# All you need are a few pixels: semantic segmentation with PixelPick

Gyungin Shin[1], Weidi Xie[1], Samuel Albanie[2]

[1]Visual Geometry Group, Department of Engineering Science University of Oxford, UK

[2]Department of Engineering, University of Cambridge, UK

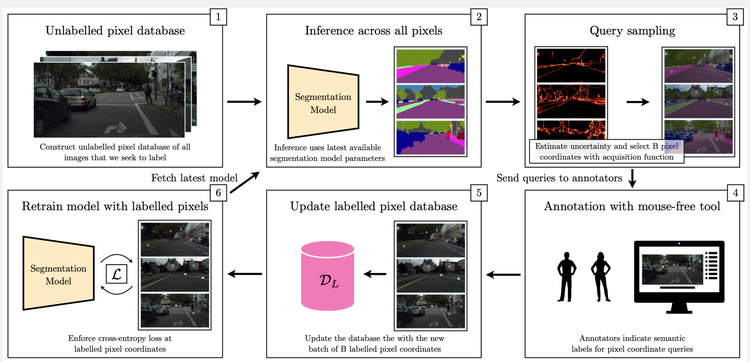
{gyungin, weidi}@robots.ox.ac.uk, sma71@cam.ac.uk

code: https://github.com/NoelShin/PixelPick

## Overview & Contribution

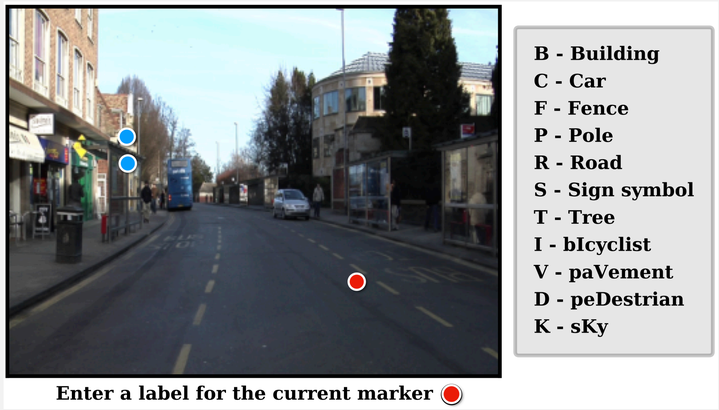
- We study the problem setting in which labels are supplied at the level of sparse pixels and show that with only a small collection of such labels, modern deep neural networks can achieve good performance.
- We show how this phenomenon can be exploited with an efficient and practical "mouse-free" annotation strategy as part of a proposed PIXELPICK active learning framework.
- We perform a series of experiments into factors that affect model performance in the low-annotation regime: annotation diversity, architectural choices and the design of the sampling mechanisms for selecting most useful pixels.
- We compare with other state of the art active learning approaches on standard segmentation benchmarks: CAMVID, CITYSCAPES and PASCAL VOC 2012, where we demonstrate comparable segmentation performance with significantly lower annotation budget.
- We assess PIXELPICK from the perspective of practical deployment, assessing its annotation efficiency and robustness.

## Proposed Method



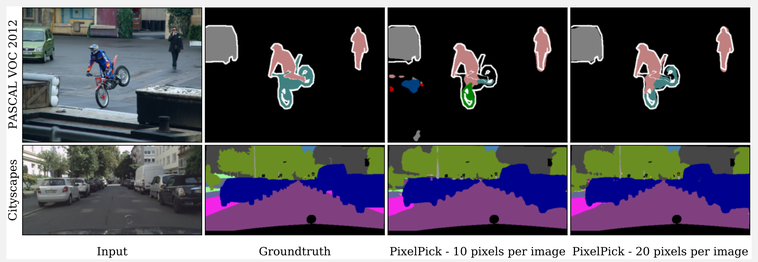

Given a database of unlabelled pixels of interest (1), each image is fed to a segmentation model to produce pixel-wise class probabilities (2), which are in turn passed to an acquisition function to estimate per-pixel uncertainties and select a batch of $B$ pixels to be labelled (3). The queries are sent to annotators (4), and the resulting labels are added to the *labelled pixel database*, $\mathcal{D}_L$ (5). Finally, the segmentation model is retrained on the expanded database (6), before the cycle repeats. To bootstrap the process and train the initial segmentation model, we randomly sample $B$ pixels and send them to be annotated.

## Mouse-free Annotation Tool

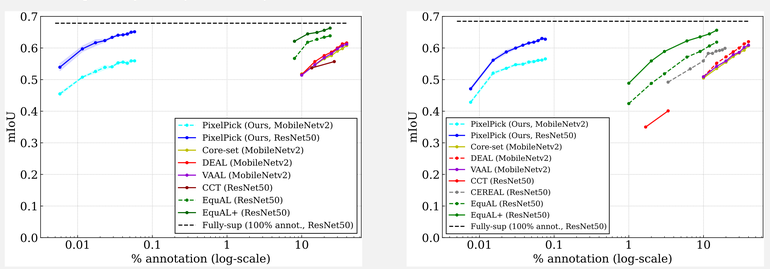



The annotator classifies the highlighted point (in red) by pressing the keyboard character of the corresponding class for the dataset. The tool then highlights the next pixel proposal and the process repeats. Note that the task requires the annnotator to perform classification, but not localisation.

## Results




Qualitative results for models trained with PIXELPICK on VOC12 (top) and Cityscapes (bottom).




PIXELPICK performs favourably against existing state-of-the-art approaches for active learning and semi-supervised learning on the CamVid (left) and Cityscapes (right) benchmarks.

| Method | Backbone | Train set (anno. type) | mIoU |
|---|---|---|---|
| **Weakly-supervised methods** | | | |
| GAIN [1] | VGG16 | 10.5K imgs (classes) | 55.3 |
| MDC [2] | VGG16 | 10.5K imgs (classes) | 60.4 |
| DSRG [3] | ResNet101 | 10.5K imgs (classes) | 61.4 |
| FickleNet [4] | ResNet101 | 10.5K imgs (classes) | 64.9 |
| BoxSup [5] | VGG16 | 10.5K imgs (boxes) | 62.0 |
| ScribbleSup [6] | VGG16 | 10.5K imgs (scribbles) | 63.1 |
| **Interactive weak supervision** | | | |
| PIXELPICK (Ours) | ResNet50 | 1.5K imgs (sparse pixels) | **65.6** |

Comparison to existing weakly-supervised methods on VOC12 validation set. PIXELPICK is competitive against existing methods, using a budget of 20 pixel annotations per image when trained on a much smaller number of images.

## Conclusion

- We proposed PIXELPICK, a framework for semantic segmentation that employs a small number of sparsely annotated pixels to train effective segmentation models.
- We showed that PIXELPICK requires considerably fewer annotations than existing state-of-the-art to achieve comparable performance.
- We showed how annotation for pixel-level active learning can be obtained efficiently with a mouse-free labelling tool, facilitating real-world deployment.

## Acknowledgements

GS is supported by AI Factory, Inc. in Korea. WX and SA are supported by Visual AI (EP/T028572/1).

## References

[1] Kunpeng Li, Ziyan Wu, Kuan-Chuan Peng, Jan Ernst, and Yun Fu. Tell me where to look: Guided attention inference network. In *Proc. CVPR*, 2018.

[2] Yunchao Wei, Huaxin Xiao, Honghui Shi, Zequn Jie, Jiashi Feng, and Thomas S. Huang. Revisiting dilated convolution: A simple approach for weakly- and semi-supervised semantic segmentation. In *Proc. CVPR*, 2018.

[3] Zilong Huang, Xinggang Wang, Jiasi Wang, Wenyu Liu, and Jingdong Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. In *Proc. CVPR*, 2018.

[4] Jungbeom Lee, Eunji Kim, Sungmin Lee, Jangho Lee, and Sungroh Yoon. Ficklenet: Weakly and semi-supervised semantic image segmentation using stochastic inference. In *Proc. CVPR*, 2019.

[5] Jifeng Dai, Kaiming He, and Jian Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In *Proc. ICCV*, 2015.

[6] Di Lin, Jifeng Dai, Jiaya Jia, Kaiming He, and Jian Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In *Proc. CVPR*, 2016.