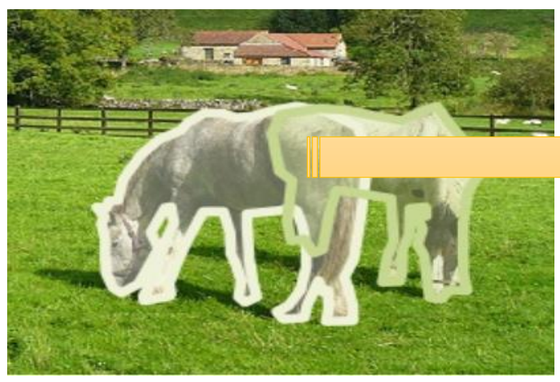
Multimodal Amodal Instance Segmentation

Peize Sun, Ke Li
(equal contribution)
Xi'an Jiaotong University, University of California Berkeley

Definition

(a) Instance Segmentation



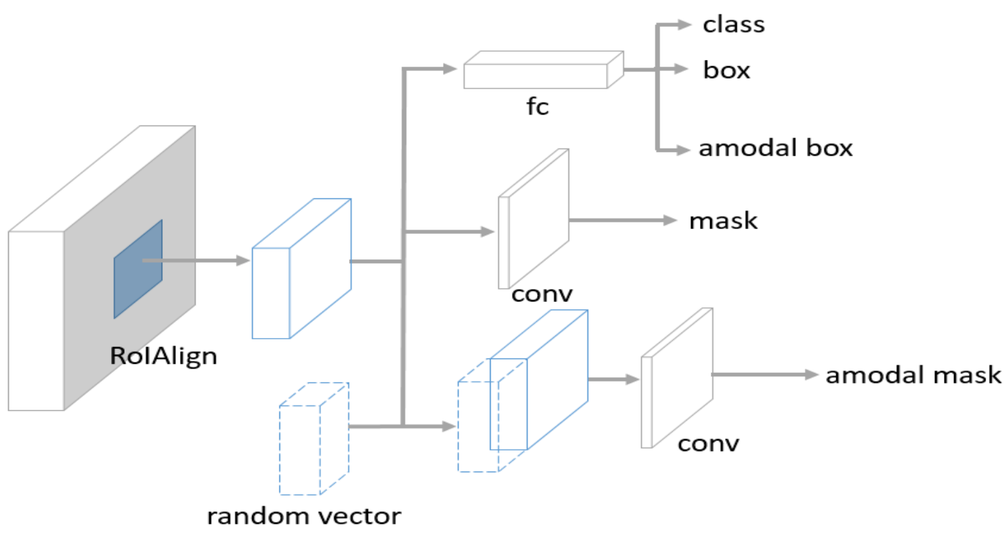
(b) Amodal Instance Segmentation

Only one possibility?

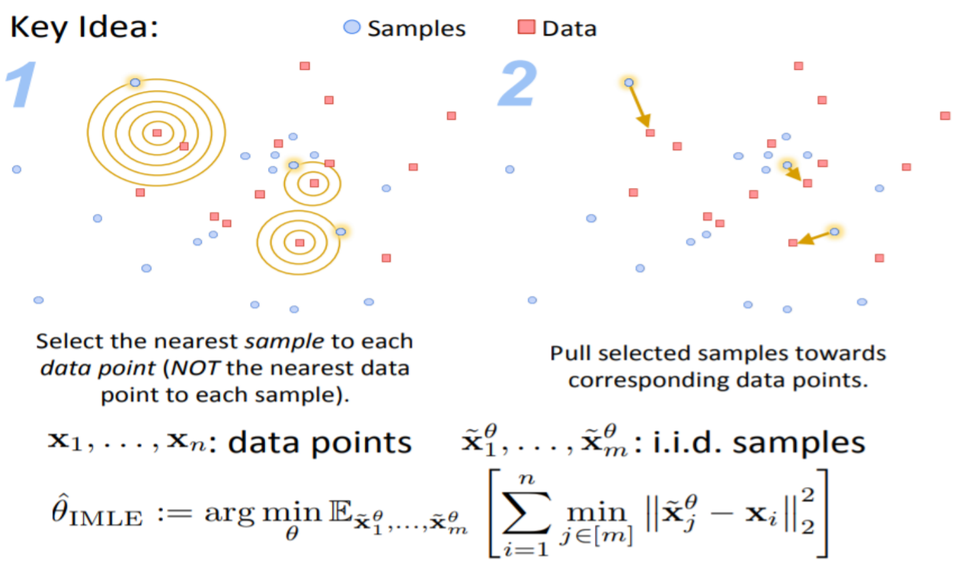
(c) Multimodal Amodal Instance Segmentation

- **Instance Segmentation:** predicting mask of (visible part of) each object instance
- **Amodal Instance Segmentation**:[1,2] predicting not only mask of visible part of object, but also that of invisible part
- **Multimodal Amodal Instance Segmentation:** predicting all possible amodal instance segmentations

Method



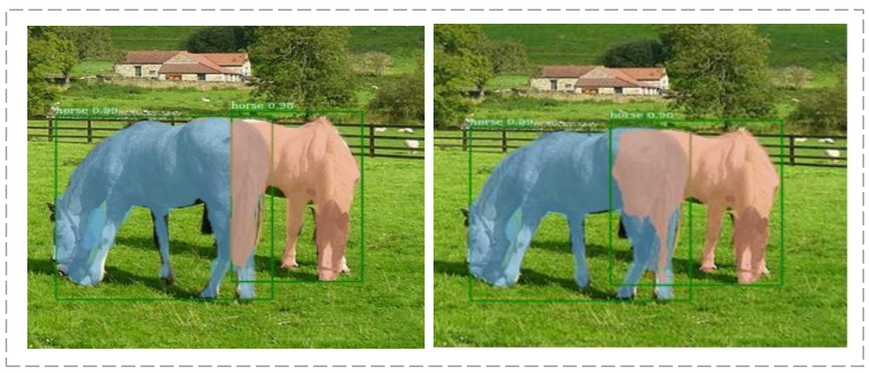
- Our model is based on Mask R-CNN[3]. We extend a new branch to predict amodal instance mask, whose input is RoIAlign feature and random vector.
- We apply IMLE[4] to **train the model to make random vector represent information about invisible part.**



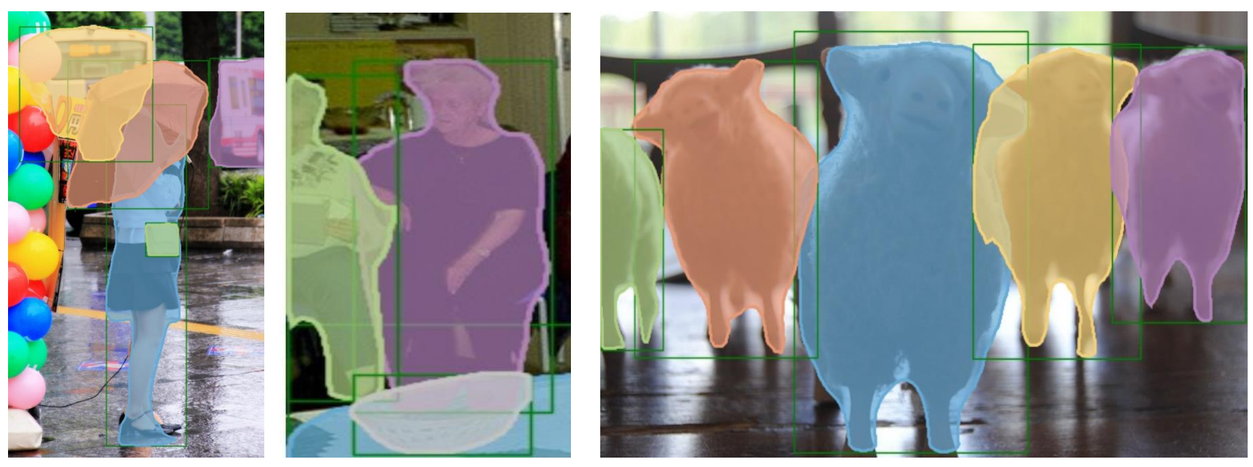
- Key idea of IMLE[4] is making sure each data point(training example) has corresponding sample(random vector) so that to avoid Mode Collapse.

Result

(a) Mask R-CNN



(b) Our result



(c) More examples

- Our model basically achieves to output multimodal amodal instance segmentation. **Future work is towards to be more realistic and smooth.**

Reference

[1] Ke Li and Jitendra Malik. Amodal Instance Segmentation. 2016. In ECCV.
[2] Yan Zhu, Yuandong Tian, Dimitris Mexatas, and Piotr Doll ár. Semantic Amodal Segmentation. 2017. In CVPR.
[3] Kaiming He, Georgia Gkioxari, Piotr Doll ár, and Ross Girshick. Mask R-CNN.2017. In ICCV.
[4] Ke Li and Jitendra Malik. Implicit Maximum Likelihood Estimation. 2018. arXiv:1809.09087.