Towards Weakly- and Semi-Supervised Object Localization and Semantic Segmentation

Lecturer: Yunchao
Wei
Image Formation and Processing (IFP) Group
University of Illinois at Urbana-Champaign
https://weiyc.github.io

BECKMAN INSTITUTE
Fully Supervised Semantic Segmentation

Fully Convolutional Networks

Hard to Collect

Masks
20K+ Categories → 100/C → 2M+ Instances → 5min/l → 19 years / person
Weakly Supervised Annotations

The simplest and the most efficient one
Our Targets

- Images
- Annotations
  - Person
  - Horse
  - Table

Object Localization Maps

Loc

Seg

- Person
- Table

BECKMAN INSTITUTE
Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Multi-dilated Convolution

Transferable Semi-supervised Network
Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Multi-dilated Convolution

Transferable Semi-supervised Network
Proposal-based Localization

Learning to Segment with Image-level Annotations. PR 2016
Proposal-based Localization

Hypotheses-CNN-Pooling

Localization Map Generation

**Weakness**
- Exhaustedly examine each proposal to generate localization
- Introducing false negative pixels (background)

Learning to Segment with Image-level Annotations. PR 2016
Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Multi-dilated Convolution

Transferable Semi-supervised Network
Simple to Complex

Simple Images

Complex Images

STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation. PAMI 2017
Simple to Complex

Simple images with the corresponding saliency maps

STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation. PAMI 2017
Simple to Complex

- **Initial-DCNN**
  \[-\frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} (\hat{p}_{ij}^c \log(p_{ij}^c) + \hat{p}_{ij}^0 \log(p_{ij}^0))\]

- **Enhanced-DCNN**

- **Powerful-DCNN**

STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation. PAMI 2017
Simple to Complex

Flickr-Clean (40K)

plane  bike  bird  boat  bottle  bus  car  cat  chair  cow

plane  bike  bird  boat  bottle  bus  car  cat  chair  cow

table  dog  horse  motor  person  plant  sheep  sofa  train  tv

STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation. PAMI 2017
## Simple to Complex

Ablation Analysis on Pascal VOC12 val

<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>

**Weakness**

- Collecting a large number of simple images
- Time consuming for training

STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation. PAMI 2017
Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Transferable Semi-supervised Network

Multi-dilated Convolution

Multi-dilated Convolution

Transferable Semi-supervised Network

Adversarial Complementary Learning
Adversarial Erasing

Previous works

Small and sparse object localization maps

Our target

Dense and integral object localization maps

Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017 (oral)
Adversarial Erasing

Motivation

Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017 (oral)
Adversarial Erasing

Our Solution

Visualizations

Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017 (oral)
Adversarial Erasing

The pipeline of weakly semantic segmentation based on AE

Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017 (oral)
Weakness of Adversarial Erasing

- Time consuming to learn several classification networks.
- Hard to determine how many AE steps should be conducted.

Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach. CVPR 2017 (oral)
Achievements
Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018

Revisiting CAM

Adversarial Complementary Learning
Adversarial Complementary Learning

Detailed Framework

Erasing
Thresholding

Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018
Adversarial Complementary Learning

Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018
Adversarial Complementary Learning

Localization Comparison

Image

CAM

Ours

(a) CUB-200-2011

(b) ILSVRC

Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018
Adversarial Complementary Learning

Localization error on ILSVRC validation set

<table>
<thead>
<tr>
<th>Methods</th>
<th>Top-1 err.</th>
<th>Top-5 err.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Backprop on GoogleLeNet</td>
<td>61.31</td>
<td>50.55</td>
</tr>
<tr>
<td>GoogLeNet-GAP (CVPR 2016)</td>
<td>56.40</td>
<td>43.00</td>
</tr>
<tr>
<td>GoogLeNet-HaS-32 (ICCV 2017)</td>
<td>54.53</td>
<td>-</td>
</tr>
<tr>
<td>GoogLeNet-ACoL (Ours)</td>
<td>53.28</td>
<td>42.58</td>
</tr>
<tr>
<td>GoogLeNet-ACoL*(Ours)</td>
<td>53.28</td>
<td>35.22</td>
</tr>
<tr>
<td>Backprop on VGGnet</td>
<td>61.12</td>
<td>51.46</td>
</tr>
<tr>
<td>VGG-GAP (CVPR 2016)</td>
<td>57.20</td>
<td>45.14</td>
</tr>
<tr>
<td>VGGnet-ACoL (Ours)</td>
<td>54.17</td>
<td>40.57</td>
</tr>
<tr>
<td>VGGnet-ACoL*(Ours)</td>
<td>54.17</td>
<td>36.66</td>
</tr>
</tbody>
</table>

Adversarial Complementary Learning for Weakly Supervised Object Localization. CVPR 2018
Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Multi-dilated Convolution

Transferable Semi-supervised Network
Multi-dilated Convolution

Motivation

Multi-dilated Convolution

- Multi-dilated Convolutional Network for Object Localization

Multi-dilated Convolution

Multi-dilated Convolution

- Weakly- and Semi-Supervised Semantic Segmentation

Multi-dilated Convolution

Achievements

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Adversarial Complementary Learning

Multi-dilated Convolution

Transferable Semi-supervised Network
Transferable Semi-supervised Network

In-category Semi-supervised Semantic Segmentation (I3S)

Image-level labels + In-category pixel-level labels → Testing image

Transferable Semi-supervised Semantic Segmentation. AAAI 2018
20K+ Categories → 100/C → 2M+ Instances → 5min/I → 19 years / person
Transferable Semi-supervised Network

In-category Semi-supervised Semantic Segmentation (I3S)

Cross-category Semi-supervised Semantic Segmentation (C3S)

Transferable Semi-supervised Semantic Segmentation. AAAI 2018
Transferable Semi-supervised Network

Label Transfer Network (L-Net)

Prediction Transfer Network (P-Net)
Transferable Semi-supervised Network

\[
\min_{\theta_L} \sum_{C_s} \mathcal{J}_b (L'_s, O_L(C_s; \theta_L))
\]

\(\mathcal{J}_b\) denotes the standard element-wise binary cross-entropy loss
Transferable Semi-supervised Network

Classification Activation Map

Strong categories $C_s$

Weak categories $C_w$

L-Net

$M$

Classification

$A^{bird}$, $A^{dog}$

Transferable Semi-supervised Semantic Segmentation. AAAI 2018
Transferable Semi-supervised Network

Random Walk based self-diffusion algorithm

\[
\min_q \frac{1}{2} \sum_{i,j} z_{ij} (q_i - q_j)^2
\]
Transferable Semi-supervised Network

Images  \( M \)  \( L_w \)  AE  Ground truth

Transferable Semi-supervised Semantic Segmentation. AAAI 2018
Transferable Semi-supervised Network

\[
\min_{\theta_S} \max_{\theta_P} \sum_I J_m(\mathcal{L}_I, \mathcal{O}_S(I; \theta_S)) - \\
\lambda \left[ J_b(1, \mathcal{O}_P(\mathcal{L}_I; \theta_P)) + J_b(0, \mathcal{O}_P(\mathcal{O}_S(I; \theta_S); \theta_P)) \right]
\]
Transferable Semi-supervised Network

\[
\min_{\theta_S} \max_{\theta_P} \sum_I J_m(\mathcal{L}_I, \mathcal{O}_S(I; \theta_S)) - \\
\lambda \left[ J_b(1, \mathcal{O}_P(\mathcal{L}_I; \theta_P)) + J_b(0, \mathcal{O}_P(\mathcal{O}_S(I; \theta_S); \theta_P)) \right]
\]

<table>
<thead>
<tr>
<th>Layer</th>
<th>Channels</th>
<th>Kernel</th>
<th>Activation</th>
</tr>
</thead>
<tbody>
<tr>
<td>conv1</td>
<td>16</td>
<td>3x3</td>
<td>ReLU</td>
</tr>
<tr>
<td>pool1</td>
<td>-</td>
<td>2x2, stride 2</td>
<td>-</td>
</tr>
<tr>
<td>conv2</td>
<td>64</td>
<td>3x3</td>
<td>ReLU</td>
</tr>
<tr>
<td>pool2</td>
<td>-</td>
<td>2x2, stride 2</td>
<td>-</td>
</tr>
<tr>
<td>conv3</td>
<td>128</td>
<td>3x3</td>
<td>ReLU</td>
</tr>
<tr>
<td>pool3</td>
<td>-</td>
<td>2x2, stride 2</td>
<td>-</td>
</tr>
<tr>
<td>fc4</td>
<td>256-d</td>
<td>-</td>
<td>tanh</td>
</tr>
<tr>
<td>fc5</td>
<td>512-d</td>
<td>-</td>
<td>tanh</td>
</tr>
<tr>
<td>fc6</td>
<td>1-d</td>
<td>-</td>
<td>sigmoid</td>
</tr>
</tbody>
</table>
IMAGENET

Vehicles

Animals

Others

Transferable Semi-supervised Semantic Segmentation. AAAI 2018
Summary

Proposal-based Localization

Simple to Complex

Adversarial Erasing

Multi-dilated Convolution

Transferable Semi-supervised Network

Loc

Seg

Object Localization Maps

images

annotations

person  horse  table

person  tv  table

...
Thanks