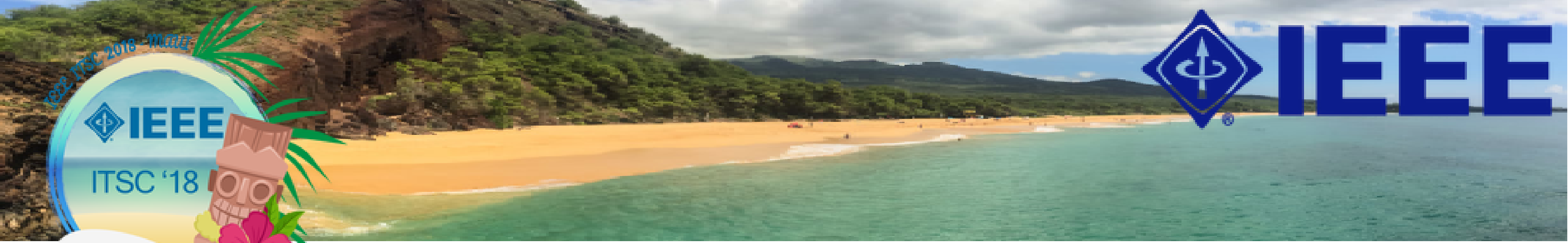


The 21st IEEE International Conference on Intelligent Transportation Systems
November 4-7, 2018, Maui, Hawaii, USA

Distance to Center of Mass Encoding for Instance Segmentation

Thomio Watanabe and Denis Wolf

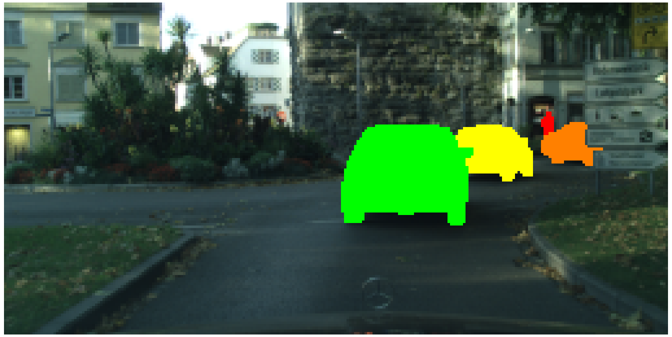
Mobile Robotics Laboratory
Institute of Mathematics and Computer Sciences, USP

ICMC USP SÃO CARLOS

USP Universidade de São Paulo

Instance Segmentation

- Image classification
- Image segmentation
- Image detection
  - Instance segmentation (object segmentation)



Instance Segmentation

- Cityscapes (Cordts et al. [1]):
  - Detection + segmentation
  - Segmentation + detection
  - Simultaneous detection and segmentation
- Pixel-level encoding (Uhrig et al. [2])
  - Proposal-based x Proposal-free

- [1] M. Cordts, M. Omran, S. 'Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele, "The cityscapes dataset for semantic urban scene understanding," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [2] J. Uhrig, M. Cordts, U. Franke, and T. Brox, "Pixel-level encoding and depth layering for instance-level semantic labeling," in German Conference on Pattern Recognition. Springer, 2016, pp. 14–25.

Instance Segmentation

- Proposal-based
  - Region proposal network
  - Network architecture
- Proposal-free
  - CNN for segmentation
  - Multi-task networks
  - Highly dependent on pre and post processing
  - Data representation (annotation problem)

DCME

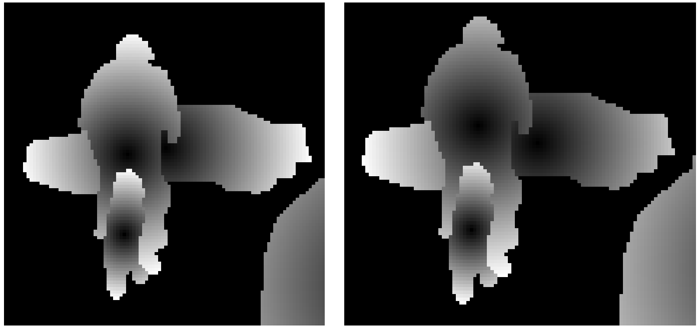
- Distance to center of mass encoding (DCME)
  - Proposal free
  - Annotation problem (not architecture)
    - Encoding x decoding
  - Independent of segmentation network
    - Single modification: softmax → regression

DCME

- Associate each instance to one point:
  - Superficial center of mass (CM)
  - Unit mass

  - Bounding box centroid are more susceptible to outliers

$$P_{CM} = (x_{cm}, y_{cm})$$
$$x_{cm} = mean(X); \quad \{X|\forall x \in I\}$$
$$y_{cm} = mean(Y); \quad \{Y|\forall y \in I\}$$

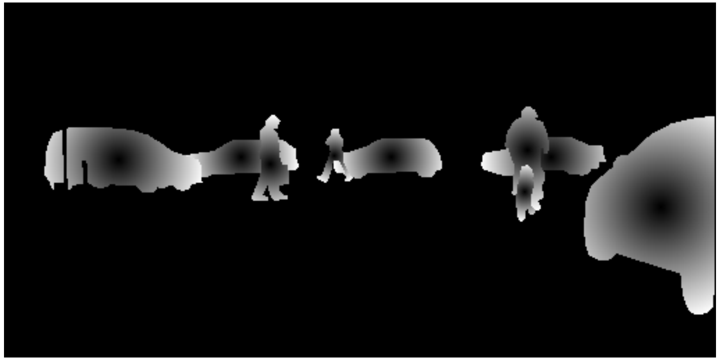


Centroid vs CM

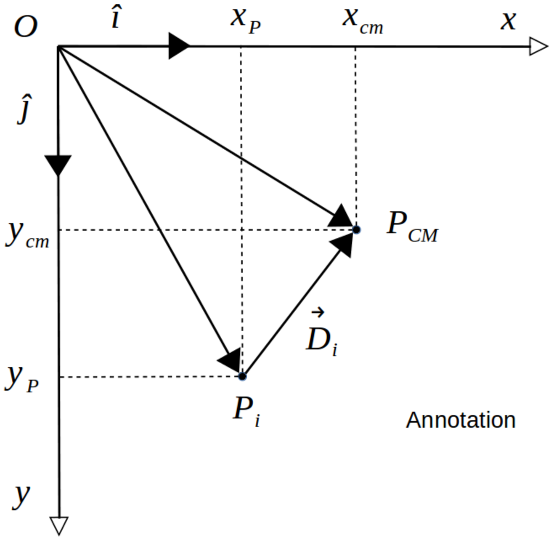
DCME

- Compute displacement vectors:
  - Each instance pixel points to its CM
  - Solves partial occlusion

$$\mathbf{D}_i = \mathbf{OP}_{CM} - \mathbf{OP}_i \quad \{\forall P_i \in I\}$$
$$\mathbf{D}_i = d_x\mathbf{i} + d_y\mathbf{j}$$
$$d_x = x_{cm} - x_p$$
$$d_y = y_{cm} - y_p$$



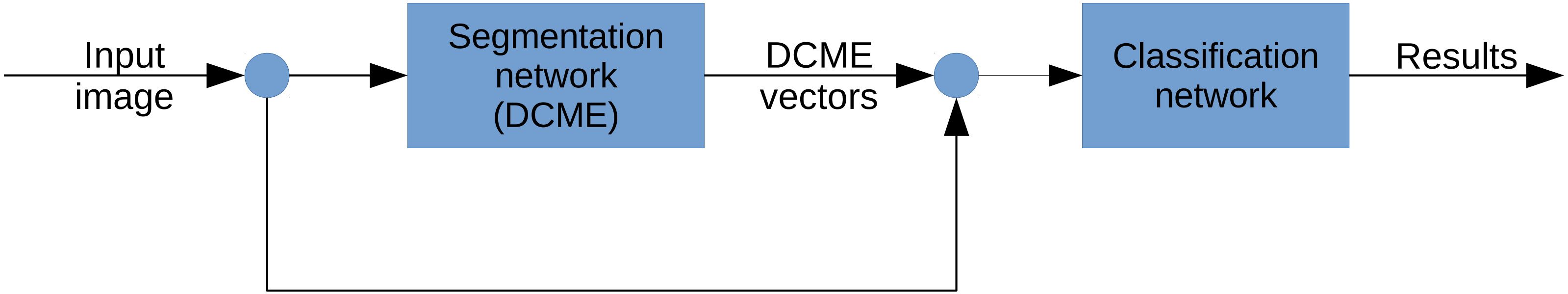
Magnitude representation of 2D vectors



DCME

- Encoding:
  - Calculate the center of mass (CM) for each instance
  - For each pixel from each instance calculate vectors that point to the CM
- Decoding:
  - A CM is pointed by several vectors
  - Pixels that point to the same CM belong to the same instance

Solution Pipeline



Experiments

- Cityscapes
  - Urban roads, high definition images
- Distance to center of mass encoding (DCME)
  - Segmentation network (FCN, SegNet)
- Classification network (inception)
- Caffe framework

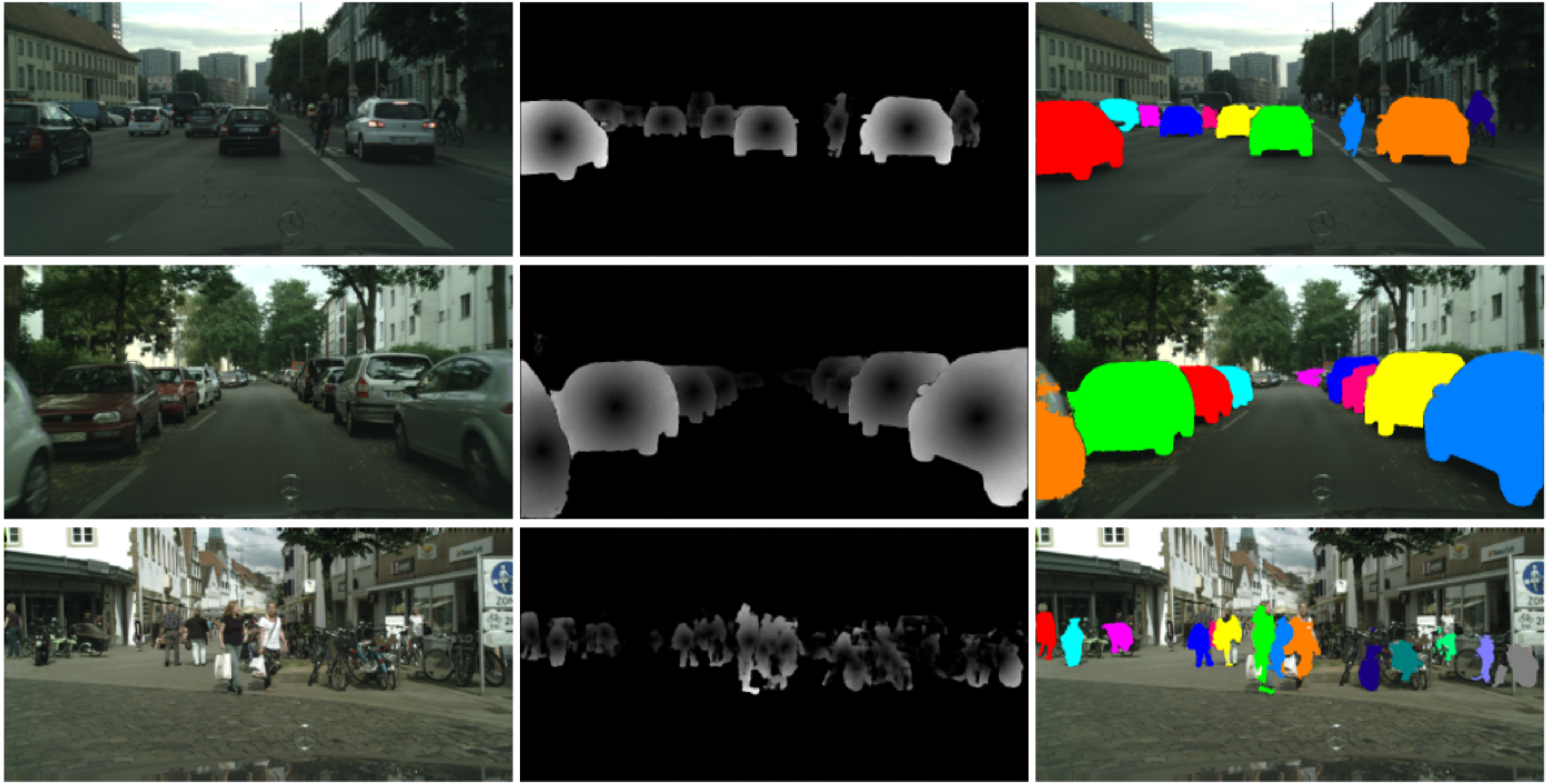
Experiments

Class	AP	AP50%	AP100m	AP50m
person	1.77	5.86	3.75	4.06
rider	0.71	3.33	1.29	1.38
car	15.53	25.65	26.60	35.54
truck	2.00	4.02	3.58	5.12
bus	4.30	8.30	8.14	14.66
train	4.57	9.98	7.73	12.86
motorcycle	0.93	3.39	1.30	1.50
bicycle	0.33	1.35	0.58	0.61
mean	3.77	7.73	6.62	9.47

DCME-SegNet

Class	AP	AP50%	AP100m	AP50m
person	1.31	5.61	2.57	2.73
rider	0.61	3.92	1.07	1.08
car	10.51	26.00	17.45	21.18
truck	6.14	13.80	10.64	13.99
bus	9.70	26.34	17.35	25.21
train	5.85	15.85	9.16	14.15
motorcycle	1.75	8.64	2.60	2.74
bicycle	0.54	3.09	0.93	0.98
mean	4.55	12.90	7.72	10.26

R-CNN + MCG convex hull



Input image, magnitude map, instances map.

Conclusion

- Solve instance segmentation
  - Network architecture → annotation representation
  - Solves partial occlusion
  - Better segmentation networks → better results

Future Work

- Training with different resolutions
  - Loss function → each pixel is an independent output
- Improve scores
  - Better segmentation network (Deeplab v4)
- Remove the classification network (single network)
  - Reduce computation costs
  - Share weights between different tasks

Thanks !

thomio.watanabe@usp.br