**What’s the Point: Semantic Segmentation with Point Supervision**

Amy Bearman\(^1\)  Olga Russakovsky\(^2\)  Vittorio Ferrari\(^3\)  Li Fei-Fei\(^1\)

\(^1\)Stanford University  \(^2\)Carnegie Mellon University  \(^3\)University of Edinburgh

---

**Contributions**

**Goal:** Obtain the most annotation cost-effective supervision for semantic image segmentation.

- Novel, cost-efficient supervision regime for semantic segmentation based on humans pointing to objects.
- Extensive human study to collect point annotations for PASCAL VOC 2012, and released annotation interfaces.
- A generic objectness prior incorporated directly in the loss to guide the training of a CNN.

**Novel supervision regime**

**Problem:** Assign one class label to every pixel in an image.

**Training:** Standard regime = costly per-pixel annotations

**Levels of supervision**

- full supervision
- image-level labels
- points
- squiggles

**Key insight:** Annotating one pixel per training image significantly improves segmentation annotation and only marginally increases the annotation cost as compared to image-level labels.

**Loss function for point-level supervision:** We have a small set of supervised pixels, and other pixels just belong to some class in \(L\).

**Model:** Fully convolutional network [Long 2015].

**Crowdsourcing point annotations**

AMT annotation UI

Example points collected

**Measuring the annotation times:**

- Points and squiggles: measured directly during data collection.
- Other types of supervision: we rely on times from literature.

**Reported annotation times:**

- Image-level labels: 20.0 sec/image
- Points: 22.1 sec/image
- Squiggles: 34.9 sec/image
- Full supervision: 239.7 sec/image

---

**Objectness prior in CNN loss**

**Purpose of the objectness prior:** Helps correctly infer the spatial extent of objects for models trained with very few supervised pixels.

**Results on PASCAL VOC 2012 dataset [Everingham 2010]**

**Effects of point supervision + objectness:** The combined effect results in a +13% mIOU over image-level labels.

**Point supervision variations:** Multiple object instances and multiple annotators achieve only modest improvements over single points.

**Segmentation on an annotation budget:** Point supervision provides the best trade-off between annotation time and segmentation accuracy.

**Bibliography**