Block Annotation: Better Image Annotation with Sub-Image Decomposition

Hubert Lin Paul Upchurch Kavita Bala

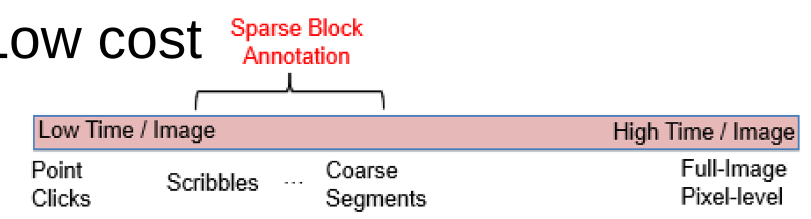
cs.cornell.edu/~hubert/block_annotation

Cornell University

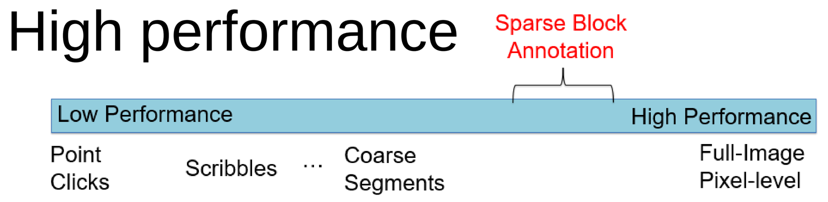
Motivation

MORE EFFICIENT ANNOTATION FOR SEGMENTATION.

Low cost



High performance



Overview of Contributions

- Crowdworker-friendly pixel-level annotation pipeline with small spatial-region annotation primitive.
- Competitive weakly-supervised semantic segmentation, outperforming existing approaches by 3-4%.
- Equivalent segmentation performance to full image pixel-level labels with half of the number of annotated pixels.
- Automatic conversion to full image pixel-level labels through label inpainting of block-annotated images.

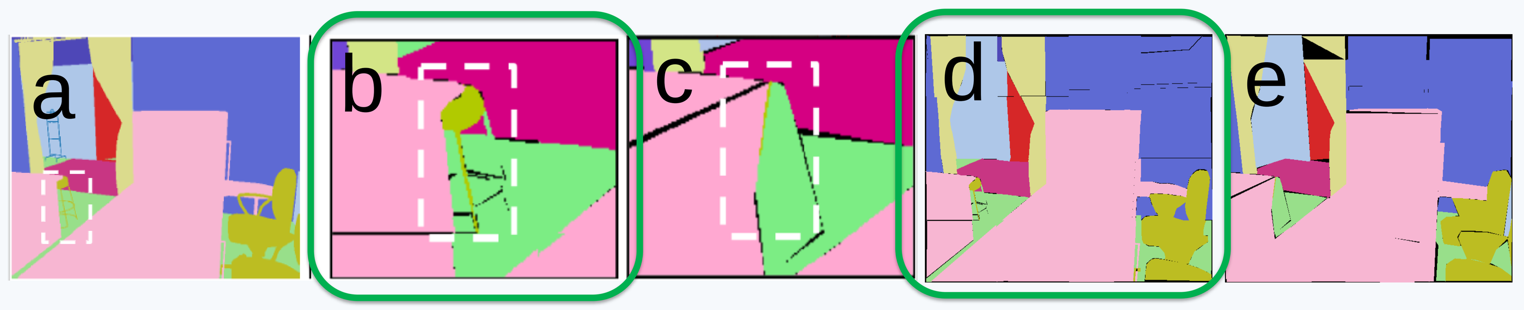
Block Annotation

Question: How do crowdworkers respond to pixel-level annotation for small spatial regions?

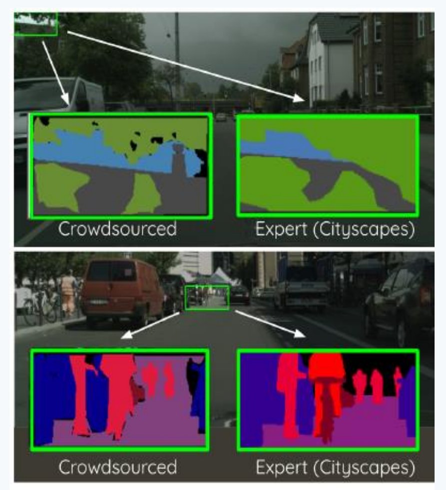
Block Annotation Task

Amazon MTurk workers are given a highlighted block region to annotate, along with the entire image as context.

Block Annotation vs. Full-Image Annotation

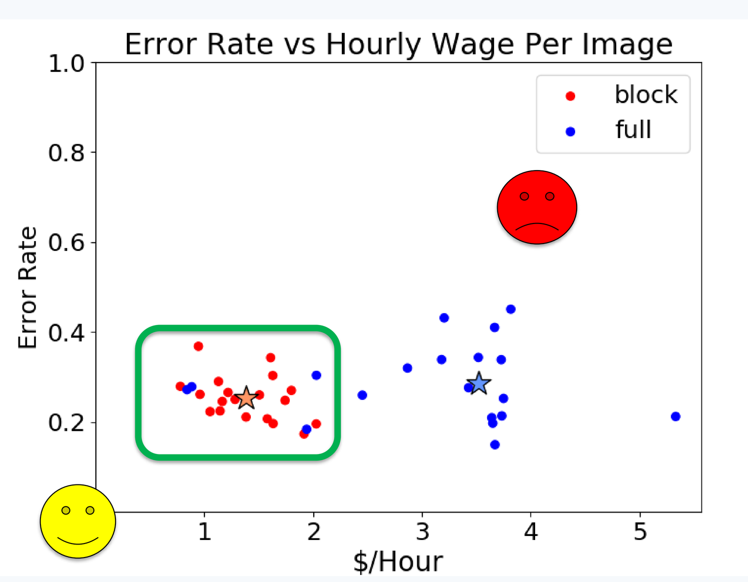


SUNCG. All segments are crowdsourced. Left to right: (a) Ground truth (b) Block annotation (zoomed-in) (c) Full-image annotation (zoomed-in) (d) Block annotation (e) Full-image annotation.
Small stool is missed by full-image annotation in this example (b vs c). The boundaries across different block tasks line up well (d vs e).



Cityscapes. Qualitative comparison of crowdsourced block annotations to expert full-image annotations.

Cost and Worker Feedback



User study on SUNCG. Crowdworkers produce higher quality annotations while demanding a lower wage.. Total cost per image is equivalent.

	"Nice" "Good" "Great"	"Fun" "Happy"	"Easy"	"Okay"	Release More HITs	Increase Pay
#	8	5	4	2	2	3

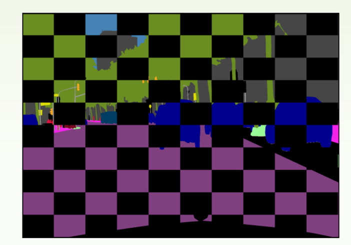
Overwhelmingly positive feedback from workers.
24 sentiments expressed by 19 worker responses over two studies on Cityscapes and SUNCG.

See our paper for more results!

Block-Supervised Semantic Segmentation

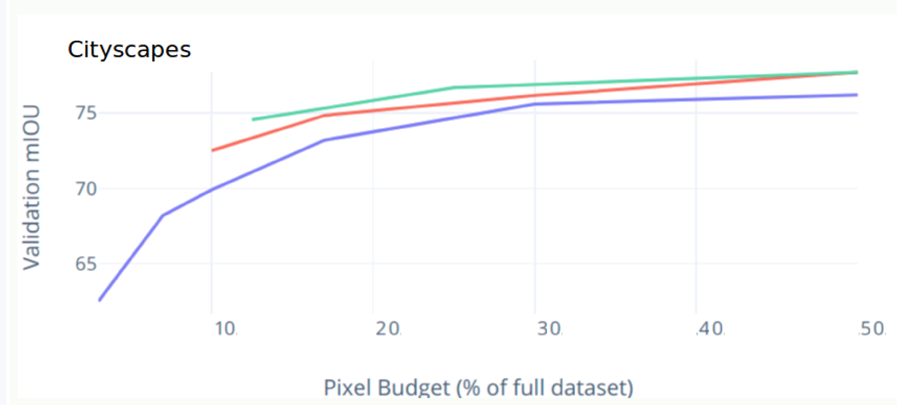
Question: How useful are block annotations for semantic segmentation?

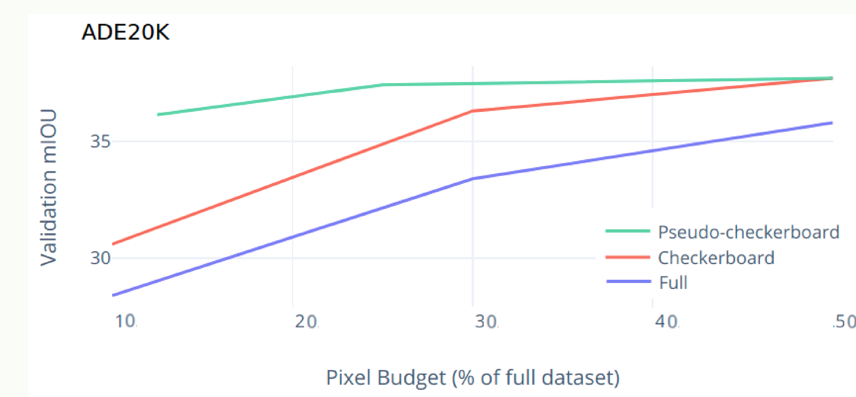
Semantic Segmentation



	Optimal (Full)	Block-50%	Block-12%
Cityscapes	77.7	77.7	74.6
ADE20K	37.4	37.2	36.1

Block annotation achieves same performance as full-image annotation with half the pixels annotated.



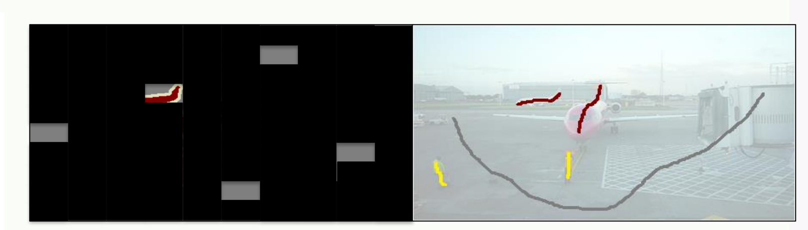


Block annotation outperforms full-image annotation given the same number of annotated pixels. *Checkerboard* refers to annotation of every other block in an image. *Pseudo-checkerboard* refers to annotation of every N blocks in *all images*.

Weakly-Supervised Semantic Segmentation

Cityscapes	**Ours:** Block (7 min)	Coarse (7 min [14])	Full Supervision (90 min [14])
mIOU (%)	**72.1**	68.8	77.7
Pascal	**Ours:** Block (25 sec)	Scribbles (25 sec [36])	Full Supervision (4 min [41])
mIOU (%)	**67.2**	63.1 [36]	69.6

Method	Annotations	mIOU (%)
MIL-FCN [46]	Image-level	25.1
WSSL [45]	Image-level	38.2
point sup. [7]	Point	46.1
ScribbleSup [36]	Point	51.6
WSSL [45]	Box	60.6
BoxSup [15]	Box	62.0
ScribbleSup [36]	Scribble	63.1
Ours: Block-1%	Pixel-level Block	61.2
Ours: Block-5%	Pixel-level Block	67.6
Ours: Block-12%	Pixel-level Block	68.4
Full Supervision	Pixel-level Image	69.6



LEFT TOP: Block annotation outperforms existing weakly-supervised methods given equivalent annotation time. Block-annotating an entire image is estimated to be 2.2x the time of full-image annotation from the user studies.

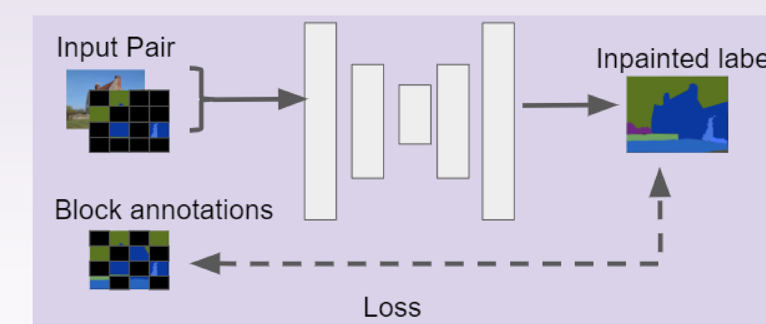
LEFT BOTTOM: Block annotation performs favorably against existing weakly-supervised methods on PASCAL VOC, even with as few as 1% to 5% of pixels annotated per image.

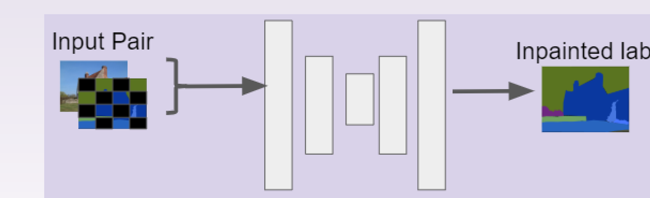
**See paper for references.

Full Image Block Inpainting

Question: Can block annotations be effectively converted into full-image annotations?

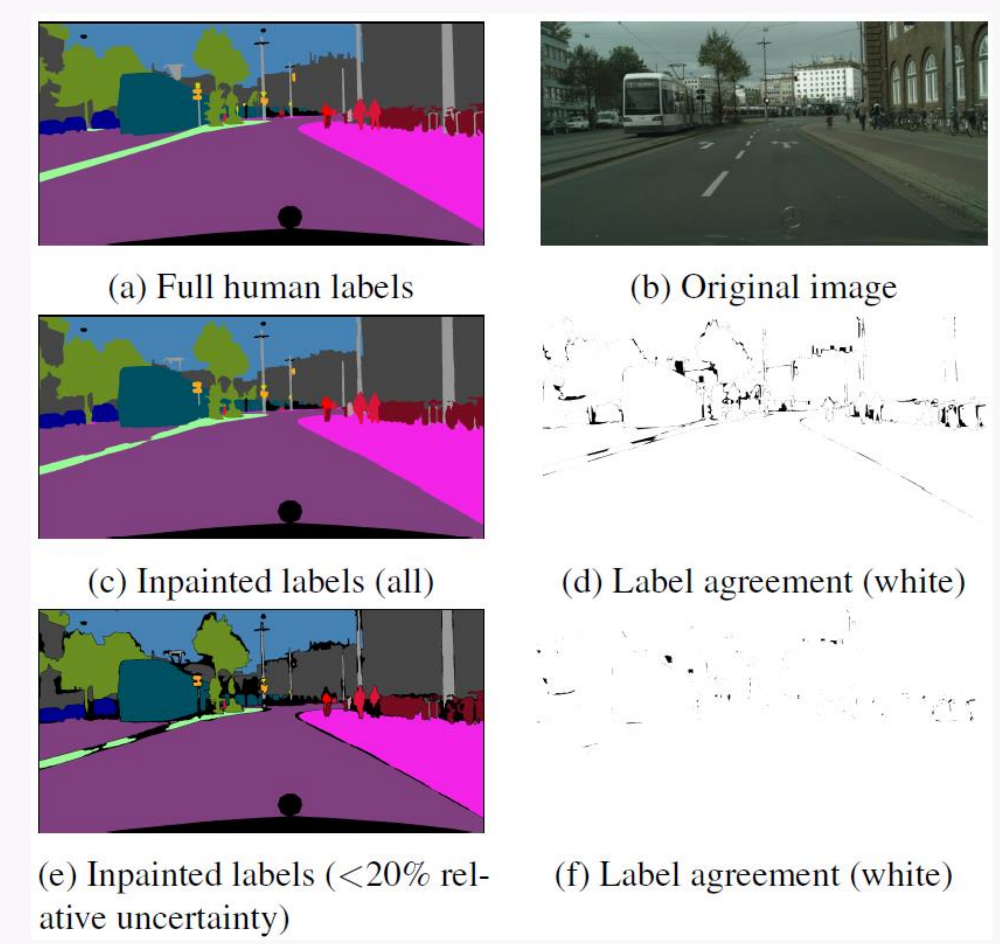
Block Inpainting Network





Input augmented with hint channel of block annotations. Train by subsampling existing block annotations as hints (left). Inference with all block annotations as hints (top).

Block-Inpainted Labels



(a) Full human labels (b) Original image (c) Inpainted labels (all) (d) Label agreement (white) (e) Inpainted labels (<20% relative uncertainty) (f) Label agreement (white)

Block inpainting experiment with Block-50% annotations (every other block of each image is annotated).

The block inpainting network is augmented with dropout so uncertainties can be estimated with Monte Carlo dropout.

After filtering out inpainted labels with >20% relative uncertainty, the block inpainting network achieves 99.8% pixel agreement with human labels on Cityscapes and has over 94% pixel coverage.

Acknowledgements

This work was supported in part by NSERC (PGS-D) and PERISCOPE MURI Contract #N00014-17-1-2699.