Weakly Supervised Semantic Segmentation

VALSE 2019 Tutorial

Yunchao Wei

University of Illinois at Urbana-Champaign
Overview: Fully-supervised Semantic Segmentation

- FCN
- Deeplab v1, v2, v3, v3+
- PSPNet
- RefineNet
- GCN
- EncNet
- Dense ASPP
- PSANet
- Non-Local Networks
- DANet, OCNet, CCNet
- ....
Overview: Manual Annotations for Object Recognition

image-level labels

points

bounding boxes

scribbles

pixel-level labels

1s/class

2.4s/instance

10s/instance

17s/instance

78s/instance

Annotation time

[Lin CVPR16, Berman ECCV16, Hakan CVPR18 Tutorial]
Overview: Heavy Cost in Labeling Pixel-level Masks
Overview: Heavy Cost in Labeling Pixel-level Masks
Overview: Weakly-supervised Semantic Segmentation

Weak Supervision

*Lower degree* (or *cheaper, simper*) annotations at *training stage* than the required outputs at the *testing stage*.

**Training Stage** (Weakly-supervised Annotations)

- image-level labels
- points
- bounding boxes
- scribbles

**Testing Stage**

- pixel-level labels

[Hakan CVPR18 Tutorial]
Current State-of-the-arts

Object Semantic Segmentation

- **VGG**
  - Tang ECCV18: 66.7
  - Tang ECCV18: 75.0
- **Res101**
  - Tang ECCV18: 65.7
- **VGG**
  - Khoreva CVPR17: 51.6
- **VGG**
  - Lin CVPR16: 61.3
- **Res101**
  - Fan ECCV18: 63.6

Scene Parsing

- **Res101**
  - Qian AAAI19: 30.0
  - Qian AAAI19: 19.6
- **VGG**
  - Lin CVPR16: 36.1

Pascal VOC

- Scribbles: 4.8
- Boxes: 1.8
- Points: 3.4
- Image-level labels: 16.9

Pascal Context

- Points: 9.6
- Scribbles: 14.3

ADE20K

- Points: 14.3
Overview: About This Tutorial

The Covered Topics

- WS Object Semantic Segmentation
- WS Scene Parsing
- WS Instance Segmentation
- Interactive Object Segmentation

Publisher

CVPR 42%
ECCV 21%
ICCV 16%
Others 21%
Overview: Researcher Distribution

- Google, Facebook, IBM
- MSR, Adobe, MIT, UCB, CMU, UIUC, Stanford, UCLA, etc.
- Tsinghua, Peking, CAS, HUST, Nankai, SYSU, BJTU, MSRA, Tencent, etc.
- UTsinghua, Peking, CAS, HUST, Nankai, SYSU, BJTU, MSRA, Tencent, etc.
- NUS, ANU, UTS
- KAIST, POSTECH, DGIST, etc.
- Oxford U, EPFL, ETH, U Amsterdam, U Edinburgh, Max Planck Institute, etc.
Outline

image-level labels

points

bounding boxes

scribbles
Outline

image-level labels

person

points

bounding boxes

scribbles
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation

image-level labels
▪ End-to-end Learning with Constraint Loss
  ▪ Learning to Produce Pseudo Pixel-level Masks
    ▪ Additional Data
    ▪ Object Proposals
    ▪ Top-down Attention
  ▪ Semi-Supervised Learning
  ▪ Instance Segmentation
Multiple Instance Learning

Input

FCN

Confidence Maps: (c+1)xhwxw

Aggregation output: (c+1)x1x1

[Pinheiro CVPR15, Pathak ICLR15 Workshop]
End-to-end Learning with Constraint Loss

Constrained Convolutional Neural Networks

Suppression Constraint
\[ \sum_{i=1}^{n} p_i(l) \leq 0 \quad \forall \, l \notin \mathcal{L}_I \]

Foreground Constraint
\[ a_l \leq \sum_{i=1}^{n} p_i(l) \quad \forall \, l \in \mathcal{L}_I \]

Background Constraint
\[ a_0 \leq \sum_{i=1}^{n} p_i(0) \leq b_0. \]

[Chen ICCV15, Pathak ICCV15]
End-to-end Learning with Constraint Loss

Distinct Class-Specific Saliency Maps

\[ \hat{M}_{i,x,y} = \sum_{c' \in \text{candidates}} \max \left( M_{i,x,y} - M'_{i,x,y}, 0 \right) [c \neq c'] \]

[Shimoda ECCV16]
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation
Learning to Produce Pseudo Pixel-level Masks

- **Standard Pipeline**

[Diagram showing a standard pipeline for producing pseudo pixel-level masks. The pipeline consists of an input image, a FCN, and a loss function. The FCN is used to produce pseudo masks that are compared with the ground truth masks to calculate the loss.]
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation
Learning to Produce Pseudo Masks--Additional Data

**Simple to Complex**

<table>
<thead>
<tr>
<th>Networks</th>
<th>Training Set</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>I-DCNN</td>
<td>Flickr-Clean</td>
<td>44.1</td>
</tr>
<tr>
<td>E-DCNN</td>
<td>Flickr-Clean</td>
<td>46.8</td>
</tr>
<tr>
<td>P-DCNN</td>
<td>Flickr-Clean+VOC</td>
<td>49.8</td>
</tr>
</tbody>
</table>

Ablation Analysis on Pascal VOC12 val

[Wei PAMI17]
Learning to Produce Pseudo Masks—Additional Data

Webly Supervised Semantic Segmentation

[Jin CVPR17]
Learning to Produce Pseudo Masks -- Additional Data

Web-Crawled Videos

Section 3.1 Learning with images

Section 3.2 Generating attention

Section 3.3 Learning with videos

Eliminating irrelevant frames

Video segmentation

[Hong CVPR17]
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation
Learning to Produce Pseudo Masks--Object Proposals

Multi-Evidence Filtering and Fusion [Ge CVPR18]
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation
Learning to Produce Pseudo Masks—Top-down Attention

Class Activation Mapping [Zhou CVPR16]

Class Activation Mapping

[Zhou CVPR16, Zhang ECCV16, Zhu ICCV17, Zhang CVPR18, Zhang ECCV18]
Learning to Produce Pseudo Masks--Top-down Attention

Seed, Expand and Constrain

[Kolesnikov ECCV16, Roy CVPR17]
Learning to Produce Pseudo Masks—Top-down Attention

Built-in Fg/Bg Prior

Tags: Horse Person

Fusion of conv4 and conv5

Pre-trained Localization Network

Generating Class Activation Map for presented classes

Score Map

Loss Function for semantic segmentation

Unary Potentials

Dense CRF with higher order potentials

Generated Labeled Mask

[Saleh ECCV16, Saleh PAMI17]
Learning to Produce Pseudo Masks---Top-down Attention

Built-in Fg/Bg Prior

[Image showing convolutional layers and masks]

[Saleh ECCV16, Saleh PAMI17]
Learning to Produce Pseudo Masks--Top-down Attention

Exploiting Saliency

dense labelling from seed + saliency
seed
saliency
guide labeller
forward backprop

dense classifier loss

segmenter convnet

(a) Foreground categories
(b) Background category

Fusion Strategies

(a) Image
(b) Ground truth
(c) Seed
(d) Saliency
(e) $G_0$
(f) $G_1$
(g) $G_2$

[Oh CVPR17]
Object Region Mining with Adversarial Erasing

[Wei CVPR17]
Learning to Produce Pseudo Masks—Top-down Attention

Object Region Mining with Adversarial Erasing

Image

AE-Step1

AE-Step2

AE-Step3

Erased Regions

[Wei CVPR17]
Learning to Produce Pseudo Masks—Top-down Attention

Two-Phase Learning

Training image → Fixed 1st network parameters → Per-class heat map → Selected class → Thresholding → Suppression mask

Two-phase learning:
1. **Fixed 1st network parameters**: These parameters are kept constant throughout the learning process.
2. **Per-class heat map**: The network produces a heat map for each class.
3. **Selected class**: A class is selected based on some criteria.
4. **Thresholding**: The output is thresholded to create a binary mask.

**Trainable 2nd network parameters**

- **Conv1, Conv2, Conv3, Conv4, Conv5, Conv6, Conv7, Conv8**: Different convolution layers.
- **Max, Max, Max**: Max pooling layers.
- **Conv, Conv, GAP**: Convolution and Global Average Pooling layers.

**Element-wise multiplication**

- **Multi-label logistic loss**: Computes the loss for each label.
- **Image-level labels**: Labels for the entire image.

**Inference-conditional feedback path**

[Kim ICCV17]
Learning to Produce Pseudo Masks—Top-down Attention

Discovering Class-Specific Pixels

Algorithm 1 Discovering Class-Specific Pixels

**Input:** Image Labels $z$; Saliency Map $S$; Attention Maps $A$; $\gamma$

1: $M = \text{zeros}(n)$, where $n$ is the number of pixels
2: for each $c \in z$ and each pixel $m$ do
3: \hspace{1em} $H(m, c) = h(A^c(m), S(m))$
4: \hspace{1em} end for
5: for each pixel $m$ do
6: \hspace{2em} if $H(m) < \gamma$ then
7: \hspace{3em} $M(m) = I_0$
8: \hspace{2em} else
9: \hspace{3em} $M(m) = \text{argmax}(H(m))$
10: \hspace{2em} end if
11: end for

**Output:** Localization cues or approximate labeling $M$

[Chaudhry BMVC17]
Learning to Produce Pseudo Masks—*Top-down Attention*

**Erasing for Mining**

**Hide-and-Seek**

*Training phase*

- W
- H
- S

Training image

- Epoch 1
- Epoch 2
- Epoch N

**Adversarial Complementary Learning**

- CNN
- **Backbone**
- **GAP**
- Classifier A
- Softmax Loss
- Classifier B
- Softmax Loss

[Singh ICCV17, Zhang CVPR18]
Learning to Produce Pseudo Masks—Top-down Attention

Guided Attention Inference Network

[Li CVPR18]
Learning to Produce Pseudo Masks—Top-down Attention

Erasing for Mining

Over Erasing

as the erasing goes on

[Wei CVPR17]
Learning to Produce Pseudo Masks--Top-down Attention

Self-Erasing Network

Image | Attention Map | Ternary Mask

[Hou NIPS18]
Learning to Produce Pseudo Masks—Top-down Attention

Self-Erasing Network

Attention Map

$$T_{A,(i,j)} = 0 \text{ if } M_{A,(i,j)} \geq \delta_h$$
$$T_{A,(i,j)} = -1 \text{ if } M_{A,(i,j)} < \delta_l$$
$$T_{A,(i,j)} = 1 \text{ otherwise}$$

Ternary Mask

background priors

object priors

[Hou NIPS18]
Learning to Produce Pseudo Masks--Top-down Attention

Self-Erasing Network

[Erasing for Mining]

[Hou NIPS18]
Learning to Produce Pseudo Masks—Top-down Attention

Deep Seeded Region Growing

\[
P(H_{u,c}, \theta_c) = \begin{cases} 
\text{TRUE} & H_{u,c} \geq \theta_c \text{ and } c = \arg \max_{c'} H_{u,c'}, \\
\text{FALSE} & \text{otherwise.}
\end{cases}
\]

[Huang CVPR18]
Learning to Produce Pseudo Masks--Top-down Attention

AffinityNet

Training AffinityNet (Section 3.2)
Generating Segmentation Labels (Section 3.3, 3.4)
Learning Segmentation Net

Generating Semantic Affinity Labels

Ahn CVPR18
Learning to Produce Pseudo Masks—*Top-down Attention*

Multi-dilated Convolution

Dilated convolution with different rates

Conv Feature Maps

Conv Kernels

Conv Kernels

Conv Kernels

Conv Kernels

Conv Kernels

[Wei CVPR18]
Learning to Produce Pseudo Masks—Top-down Attention

Multi-dilated Convolution

[Wei CVPR18]
▪ End-to-end Learning with Constraint Loss
▪ Learning to Produce Pseudo Pixel-level Masks
  ▪ Additional Data
  ▪ Object Proposals
  ▪ Top-down Attention
▪ Semi-Supervised Learning
▪ Instance Segmentation
Semi-Supervised Learning

EM Adapt/Fix

- Pixel Annotations
- Deep Convolutional Neural Network
- Loss
- Image-level Annotation
- Score maps
- FG/BG Bias
- argmax
- Weakly-Supervised E-step

Multi-dilated Convolution

- Dilated Convolution for Object Localization
- Segmentation Mask Generation
- Argmax
- Loss

1. Car
2. Person
3. Horse

[Chen ICCV15, Wei CVPR18]
Semi-Supervised Learning

Guided Attention Inference Network

[Li CVPR18]
Transferable Semi-supervised Semantic Segmentation

**In-category** Semi-supervised Semantic Segmentation

- **Image-level labels**
  - Dog
  - Cat
  - In-category pixel-level labels
  - Dog
  - Cat

  Testing image

**Cross-category** Semi-supervised Semantic Segmentation

- **Image-level labels**
  - Dog
  - Cat
- **Cross-category pixel-level labels**
  - Horse
  - Aeroplane

  Testing image

[Xiao AAAI18]
Semi-Supervised Learning

Transferable Semi-supervised Semantic Segmentation

Label Transfer Network (L-Net)

Prediction Transfer Network (P-Net)

[Xiao AAAI18]
Semi-Supervised Learning

Transferable Semi-supervised Semantic Segmentation

Vehicles

Animals

Others

[Xiao AAAI18]
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation
Instance Segmentation

Peak Response Maps

Input Image

Feature Encoding

Class Response Map

(bird)

[Zhou CVPR18]
Instance Segmentation

Peak Response Maps

Input Image → Feature Encoding → Class Response Map (bird)

[Zhou CVPR18]
Instance Segmentation

End-to-End training with standard classification settings

Class Response Maps

Scores

Cross Entropy Loss

[Zhou CVPR18]
Peak Response Maps

\[ f(X^{l-1}) \ast w = X^l \]

Model \textit{Top-down relevance} between locations

[Zhou CVPR18]
Peak Response Maps

Model Top-down relevance between locations

$P$ transition probability

[Zhou CVPR18]
Instance Segmentation

Peak Response Maps

Pipeline

Peak Backprop

Class Response Maps

Input Image

[Zhou CVPR18]
Outline

image-level labels

points

bounding boxes

scribbles
- PointSup: Object Semantic Segmentation
- PointSup: Scene Parsing
- PointSup: Object Semantic Segmentation

- PointSup: Scene Parsing
PointSup: Object Semantic Segmentation

What's the Point

Image-Level Supervision:

$$\mathcal{L}_{img}(S, L, L') = -\frac{1}{|L|} \sum_{c \in L} \log(S_{t,c}) - \frac{1}{|L'|} \sum_{c \in L'} \log(1 - S_{t,c})$$

Point-Level Supervision:

$$\mathcal{L}_{point}(S, G, L, L') = \mathcal{L}_{img}(S, L, L') - \sum_{i \in \mathcal{I}_s} \alpha_i \log(S_{i,G_i})$$

[Bearman ECCV16]
- PointSup: Object Semantic Segmentation

- PointSup: Scene Parsing
PointSup: Scene Parsing

Point-based Distance Metric Learning

Image

FullSup

PointSup

The Number of Annotated Pixels

170K → 12.26

[Qian AAAI19]
PointSup: Scene Parsing

Confident Constraint

\[ M_{a \text{-score}} = \bigcup_u \{ \text{max}(g_{u,c} \in C(f(I_a; \theta_{f1}); \theta_{g1})) > \text{thr} \} \]

Spatial Constraint

\[ M_{a \text{-region}} \text{ is a 5X5 square around the ground-truth point} \]

\[ M_{a \text{-extend}} = M_{a \text{-score}} \cap M_{a \text{-region}} \]

[Qian AAAI19]
PointSup: Scene Parsing

Point-based Distance Metric Learning

Qualitative Comparison on PASCAL-Context

Results on PASCAL-Context and ADE 20K

[Qian AAAI19]
Outline

image-level labels  
points  
bounding boxes  
scribbles

horse  
person  
horse  
person
- BoxSup: Iterative Feedback
- BoxSup: Mining Strategies for Mask Generation
- BoxSup: Semi-supervised Learning
- **BoxSup: Iterative Feedback**
  - BoxSup: Mining Strategies for Mask Generation
  - BoxSup: Semi-supervised Learning

**Bounding boxes**
BoxSup: Iterative Feedback

$$\min_{\theta, \{l_S\}} \sum_i (E_o + \lambda E_r)$$

$$E_o = \frac{1}{N} \sum_S (1 - IoU(B, S)) \delta(l_B, l_S)$$

$$E_r = \sum_p e(X_\theta(p), l_S(p)).$$

[Dai ICCV15]
- BoxSup: Iterative Feedback
- BoxSup: Mining Strategies for Mask Generation
- BoxSup: Semi-supervised Learning
BoxSup: Mining Strategies for Mask Generation

[Chen ICCV15, Khoreva CVPR17]
BoxSup: Mining Strategies for Mask Generation

<table>
<thead>
<tr>
<th>Method</th>
<th>val. mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Fast-RCNN</td>
<td>44.3</td>
</tr>
<tr>
<td>GT Boxes</td>
<td>62.2</td>
</tr>
<tr>
<td>Box</td>
<td>61.2</td>
</tr>
<tr>
<td>Box¹</td>
<td>62.7</td>
</tr>
<tr>
<td>Weakly supervised</td>
<td></td>
</tr>
<tr>
<td>MCG</td>
<td>62.6</td>
</tr>
<tr>
<td>GrabCut+</td>
<td>63.4</td>
</tr>
<tr>
<td>GrabCut+¹</td>
<td>64.3</td>
</tr>
<tr>
<td>M ∩ G+</td>
<td>65.7</td>
</tr>
<tr>
<td>Fully supervised</td>
<td></td>
</tr>
<tr>
<td>DeepLab_{ours} [5]</td>
<td>69.1</td>
</tr>
</tbody>
</table>

[Chen ICCV15, Khoreva CVPR17]
- BoxSup: Iterative Feedback
- BoxSup: Mining Strategies for Mask Generation
- BoxSup: Semi-supervised Learning
BoxSup: Semi-supervised Learning

Green Boxes: Instance mask annotations
Red Boxes: Bounding box annotations

[Hu CVPR18]
BoxSup: Semi-supervised Learning

Mask$^X$ R-CNN

[Hu CVPR18]
- ScribbleSup: Learning with Alternative Optimization

- ScribbleSup: Learning with Joint Optimization
- ScribbleSup: Learning with Alternative Optimization

- ScribbleSup: Learning with Joint Optimization
ScribbleSup: Learning with Alternative Optimization

\[
\sum_i \psi_i(y_i | X, S) + \sum_{i,j} \psi_{ij}(y_i, y_j | X)
\]

\[
\psi_i^{scr}(y_i) = \begin{cases} 
0 & \text{if } y_i = c_k \text{ and } x_i \cap s_k \neq \emptyset \\
-\log \left( \frac{1}{|\{c_k\}|} \right) & \text{if } y_i \in \{c_k\} \text{ and } x_i \cap S = \emptyset \\
\infty & \text{otherwise}
\end{cases}
\]

\[
\psi_i^{net}(y_i) = -\log P(y_i | X, \Theta)
\]

\[
\psi_{ij}(y_i, y_j | X) = [y_i \neq y_j] \exp \left\{ - \frac{||h_c(x_i) - h_c(x_j)||_2^2}{\delta_c^2} - \frac{||h_t(x_i) - h_t(x_j)||_2^2}{\delta_t^2} \right\}
\]

[Lin CVPR16]
• ScribbleSup: Learning with Alternative Optimization

• ScribbleSup: Learning with Joint Optimization
ScribbleSup: Learning with Joint Optimization

Semi-supervised Optimization Problem:

\[
\min_{\theta} \left\{ \ell(f_{\theta}(I), Y) + \lambda \cdot R(f_{\theta}(I)) \right\}
\]

\[\sum_{p \in \Omega_L} H(Y_p, S_p) + \lambda \cdot R(S)\]

Ground-truth Loss
Regularization Loss

[R_{NC}(S) = \sum_{k} \frac{S^{k'} \hat{W}(1 - S^k)}{d'S^k}]

[R_{KC}(S) = \sum_{k} S^{k'} \hat{W}(1 - S^k) + \gamma \sum_{k} \frac{S^{k'} \hat{W}(1 - S^k)}{d'S^k}]

[Tang CVPR18, Tang ECCV18]
Outline

image-level labels

points

bounding boxes

scribbles

Beyond Weakly Supervised Semantic Segmentation
Interactive Image Segmentation

- Interactive Image Segmentation
  - Semi-automated, class-agnostic segmentation
  - Target segmentation depends on the user inputs (e.g. bounding box, points)
  - Allows iterative refinement until result is satisfactory
Interactive Image Segmentation

- Common types of Inputs
  - Regional Scribbles/points
  - Boundary points
  - Bounding box

(a) Regional points  (b) Boundary points  (c) Bounding box
Interactive Image Segmentation

- **Standard pipeline**
  - The sparse inputs are transformed using either:
    - Euclidean distance transform (Xu CVPR 16, Liew ICCV 17)
    - Gaussian transform (Maninis CVPR 18, Mahadevan BMVC 18)
  - Train end-to-end with FCNs (*e.g.*, FCN-8s, DeepLabv2-PSP, DeepLabv3+)

![Diagram of image segmentation process](image.png)
Interactive Image Segmentation

- [Xu CVPR16]
  - First **deep** interactive image segmentation work
  - Simulate user clicks with different clicks sampling strategies
  - Encode user inputs with truncated Euclidean distance maps
  - FCN-8s as the backbone architecture
  - Apply graph cut optimization for better boundary localization
Interactive Image Segmentation

- RIS-Net [Liew ICCV 17]
  - A local branch to focus on the local region around each (+ve, -ve) click pairs
  - Append PSP feature as multiscale global context
  - Click discounting factor to enforce the network to use minimal amount of user inputs for refinement

[Liew ICCV17]
Interactive Image Segmentation

- **DEXTR [Maninis CVPR18]**
  - Take 4 extreme points (leftmost, rightmost, top and bottom pixels)
  - Relax the bounding box before cropping to include context
  - Encode the extreme points with Gaussian maps
  - DeepLabv2-PSP as the backbone architecture
  - Train with balanced binary cross-entropy loss
Interactive Image Segmentation

- [Le ECCV 18]
  - Predict object boundaries using boundary clicks (control points)
  - Encode the control points with multiscale “Gaussian” \( S_{c_i}^\sigma(p) = \exp \left( \frac{-d(p,c_i)^2}{2(\sigma \cdot L)^2} \right) \)
  - Encoder-decoder with skip connections as backbone
  - Train with 3 types of losses:
    - (1) \( L_{global} \): BCE loss for predicted edge
    - (2) \( L_{local} \): Similar to (1) but focuses on local prediction around each control point
    - (3) \( L_{segment} \): BCE loss for predicted mask (auxiliary branch)
  - Predicted boundary is used to extract final boundary using a minimal path solver
Interactive Image Segmentation

- Deep GrabCut [Xu BMVC 17]

- Polygon-RNN

[Xu BMVC17, Castrejon CVPR17]
Interactive Image Segmentation

- Large-scale Interactive Object Segmentation [Arxiv19]

2.5M new instances on the OpenImages dataset
Conclusion

▪ Weakly Supervised Learning is developing rapidly!!

▪ Future Targets
  ▪ Develop better end-to-end learning methods
  ▪ Develop better annotation tools
  ▪ Use large-scale dataset for training
  ▪ Outperform fully-supervised counterparts
References


References

Weakly Supervised Learning for Real-World Computer Vision Applications & The 1st Learning from Imperfect Data (LID) Challenge

CVPR 2019 Workshop, Long Beach, CA

https://lidchallenge.github.io/

Organizer
Weakly Supervised Learning for Real-World Computer Vision Applications & The 1st Learning from Imperfect Data (LID) Challenge

CVPR 2019 Workshop, Long Beach, CA

https://lidchallenge.github.io/

Challenge

Object Segmentation on ILSVRC DET (Image-level Supervision)  
Scene Parsing on ADE20K (Point Supervision)
Thanks

Q & A

All the materials will be made available on my homepage: https://weiyc.github.io