~skong2/SMMMSG.html

Acknowledgements: This work was supported by NSF grants DBI-1262547, IIS-1253538, hardware donation from NVIDIA.