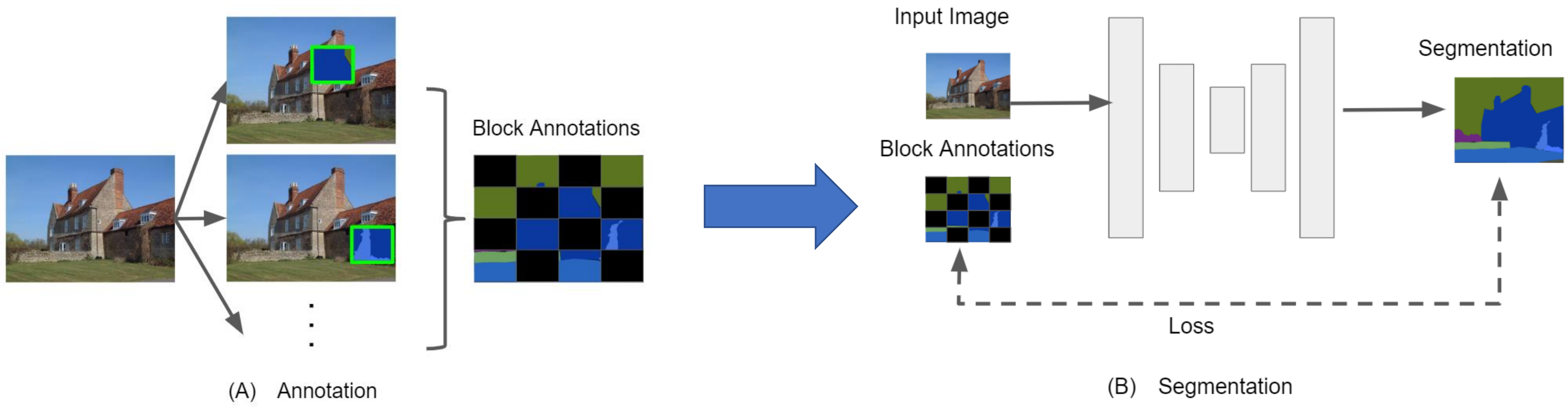
Efficient Image Annotation for Semantic Segmentation



Hubert Lin Paul Upchurch Kavita Bala

[*Block Annotation: Better Image Annotation with Sub-Image Decomposition.* Lin et al, ICCV 2019]

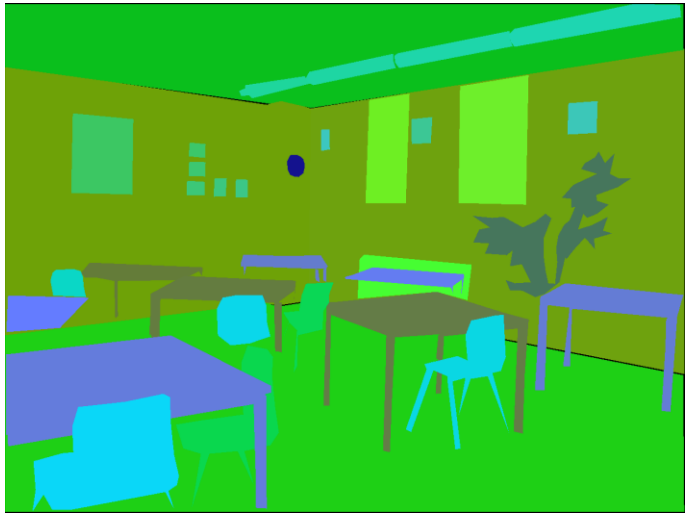
Goal: More efficient annotation for semantic segmentation.

Goal: More efficient annotation for semantic segmentation.

- Low cost
  - Low annotation time / monetary cost.
  - Depends on task difficulty and worker skill.
- High performance
  - Good training data for segmentation.
  - Depends on label quality (label completeness / label noise).

Existing Approaches

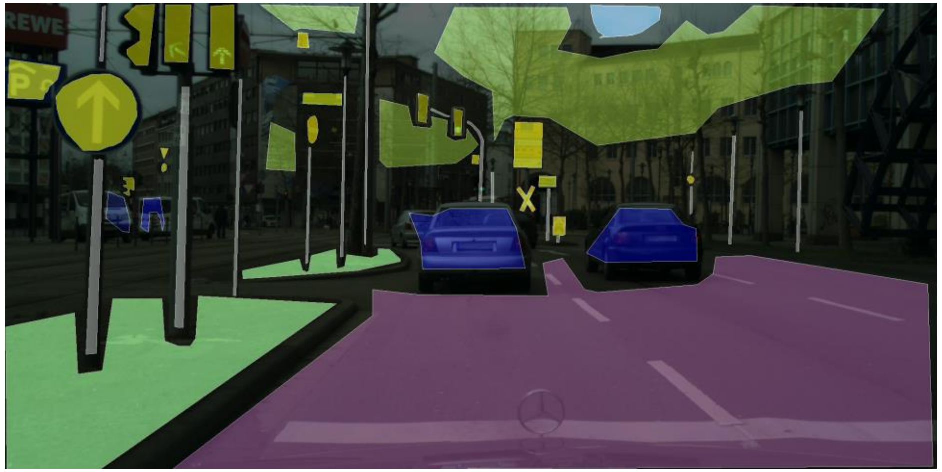
- Strong supervision: dense pixel-level labels
  - Gold standard, highest performance.
  - Expensive to collect (e.g. 90 min per images for Cityscapes).
  - Difficult task – many standard datasets utilize expert workers for annotation (Cityscapes, ADE20k, Pascal Context, Mapillary Vistas…).



[Cordts et al 2016; Zhou et al 2017]

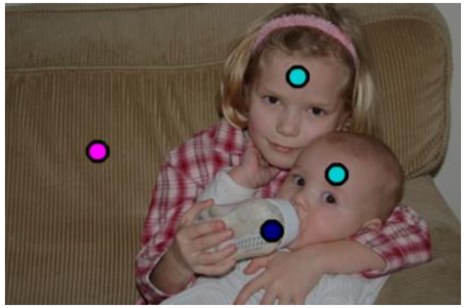
Existing Approaches

- Weak supervision
  - Dense pixel-level labels are expensive, so a lot of work on leveraging weak supervision for segmentation.
  - Weak supervision can be used stand-alone or in semi-supervised setting.

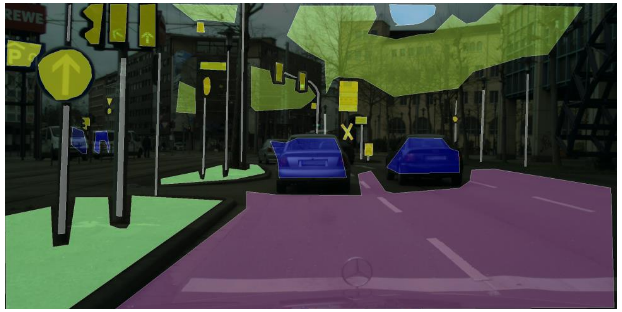


Existing Approaches

- Weak supervision



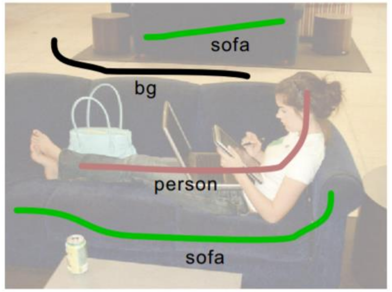
Point Clicks



Coarse Segments

Bounding Boxes

Image-level Labels



Scribbles

[Cordts et al 2016; Bearman et al 2015; Lin et al 2016; Dai et al 2015]

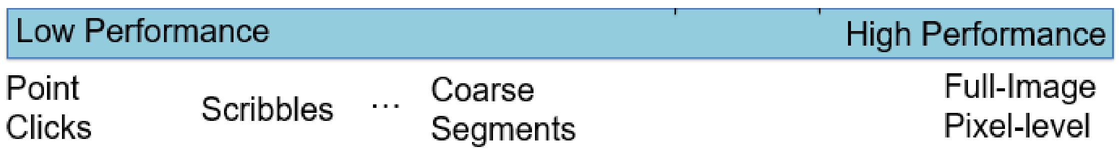
Existing Approaches

- What about interactive segmentation for annotation?
  - Deep Extreme Cut
  - Deep Object Selection
  - Fluid Annotation
  - Interactive Full Image Segmentation by Considering All Regions Jointly
  - PolygonRNN++
  - Curve-GCN
  - …
- Modern deep learning based approaches require seed training data. Manual annotation required for high quality seed training data.

Goal: More efficient annotation for semantic segmentation.

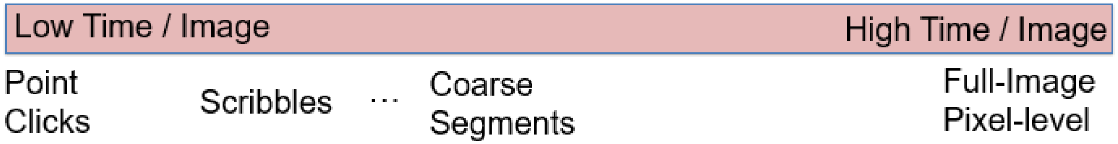
- Low cost

- High performance



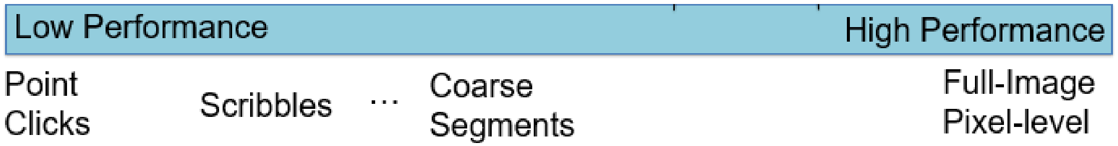
Goal: More efficient annotation for semantic segmentation.

- Low cost



?

- High performance



?

Block Annotation

- We propose spatially-constrained dense pixel-level annotation for a small block region in an image.

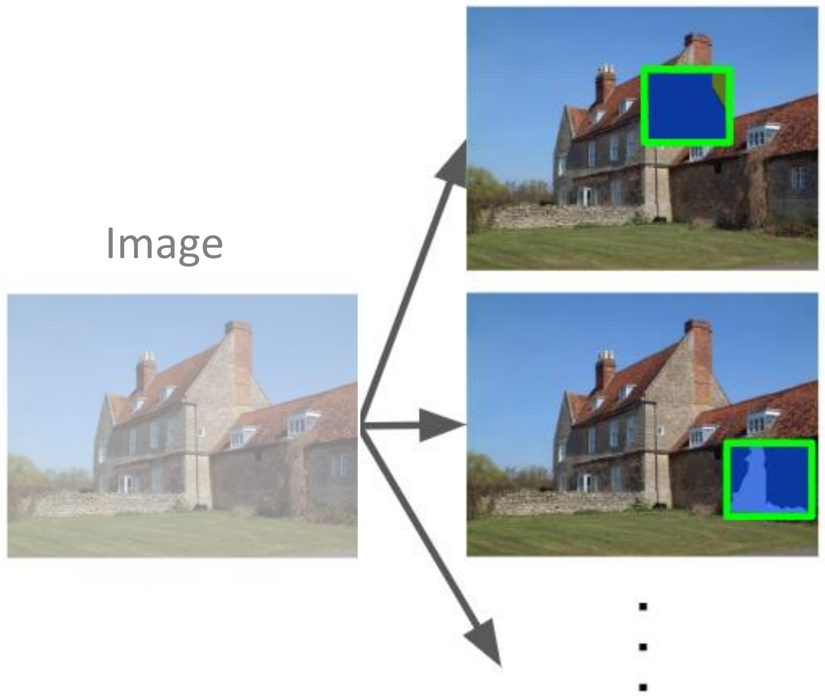
Block Annotation

- We propose spatially-constrained dense pixel-level annotation for a small block region in an image.

Image

Block Annotation

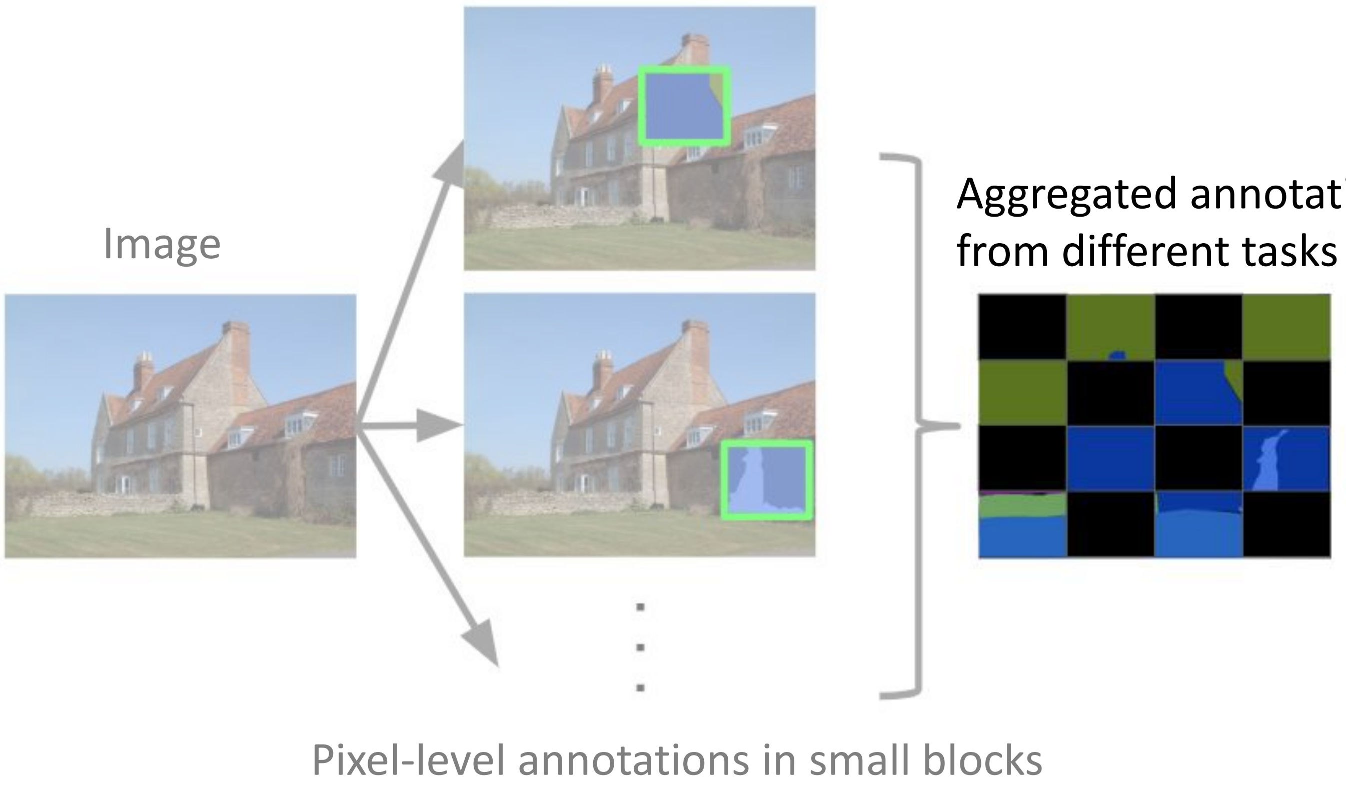
- We propose spatially-constrained dense pixel-level annotation for a small block region in an image.



Pixel-level annotations in small blocks

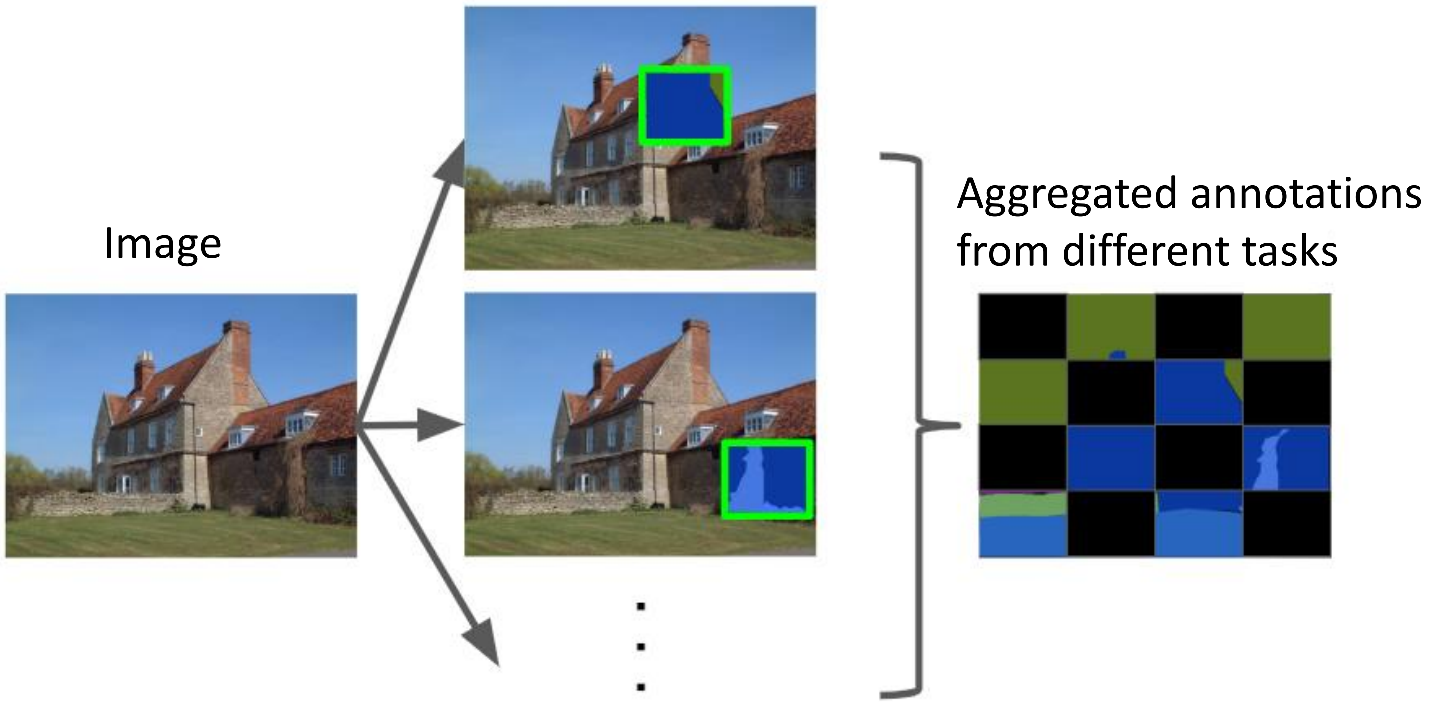
Block Annotation

- We propose spatially-constrained dense pixel-level annotation for a small block region in an image.



Block Annotation

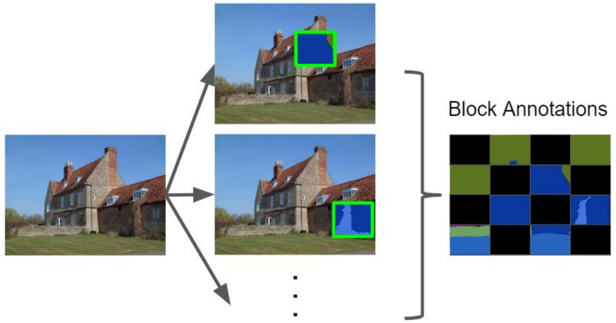
- We propose spatially-constrained dense pixel-level annotation for a small block region in an image.



Pixel-level annotations in small blocks

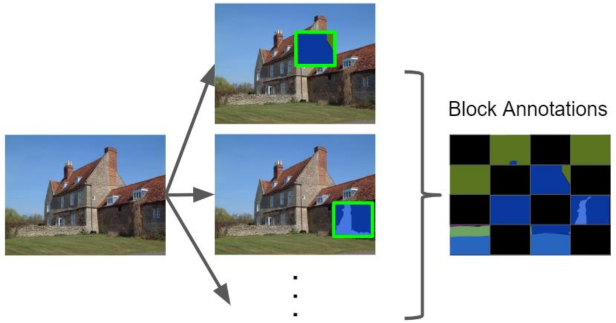
Block Annotation

- At first glance, this approach has several appealing properties:



Block Annotation

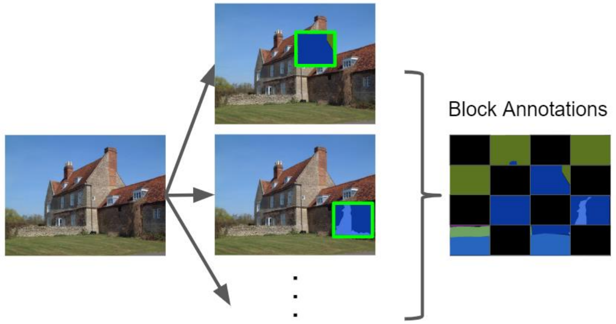
- At first glance, this approach has several appealing properties:
  - Compatibility with existing annotation tools.



Block Annotation

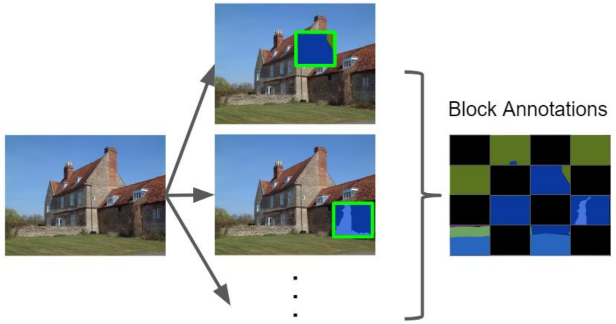
- At first glance, this approach has several appealing properties:

  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.



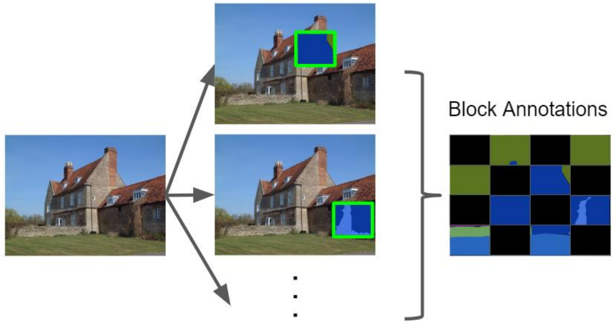
Block Annotation

- At first glance, this approach has several appealing properties:
  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.
  - Low difficulty.



Block Annotation

- At first glance, this approach has several appealing properties:

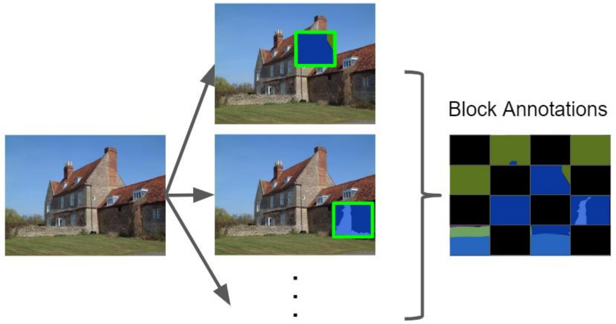


  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.

  - Low difficulty.
    - Workers only need to focus attention on small region.

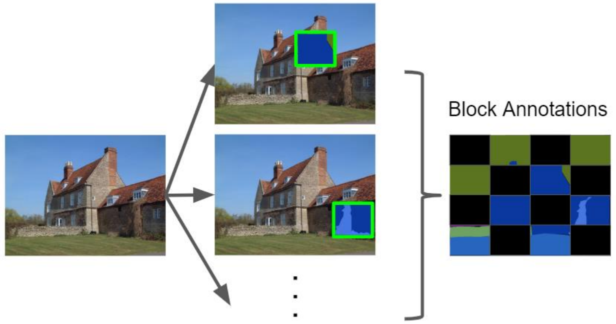
Block Annotation

- At first glance, this approach has several appealing properties:
  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.
  - Low difficulty.
    - Workers only need to focus attention on small region.
  - Annotated semantic boundaries.



Block Annotation

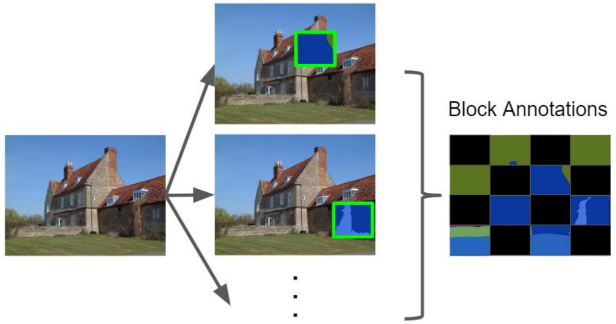
- At first glance, this approach has several appealing properties:



  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.
  - Low difficulty.
    - Workers only need to focus attention on small region.
  - Annotated semantic boundaries.
    - Pixel-level labels within annotated regions, in contrast to other forms of weak supervision.
    - Useful for training semantic segmentation networks.

Block Annotation

- At first glance, this approach has several appealing properties:
  - Compatibility with existing annotation tools.
    - Small image region is equivalent to a small image.
    - Tools and required skillsets are familiar to crowdworkers.
    - Negligible engineering required to deploy in practice.
  - Low difficulty.
    - Workers only need to focus attention on small region.
  - Annotated semantic boundaries.
    - Pixel-level labels within annotated regions, in contrast to other forms of weak supervision.
    - Useful for training semantic segmentation networks.



- This leads to several questions…

Block Annotation

- Research questions to explore:

Block Annotation

- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?

Block Annotation

- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?
  - What is the annotation cost / quality? How do workers respond to this task?

Block Annotation

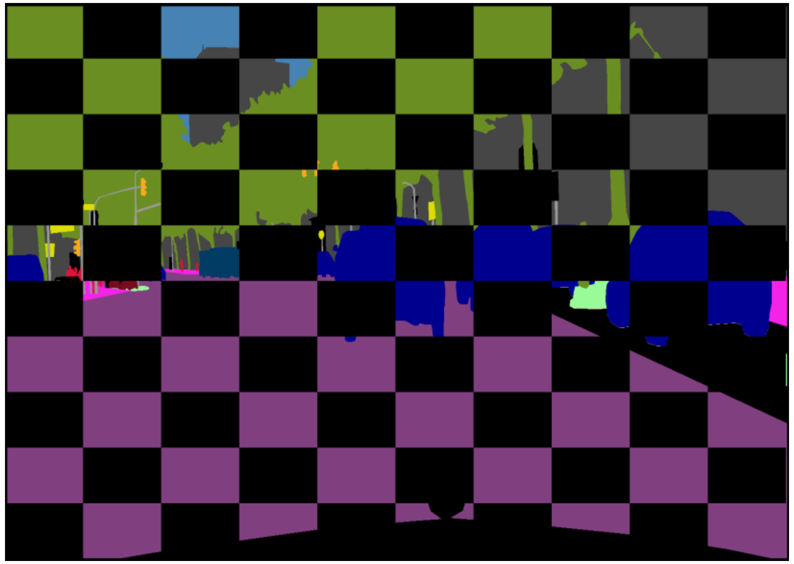
- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?
  - What is the annotation cost / quality? How do workers respond to this task?
  - Can a partially labeled image be converted to a fully labeled image automatically?

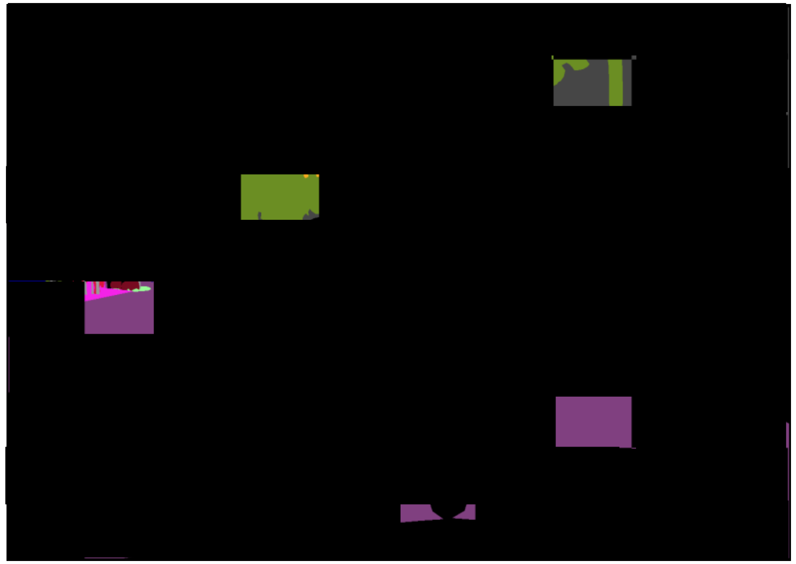
Block Annotation

- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?
  - What is the annotation cost / quality? How do workers respond to this task?
  - Can a partially labeled image be converted to a fully labeled image automatically?

Block Annotation: Semantic Segmentation

- Experimental set-up:
  - Datasets: Cityscapes, ADE20K. Chosen for variety in # classes, types of classes, environments.
  - Network: DeepLabv3+ w/ Xception backbone.
  - Block annotation: image divided into 10x10 grid; # of labeled blocks varies.



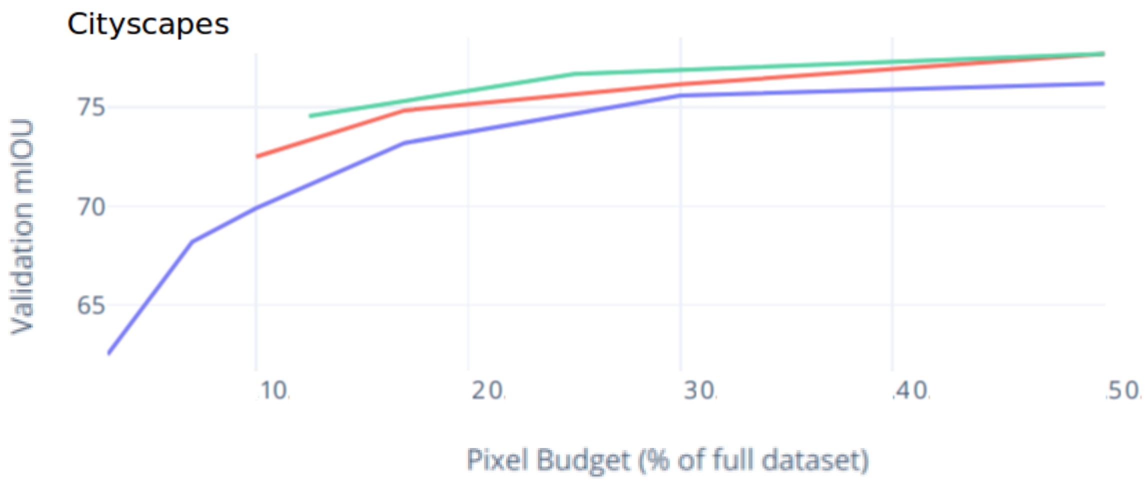


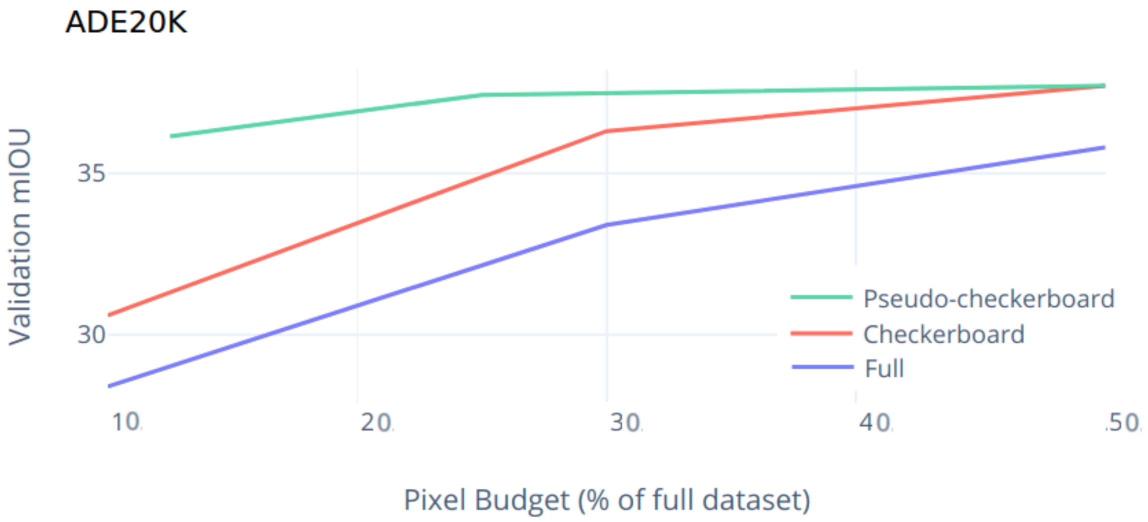
Block Annotation: Semantic Segmentation

- *Q: How does block annotation compare to full-image annotation?*

Block Annotation: Semantic Segmentation

- *Q: How does block annotation compare to full-image annotation?*
- Block annotation outperforms full-image annotation given the same number of annotated pixels.





Block Annotation: Semantic Segmentation

- *Q: How does block annotation compare to full-image annotation?*
- Block annotation achieves same performance as full-image annotation with half the pixels annotated.

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

Block Annotation: Semantic Segmentation

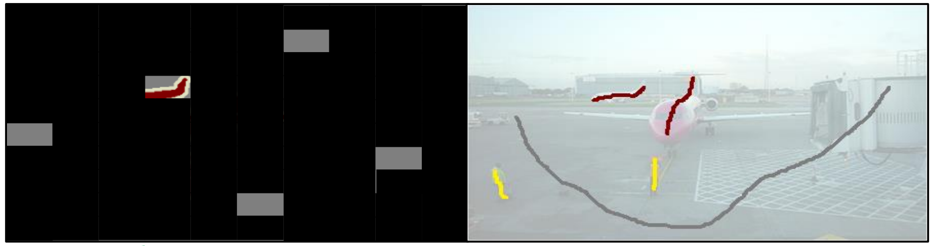
- *Q: How does block annotation perform against other forms of weak supervision?*

Block Annotation: Semantic Segmentation

- *Q: How does block annotation perform against other forms of weak supervision?*
- Block annotation outperforms existing weakly-supervised methods given equivalent annotation time.

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

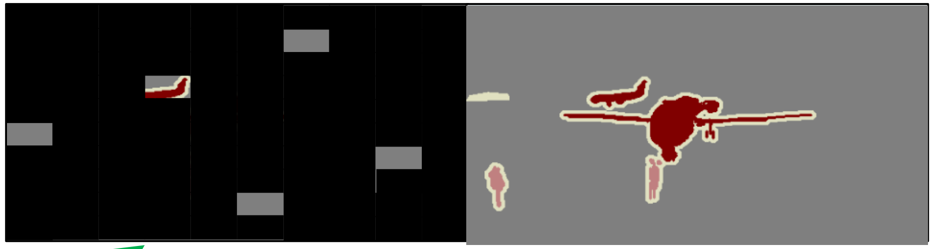


Block Annotation: Semantic Segmentation

- *Q: How does block annotation perform against other forms of weak supervision?*
- Block annotation achieves up to 97% of strong supervision with 1/10th annotation time.

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



Block Annotation: Semantic Segmentation

- *Q: How does block annotation perform against other forms of weak supervision?*
- Block annotation's performance does not depend on additional loss functions or label propagation (e.g. scribble/box methods)

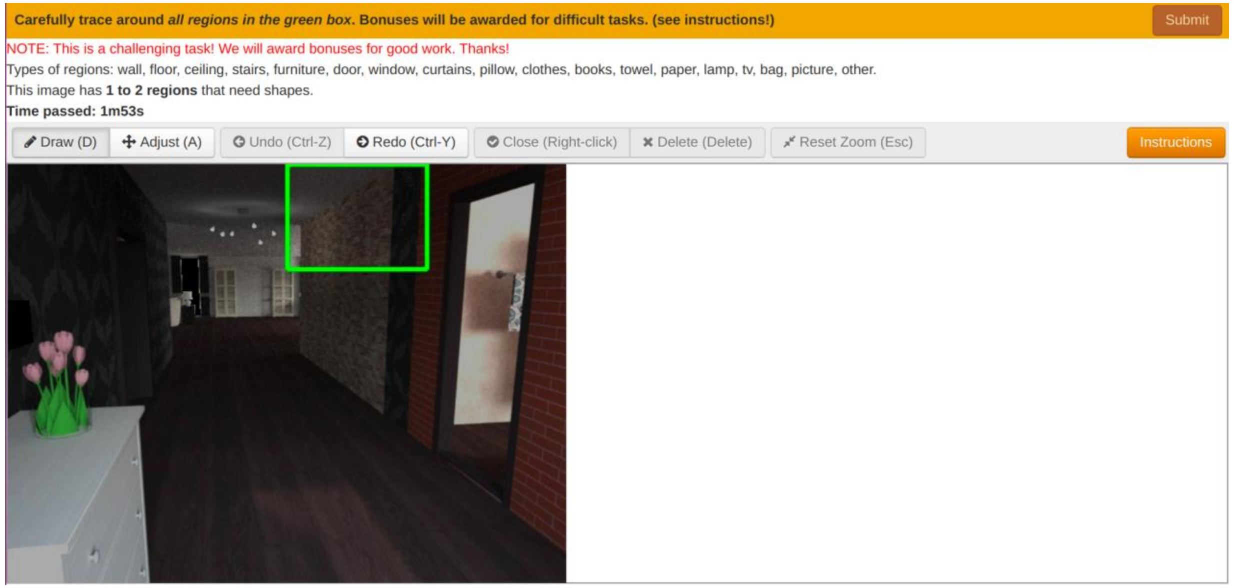
Method	Annotations	mIOU (%)
MIL-FCN [46]	Image-level	25.1
WSSL [45]	Image-level	38.2
point sup. [7]	Point	46.1
ScribbleSup [36]	Point	51.6
WSSL [45]	Box	60.6
BoxSup [15]	Box	62.0
ScribbleSup [36]	Scribble	63.1
Ours: Block-1%	Pixel-level Block	61.2
Ours: Block-5%	Pixel-level Block	67.6
Ours: Block-12%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

Block Annotation

- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?
  - What is the annotation cost / quality? How do workers respond to this task?
  - Can a partially labeled image be converted to a fully labeled image automatically?

Block Annotation: Annotation Quality / Cost

- Experimental setup:
  - Datasets: Cityscapes (representative of 'hard' dataset); SUNCG/CGIntrinsics (synthetic, has ground truth labels).
  - Interface based on OpenSurfaces
  - User study performed on Amazon Mechanical Turk

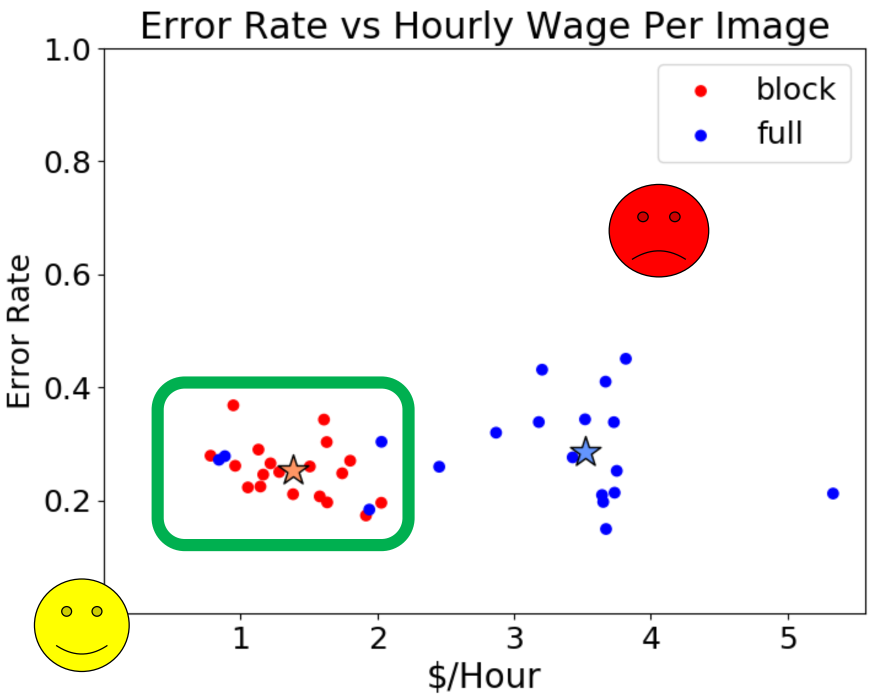


Block Annotation: Annotation Quality / Cost

- *Q: What is the cost of annotation?*

Block Annotation: Annotation Quality / Cost

- *Q: What is the cost of annotation?*
- We find that workers produce quality annotations while demanding lower wage. Quality (error rate) measured vs ground truth in SUNCG.

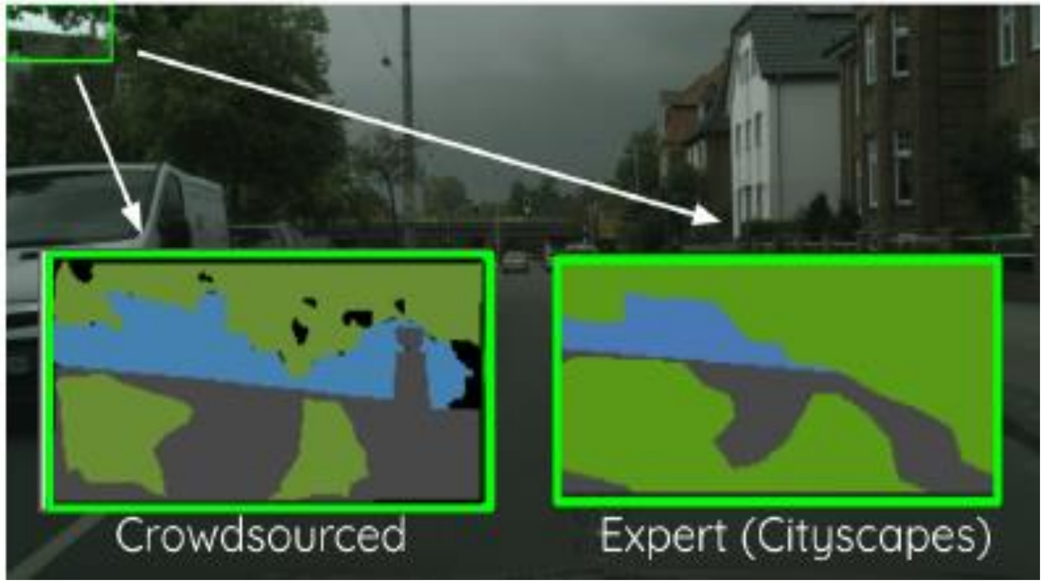


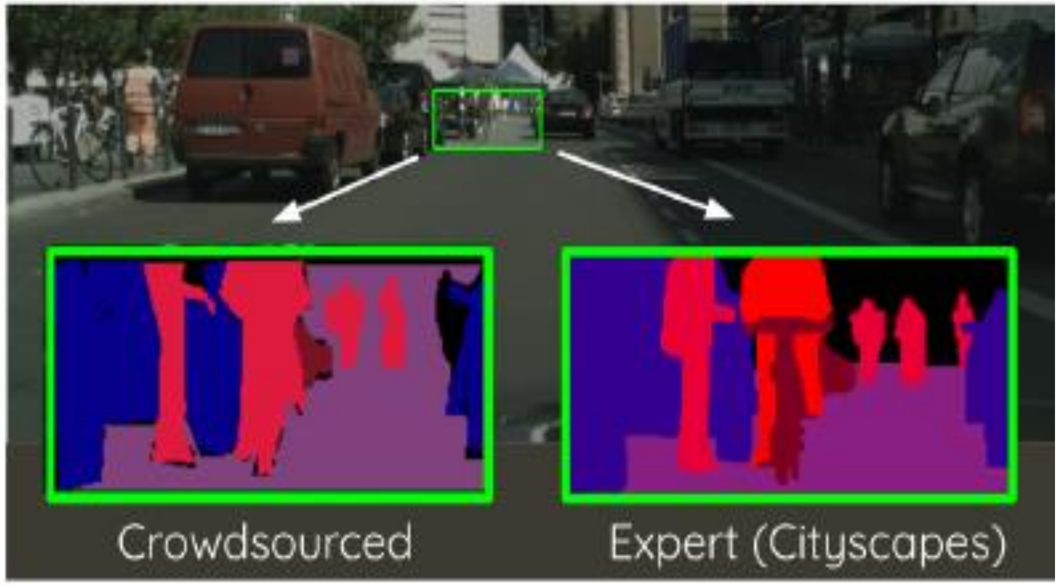
Block Annotation: Annotation Quality / Cost

- *Q: What is the quality of annotation?*

Block Annotation: Annotation Quality / Cost

- *Q: What is the quality of annotation?*
- We find crowdworkers produce work that is qualitatively comparable to work by expert workers on Cityscapes.





Block Annotation: Annotation Quality / Cost

- *Q: How do workers feel about the task?*

Block Annotation: Annotation Quality / Cost

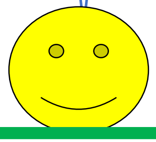
- *Q: How do workers feel about the task?*
- We receive overwhelmingly positive feedback from workers across both studies on SUNCG and Cityscapes.

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

Block Annotation: Annotation Quality / Cost

- *Q: How do workers feel about the task?*
- We receive overwhelmingly positive feedback from workers across both studies on SUNCG and Cityscapes.

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

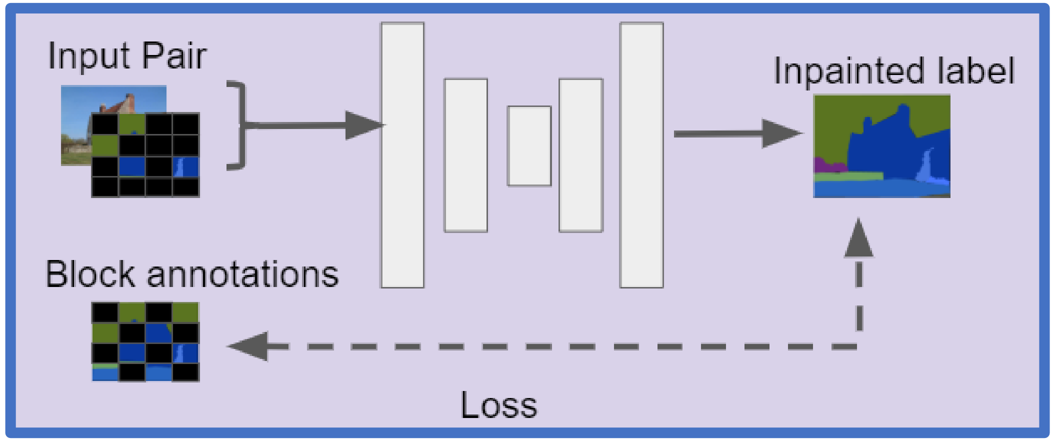


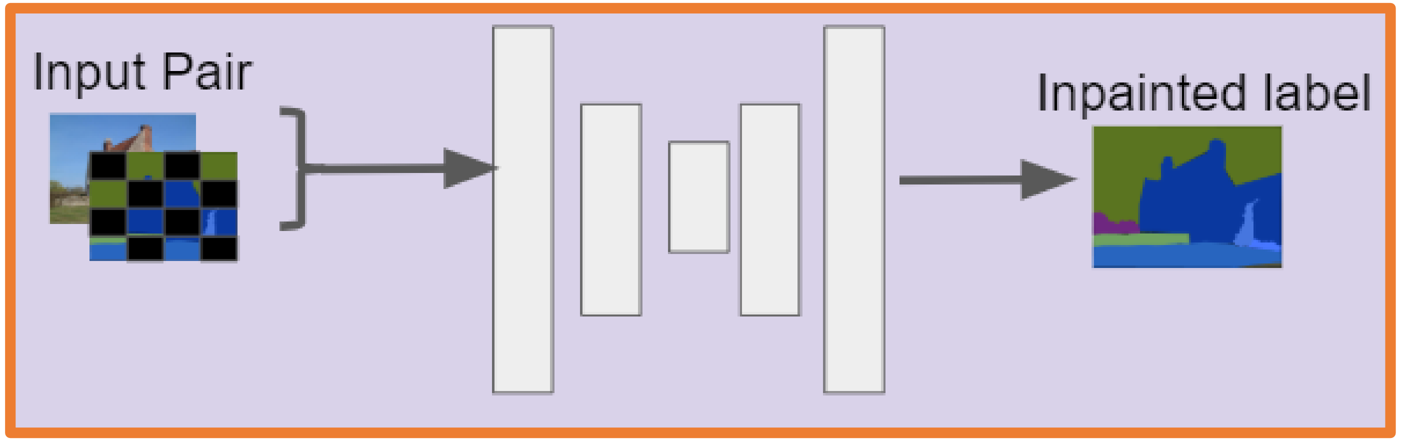
Block Annotation

- Research questions to explore:
  - Are these types of annotations effective for learning semantic segmentation, especially if only a small number of blocks are annotated?
  - What is the annotation cost / quality? How do workers respond to this task?
  - Can a partially labeled image be converted to a fully labeled image automatically?

Block Annotation: Sparse-to-Dense

- Experimental setup:
  - Datasets: Cityscapes, ADE20K
  - Network: DeepLabv3+ modified with input channel of labeled blocks.
  - Train by sampling annotated blocks; inference with all annotated blocks.





Block Annotation: Sparse-to-Dense

- *Q: What are the quality of the inpainted labels?*

Block Annotation: Sparse-to-Dense

- *Q: What are the quality of the inpainted labels?*

(a) Full human labels

(b) Original image

(c) Inpainted labels (all)

(d) Label agreement (white)

Block Annotation: Sparse-to-Dense

- *Q: What are the quality of the inpainted labels?*

(a) Full human labels

(b) Original image

(c) Inpainted labels (all)

(d) Label agreement (white)

mIOU: 92% (vs 78% from automatic segmentation)

Future Directions

- Adaptive block size, shape, and distribution
  - Estimate annotation difficulty and training signal from image content

Future Directions

- Adaptive block size, shape, and distribution
  - Estimate annotation difficulty and training signal from image content
- Active learning pipeline
  - Per-block labels instead of per-image labels
  - Can drive size/shape/distribution of blocks

Future Directions

- Adaptive block size, shape, and distribution
  - Estimate annotation difficulty and training signal from image content
- Active learning pipeline
  - Per-block labels instead of per-image labels
  - Can drive size/shape/distribution of blocks
- Improved label inpainting / label propagation
  - Leverage structure in annotated samples

Future Directions

- Adaptive block size, shape, and distribution
  - Estimate annotation difficulty and training signal from image content
- Active learning pipeline
  - Per-block labels instead of per-image labels
  - Can drive size/shape/distribution of blocks
- Improved label inpainting / label propagation
  - Leverage structure in annotated samples
- Instance-level annotation
  - Merge segments across annotated blocks

Future Directions

- Adaptive block size, shape, and distribution
  - Estimate annotation difficulty and training signal from image content
- Active learning pipeline
  - Per-block labels instead of per-image labels
  - Can drive size/shape/distribution of blocks
- Improved label inpainting / label propagation
  - Leverage structure in annotated samples
- Instance-level annotation
  - Merge segments across annotated blocks

Conclusions

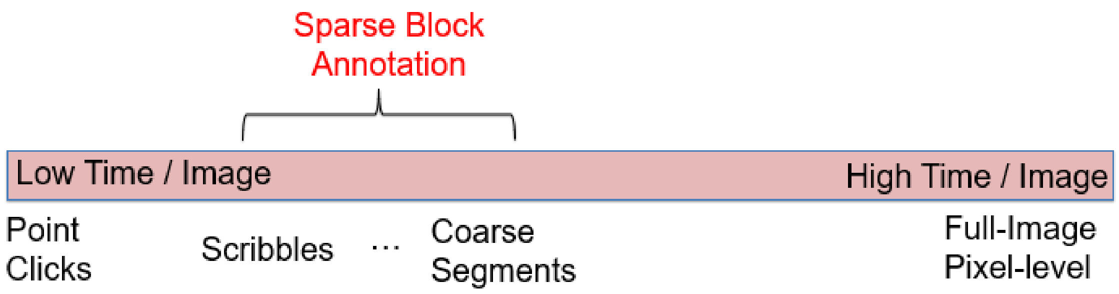
Sparse block annotations:

[*Block Annotation: Better Image Annotation with Sub-Image Decomposition.* Lin et al, ICCV 2019]

Conclusions

Sparse block annotations:

- are **scalable**, **cost-effective**, and **easy to implement**.

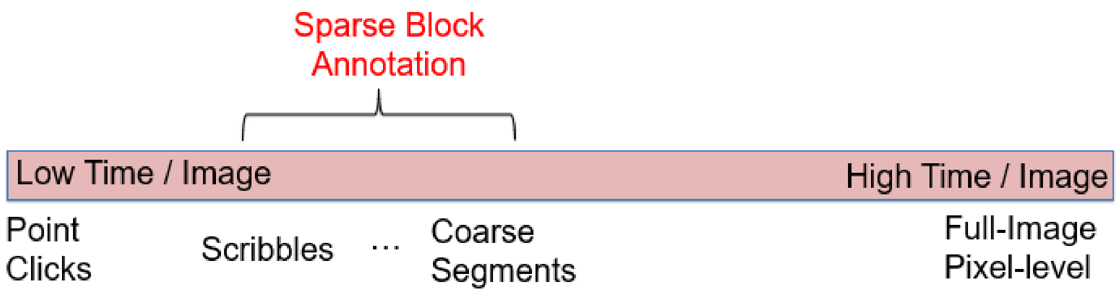


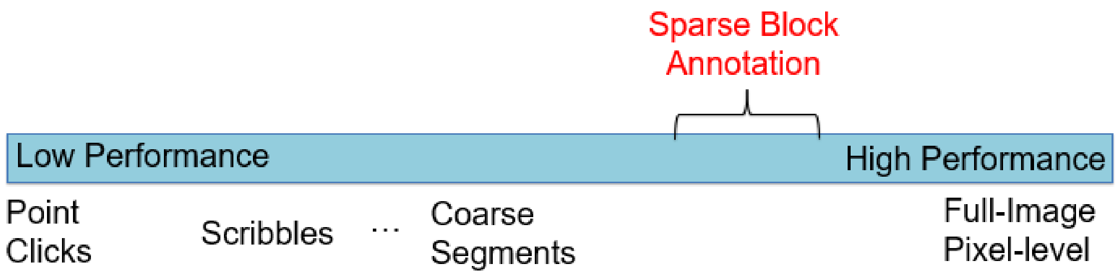
[*Block Annotation: Better Image Annotation with Sub-Image Decomposition.* Lin et al, ICCV 2019]

Conclusions

Sparse block annotations:

- are **scalable**, **cost-effective**, and **easy to implement**.
- enable **high semantic segmentation performance in weakly-supervised settings** and **scales to strongly-supervised performance**.



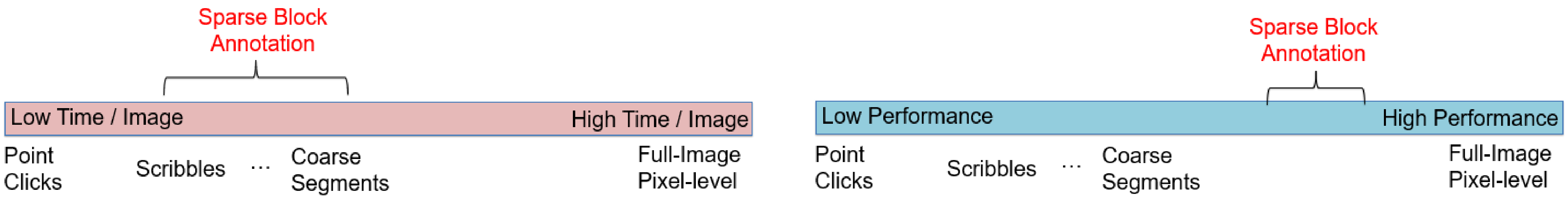


[*Block Annotation: Better Image Annotation with Sub-Image Decomposition.* Lin et al, ICCV 2019]

Conclusions

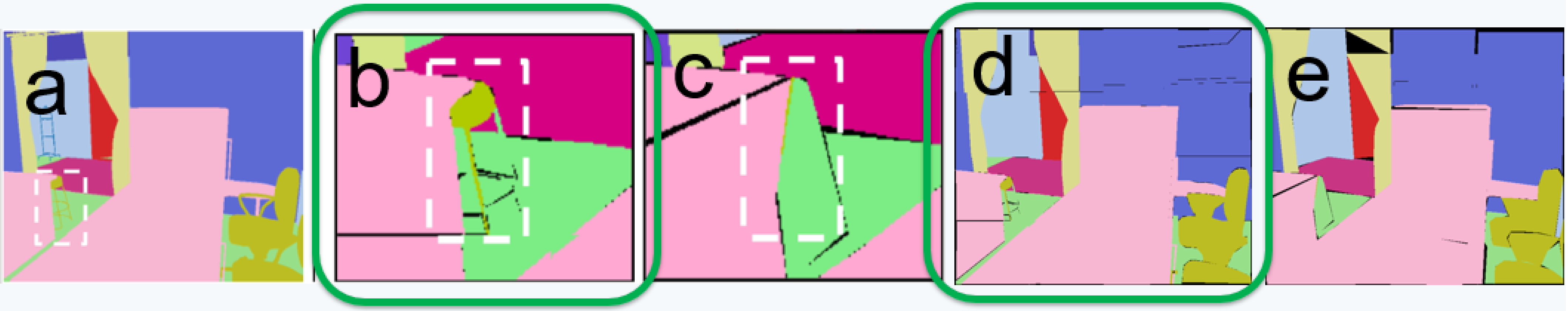
Sparse block annotations:

- are **scalable**, **cost-effective**, and **easy to implement**.
- enable **high semantic segmentation performance in weakly-supervised settings** and **scales to strongly-supervised performance**.
- can be converted to high quality full-image pixel-level annotations.



[*Block Annotation: Better Image Annotation with Sub-Image Decomposition.* Lin et al, ICCV 2019]

Thank you!



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).