Weakly & Interactive Image Segmentation

Yao Zhao

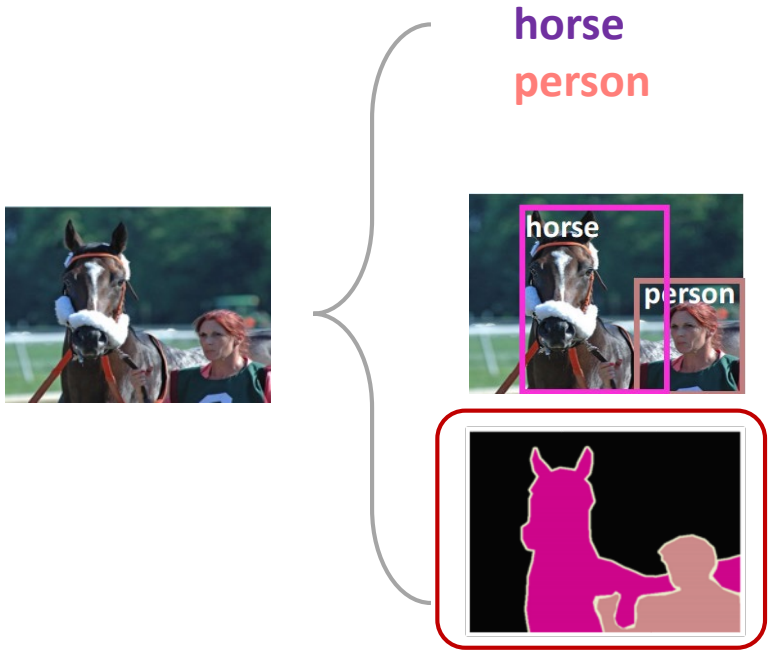
MePro, Institute of Information Science

Beijing Jiaotong University

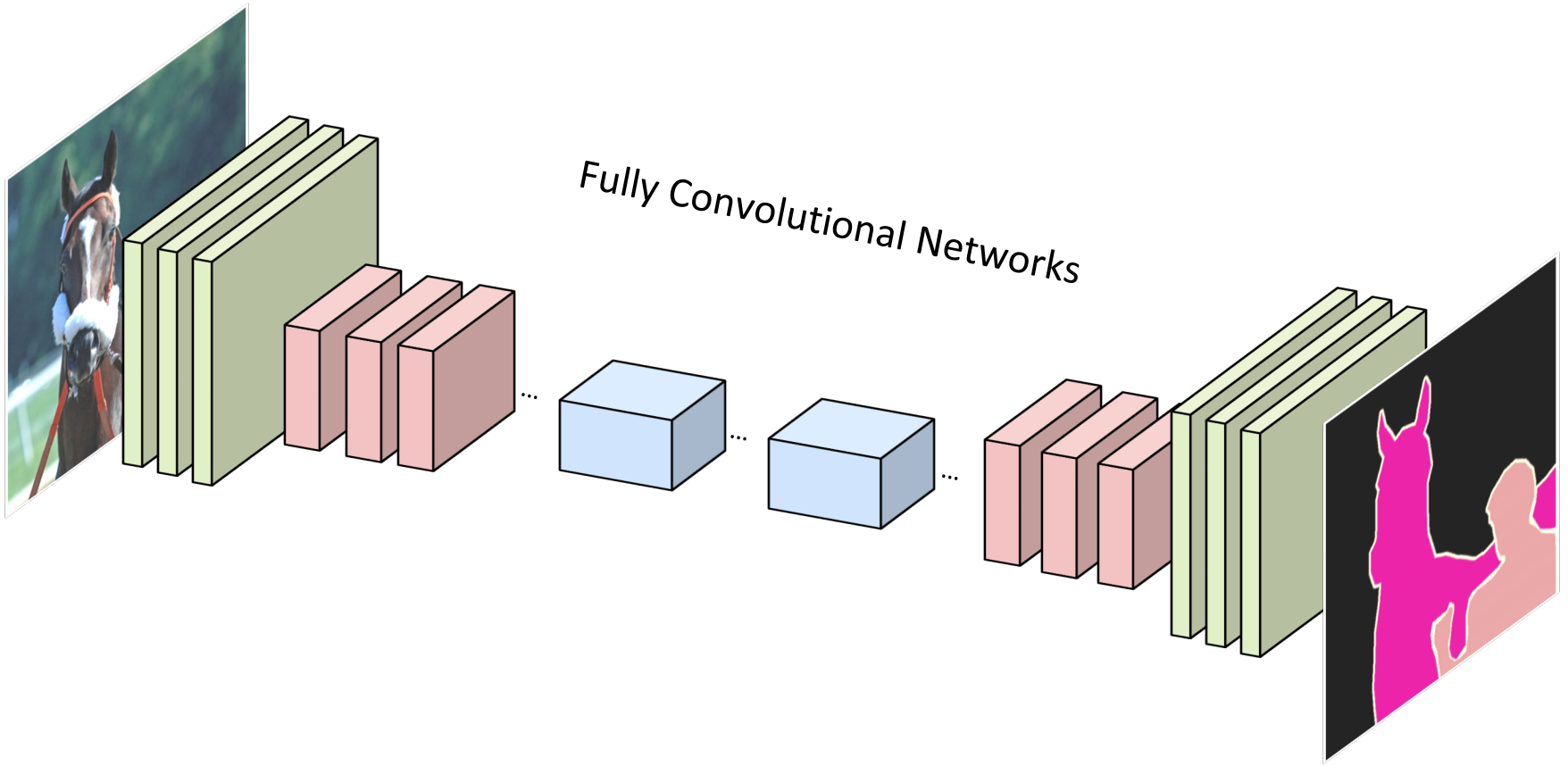
http://mepro.bjtu.edu.cn/

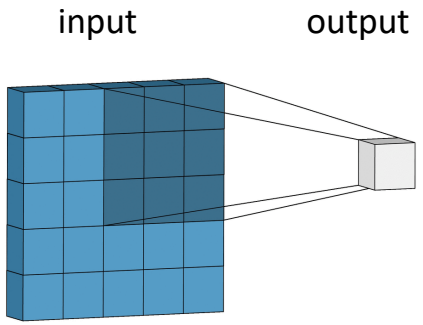
What is semantic segmentation?

- Three fundamental tasks in computer vision community
  - Image classification
  - Object detection
  - Semantic Segmentation



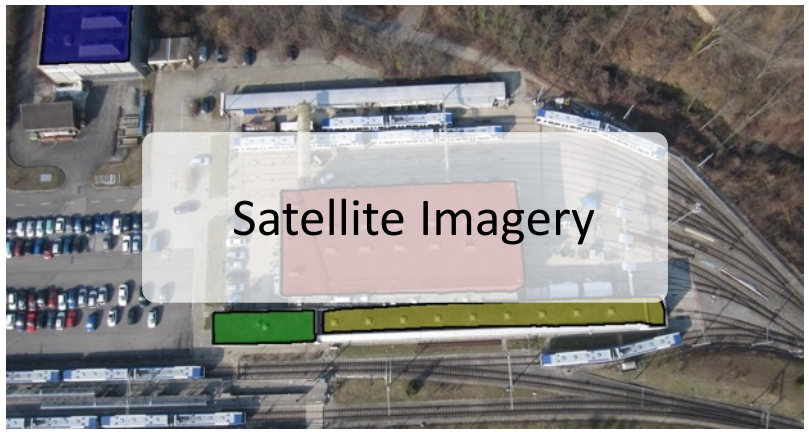
Overview: Fully-supervised Semantic Segmentation

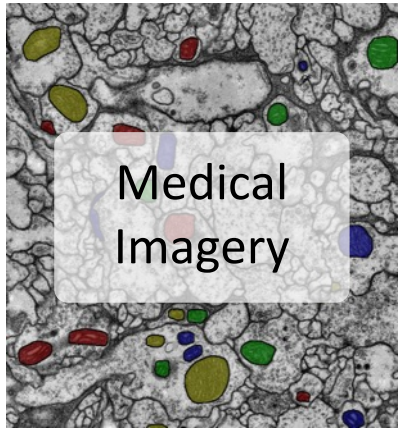


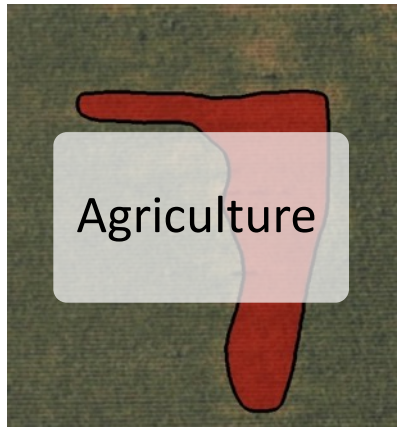


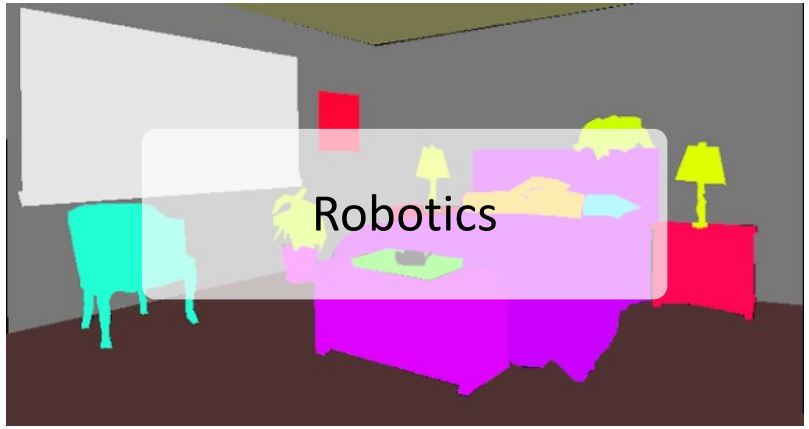
Convolutional operation

The importance of semantic segmentation









Challenge: Pixel-level annotation is time-consuming

pixel-level labels

78s/instance

Solution

Weakly supervised semantic segmentation

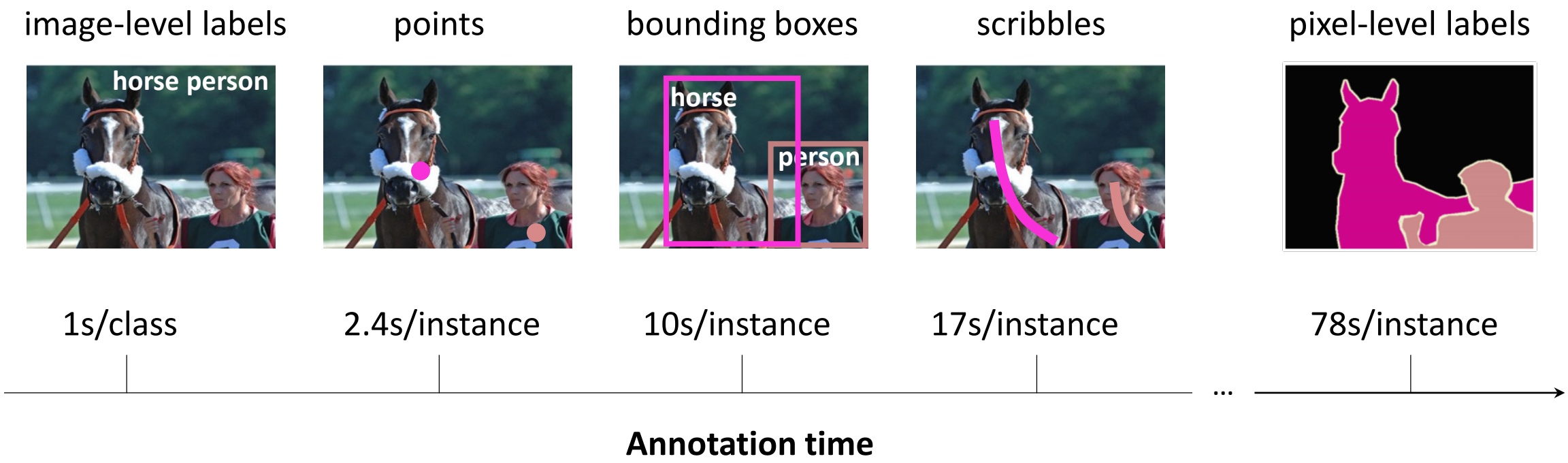
Interactive object segmentation

Solution

Weakly supervised semantic segmentation

Interactive object segmentation

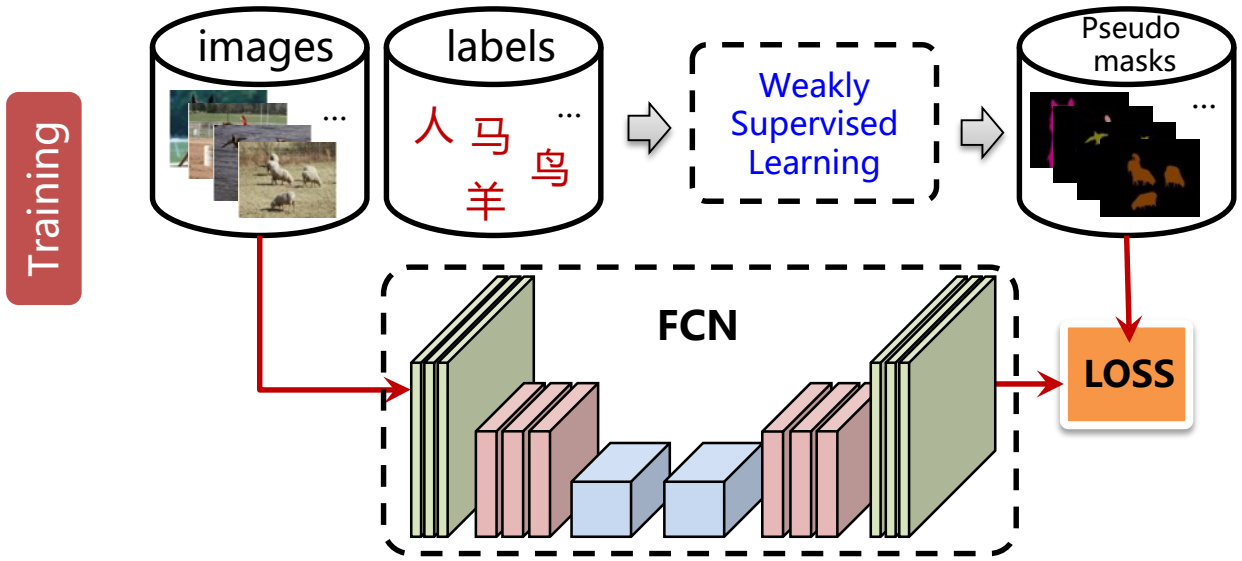
Weak annotations



[Lin CVPR16, Berman ECCV16, Hakan CVPR18 Tutorial]

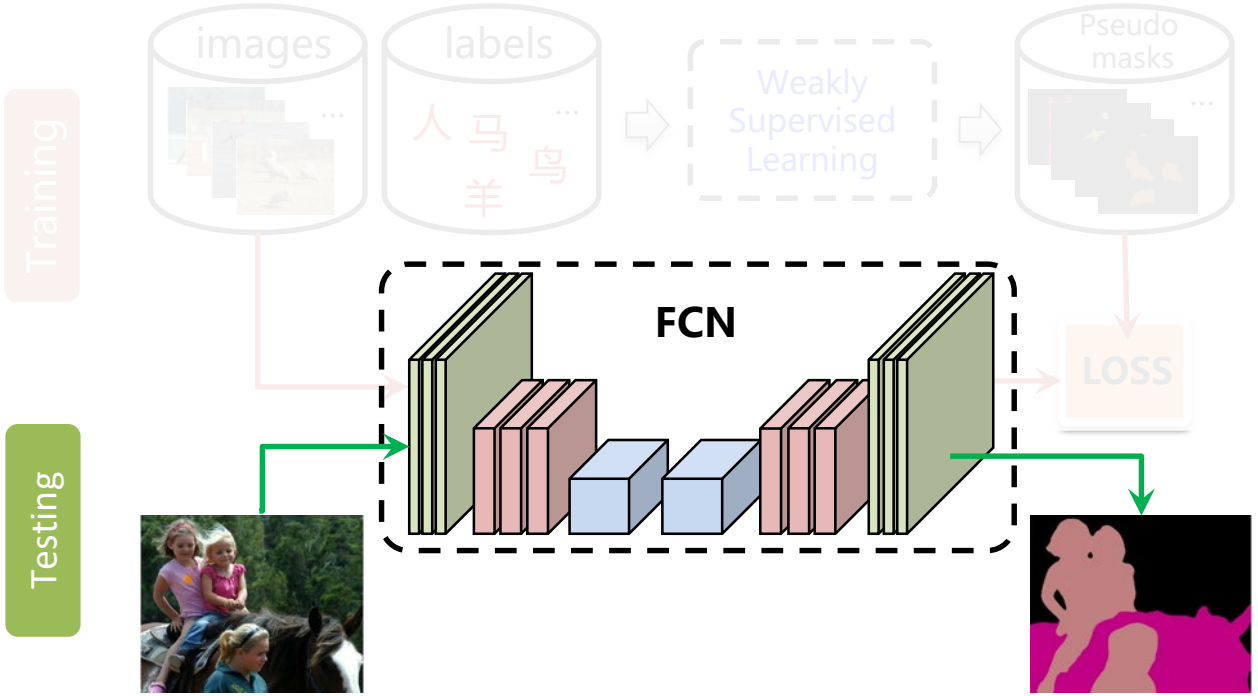
Weakly-supervised Semantic Segmentation

- Our target
  - Using image-level labels as supervision

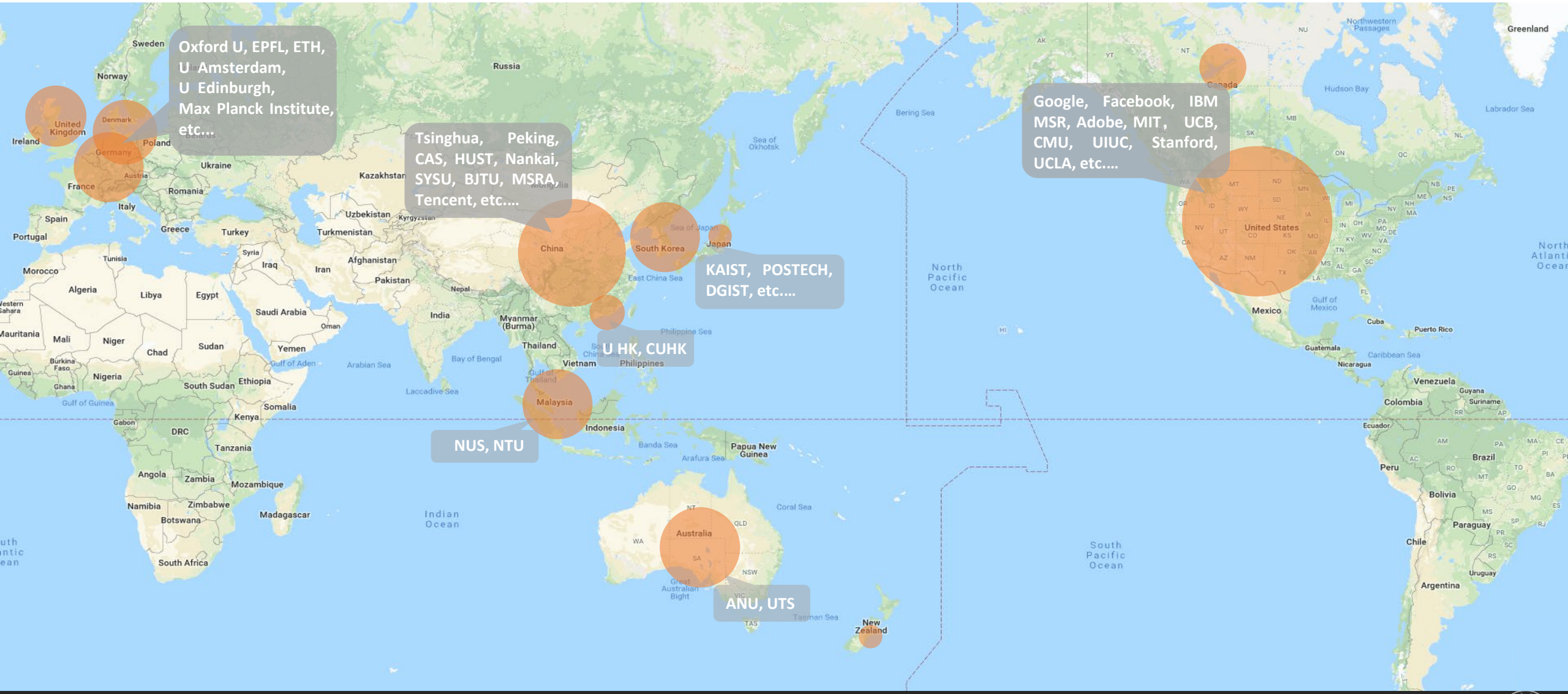


Weakly-supervised Semantic Segmentation

- Our target
  - Using image-level labels as supervision



Researcher Distribution



Method-1: Simple to complex

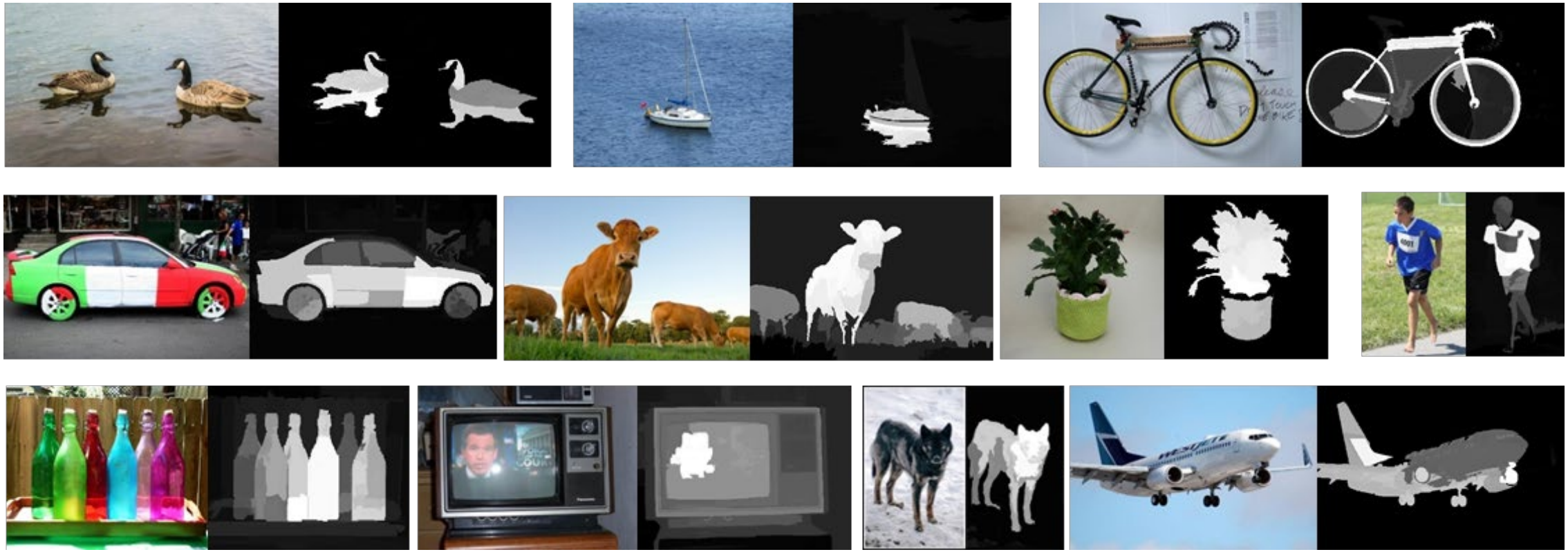
Simple Images

Complex Images

Y. Wei, Y. Zhao, etc. **STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation.** T-PAMI, 2017

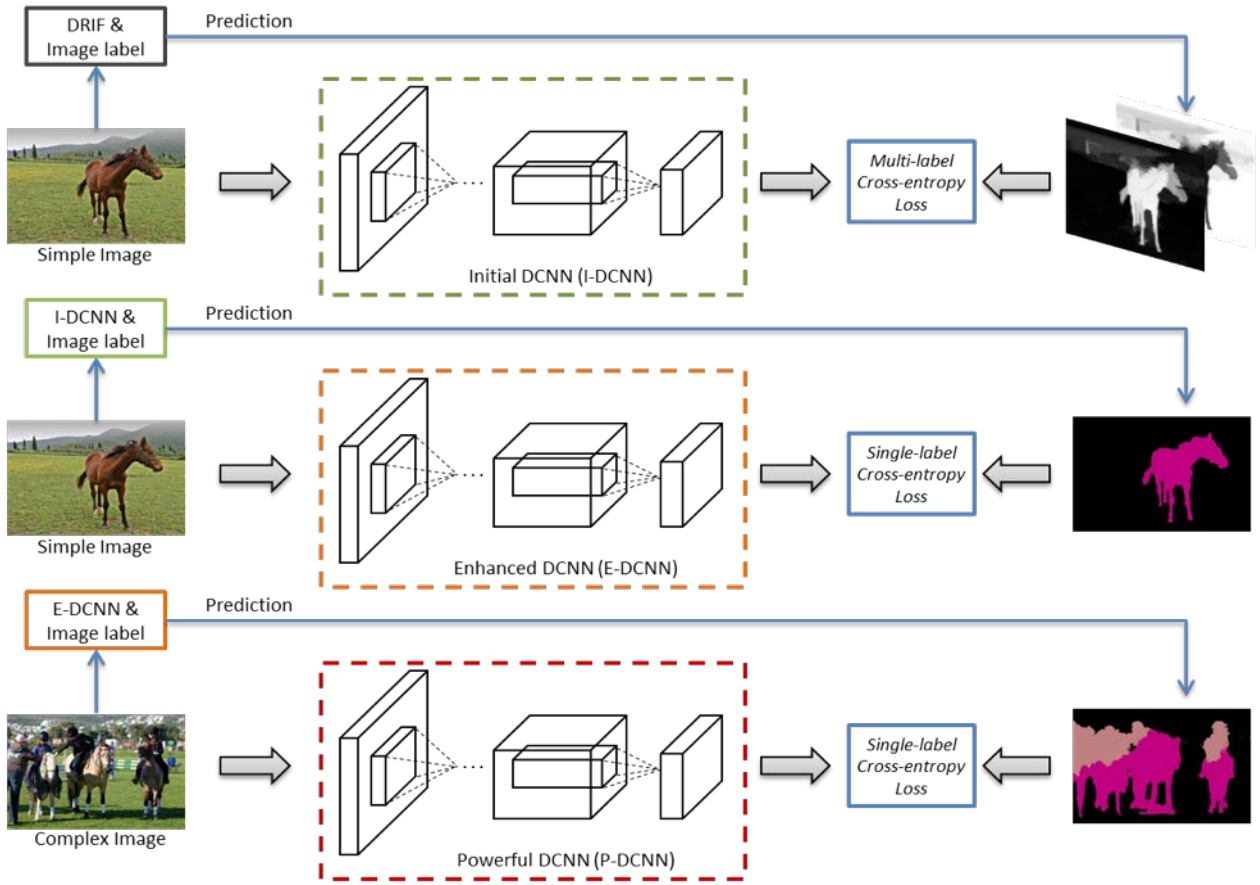
Method-1: Simple to complex

- Simple images with the corresponding saliency maps



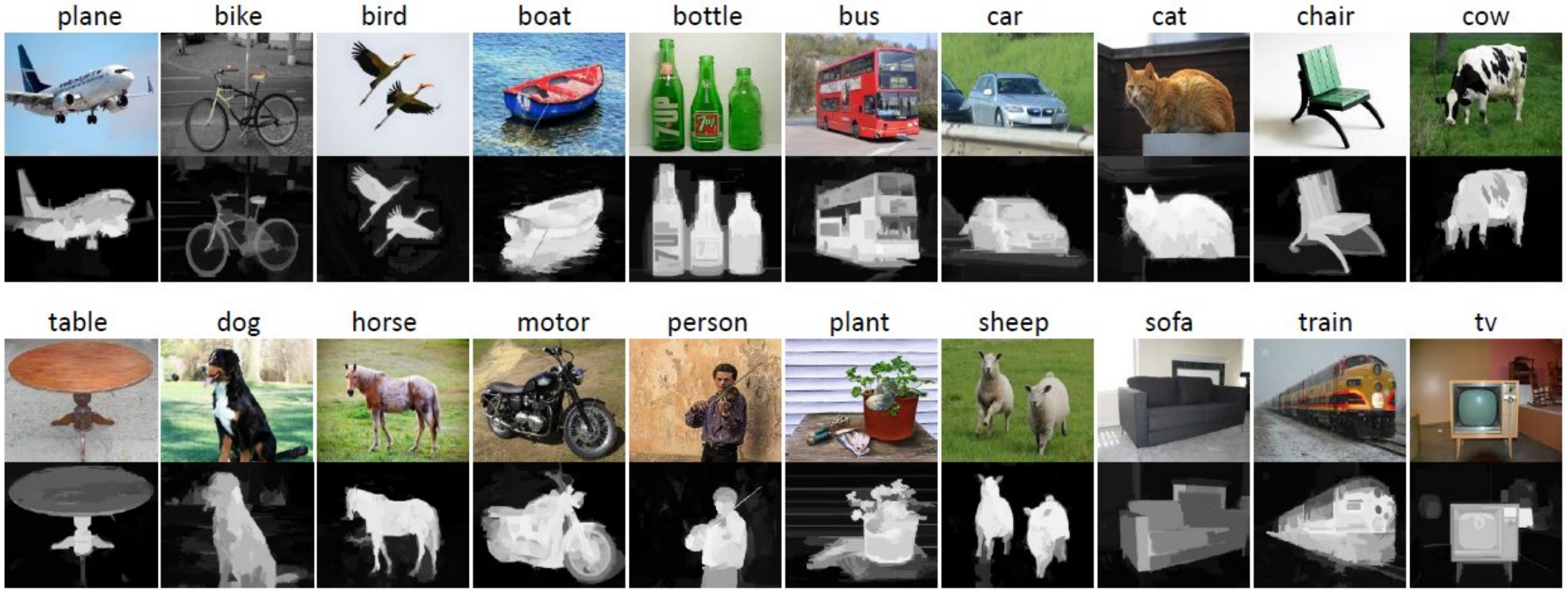
Method-1: Simple to complex

- Overview of STC
  - Initial-DCNN
  - Enhanced-DCNN
  - Powerful-DCNN



Method-1: Simple to complex

- Dataset with simple images: Flickr-Clean (40K)



Y. Wei, Y. Zhao, etc. **STC: A Simple to Complex Framework for Weakly-supervised Semantic Segmentation.** T-PAMI, 2017

Method-1: Simple to complex

- Results

Ablation Analysis on Pascal VOC12 val

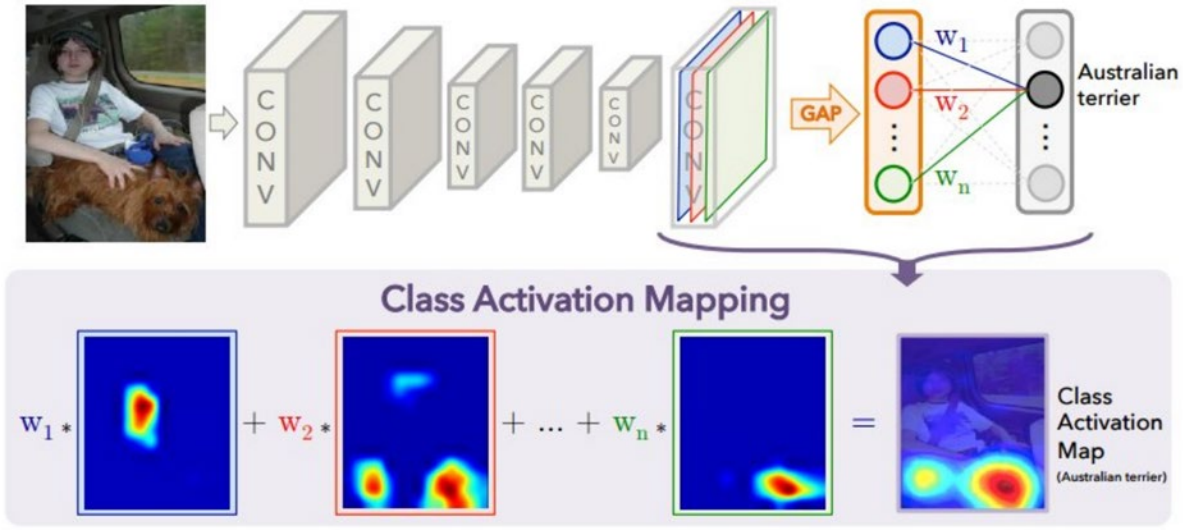
Networks	Training Set	mIoU
I-DCNN	Flickr-Clean	44.1
E-DCNN	Flickr-Clean	46.8
P-DCNN	Flickr-Clean+VOC	49.8

Comparisons on Pascal VOC12 test

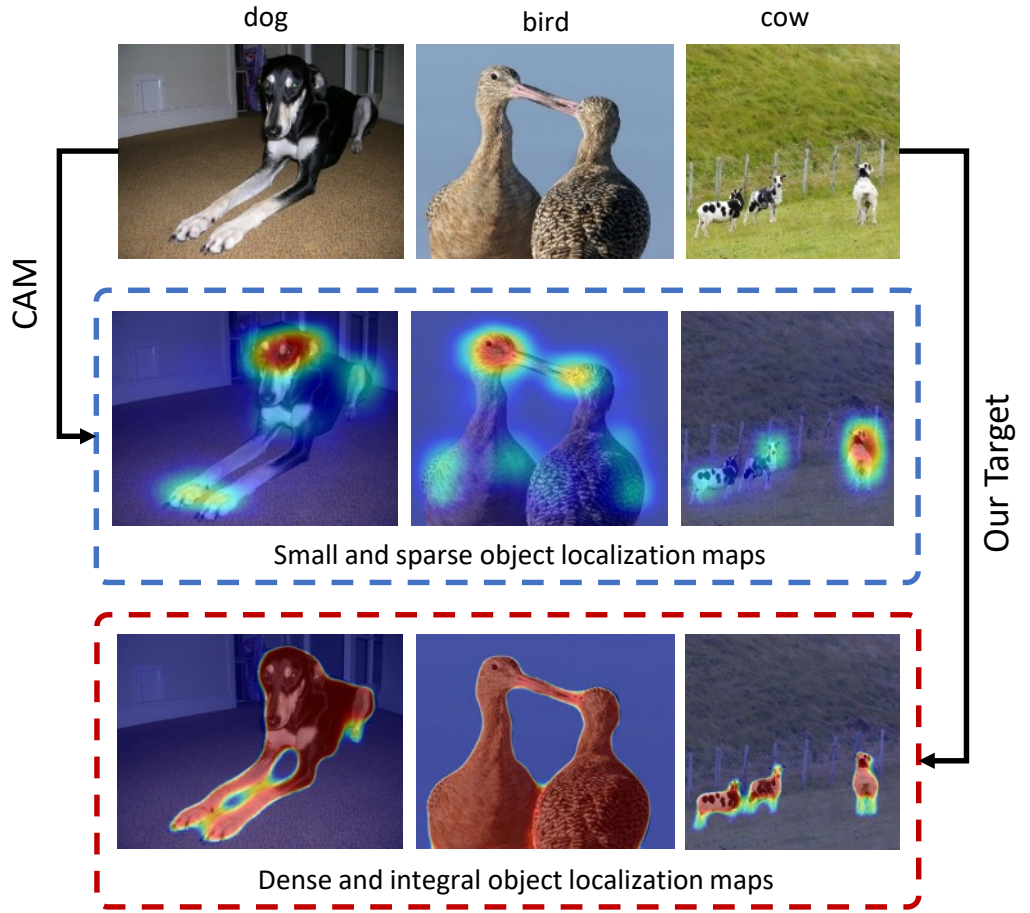
Methods	mIoU
MIL-FCN (ICLR 2015)	24.9
CCNN (ICCV 2015)	35.5
EM-Adapt (ICCV 2015)	39.6
MIL-ILP-Seg (CVPR 2015)	40.6
STC (ours)	**51.2**

Method-2: Adversarial Erasing

- Problem of Class Activation Mapping (CAM)

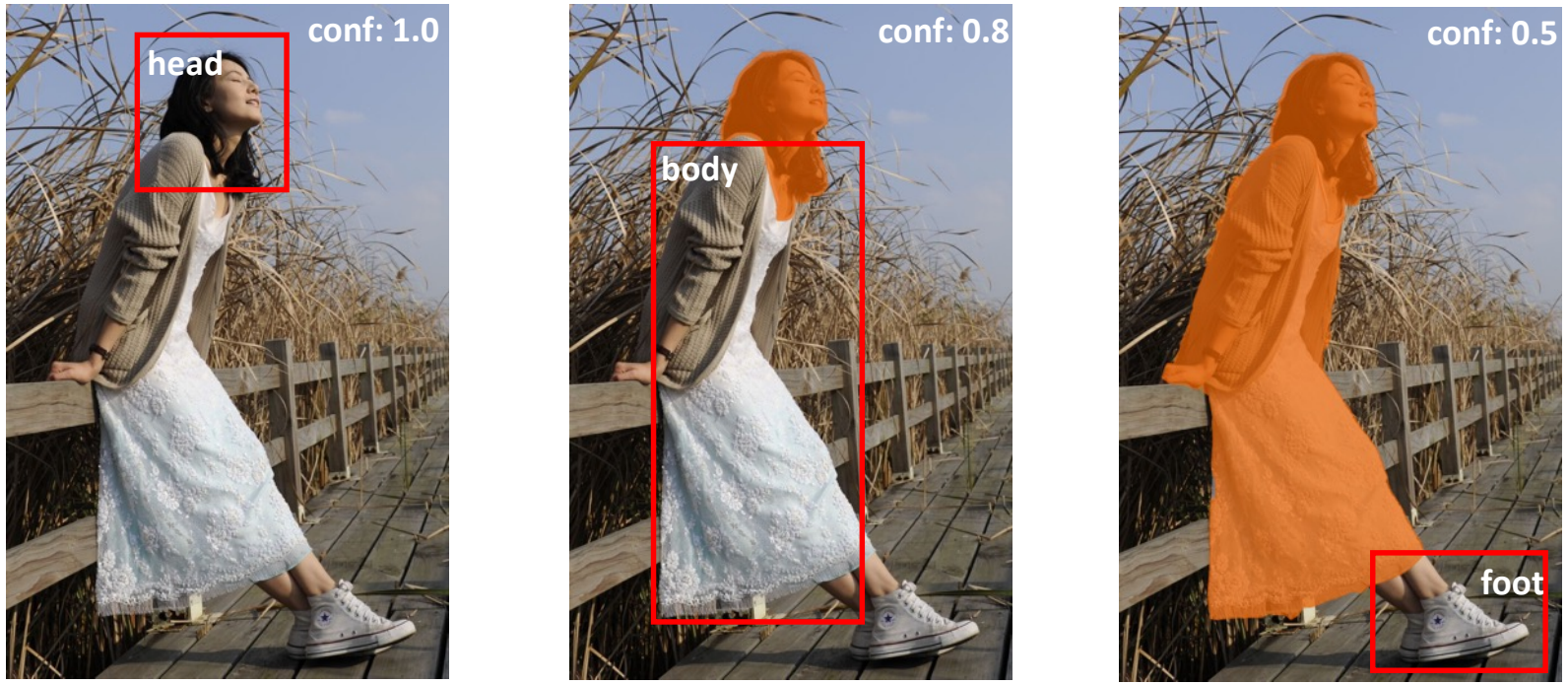


[Zhou et al. CVPR 2016]



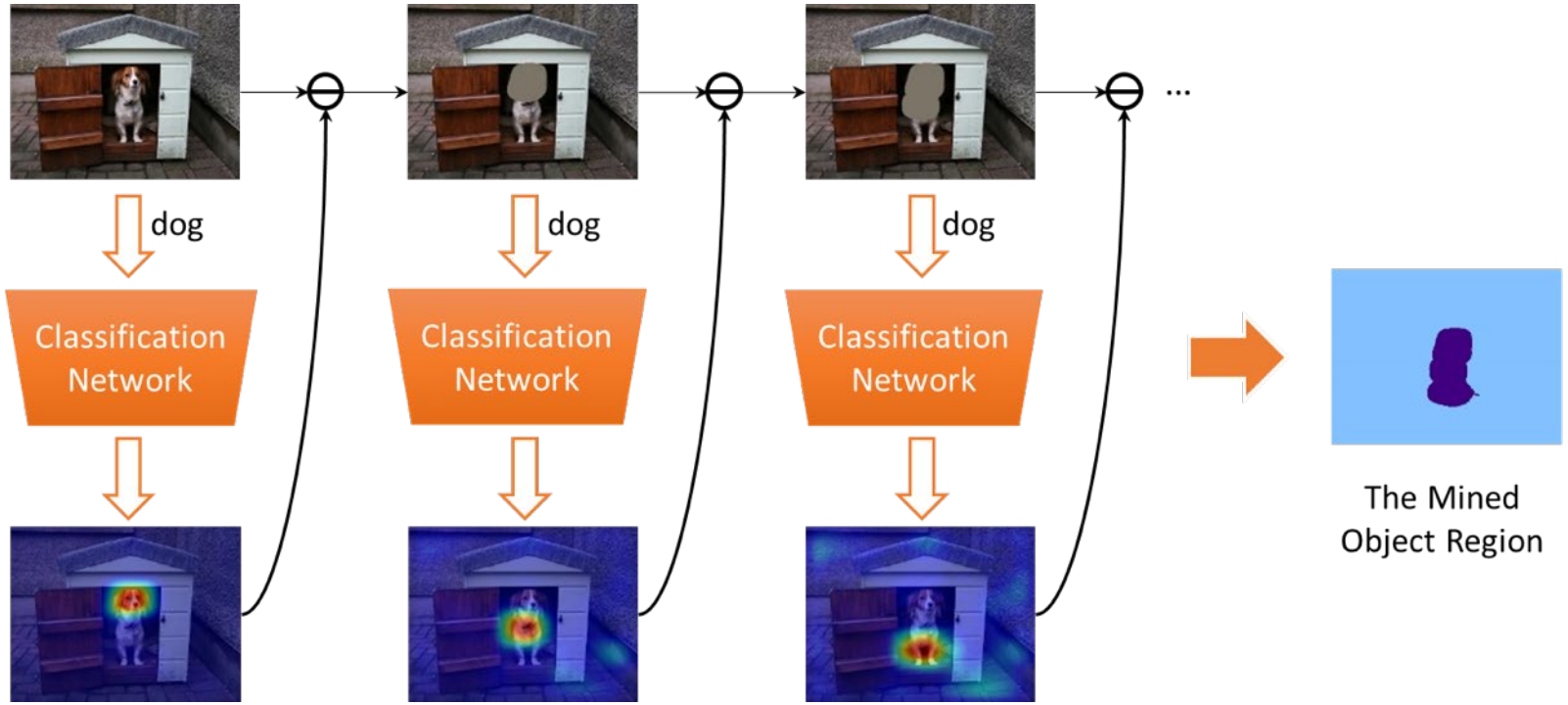
Method-2: Adversarial Erasing

- Motivation



Method-2: Adversarial Erasing

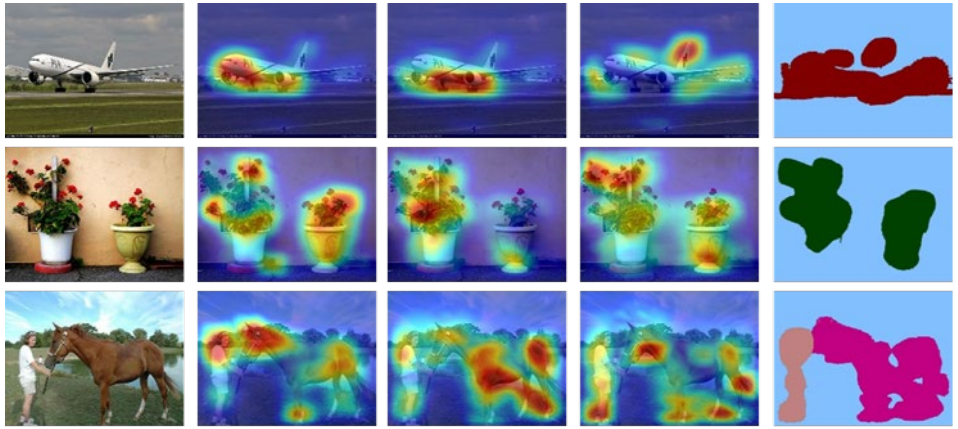
- Overview of the framework

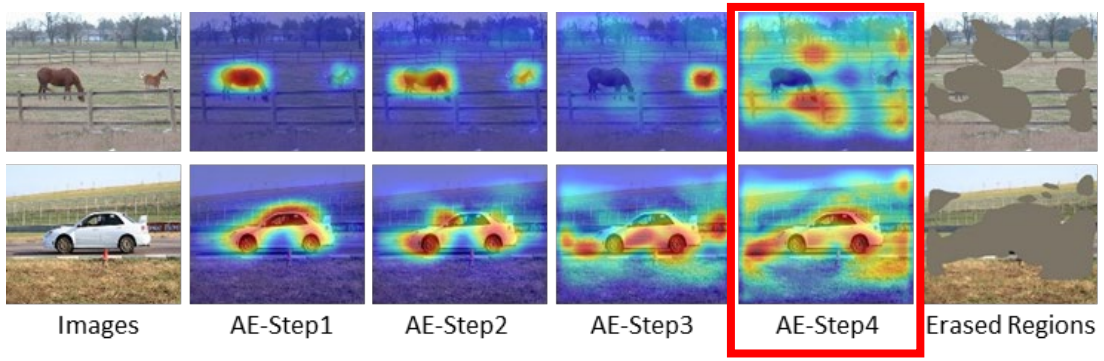


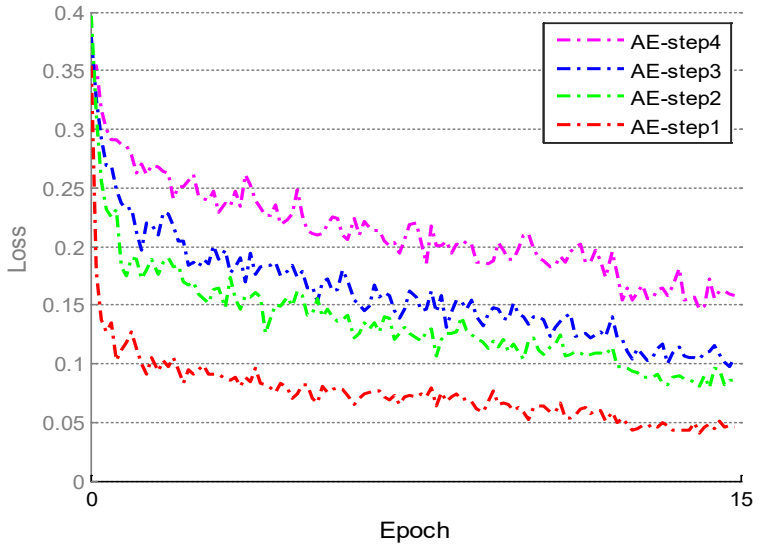
Y. Wei, Y. Zhao, etc. **Object Region Mining with Adversarial Erasing: A Simple Classification to Semantic Segmentation Approach.** CVPR 2017 (oral)

Method-2: Adversarial Erasing

- Results







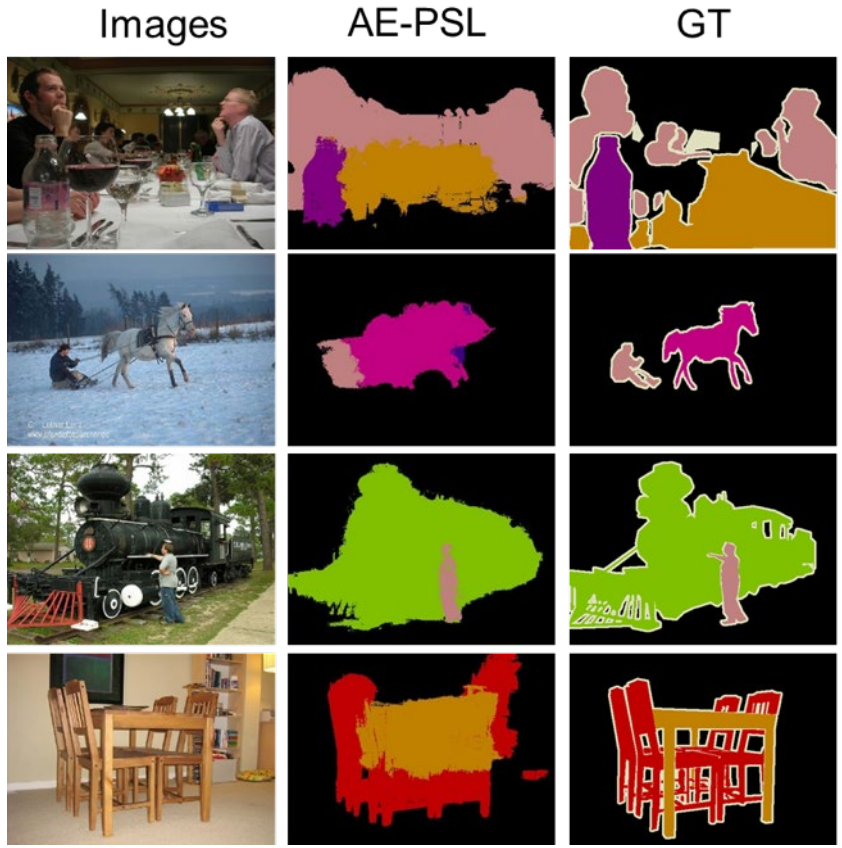
AE-Steps	mIoU
AE-step1	44.9
AE-step2	49.5
AE-step3	**50.9**
AE-step4	48.8

Method-2: Adversarial Erasing

- Results

Comparisons on Pascal VOC12 test

Methods	mIoU
MIL-FCN (ICLR 2015)	24.9
CCNN (ICCV 2015)	35.5
EM-Adapt (ICCV 2015)	39.6
MIL-ILP-Seg (CVPR 2015)	40.6
STC (PAMI 2016)	51.2
DCSM (ECCV 2016)	45.1
BFBP (ECCV 2016)	48.0
SEC (ECCV 2016)	51.7
AF-SS(ECCV 2016)	52.7
AE-PSL (ours)	**55.7**



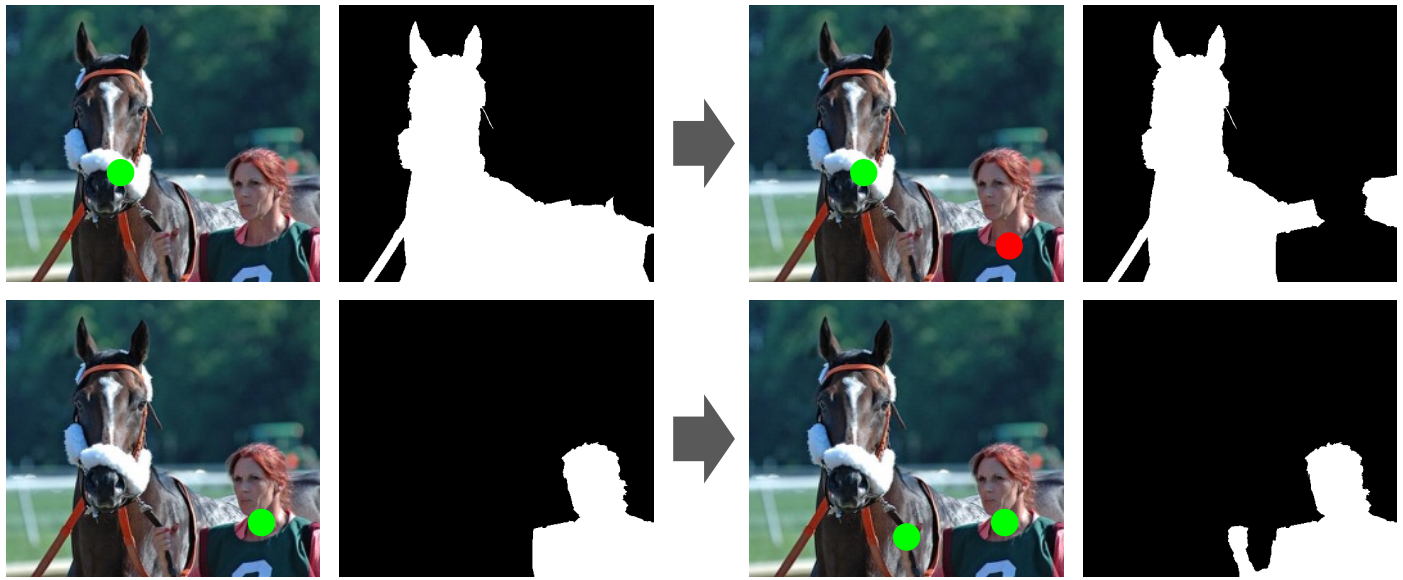
Solution

Weakly supervised semantic segmentation

Interactive object segmentation

What is Interactive Object Segmentation?

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

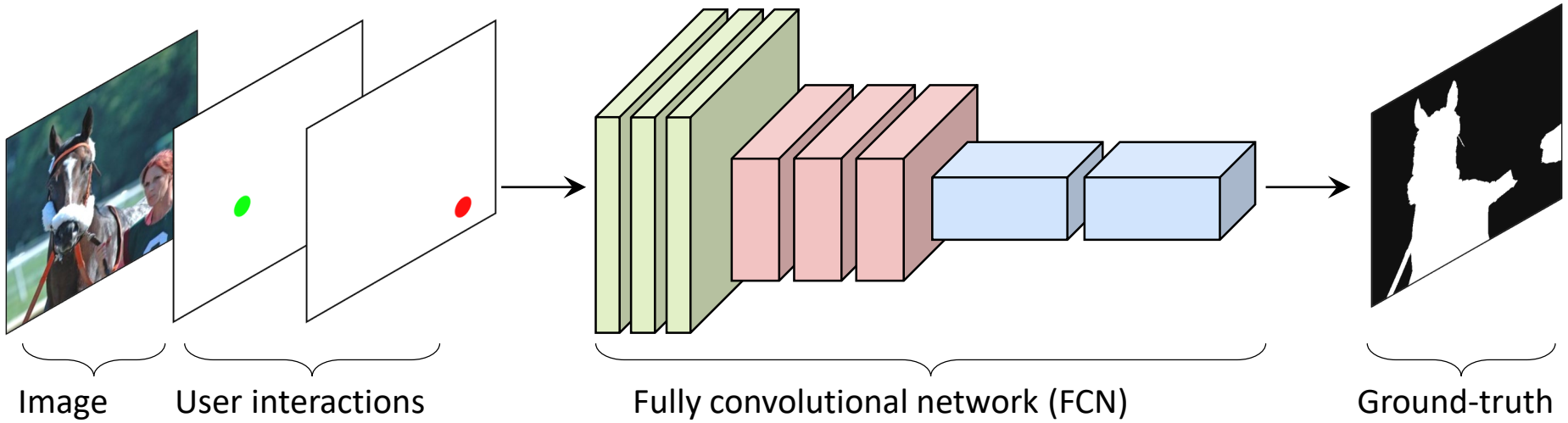


Target object

Unrelated region

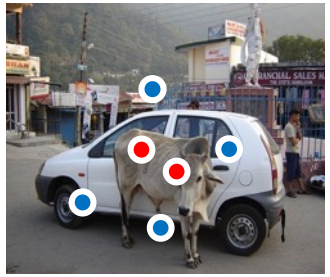
Standard pipeline

- RGB image and user interactions are used as the network input
- Train end-to-end with FCNs (e.g. Deeplab series, PSPNet)



Common types of user interaction

- Sparse clicks



- Bounding box

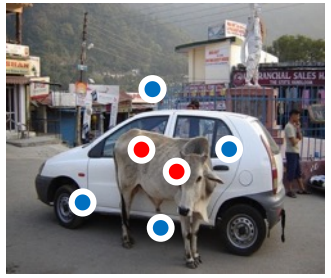
- Scribbles

Common types of user interaction

- Sparse clicks
- Bounding box
- Scribbles

Common types of user interaction

- Sparse clicks



≈ 2s per instance

- Bounding box

≈ 7s per instance

- Scribbles

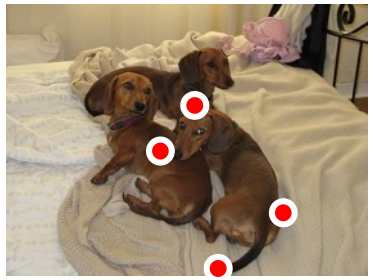
≈ 17s per instance

Manual annotation

≈ 60s per instance

Existing State-of-the-Art Method: DEXTR

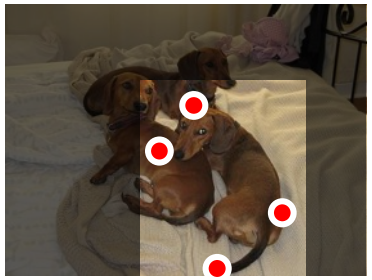
- DEXTR (Deep Extreme Cut)
  - Take 4 extreme points (top, bottom, leftmost and rightmost pixels) as inputs

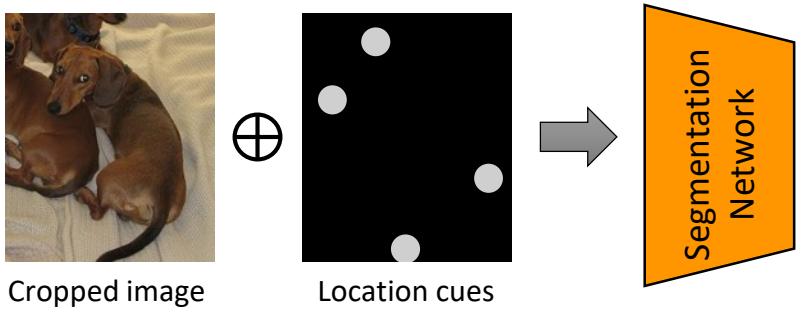


Maninis et al. “Deep Extreme Cut: From Extreme Points to Object Segmentation”, CVPR 2018.

Existing State-of-the-Art Method: DEXTR

- DEXTR (Deep Extreme Cut)
  - Take 4 extreme points (top, bottom, leftmost and rightmost pixels) as inputs

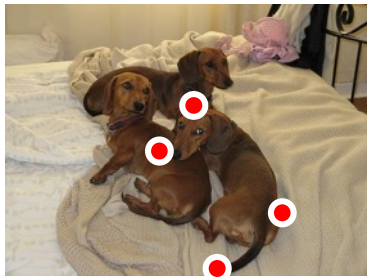




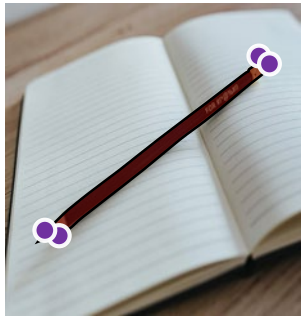
Maninis et al. “Deep Extreme Cut: From Extreme Points to Object Segmentation”, CVPR 2018.

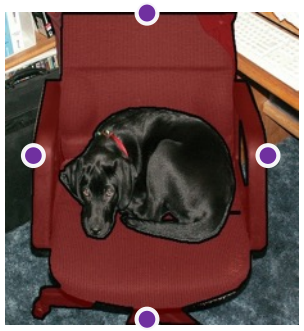
Existing State-of-the-Art Method: DEXTR

- DEXTR (Deep Extreme Cut)
  - Take 4 extreme points (top, bottom, leftmost and rightmost pixels) as inputs



- Problems
  - **Multiple extreme points appear at similar location**
  - **Unrelated object lying inside the target object**

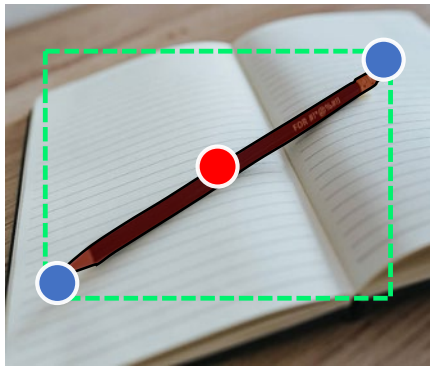




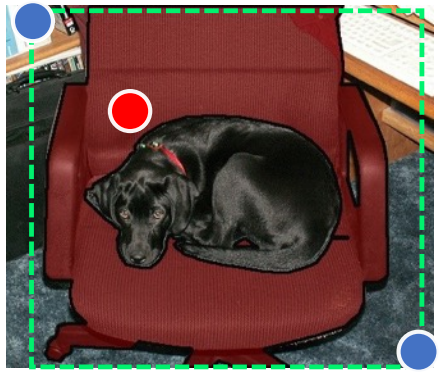
Maninis et al. “Deep Extreme Cut: From Extreme Points to Object Segmentation”, CVPR 2018.

Inside-Outside Guidance (IOG)

- Inside guidance (1 click)
  - Interior point located roughly at the object center
  - Disambiguate the segmentation target



- Outside guidance (2 clicks)
  - 2 corner clicks of a box enclosing the object
  - Indicate the background region
  - The remaining 2 corners can be inferred automatically



S. Zhang, J. Liew, Y. Wei, S. Wei, Y. Zhao. **Inside-Outside Guidance for Interactive Object Segmentation.** CVPR, 2019 (oral)

Clicking Paradigm

- Click on a corner point
- Click on the symmetrical corner
- Click on the object center

The vertical and horizontal guided lines are used to make the box visible

Clicking Paradigm

- Click on a corner point
- Click on the symmetrical corner
- Click on the object center

The vertical and horizontal guided lines are used to make the box visible

Clicking Paradigm

- Click on a corner point
- Click on the symmetrical corner
- Click on the object center

Clicking Paradigm

- Click on a corner point
- Click on the symmetrical corner
- Click on the object center

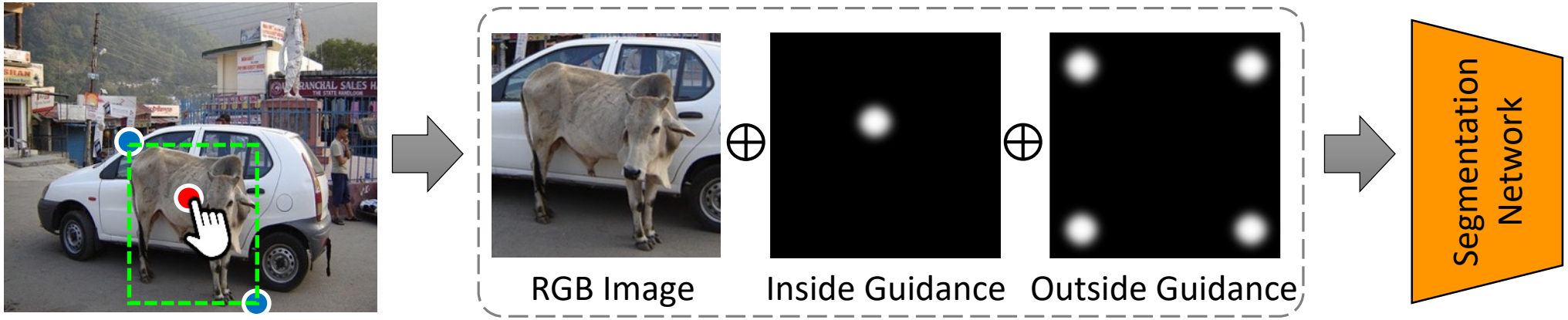
Clicking Paradigm

- Click on a corner point
- Click on the symmetrical corner
- Click on the object center

Clicks	Time
Outside clicks	6.7s
Inside click	1.5s

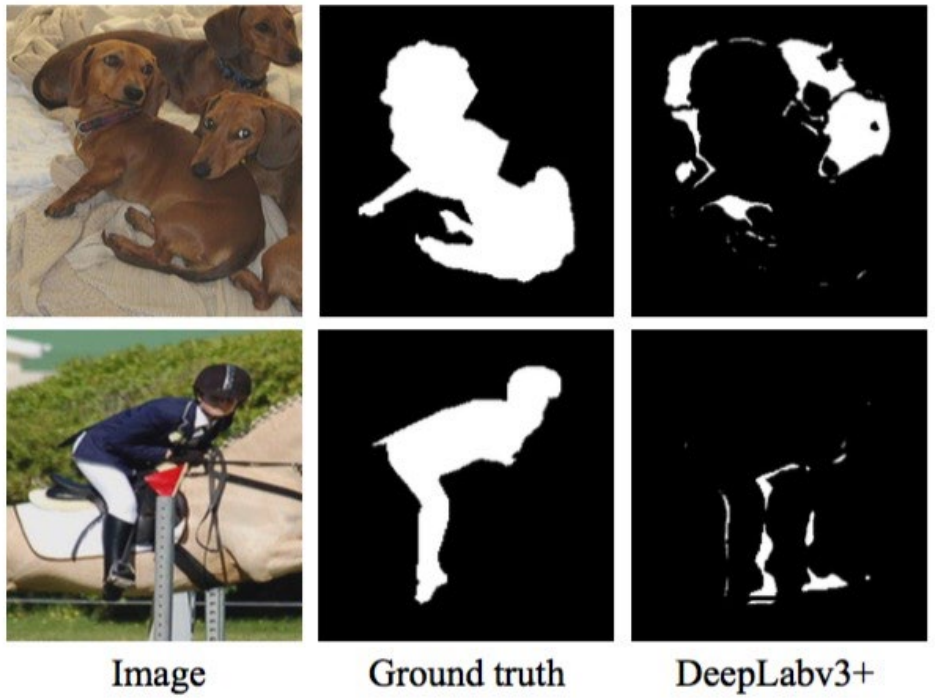
Input Representation

- Follow the practice of DEXTR
  - Enlarge the bounding box by 10 pixels to include context
  - Crop and resize the inputs to 512x512
- Input representation
  - 2 separate Gaussian heatmaps for the inside and outside clicks



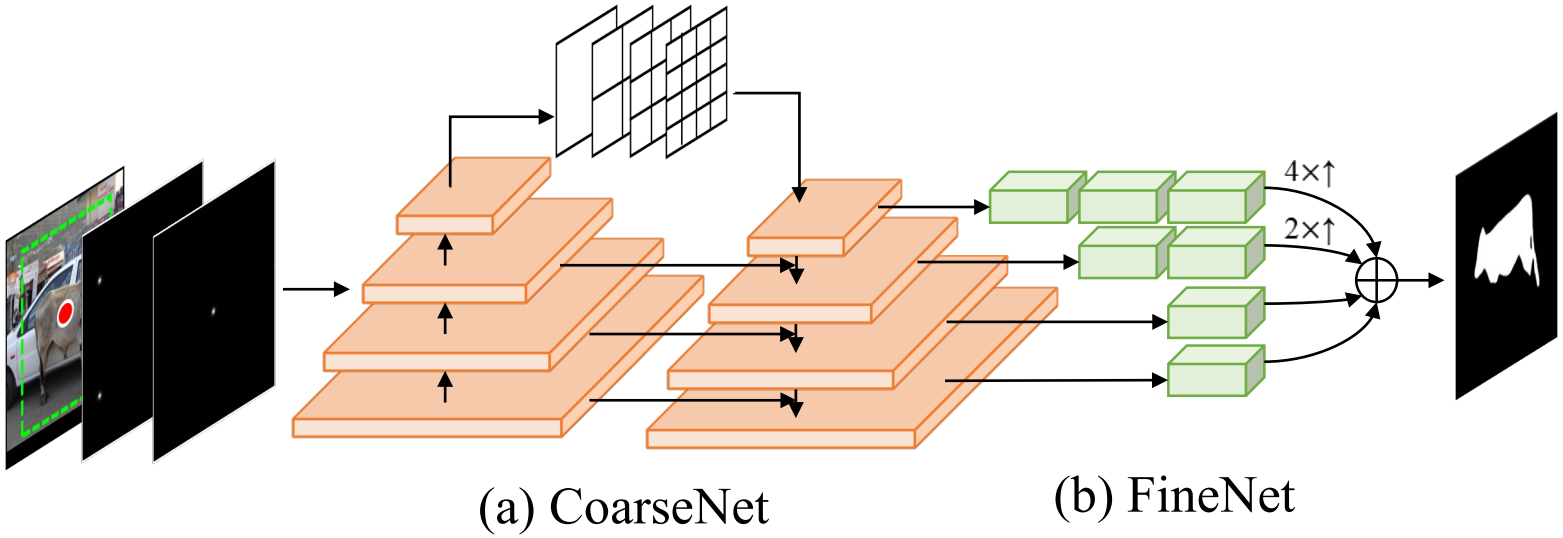
Network Architecture

- Segmentation errors mostly occur around the object boundaries



Network Architecture

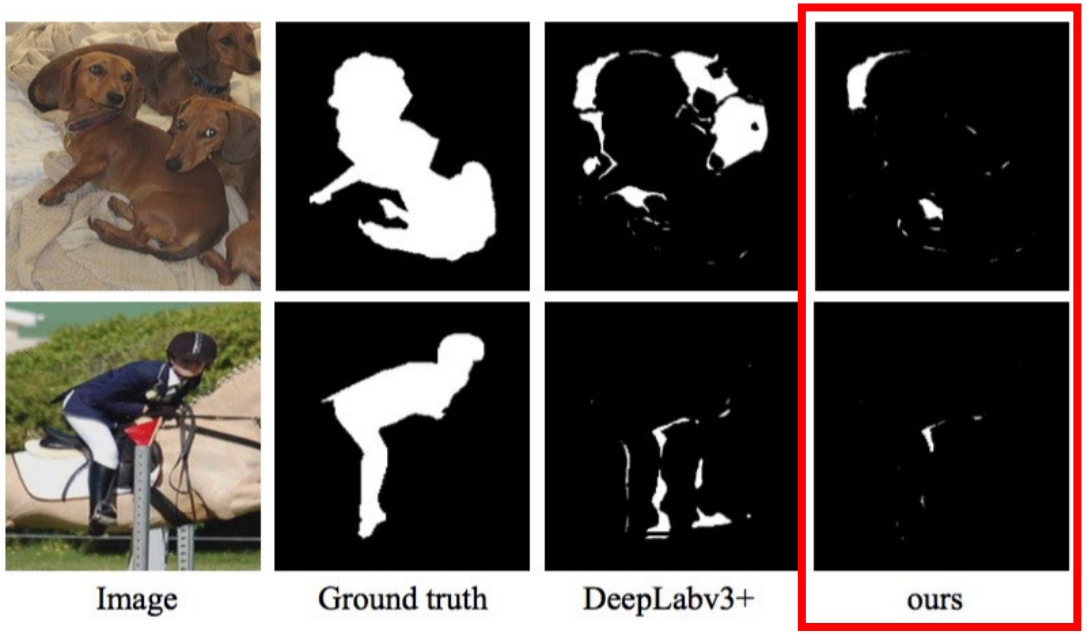
- Segmentation errors mostly occur around the object boundaries
- Use a coarse-to-fine network structure



The coarse-to-fine structure is similar to : Yilun Chen et al. “Cascaded pyramid network for multi-person pose estimation”, CVPR 2018.

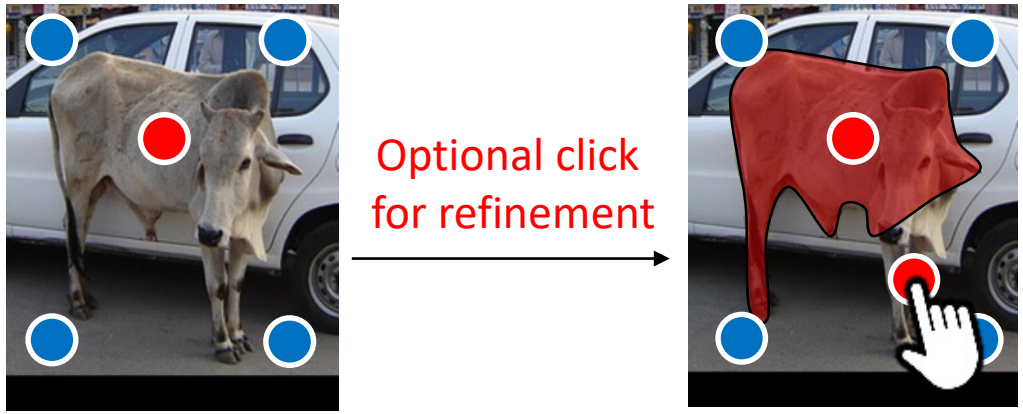
Network Architecture

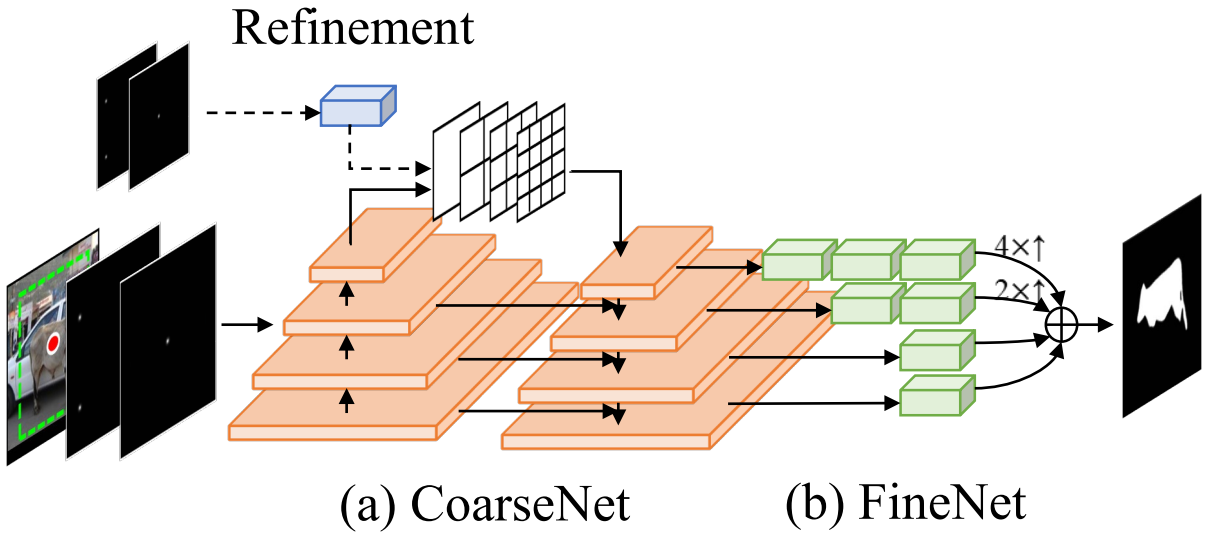
- Segmentation errors mostly occur around the object boundaries
- Use a coarse-to-fine network structure



Beyond Three Clicks

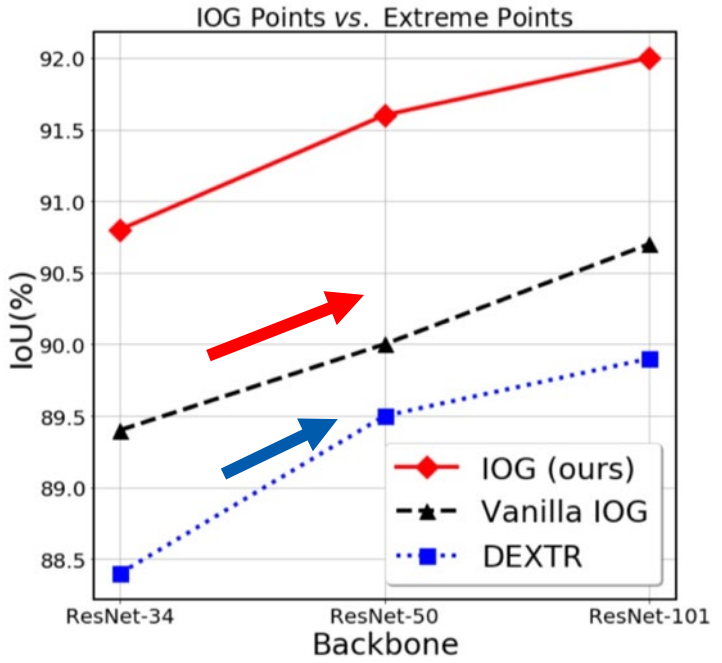
- Our IOG naturally supports interactive adding of new clicks
- Add a lightweight branch to accept additional inputs
- Train with iterative training strategy





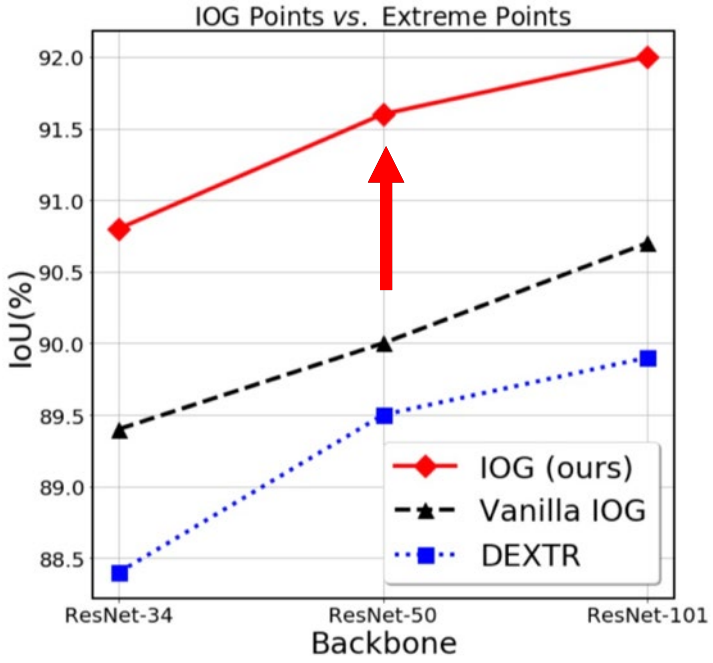
IOG vs. Extreme Clicks

- Observation
  - IOG is more effective than extreme points across different backbone

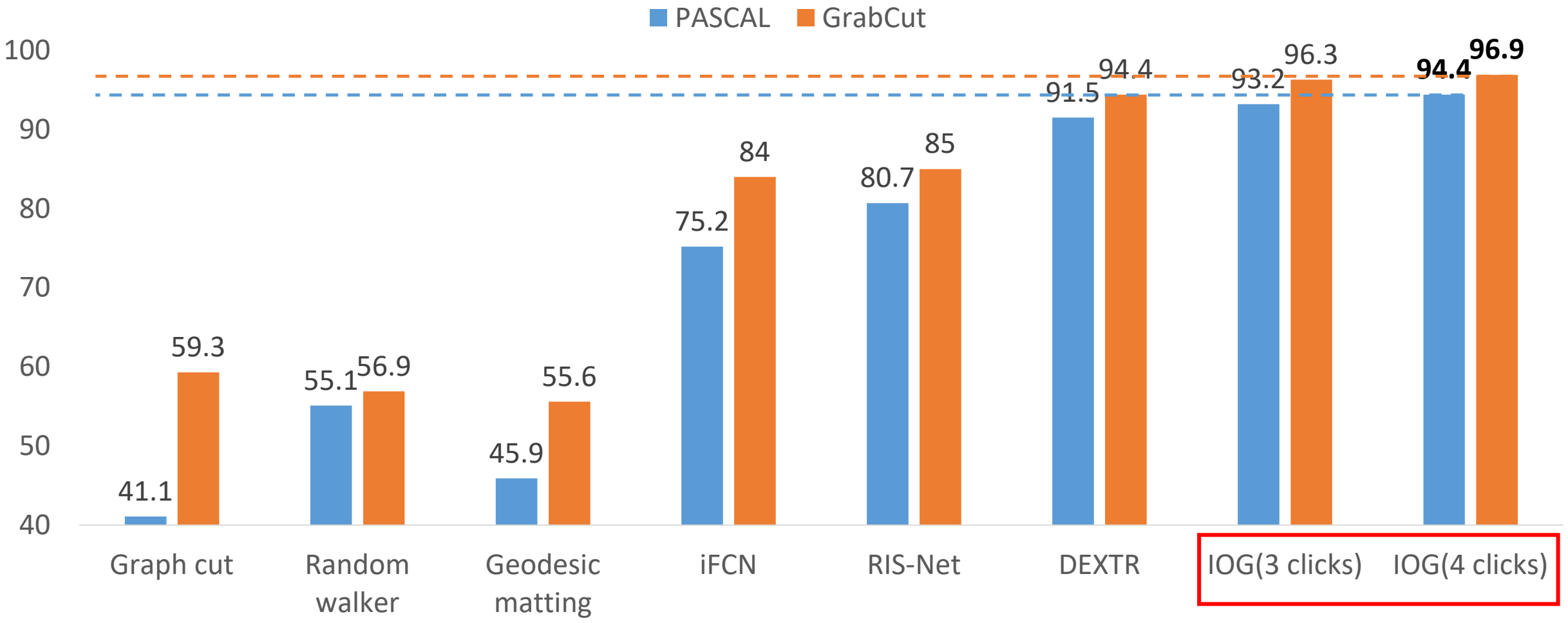


IOG vs. Extreme Clicks

- Observation
  - IOG is more effective than extreme points across different backbone
  - Using a coarse-to-fine network structure further improves the performance

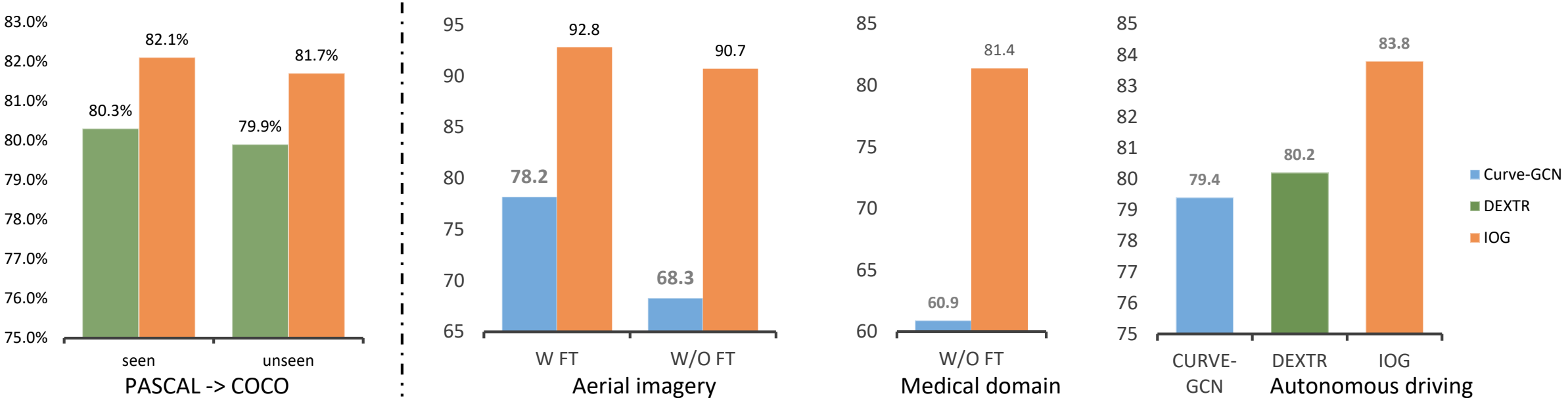


Comparison with SOTA



Generalization

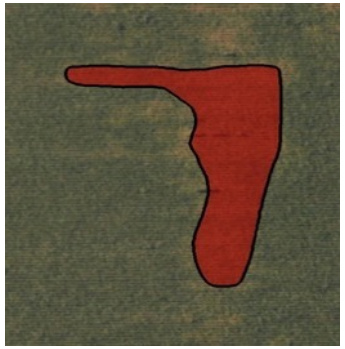
- Our IOG performs well even on ***unseen*** categories
- Performs well across different domain even without fine-tuning
- Can be further improved using 10% domain data for fine-tuning



Qualitative Results

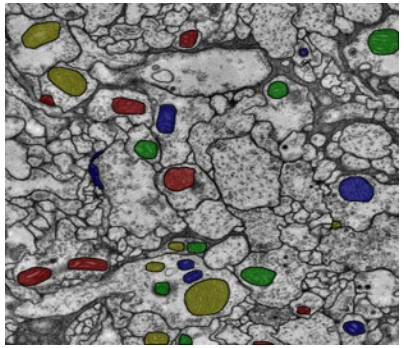
General object scenes

Cityscapes



Agriculture-Vision

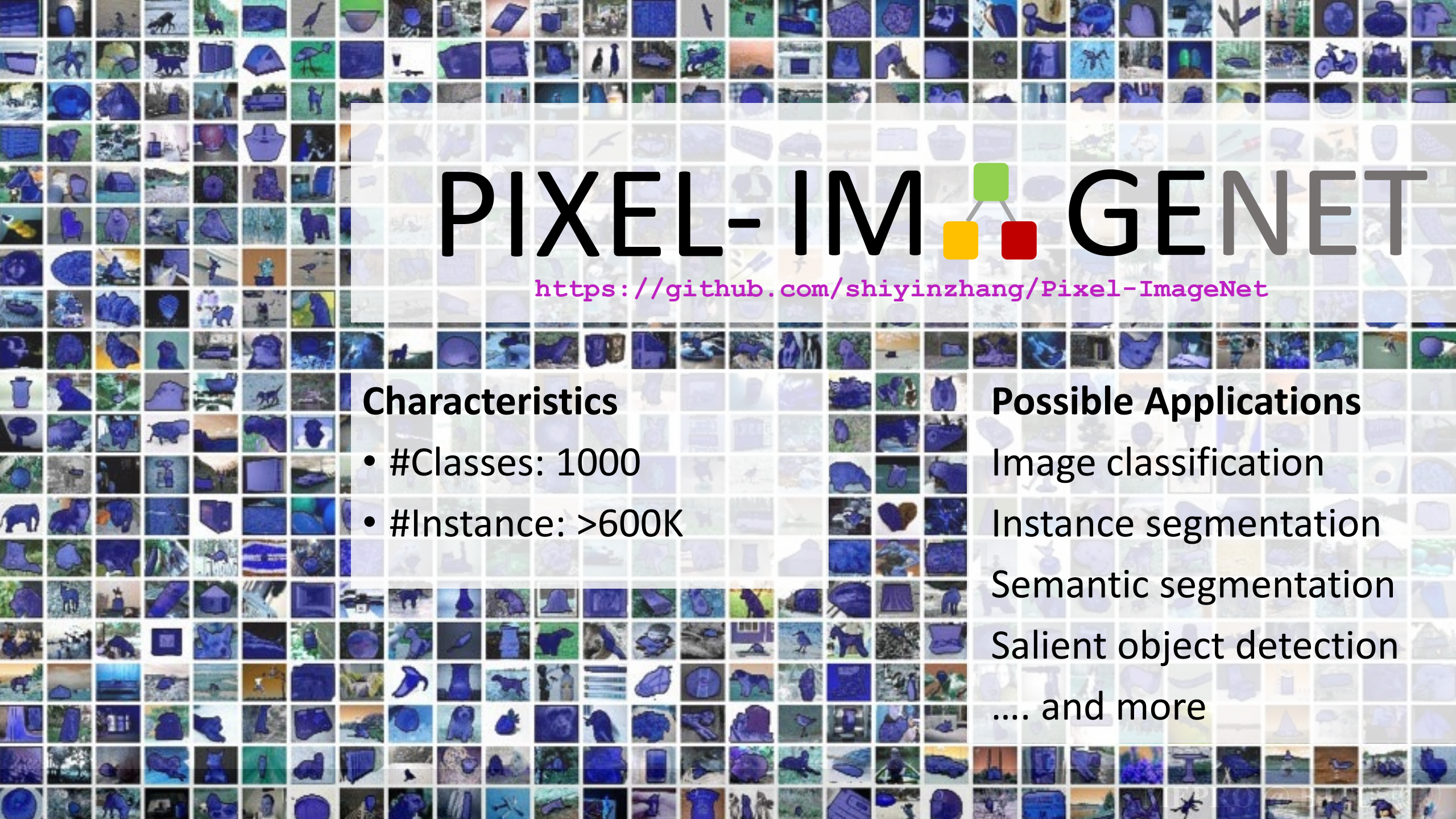
Rooftop



ssTEM


IMAGENET


PIXEL-IMAGENET



PIXEL-IMAGENET

https://github.com/shiyinzhang/Pixel-ImageNet

Characteristics

- #Classes: 1000
- #Instance: >600K

Possible Applications

Image classification

Instance segmentation

Semantic segmentation

Salient object detection

…. and more

Thanks

Q & A