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

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
  - Low Time / Image
    - Point Clicks
    - Scribbles
    - Coarse Segments
  - High Time / Image
    - Full-Image
    - Pixel-level

- High performance
  - Low Performance
    - Point Clicks
    - Scribbles
    - Coarse Segments
  - High Performance
    - Full-Image
    - Pixel-level
Goal: **More efficient** annotation for semantic segmentation.

- **Low cost**
  - Low Time / Image
    - Point Clicks
    - Scribbles
    - Coarse Segments

- **High performance**
  - High Time / Image
    - Full-Image Pixel-level
  - High Performance
    - Full-Image Pixel-level

?
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.
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.
Block Annotation

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

• 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

• 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.

This leads to several questions…
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.

• This leads to several questions…
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.

<table>
<thead>
<tr>
<th></th>
<th>Optimal (Full)</th>
<th>Block-50%</th>
<th>Block-12%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cityscapes</td>
<td>77.7</td>
<td>77.7</td>
<td>74.6</td>
</tr>
<tr>
<td>ADE20K</td>
<td>37.4</td>
<td>37.2</td>
<td>36.1</td>
</tr>
</tbody>
</table>
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.**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Ours: Block (7 min)</th>
<th>Coarse (7 min [14])</th>
<th>Full Supervision (90 min [14])</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cityscapes</td>
<td>72.1</td>
<td>68.8</td>
<td>77.7</td>
</tr>
<tr>
<td>mIOU (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pascal</td>
<td>Ours: Block (25 sec)</td>
<td>Scribbles (25 sec [36])</td>
<td>Full Supervision (4 min [41])</td>
</tr>
<tr>
<td>mIOU (%)</td>
<td>67.2</td>
<td>63.1 [36]</td>
<td>69.6</td>
</tr>
</tbody>
</table>
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.*

<table>
<thead>
<tr>
<th></th>
<th>Cityscapes</th>
<th>Pascal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><em>Ours:</em> Block (7 min)</td>
<td>Coarse (7 min [14])</td>
</tr>
<tr>
<td>mIOU (%)</td>
<td><strong>72.1</strong></td>
<td>68.8</td>
</tr>
<tr>
<td></td>
<td><em>Ours:</em> Block (25 sec)</td>
<td>Scribbles (25 sec [36])</td>
</tr>
<tr>
<td>mIOU (%)</td>
<td><strong>67.2</strong></td>
<td>63.1 [36]</td>
</tr>
</tbody>
</table>
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)

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

<table>
<thead>
<tr>
<th></th>
<th>&quot;Nice&quot;</th>
<th>&quot;Fun&quot;</th>
<th>&quot;Easy&quot;</th>
<th>&quot;Okay&quot;</th>
<th>Release More HITs</th>
<th>Increase Pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th></th>
<th>“Nice”</th>
<th>“Good”</th>
<th>“Fun”</th>
<th>“Easy”</th>
<th>“Okay”</th>
<th>Release More HITs</th>
<th>Increase Pay</th>
</tr>
</thead>
<tbody>
<tr>
<td>#</td>
<td>8</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
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?
Block Annotation: Sparse-to-Dense

Q: What are the quality of the inpainted labels?

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).