Weakly Supervised Semantic Segmentation

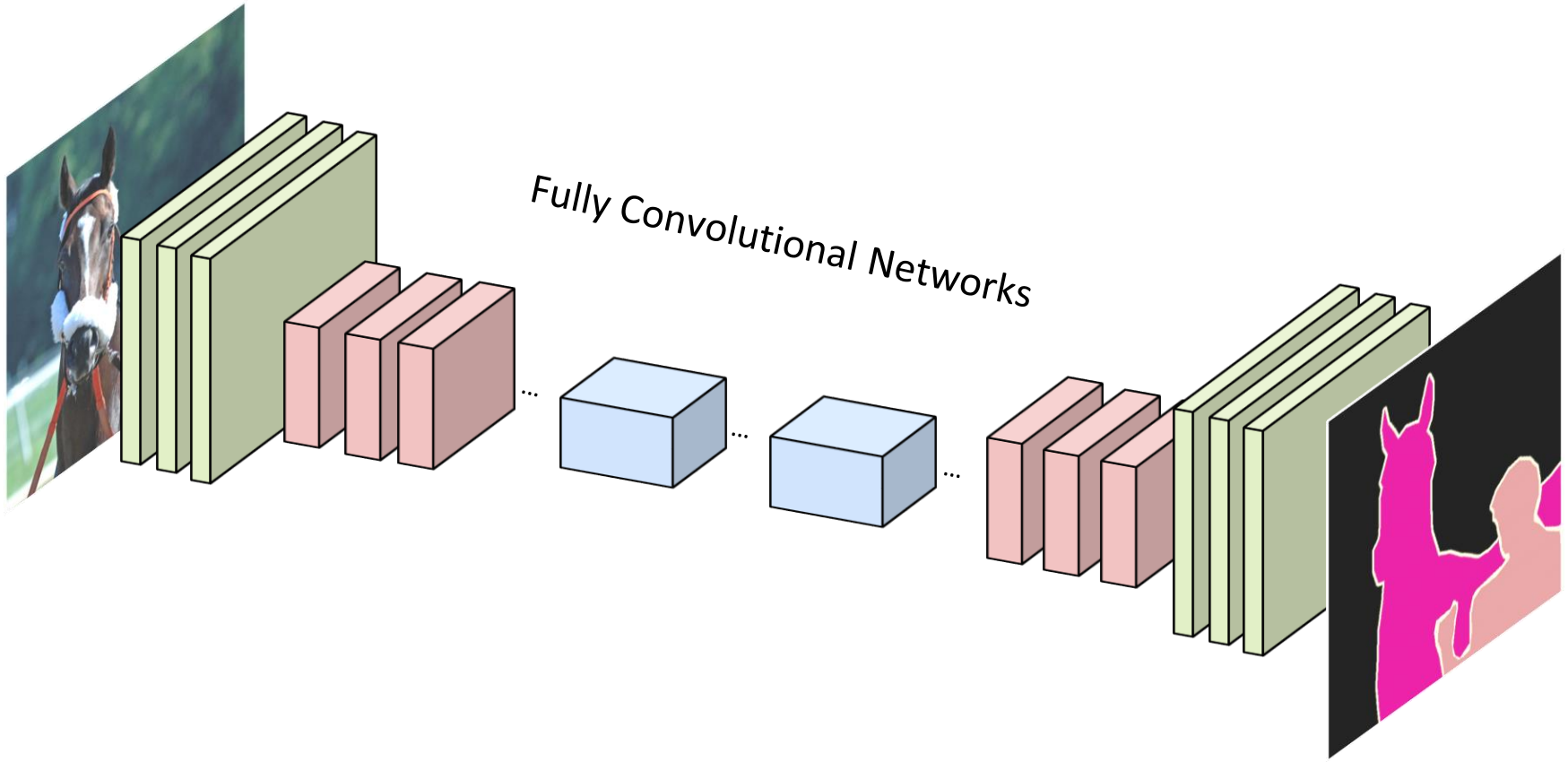
VALSE 2019 Tutorial

Yunchao Wei

University of Illinois at Urbana-Champaign

ILLINOIS

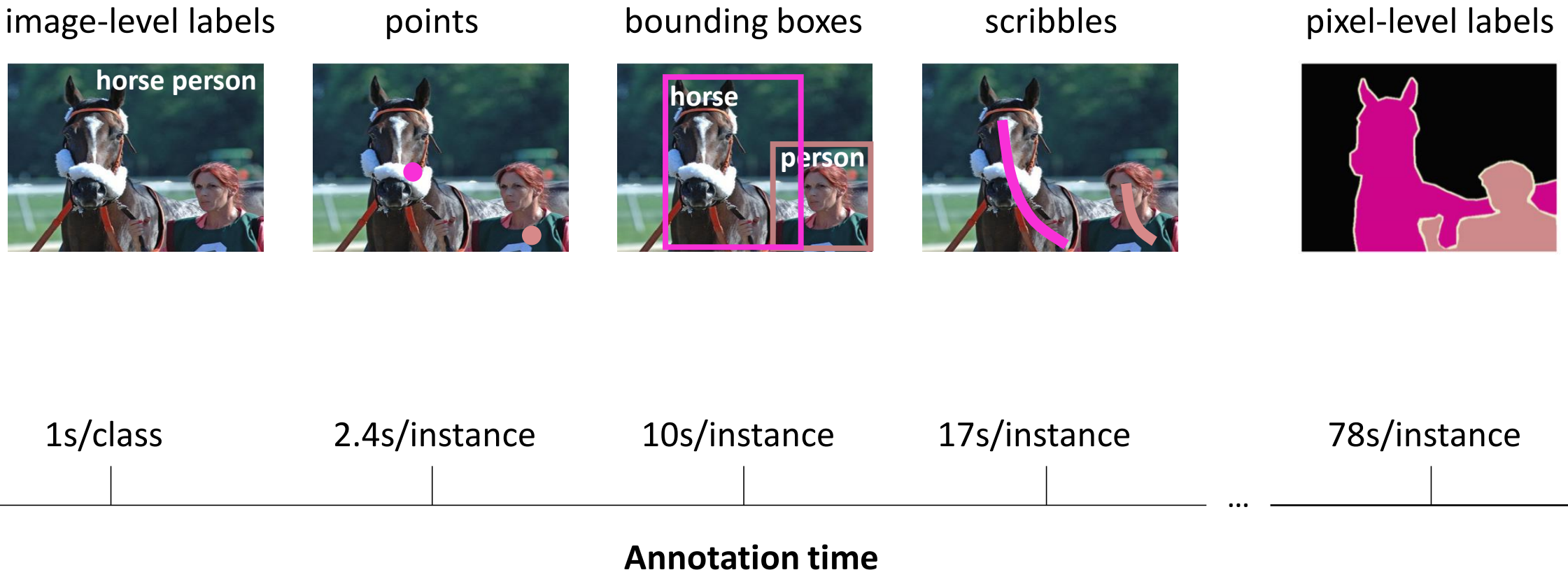
Overview: Fully-supervised Semantic Segmentation



- FCN
- Deeplab v1,v2,v3, v3+
- PSPNet
- RefineNet
- GCN
- EncNet
- Dense ASPP
- PSANet
- Non-Local Networks
- DANet, OCNet, CCNet

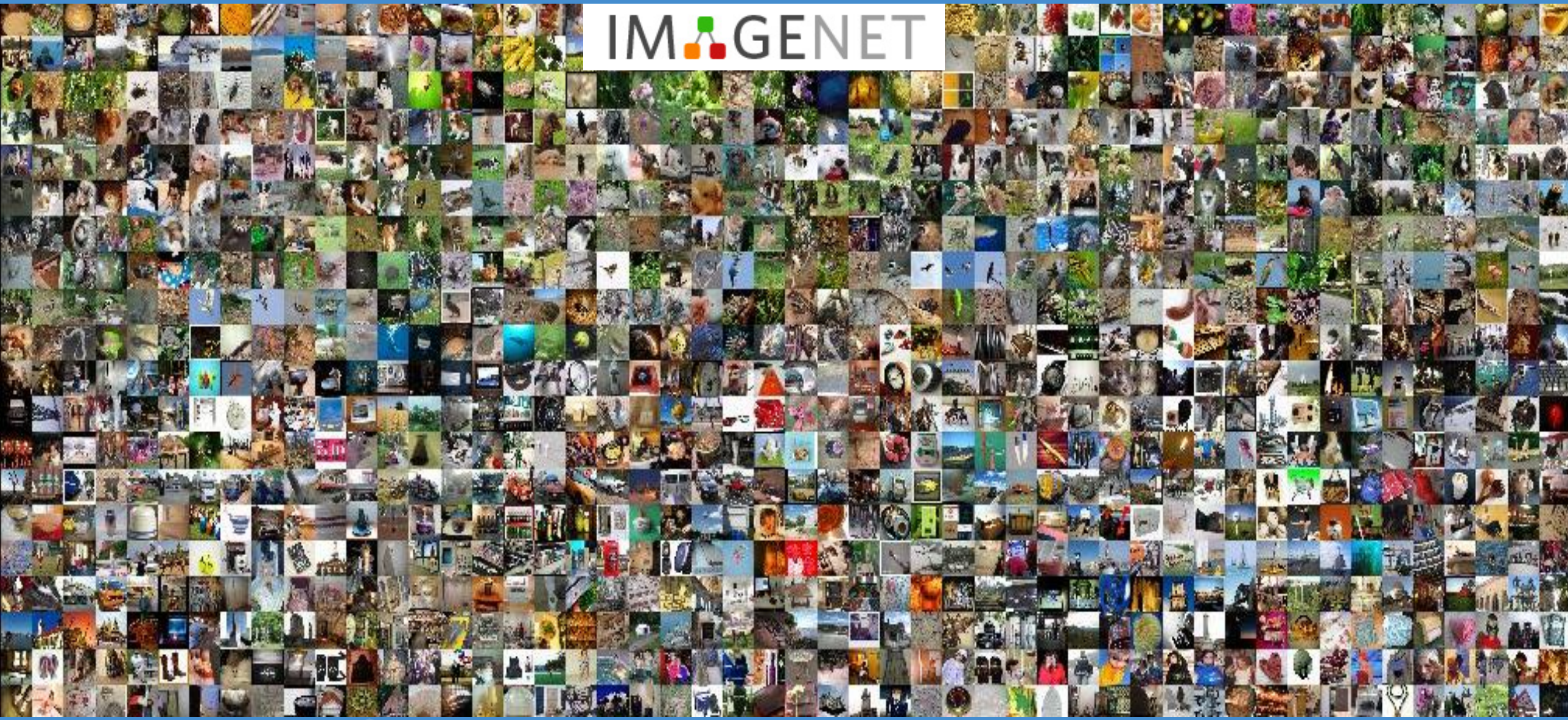
....

Overview: Manual Annotations for Object Recognition



[Lin CVPR16, Berman ECCV16, Hakan CVPR18 Tutorial]

Overview: Heavy Cost in Labeling Pixel-level Masks



Overview: Heavy Cost in Labeling Pixel-level Masks

IMAGENET

Overview: Weakly-supervised Semantic Segmentation

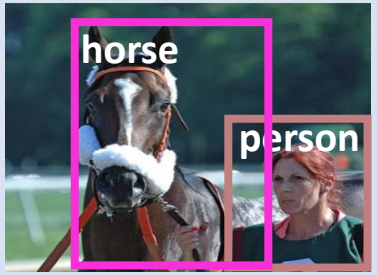
Weak Supervision

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

image-level labels

points

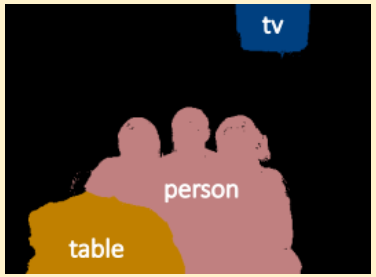
bounding boxes



scribbles

Training Stage (Weakly-supervised Annotations)

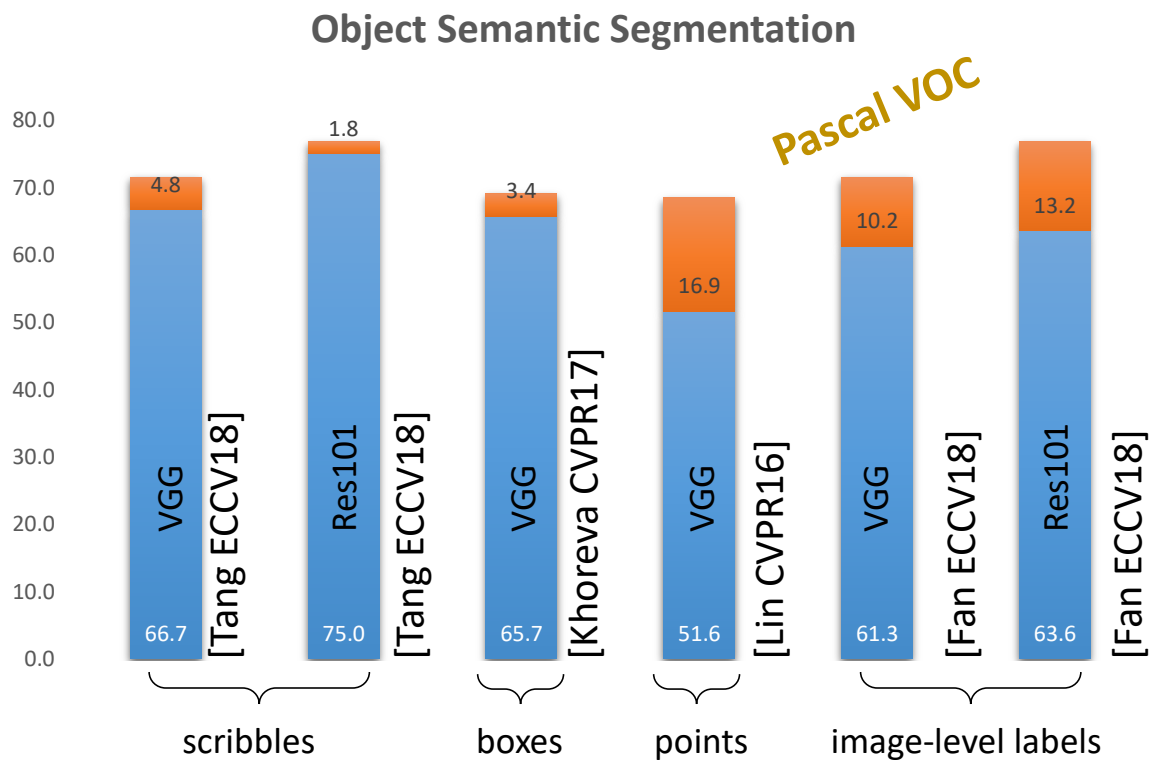
pixel-level labels

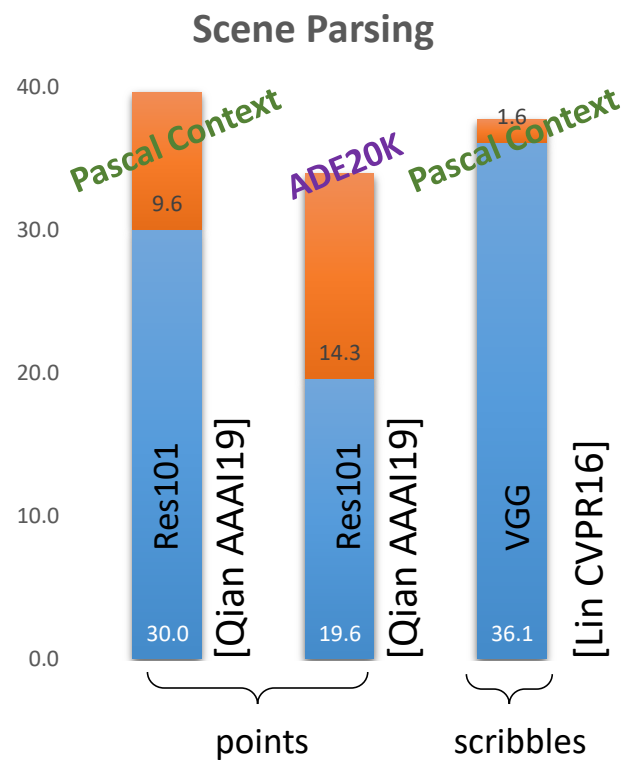


Testing Stage

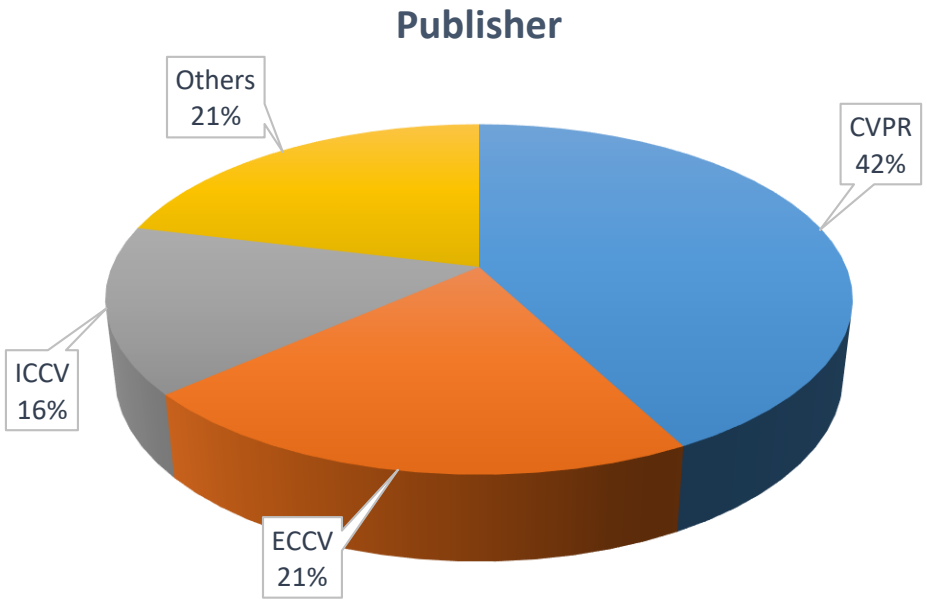
[Hakan CVPR18 Tutorial]

Current State-of-the-arts





Overview: About This Tutorial



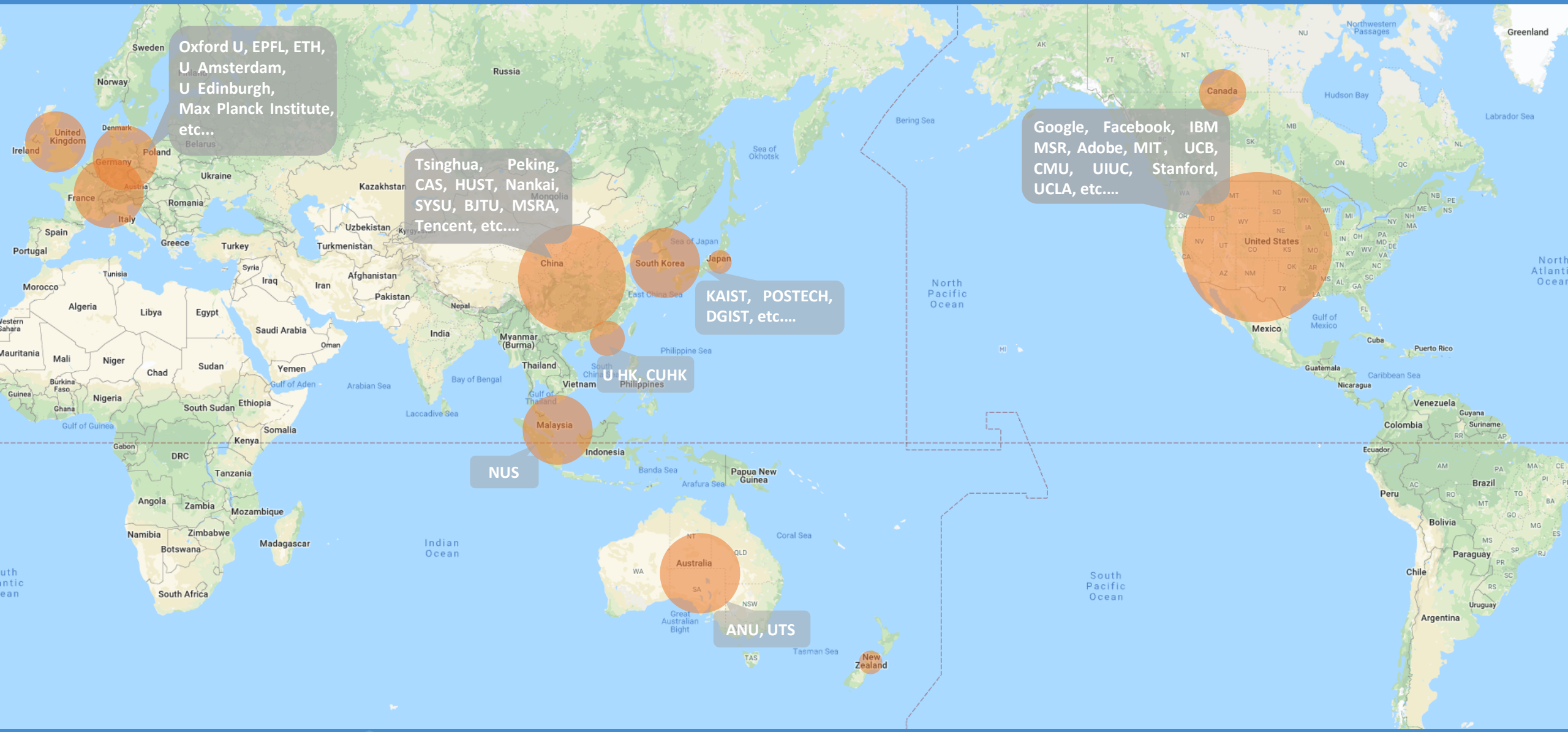
The Covered Topics

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

AAAI18-Transferable Semi-supervised Semanti...
Arxiv19-Large-scale interactive object segment...
BMVC17-Discovering Class-Specific Pixels for ...
CVPR15-From Image-level to Pixel-level Labeli...
CVPR16-Learning Deep Features for Discrimina...
CVPR17-Combining Bottom-Up Top-Down an...
CVPR17-Object Region Mining with Adversaria...
CVPR17-Weakly Supervised Semantic Segment...
CVPR18-Adversarial Complementary Learning ...
CVPR18-Interactive Image Segmentation with ...
CVPR18-Learning to Segment Every Thing.pdf
CVPR18-Normalized Cut Loss for Weakly-supe...
CVPR18-Tell Me Where to Look Guided Attenti...
CVPR18-Weakly-Supervised Semantic Segmen...
ECCV16-Augmented Feedback in Semantic Se...
ECCV16-Distinct Class-specific Saliency Maps f...
ECCV16-Seed, Expand and Constrain.pdf
ECCV18-Associating Inter-Image Salient Instan...
ECCV18-On Regularized Losses for Weakly-sup...
ECCV18-Weakly- and Semi-Supervised Panopti...
ICCV15-Constrained Convolutional Neural Net...
ICCV17-Bringing Background into the Foregro...
ICCV17-Regional Interactive Image Segmentati...
ICCV17-Two-Phase Learning for Weakly Super...
NIPS18-Self-Erasing Network for Integral Objec...
PAMI17-STC A Simple to Complex Framework ...

AAAI19-Weakly Supervised Scene Parsing with ...
BMVC17-Deep GrabCut for Object Selection.pdf
BMVC18-Iteratively Trained Interactive Segmen...
CVPR16-Deep Interactive Object Selection.pdf
CVPR16-ScribbleSup Scribble-Supervised Conv...
CVPR17-Exploiting saliency for object segment...
CVPR17-Simple Does It Weakly Supervised Inst...
CVPR17-Webly Supervised Semantic Segmenta...
CVPR18-Deep Extreme Cut From Extreme Point...
CVPR18-Learning Pixel-level Semantic Affinity ...
CVPR18-Multi-Evidence Filtering and Fusion fo...
CVPR18-Revisiting Dilated Convolution.pdf
CVPR18-Weakly Supervised Instance Segmenta...
CVPR18-Weakly-supervised semantic segment...
ECCV16-Built-in Foreground Background Prior...
ECCV16-ExcitationBackprop.pdf
ECCV16-What s the Point Semantic Segmentat...
ECCV18-Interactive boundary prediction for ob...
ECCV18-Self-produced Guidance for Weakly-s...
ICCV15-BoxSup Exploiting Bounding Boxes to ...
ICCV15-Weakly- and Semi-Supervised Learnin...
ICCV17-Hide-and-Seek Forcing a Network to b...
ICCV17-Soft Proposal Networks for Weakly Su...
ICLR15-Fully Convolutional Multi-Class Multip...
PAMI17-Incorporating Network Built-in Priors i...
PR16-Learning to segment with image-level an...

Overview: Researcher Distribution



Outline

image-level labels

points

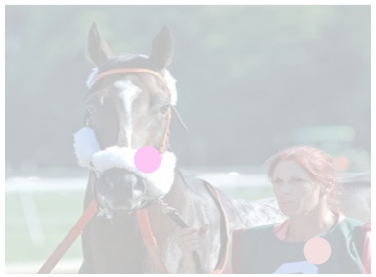
bounding boxes

scribbles

Outline

image-level labels

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

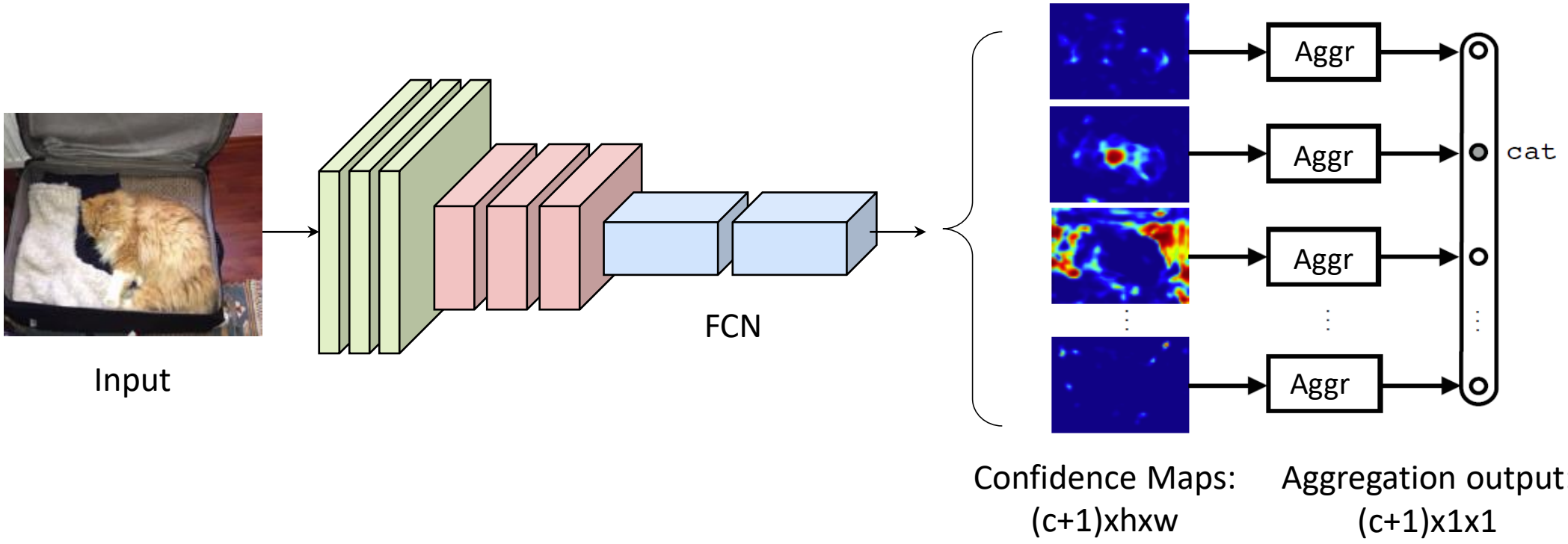
image-level labels

- 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

image-level labels

End-to-end Learning with Constraint Loss

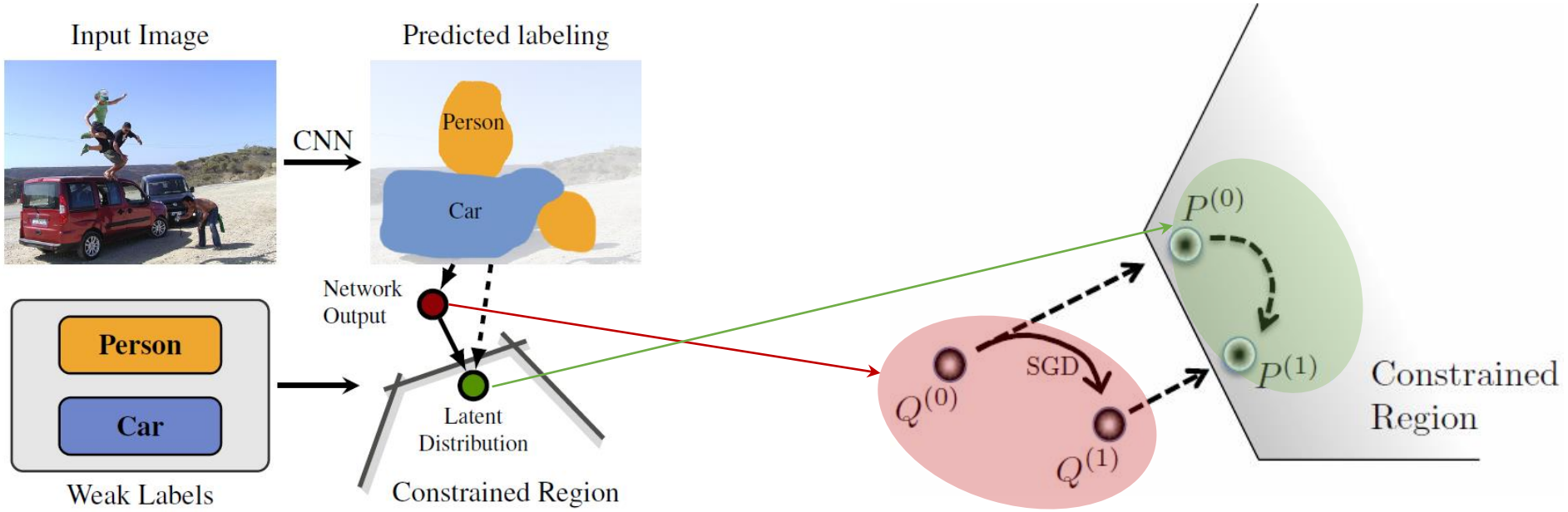
Multiple Instance Learning



[Pinheiro CVPR15, Pathak ICLR15 Workshop]

End-to-end Learning with Constraint Loss

Constrained Convolutional Neural Networks



Suppression Constraint

$$\sum_{i=1}^{n} p_i(l) \leq 0 \qquad \forall\, l \notin \mathcal{L}_I$$

Foreground Constraint

$$a_l \leq \sum_{i=1}^{n} p_i(l) \qquad \forall\, l \in \mathcal{L}_I$$

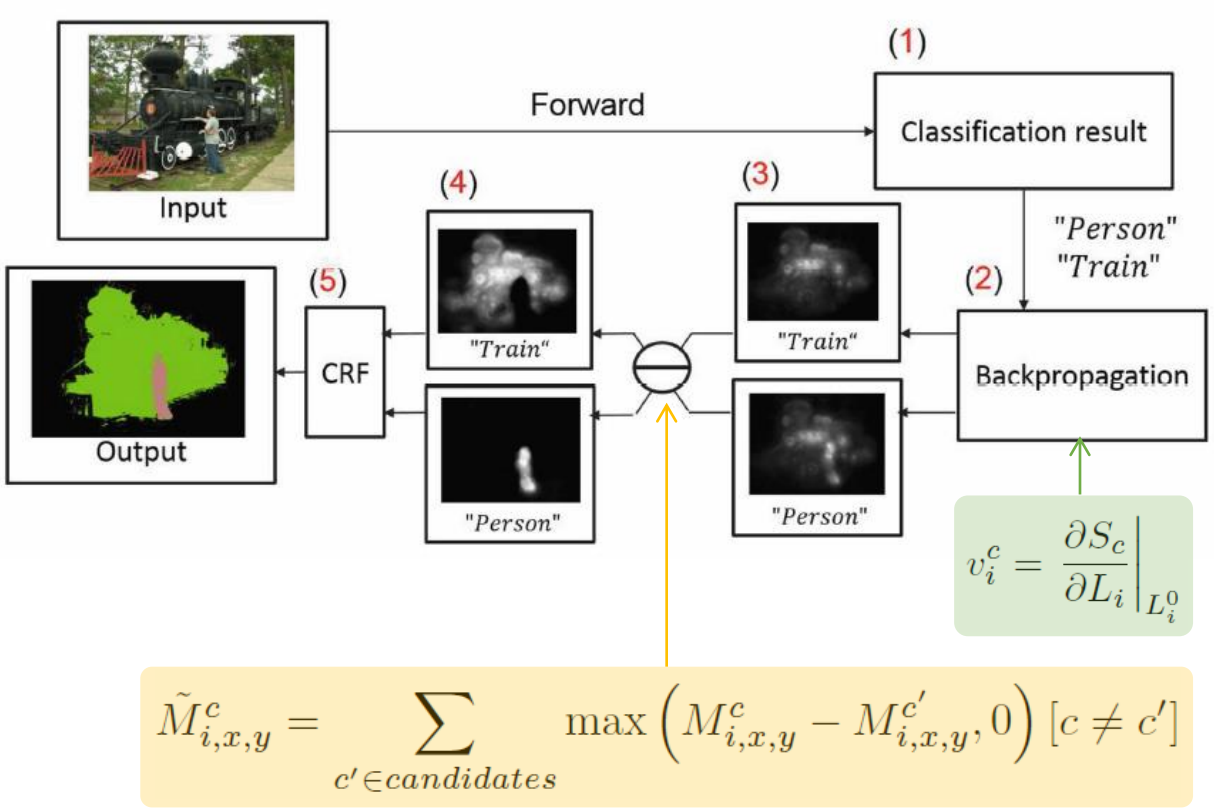
Background Constraint

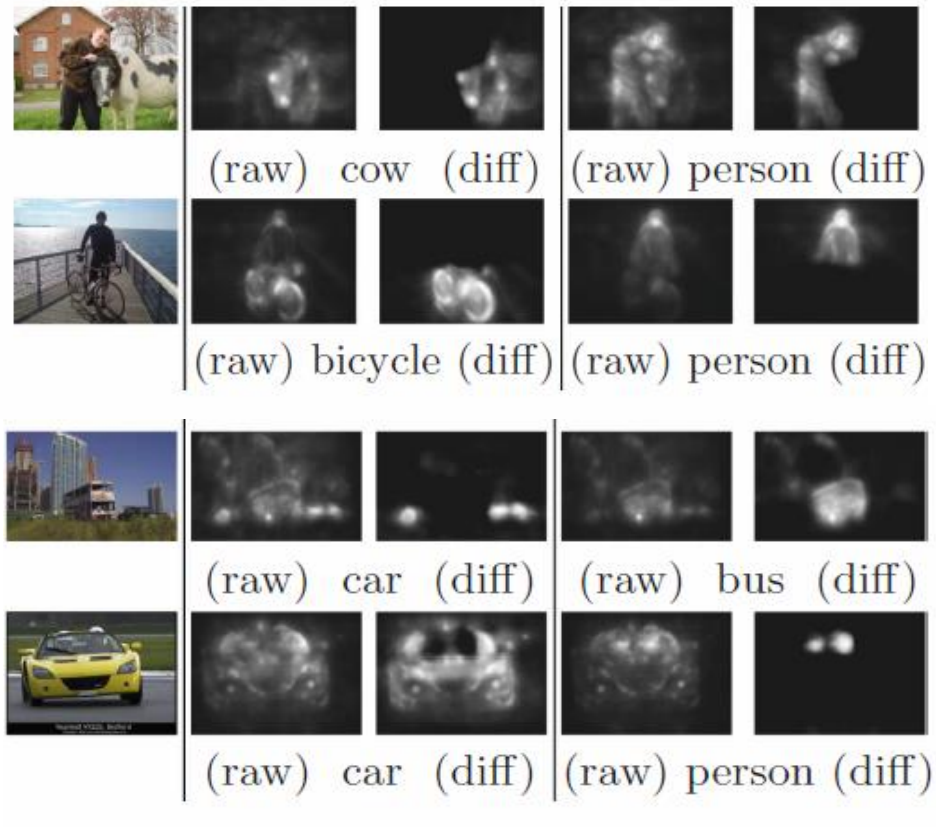
$$a_0 \leq \sum_{i=1}^{n} p_i(0) \leq b_0$$

[Chen ICCV15, Pathak ICCV15]

End-to-end Learning with Constraint Loss

Distinct Class-Specific Saliency Maps





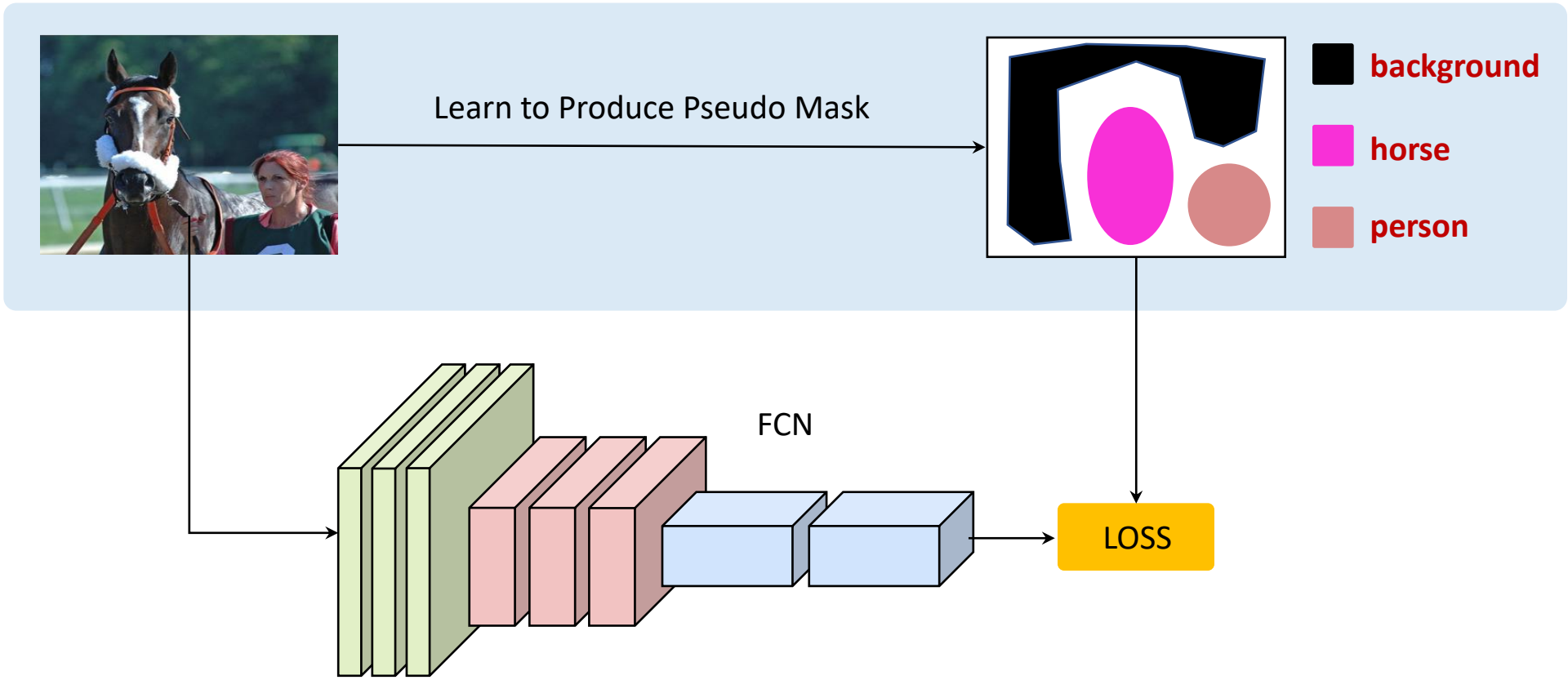
[Shimoda ECCV16]

- 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

image-level labels

Learning to Produce Pseudo Pixel-level Masks

- Standard Pipeline

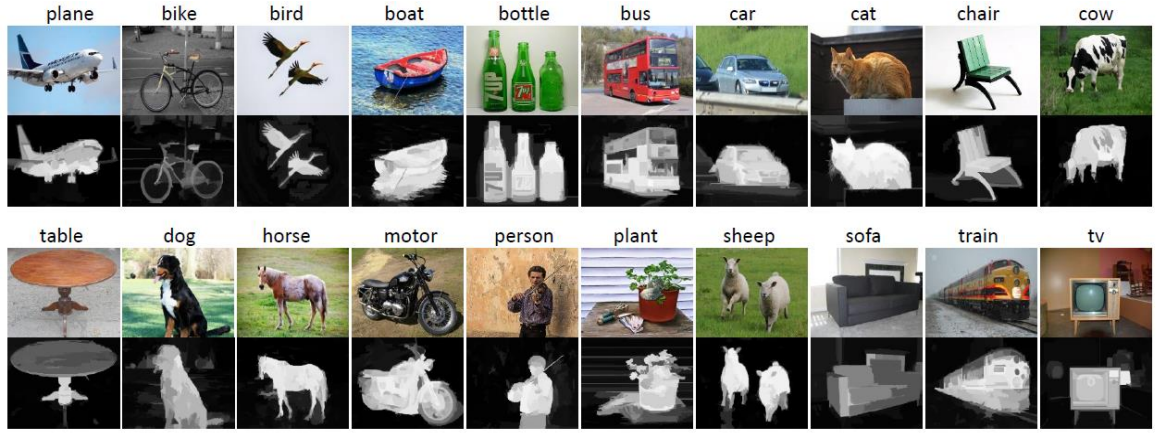


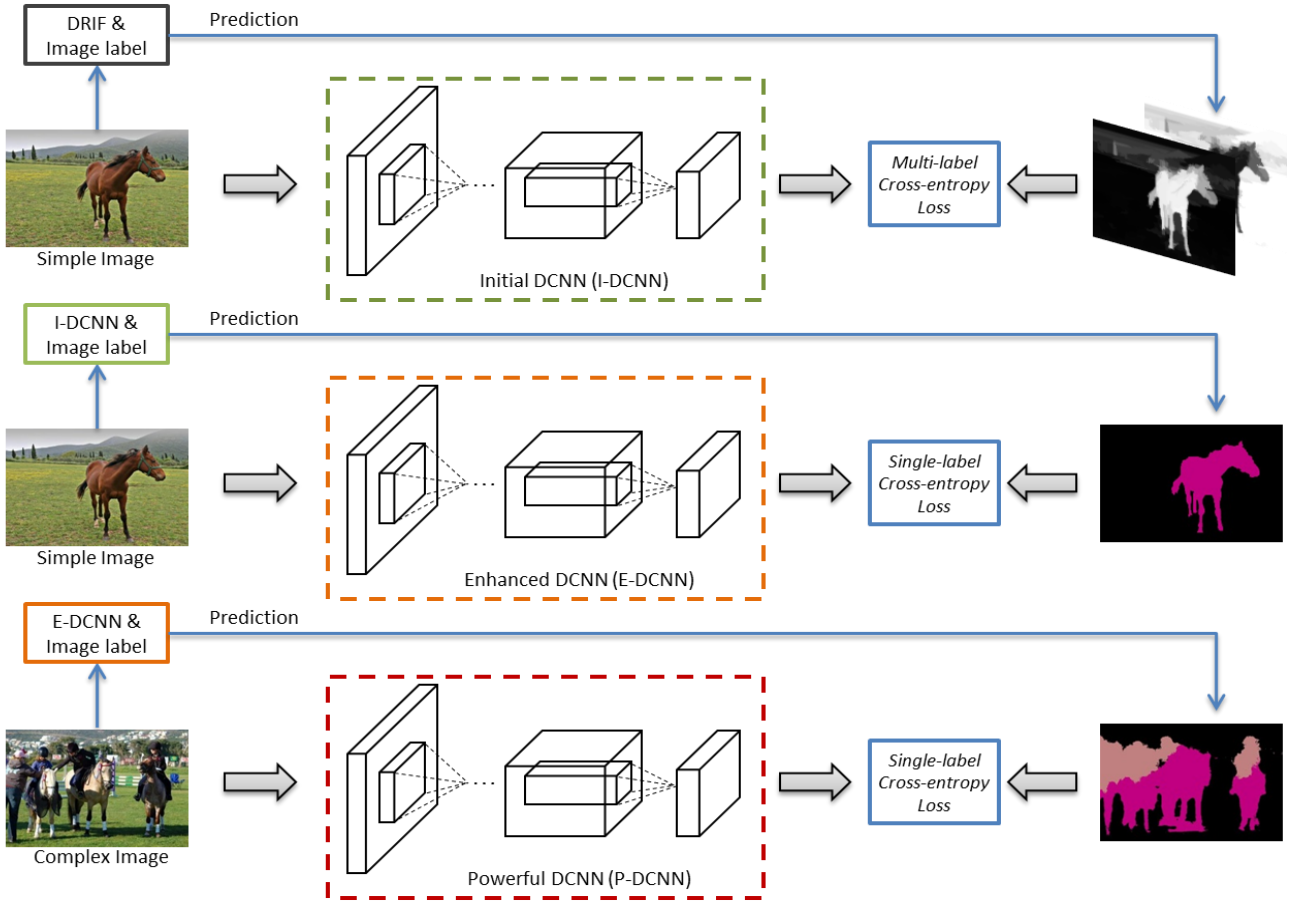
- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
    - Object Proposals
    - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation

image-level labels

Learning to Produce Pseudo Masks--Additional Data

Simple to Complex





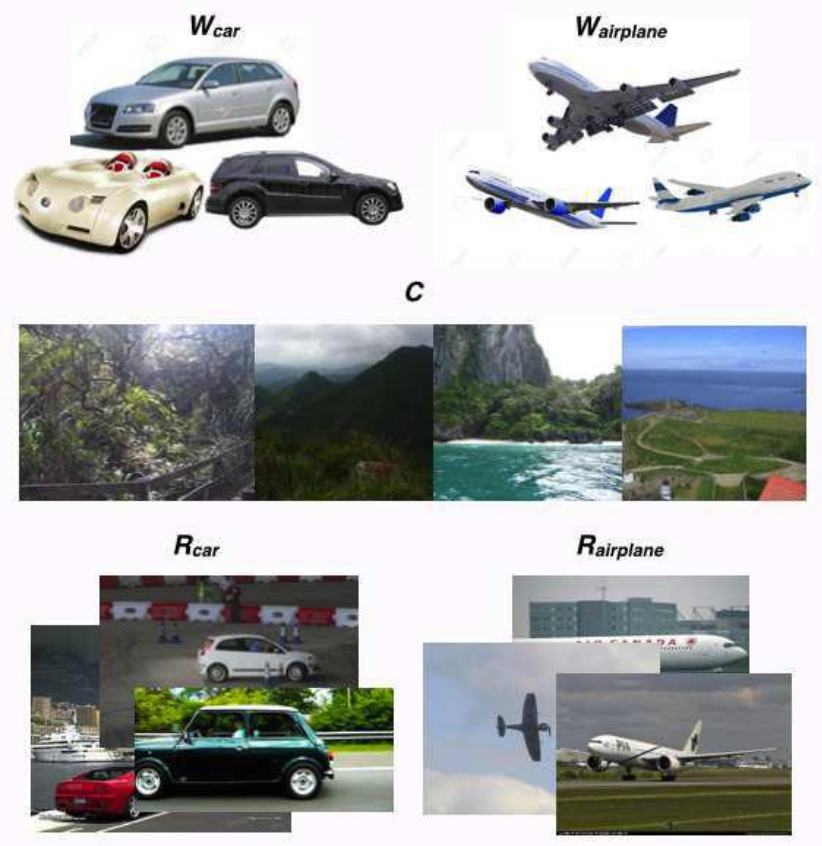
Ablation Analysis on Pascal VOC12 val

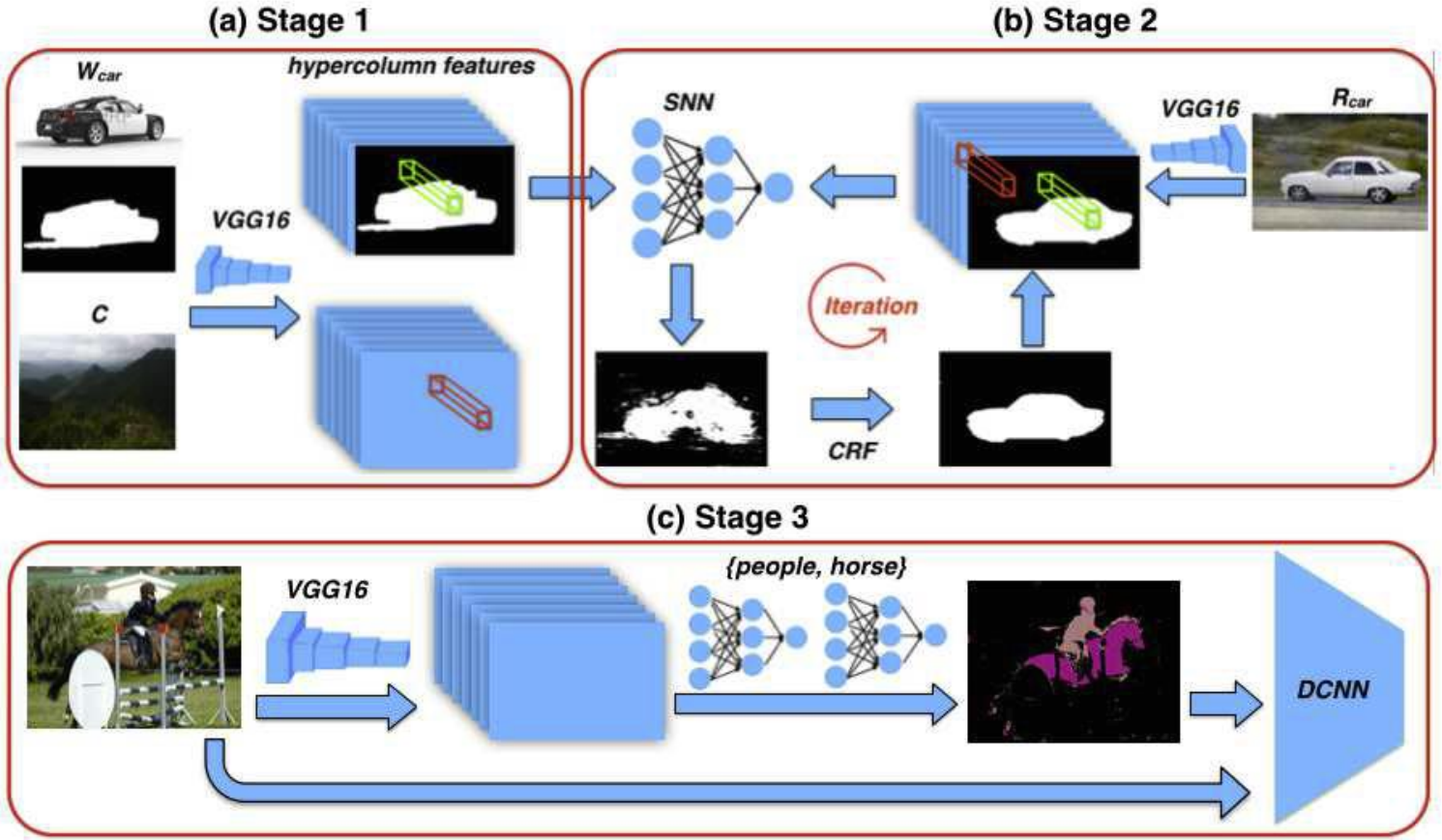
Networks	Training Set	mIoU
I-DCNN	Flickr-Clean	44.1
E-DCNN	Flickr-Clean	46.8
P-DCNN	Flickr-Clean+VOC	49.8

[Wei PAMI17]

Learning to Produce Pseudo Masks--Additional Data

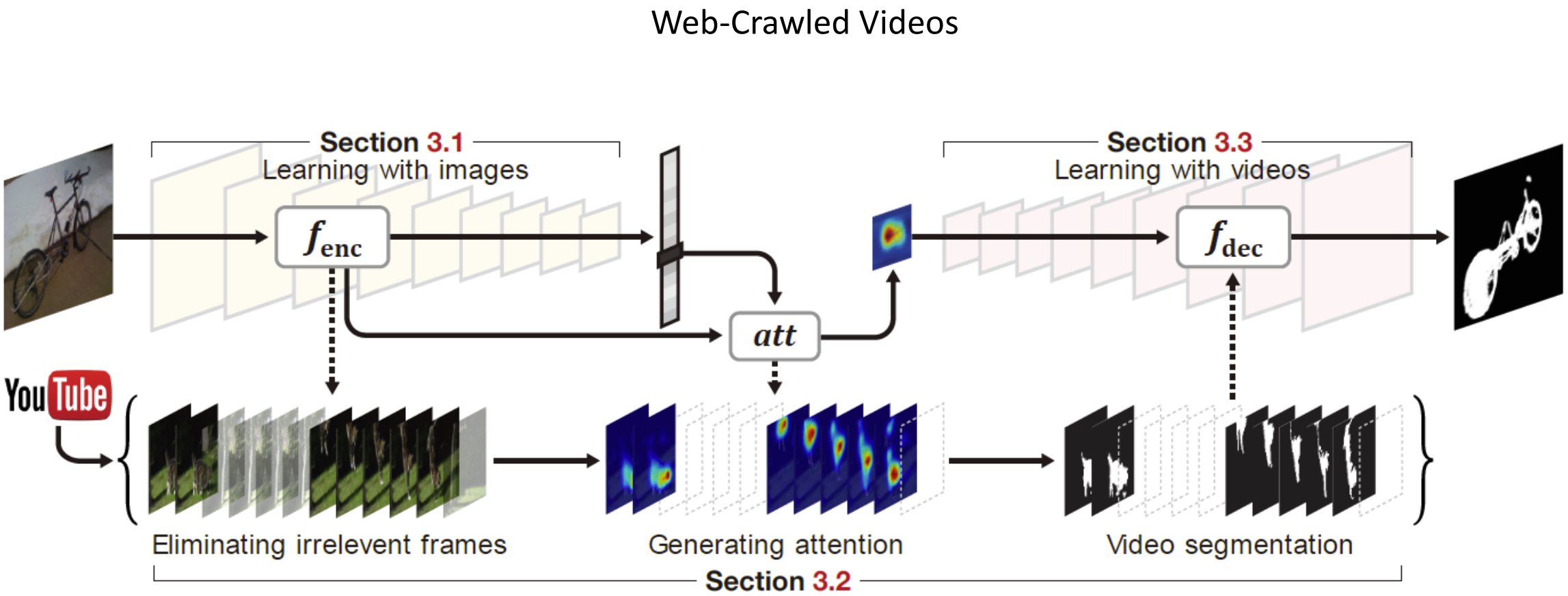
Webly Supervised Semantic Segmentation





[Jin CVPR17]

Learning to Produce Pseudo Masks--Additional Data



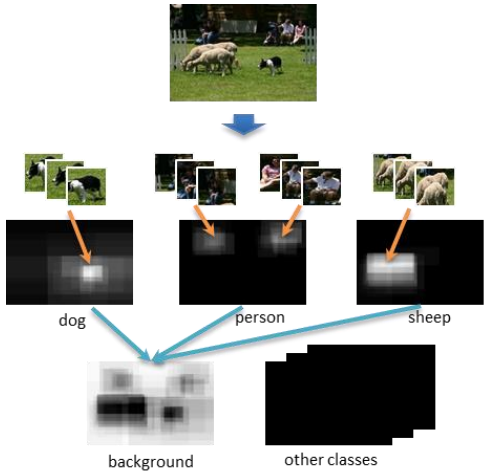
[Hong CVPR17]

- 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

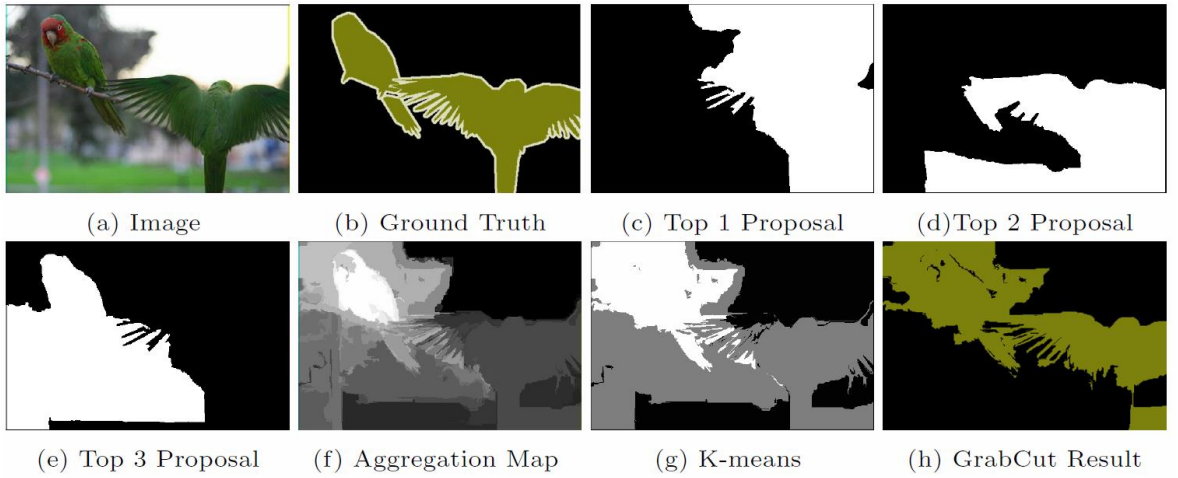
image-level labels

Learning to Produce Pseudo Masks--Object Proposals

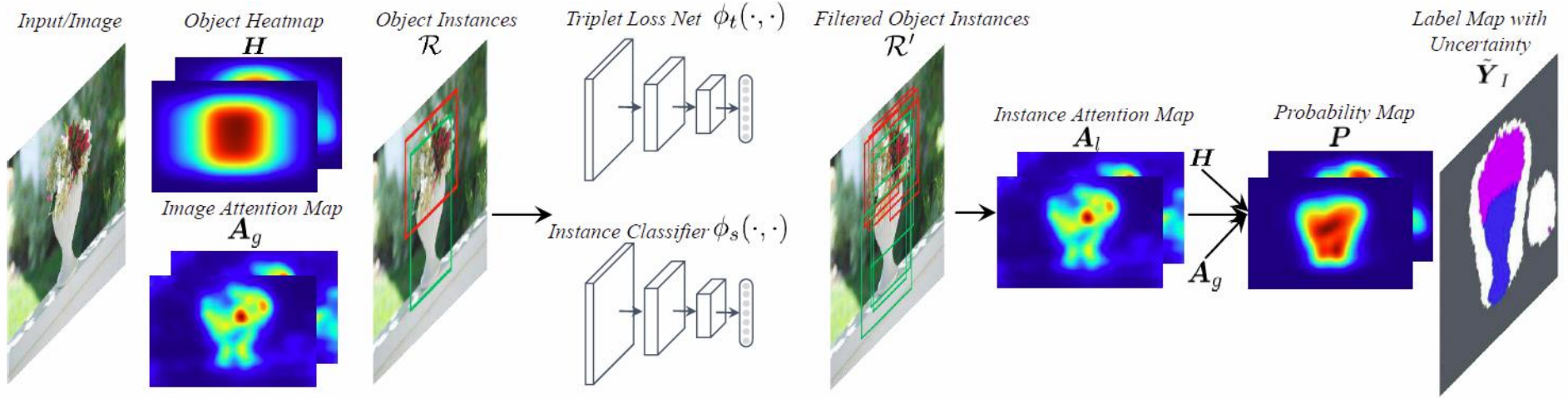
[Wei PR16]



Augmented Feedback [Qi ECCV16]



Multi-Evidence Filtering and Fusion [Ge CVPR18]

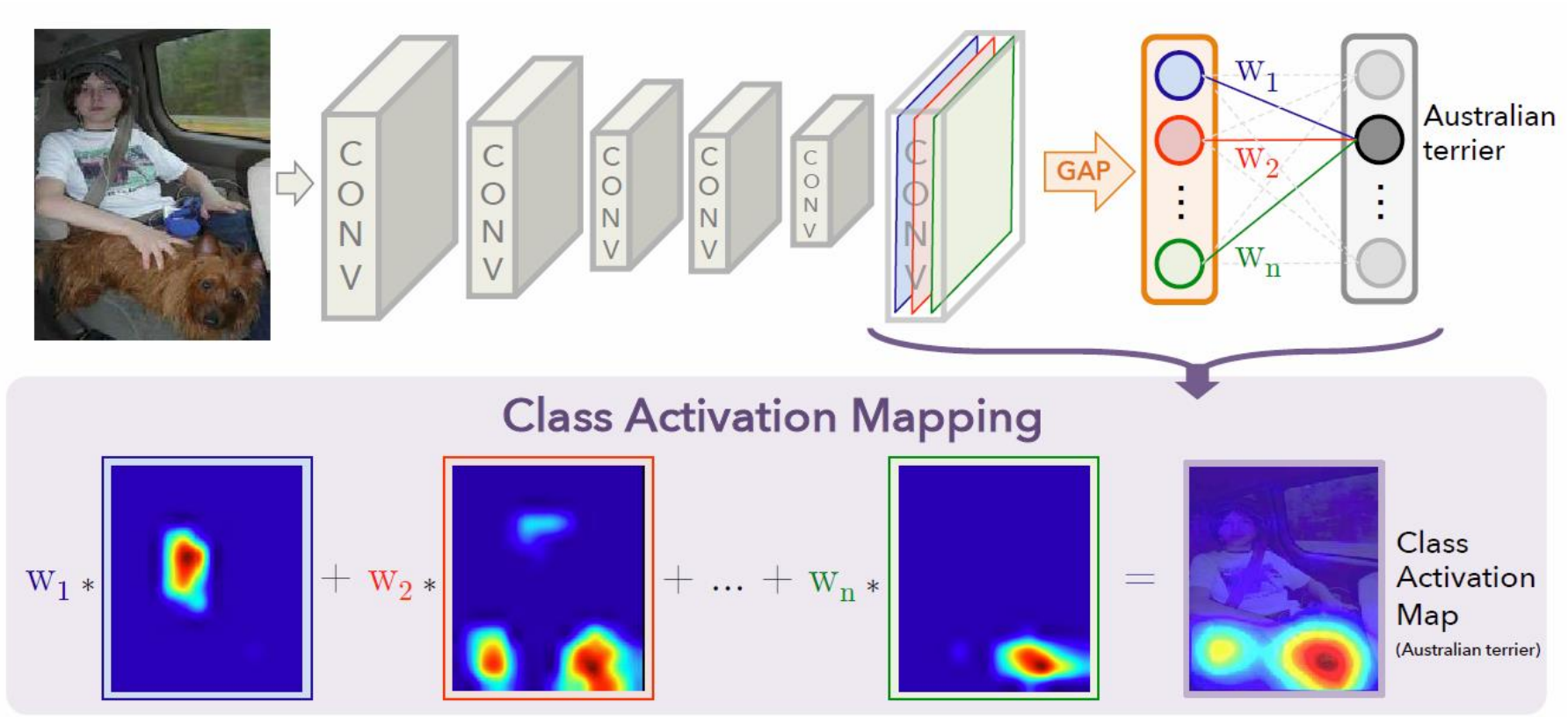


- End-to-end Learning with Constraint Loss
- Learning to Produce Pseudo Pixel-level Masks
  - Additional Data
  - Object Proposals
  - Top-down Attention
- Semi-Supervised Learning
- Instance Segmentation

image-level labels

Learning to Produce Pseudo Masks--Top-down Attention

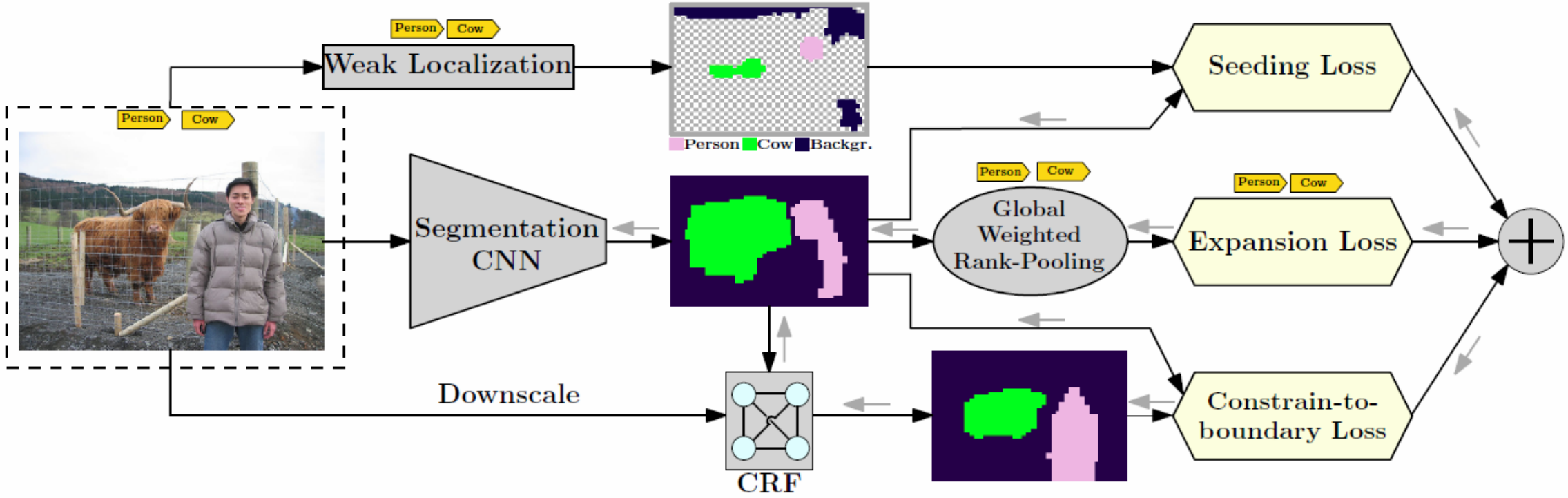
Class Activation Mapping [Zhou CVPR16]



[Zhou CVPR16, Zhang ECCV16, Zhu ICCV17, Zhang CVPR18, Zhang ECCV18]

Learning to Produce Pseudo Masks--Top-down Attention

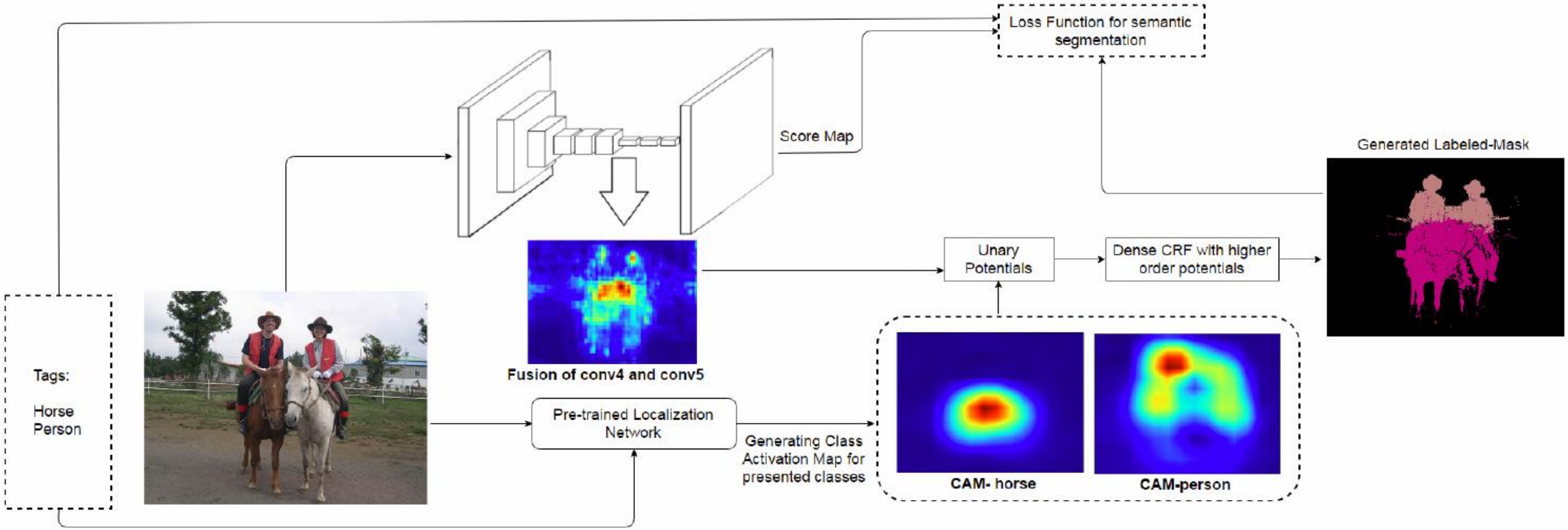
Seed, Expand and Constrain



[Kolesnikov ECCV16, Roy CVPR17]

Learning to Produce Pseudo Masks--Top-down Attention

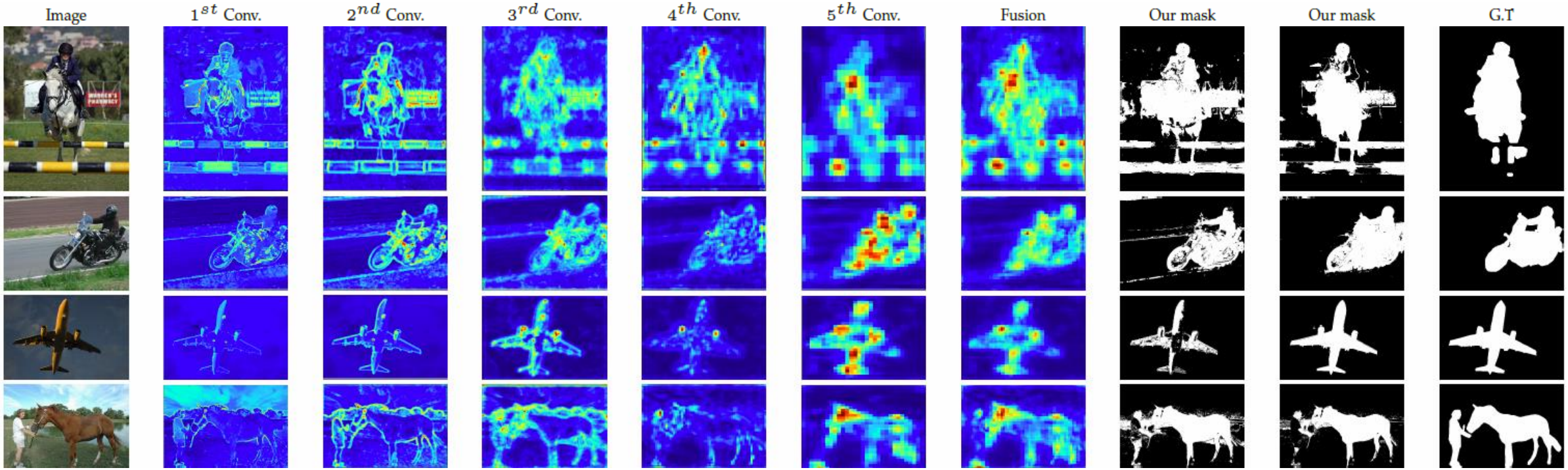
Built-in Fg/Bg Prior



[Saleh ECCV16, Saleh PAMI17]

Learning to Produce Pseudo Masks--Top-down Attention

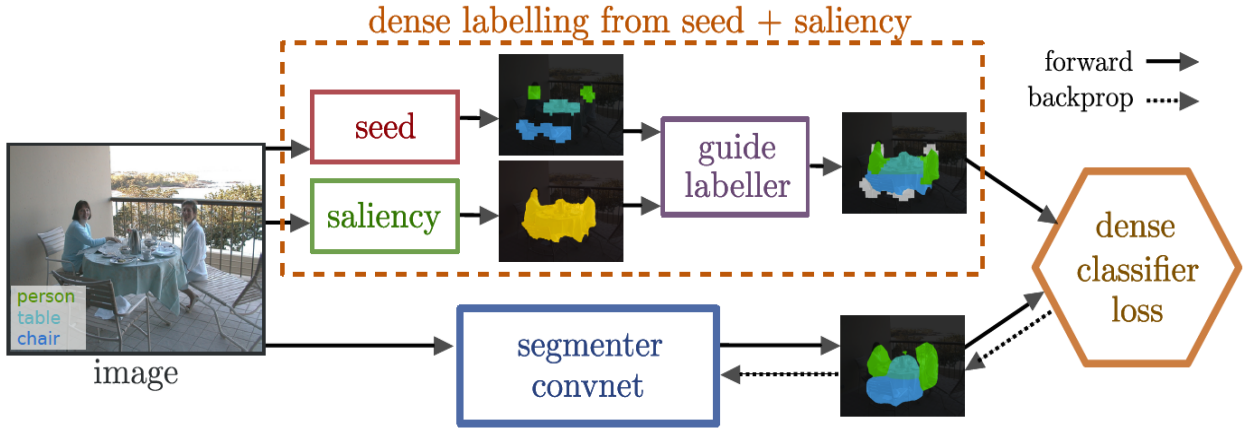
Built-in Fg/Bg Prior

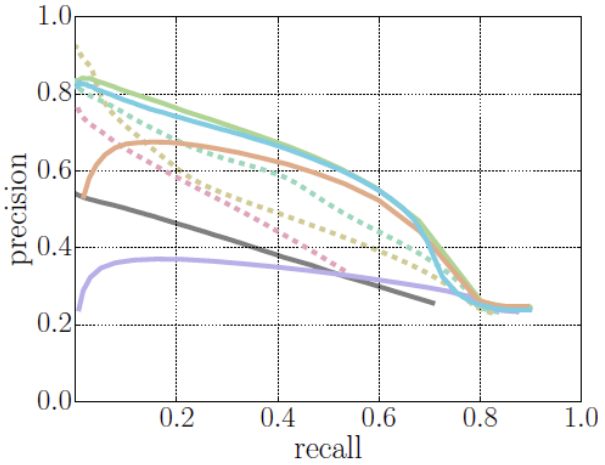


[Saleh ECCV16, Saleh PAMI17]

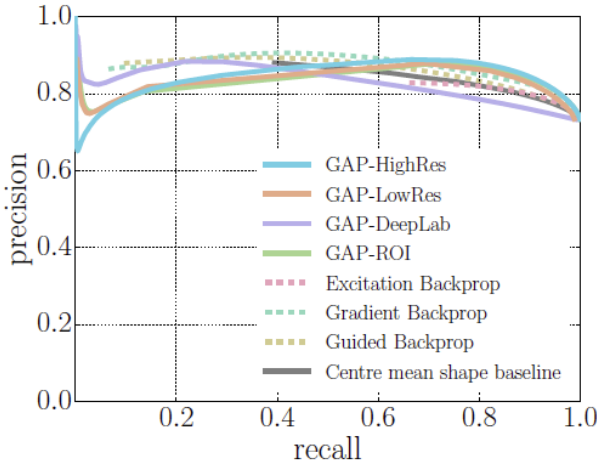
Learning to Produce Pseudo Masks--Top-down Attention

Exploiting Saliency





(a) Foreground categories



(b) Background category

Fusion Strategies

(a) Image

(b) Ground truth

(c) Seed

(d) Saliency

(e) $\mathcal{G}_0$

(f) $\mathcal{G}_1$

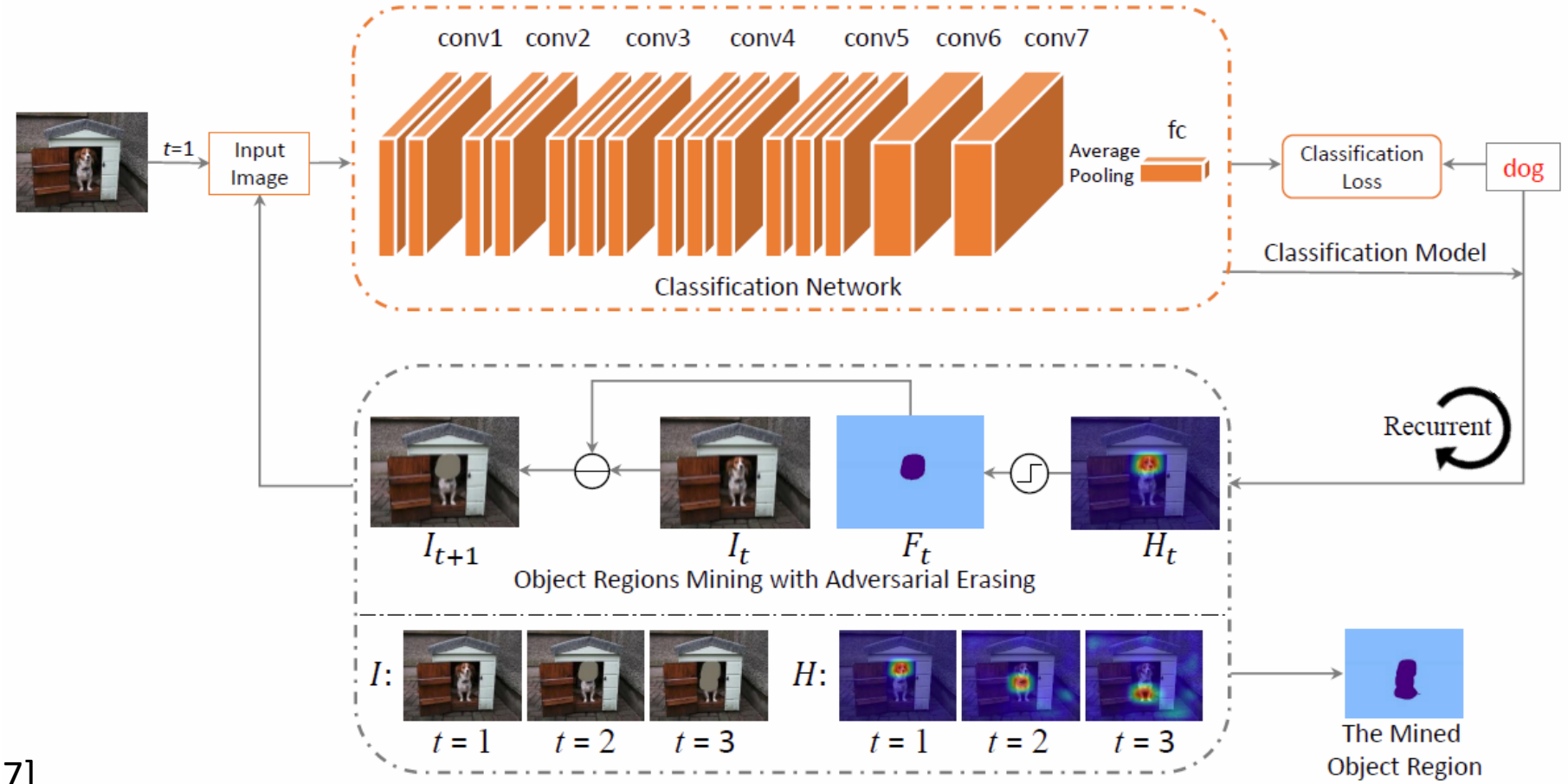
(g) $\mathcal{G}_2$

[Oh CVPR17]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Object Region Mining with Adversarial Erasing

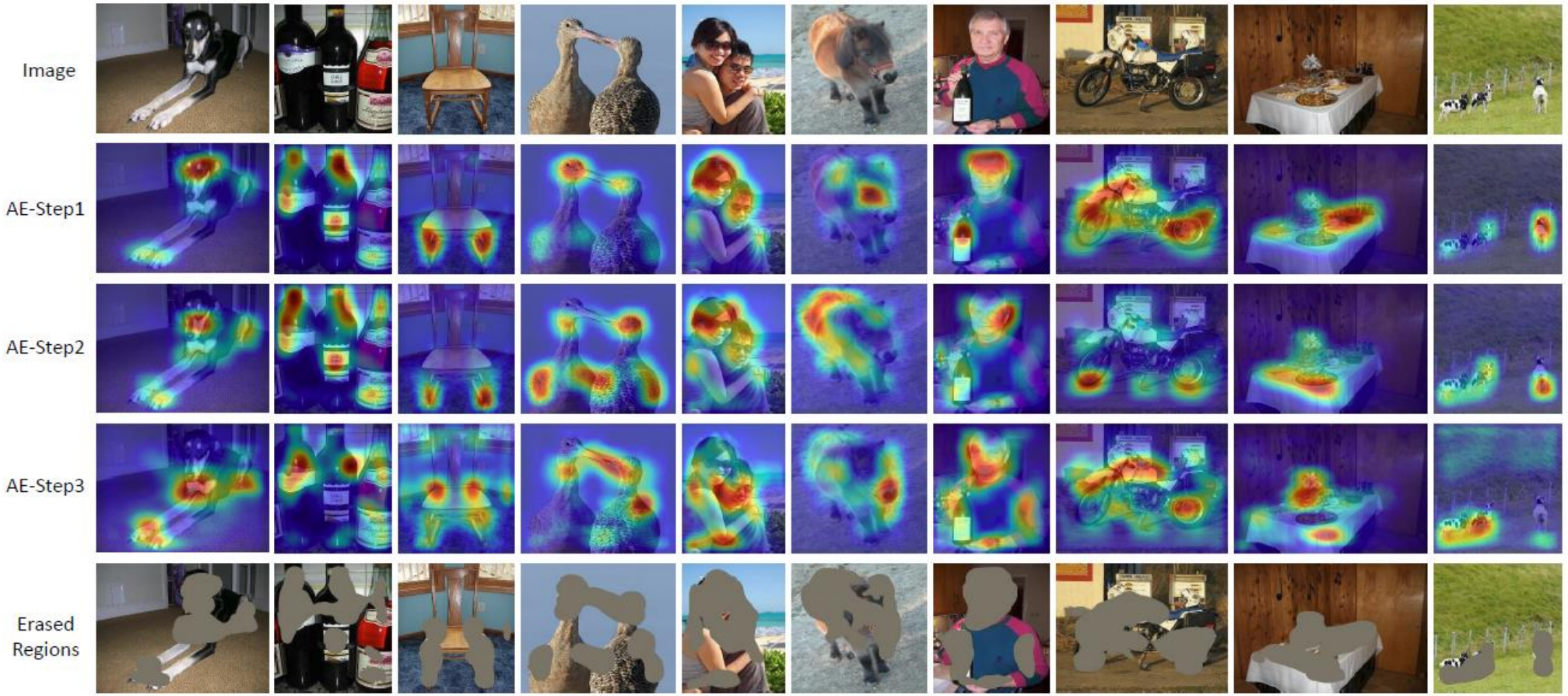


[Wei CVPR17]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

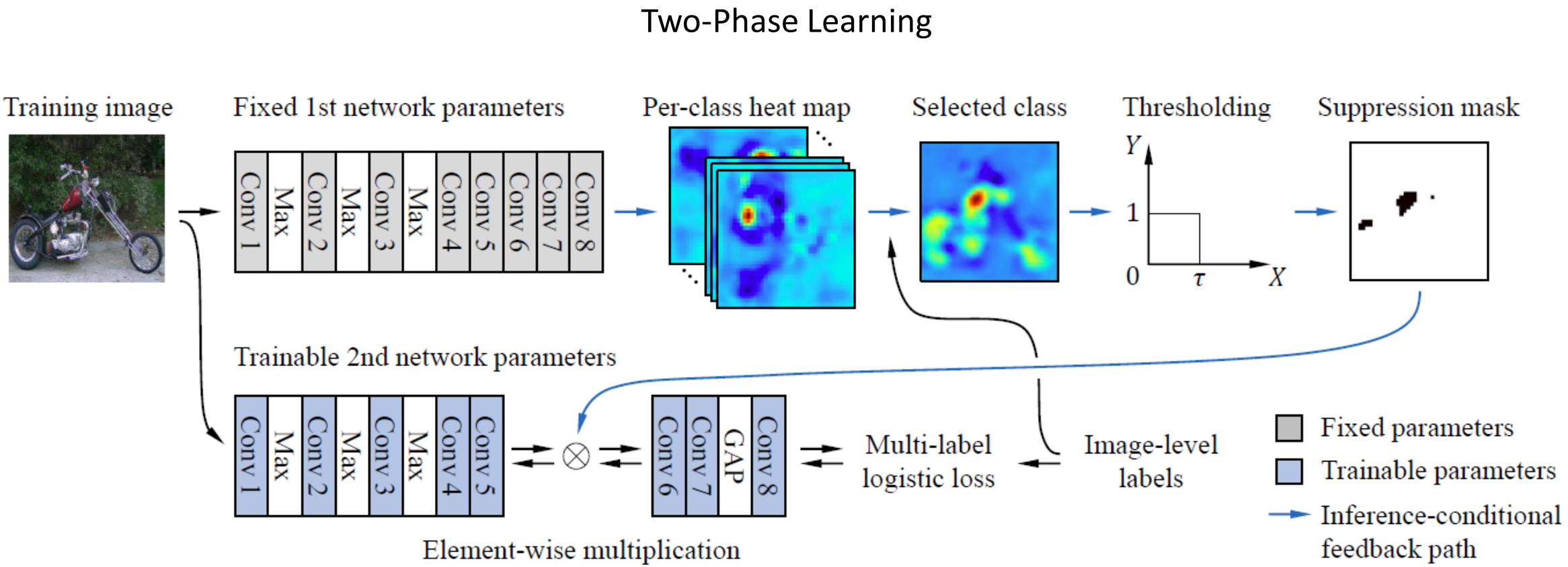
Object Region Mining with Adversarial Erasing



[Wei CVPR17]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

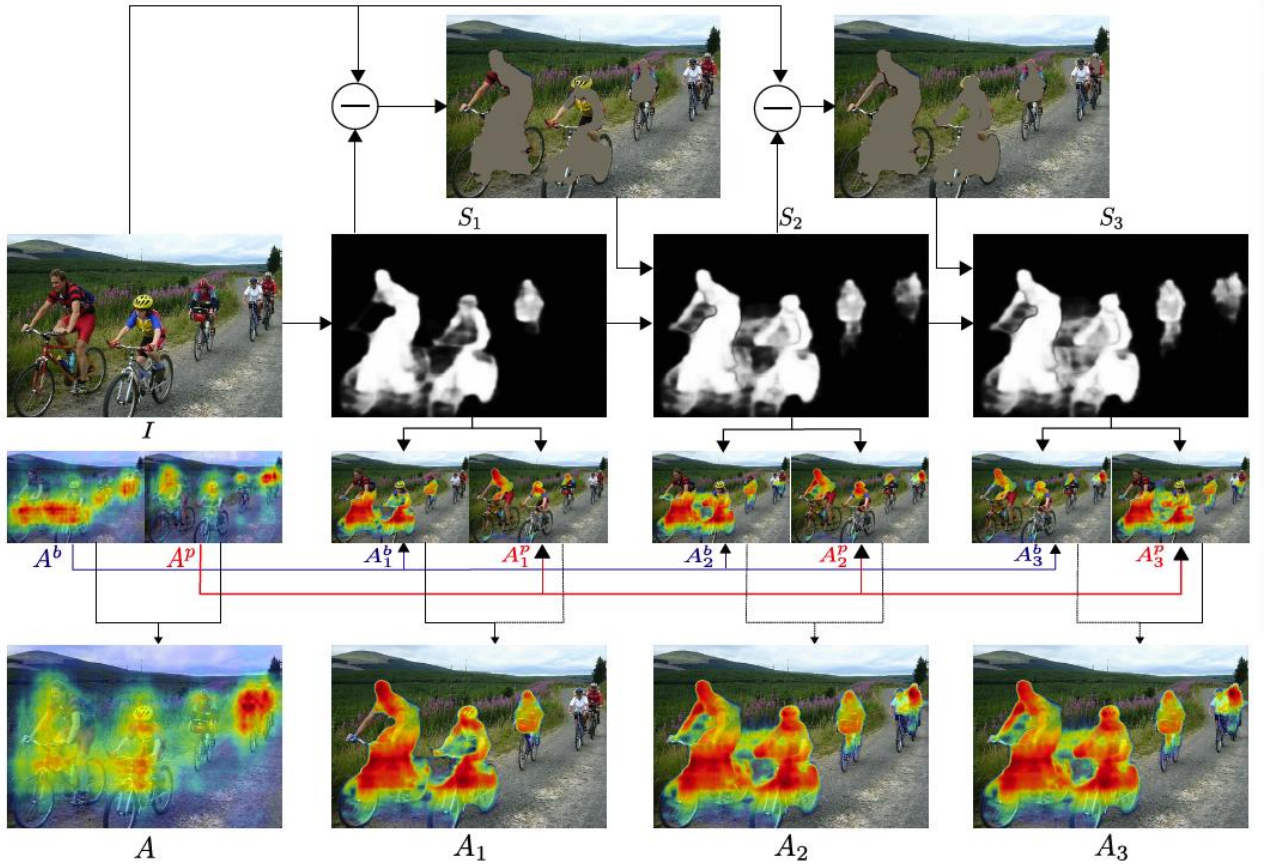


[Kim ICCV17]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Discovering Class-Specific Pixels



[Chaudhry BMVC17]

Algorithm 1 Discovering Class-Specific Pixels

Input: Image Labels $\mathbf{z}$; Saliency Map S; Attention Maps A; γ

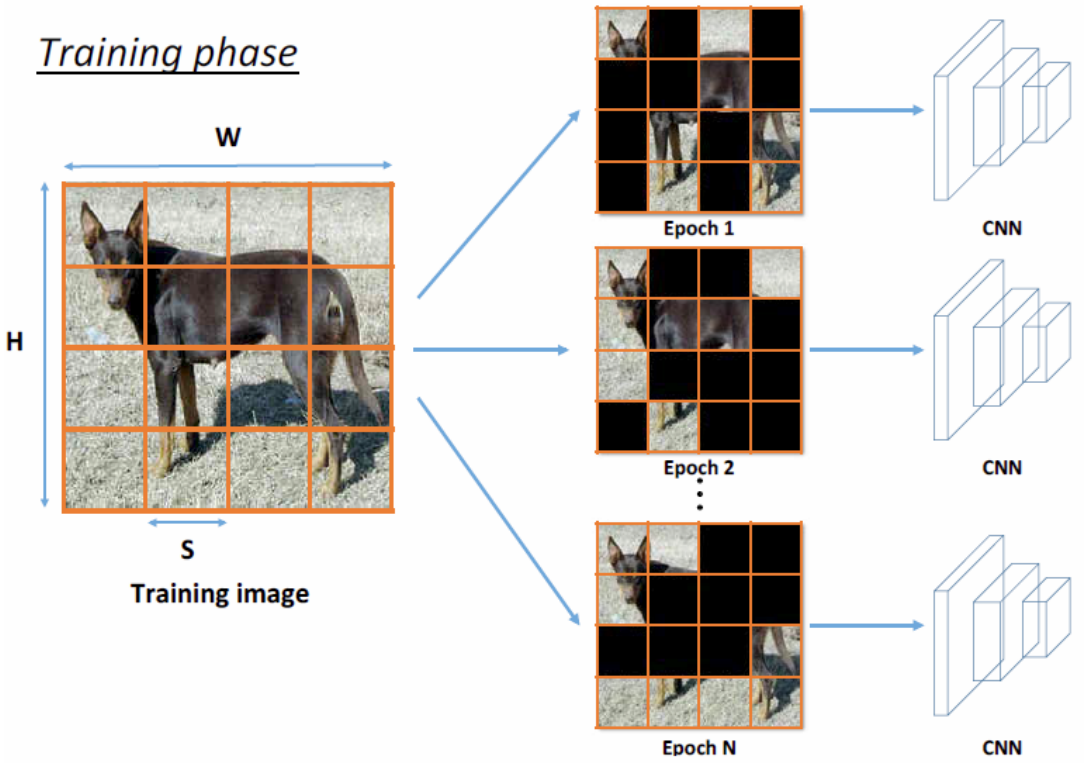
$M = zeros(n)$, where n is the number of pixels
for for each $c \in \mathbf{z}$ and each pixel m **do**
    $H(m,c) = h(A^c(m), S(m))$
end for
for for each pixel m **do**
    if $H(m) < \gamma$ **then**
        $M(m) = l_0$
    else
        $M(m) = \text{argmax}(H(m))$
    end if
end for

Output: Localization cues or approximate labeling M

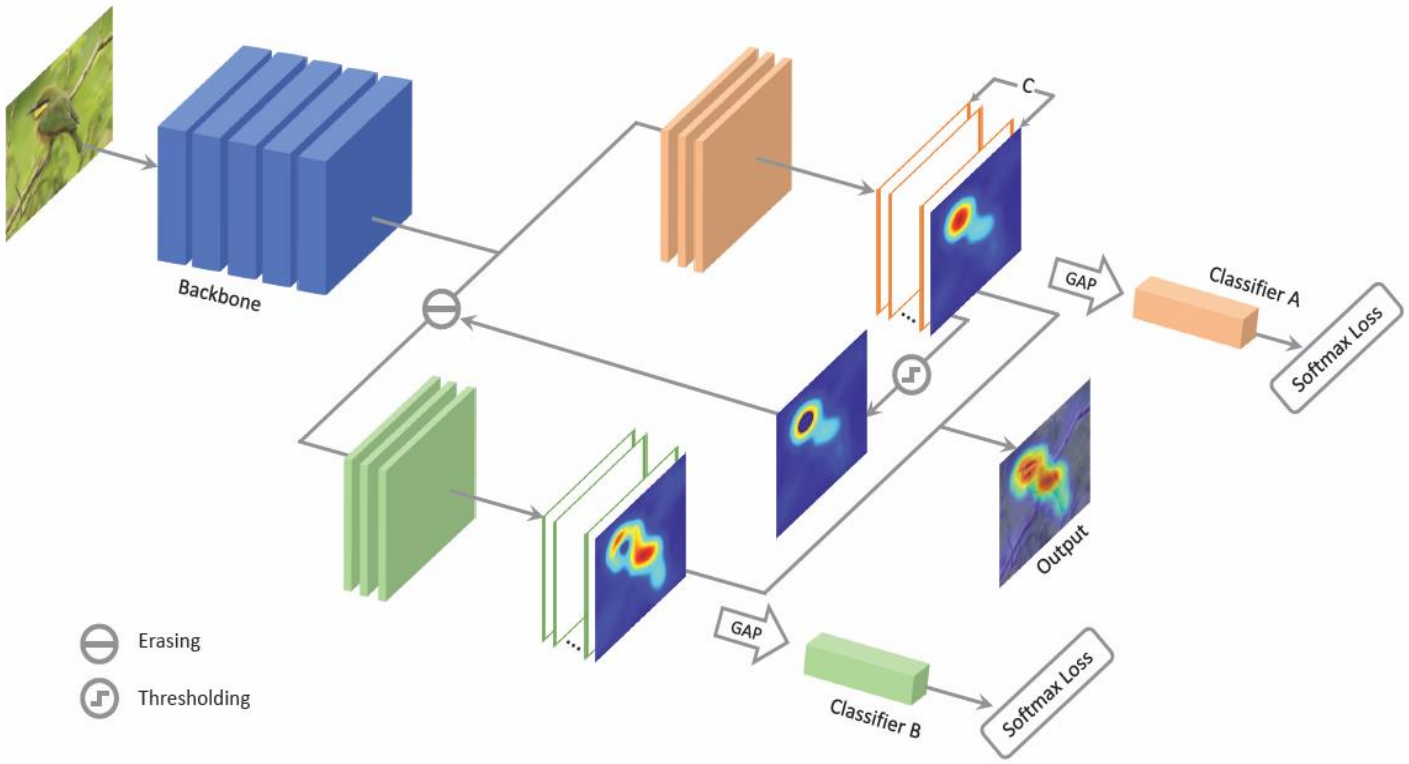
Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Hide-and-Seek



Adversarial Complementary Learning

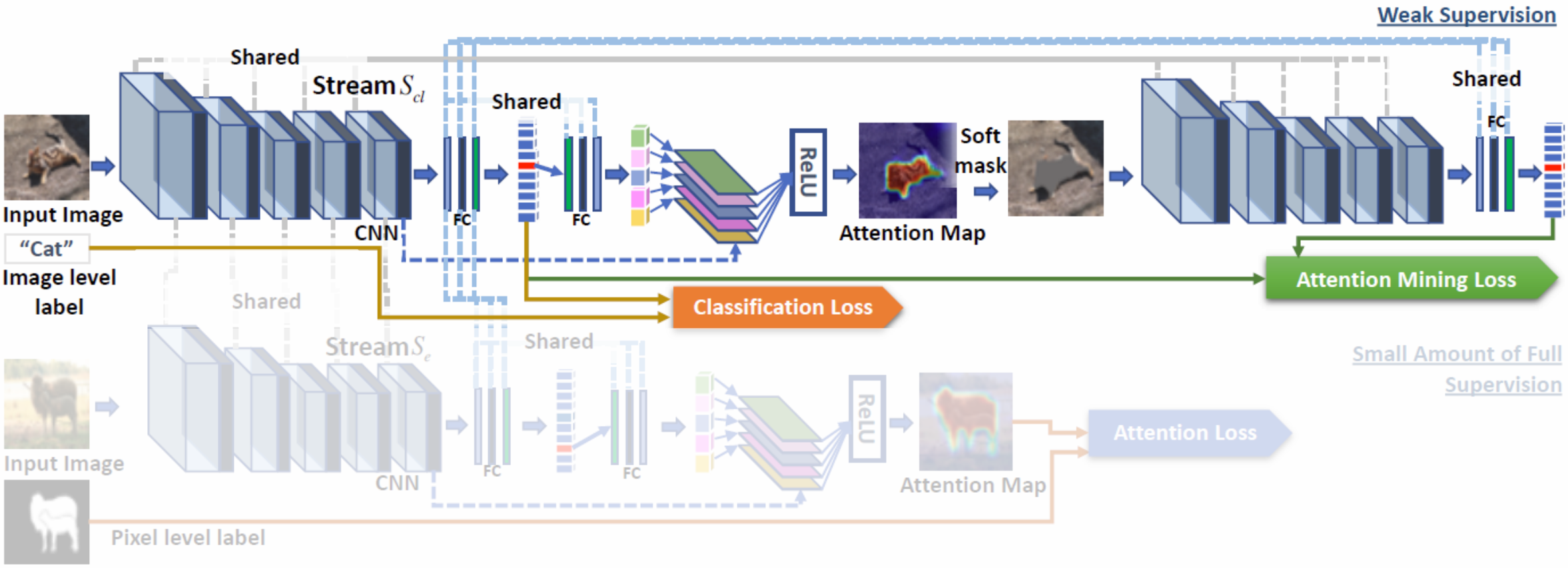


[Singh ICCV17, Zhang CVPR18]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Guided Attention Inference Network

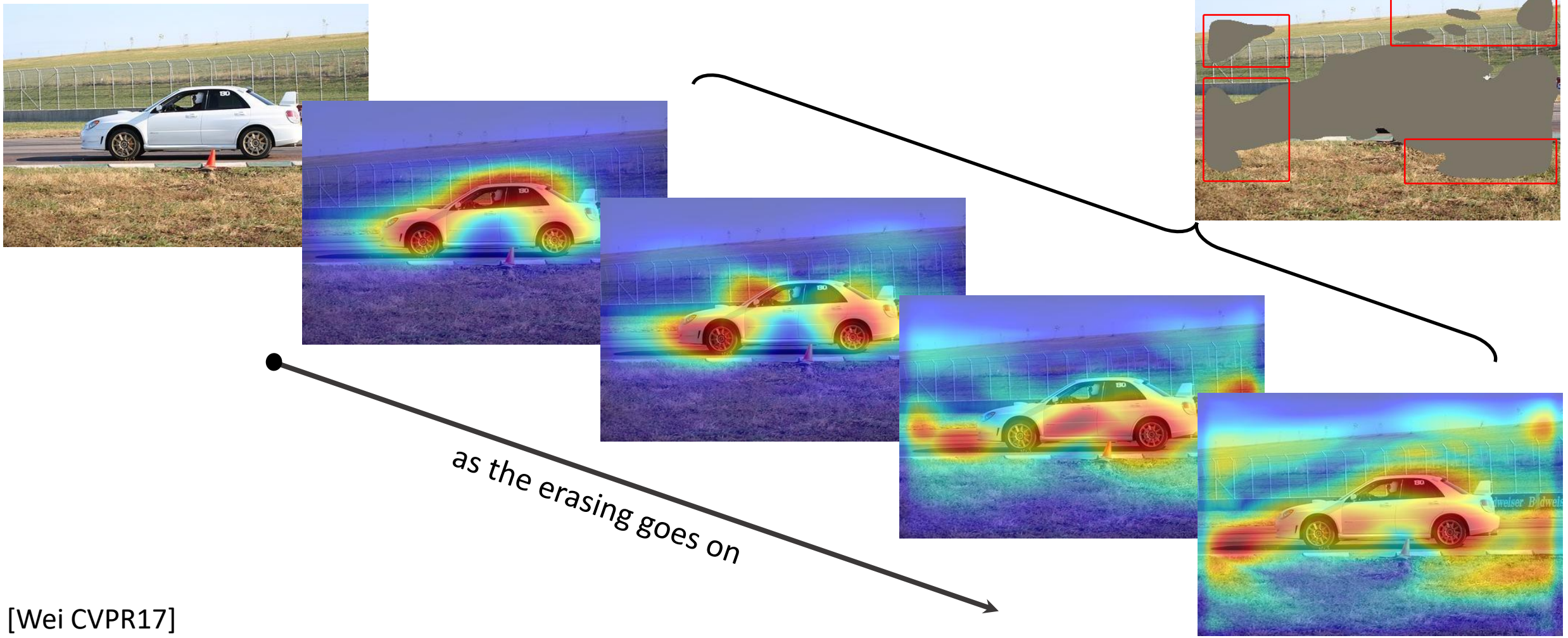


[Li CVPR18]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Over Erasing



as the erasing goes on

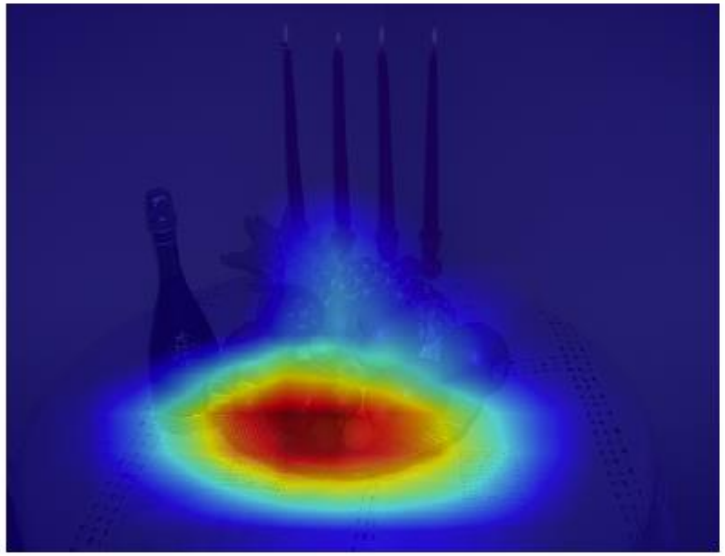
[Wei CVPR17]

Learning to Produce Pseudo Masks--Top-down Attention

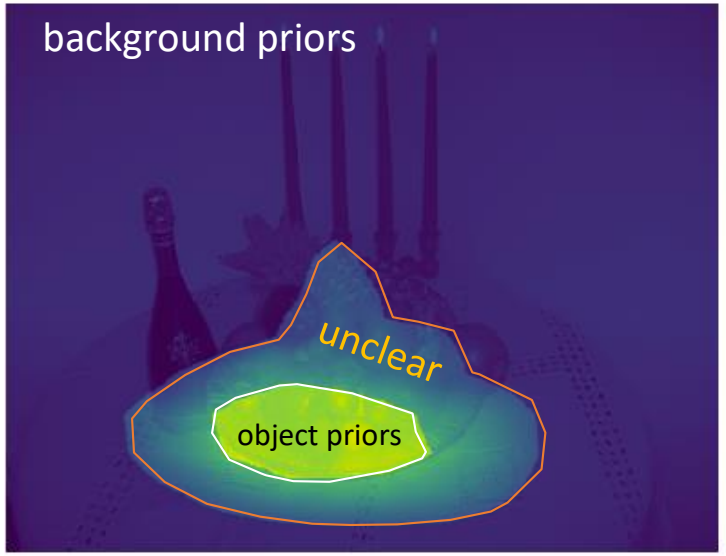
Self-Erasing Network

Image

Attention Map



Ternary Mask

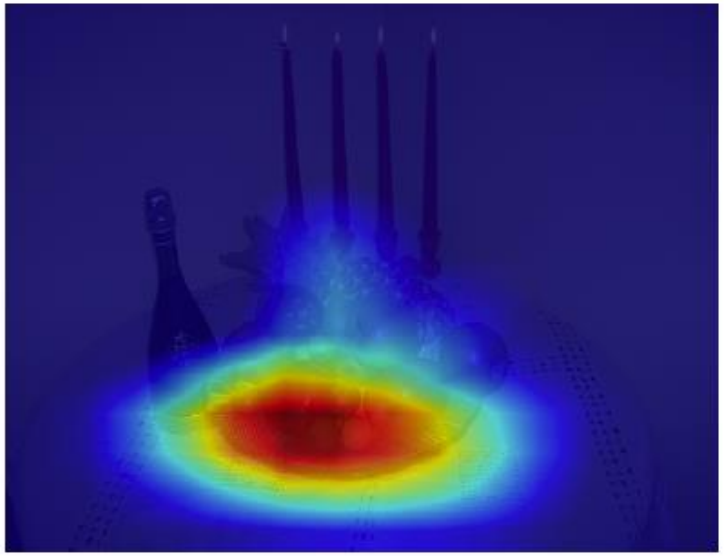


[Hou NIPS18]

Learning to Produce Pseudo Masks--Top-down Attention

Self-Erasing Network

Attention Map



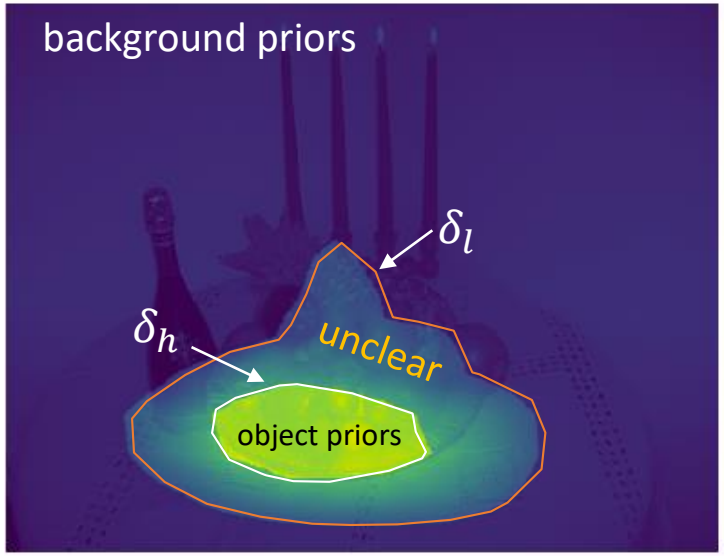
M_A

$$T_{A,(i,j)} = 0 \text{ if } M_{A,(i,j)} \geq \delta_h$$

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

$$T_{A,(i,j)} = 1 \text{ otherwise}$$

Ternary Mask



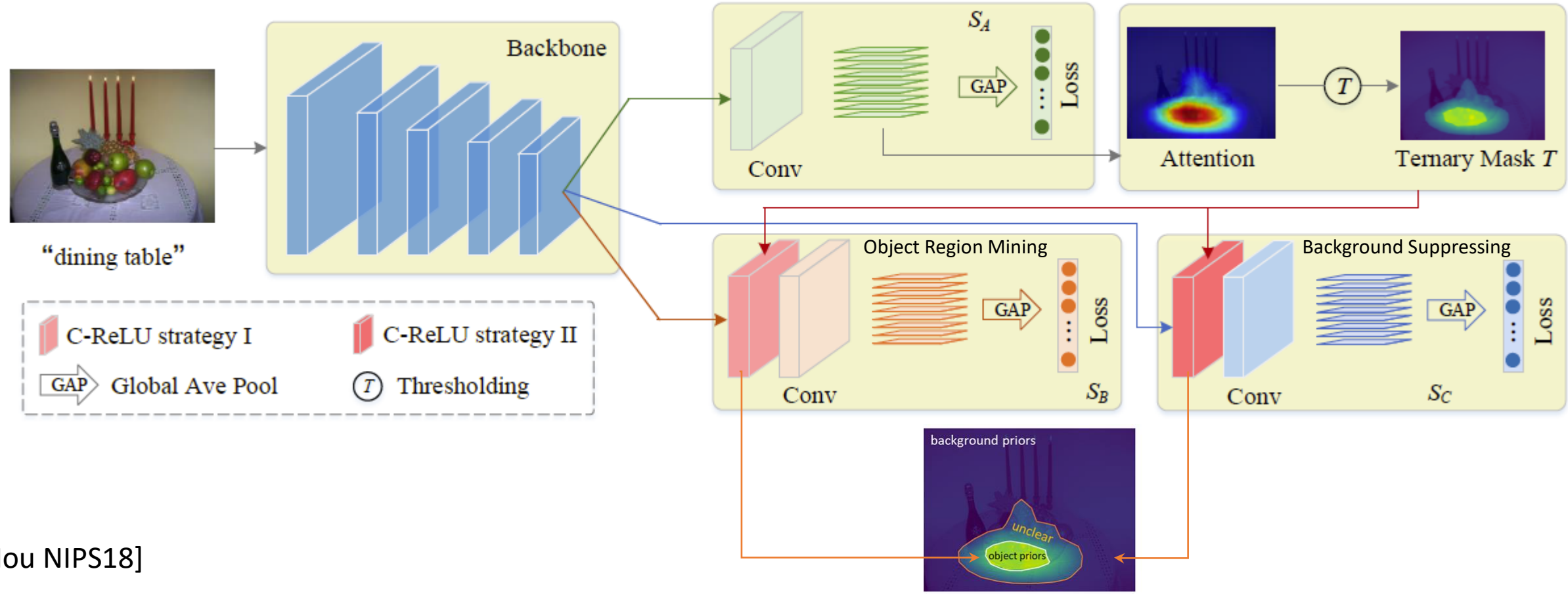
T_A

[Hou NIPS18]

Learning to Produce Pseudo Masks--Top-down Attention

Erasing for Mining

Self-Erasing Network

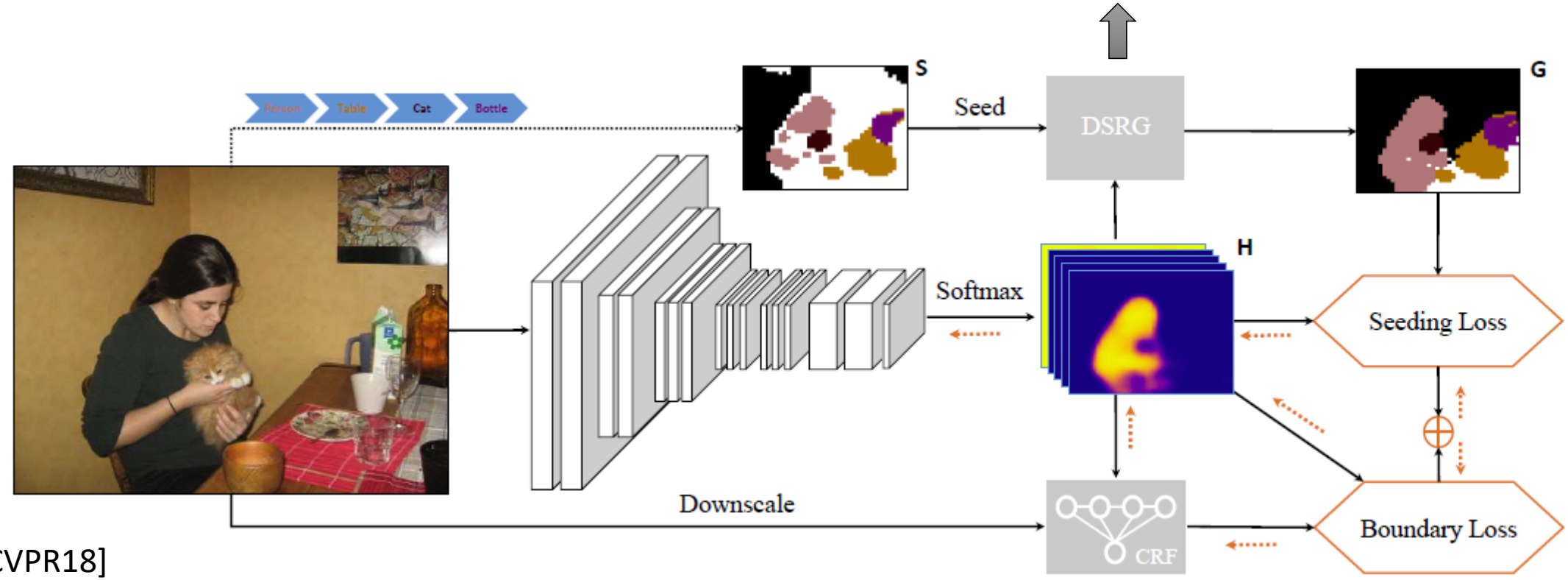


[Hou NIPS18]

Learning to Produce Pseudo Masks--Top-down Attention

Deep Seeded Region Growing

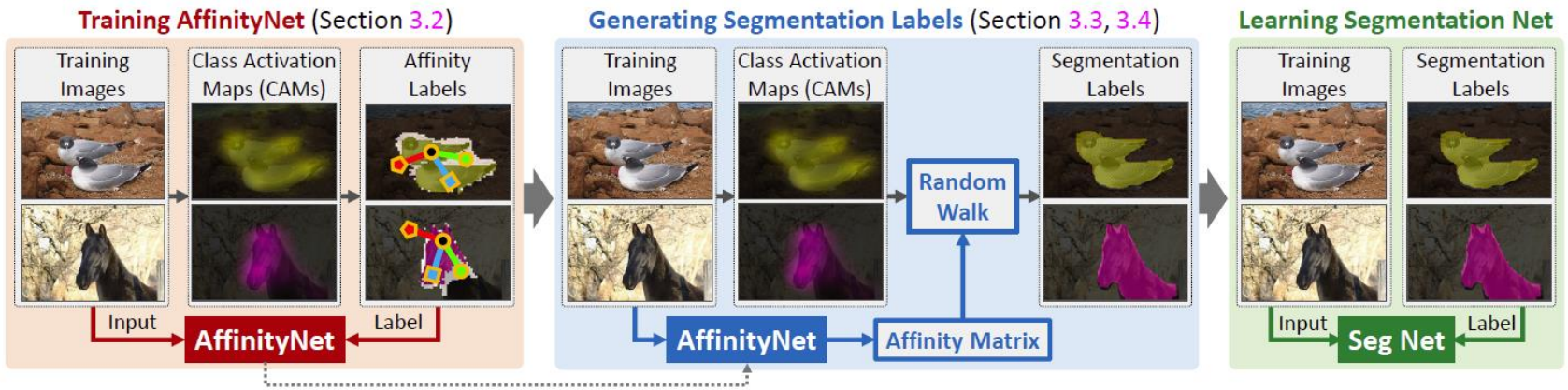
$$P(H_{u,c}, \theta_c) = \begin{cases} \text{TRUE} & H_{u,c} \geq \theta_c \text{ and} \\ & c = \arg\max_{c'} H_{u,c'}, \\ \text{FALSE} & \text{otherwise.} \end{cases}$$

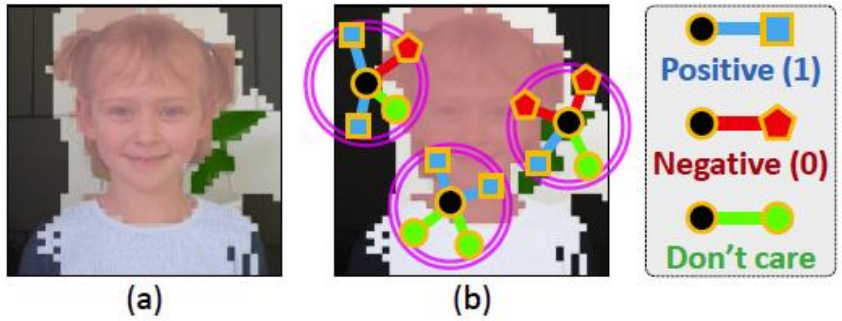


[Huang CVPR18]

Learning to Produce Pseudo Masks--Top-down Attention

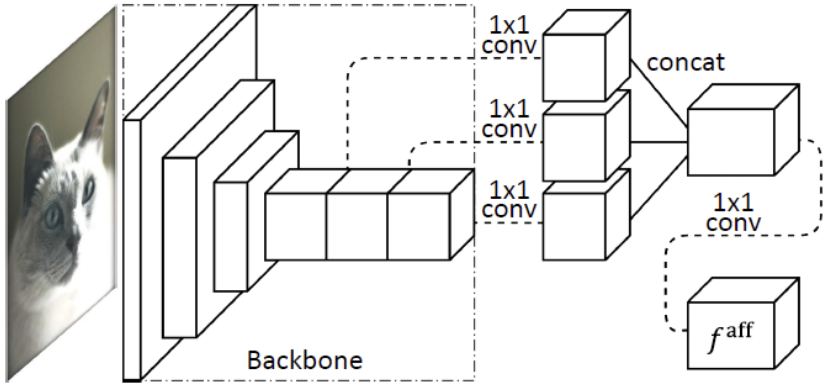
AffinityNet





Generating Semantic Affinity Labels

[Ahn CVPR18]



AffinityNet

$$W_{ij} = \exp\left\{ -\left\| f^{\text{aff}}(x_i, y_i) - f^{\text{aff}}(x_j, y_j) \right\|_1 \right\}$$

$$\mathcal{P} = \left\{ (i,j) \mid \mathrm{d}\big((x_i, y_i), (x_j, y_j)\big) < \gamma, \forall i \neq j \right\}$$

$$\mathcal{P}^+ = \left\{ (i,j) \mid (i,j) \in \mathcal{P}, W^*_{ij} = 1 \right\}$$

$$\mathcal{P}^- = \left\{ (i,j) \mid (i,j) \in \mathcal{P}, W^*_{ij} = 0 \right\}$$

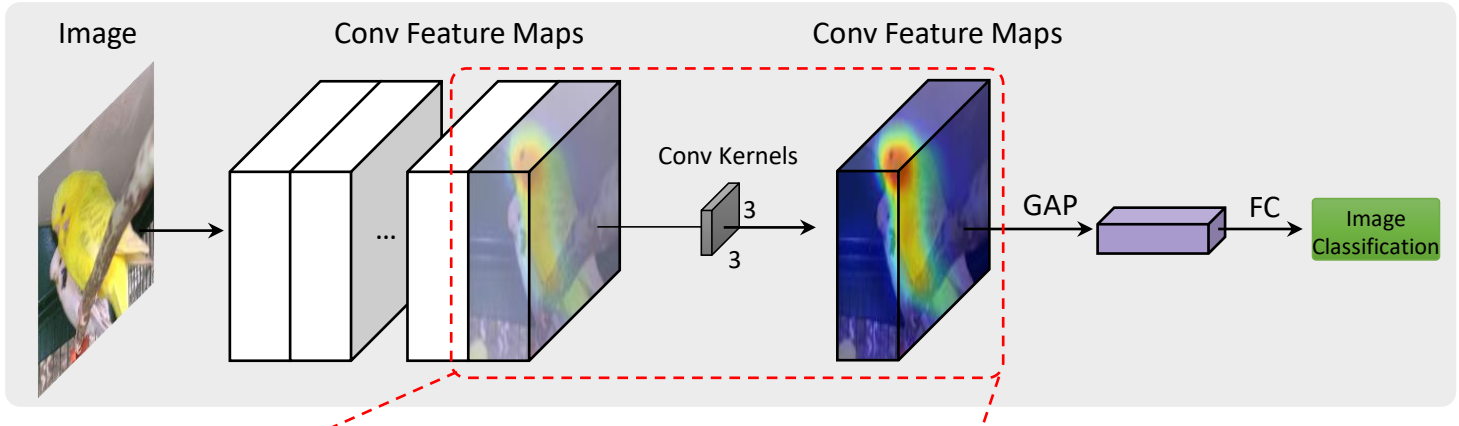
$$\mathcal{L}^+_{\text{fg}} = -\frac{1}{|\mathcal{P}^+_{\text{fg}}|} \sum_{(i,j) \in \mathcal{P}^+_{\text{fg}}} \log W_{ij},$$

$$\mathcal{L}^+_{\text{bg}} = -\frac{1}{|\mathcal{P}^+_{\text{bg}}|} \sum_{(i,j) \in \mathcal{P}^+_{\text{bg}}} \log W_{ij},$$

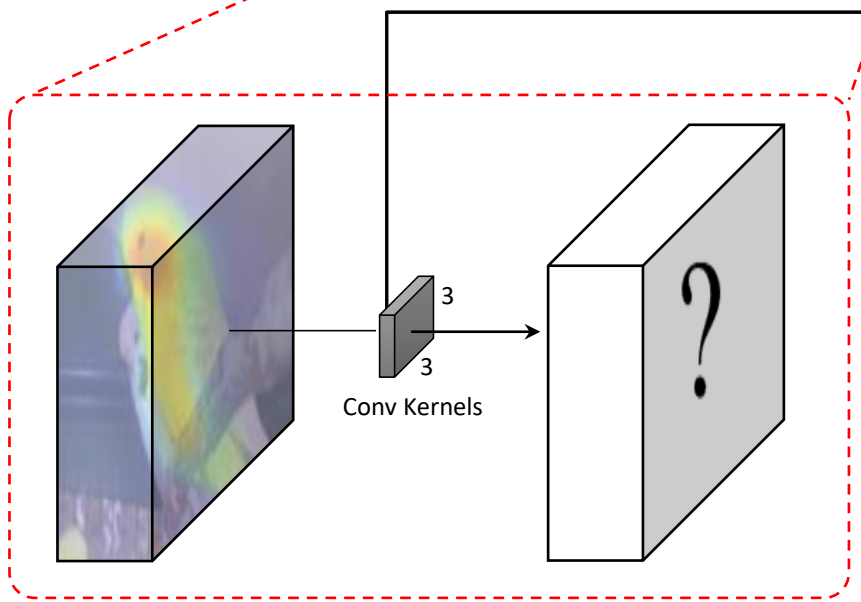
$$\mathcal{L}^- = -\frac{1}{|\mathcal{P}^-|} \sum_{(i,j) \in \mathcal{P}^-} \log(1 - W_{ij}).$$

$$\mathcal{L} = \mathcal{L}^+_{\text{fg}} + \mathcal{L}^+_{\text{bg}} + 2\mathcal{L}^-$$

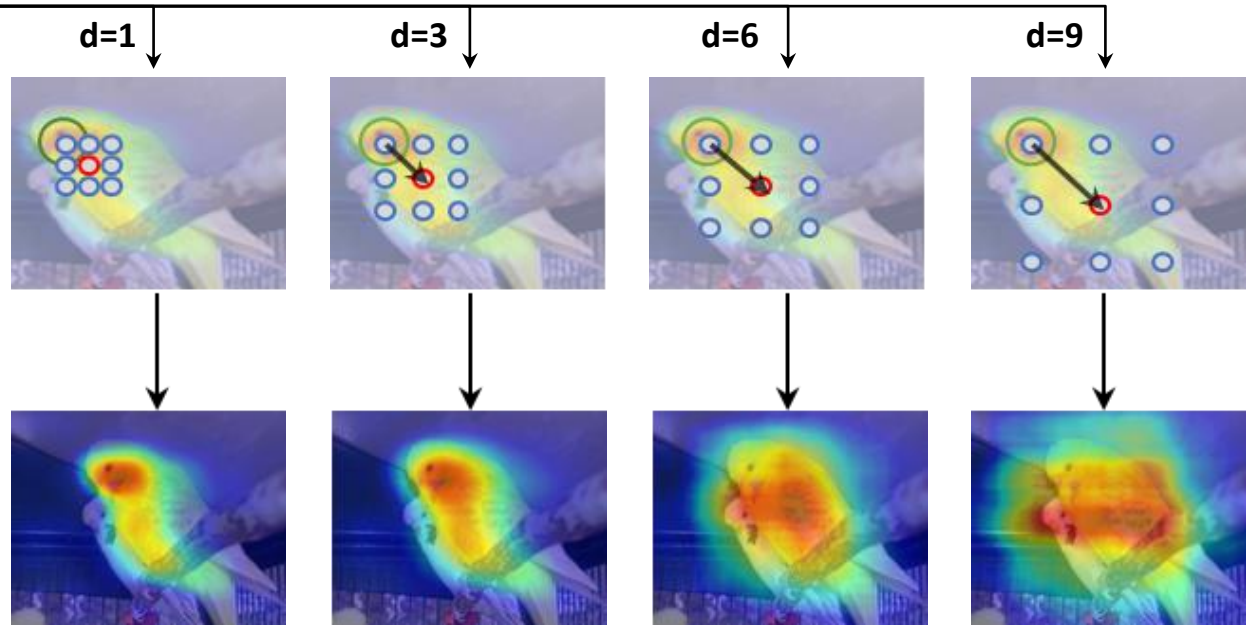
Learning to Produce Pseudo Masks--Top-down Attention



Multi-dilated Convolution



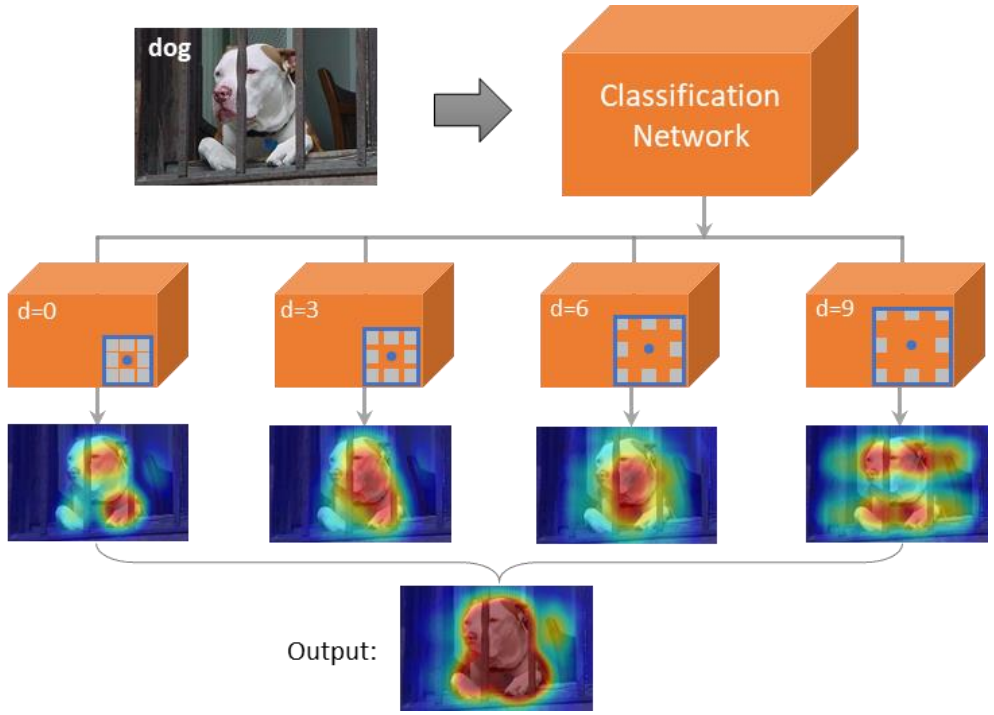
Dilated convolution with different rates

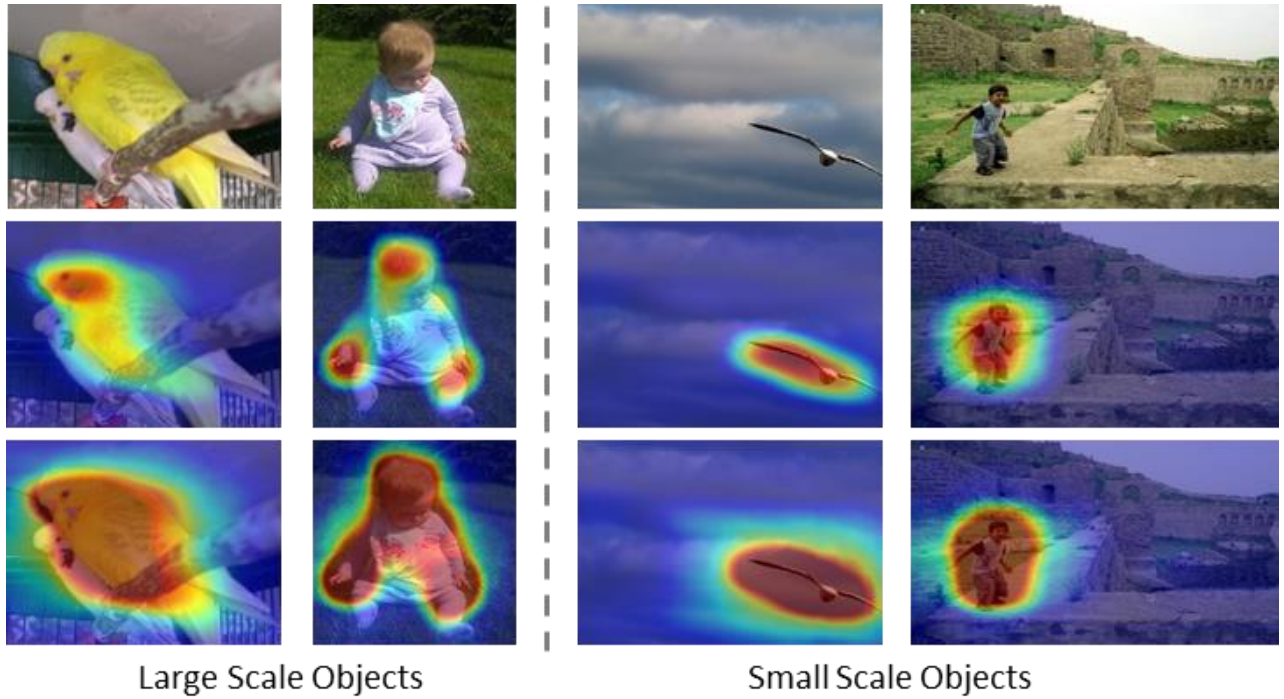


[Wei CVPR18]

Learning to Produce Pseudo Masks--Top-down Attention

Multi-dilated Convolution





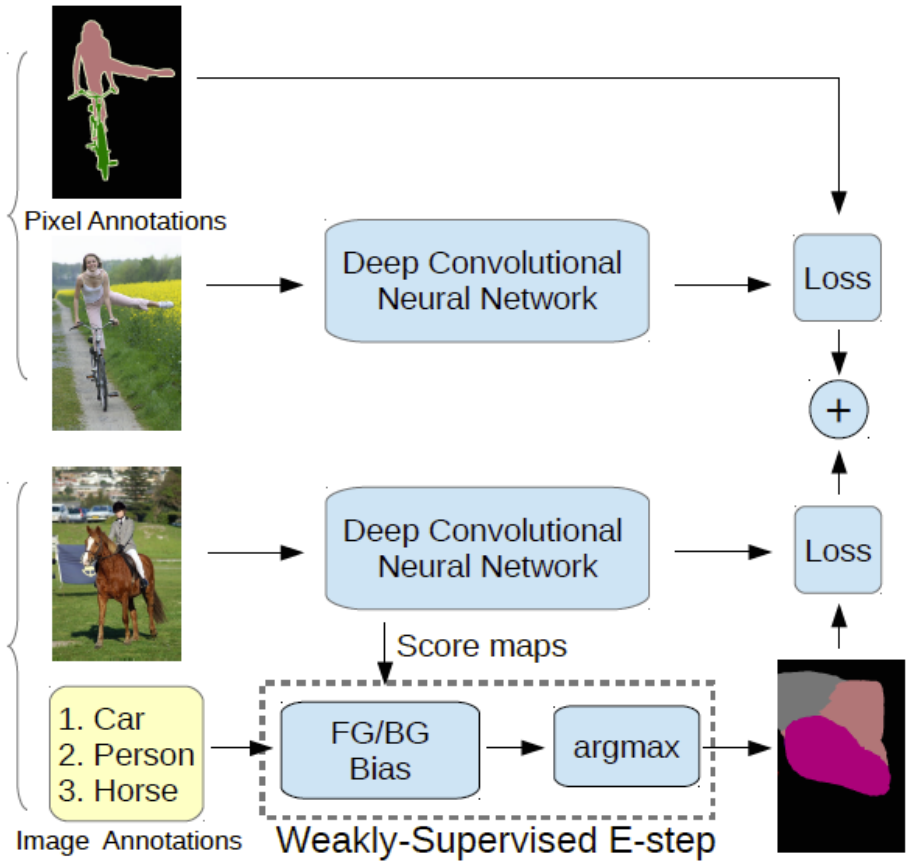
[Wei CVPR18]

- 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

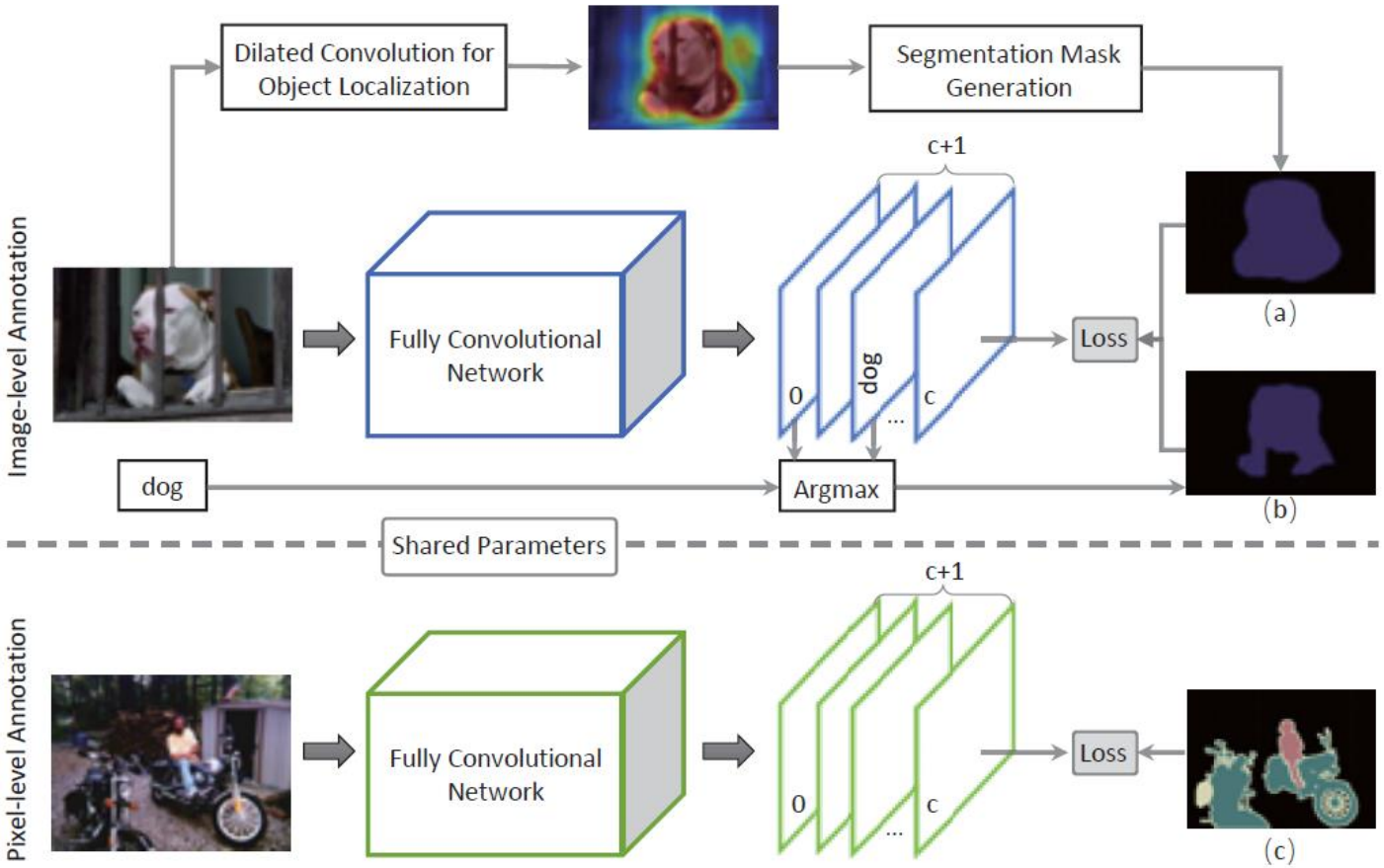
image-level labels

Semi-Supervised Learning

EM Adapt/Fix



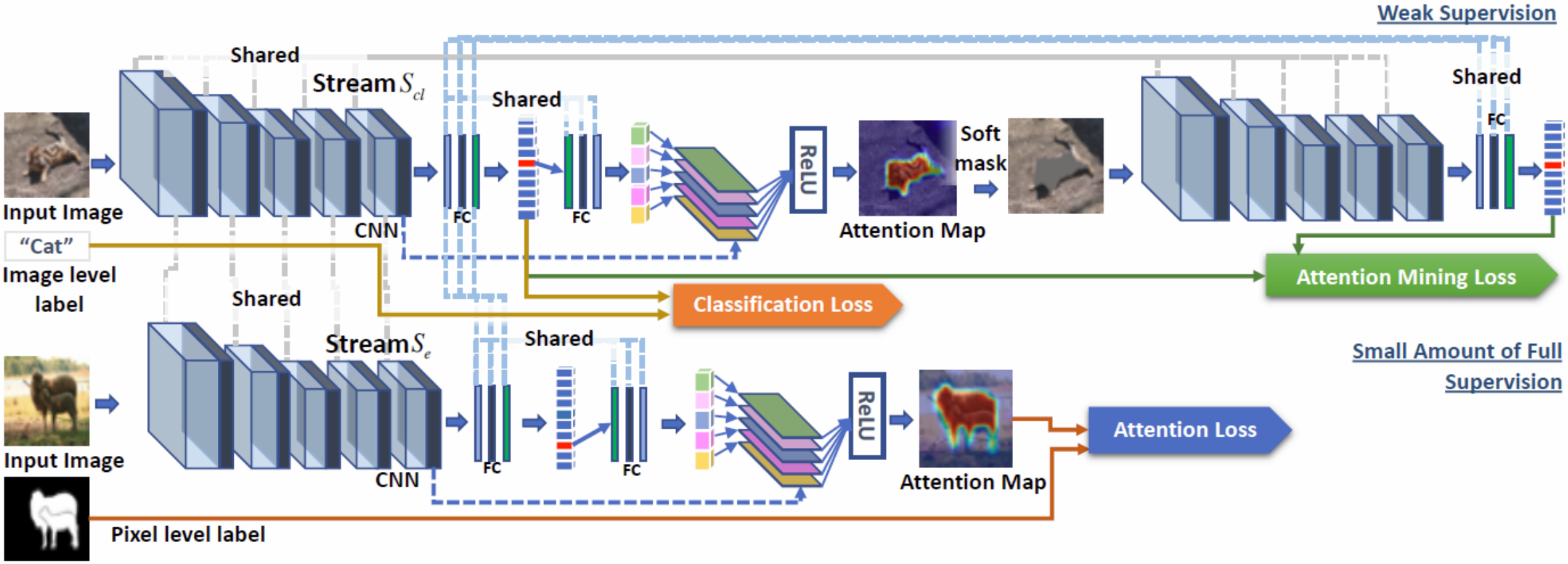
Multi-dilated Convolution



[Chen ICCV15, Wei CVPR18]

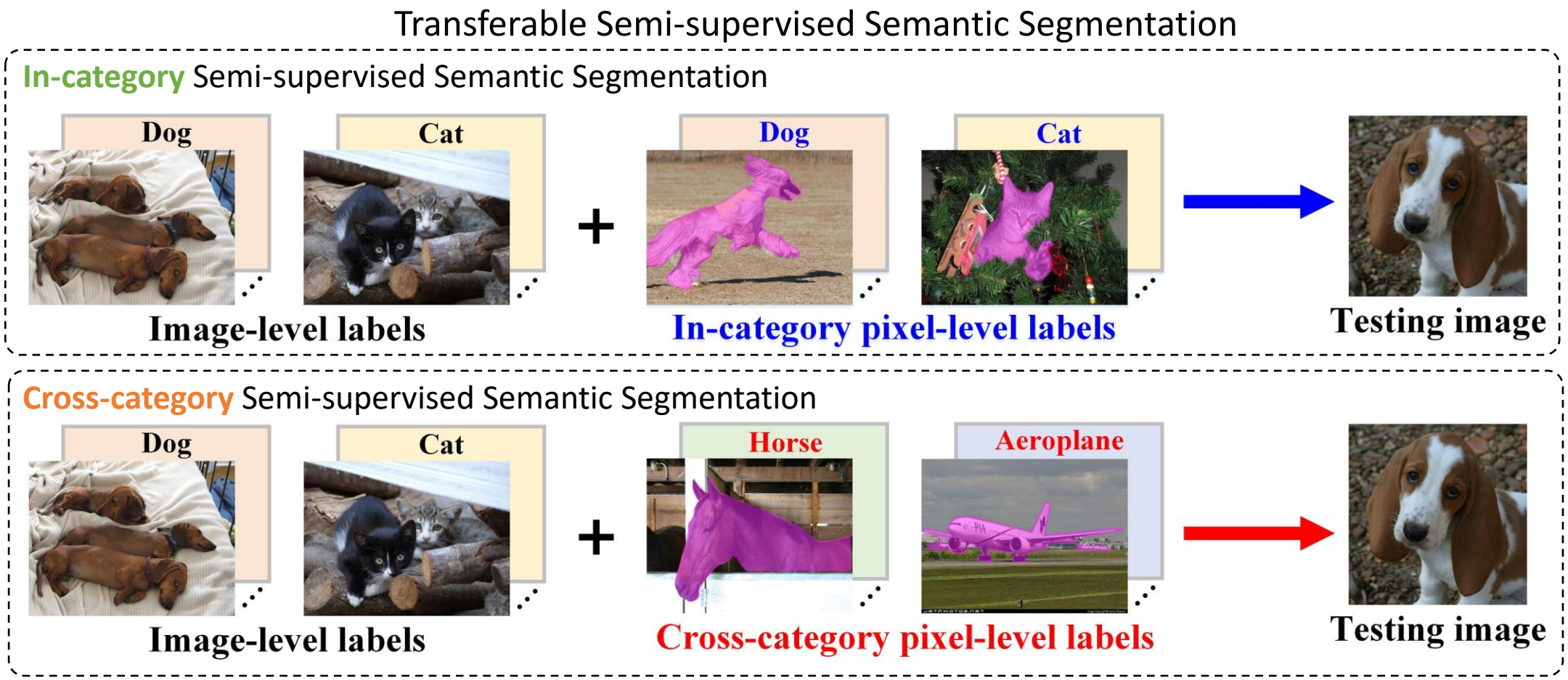
Semi-Supervised Learning

Guided Attention Inference Network



[Li CVPR18]

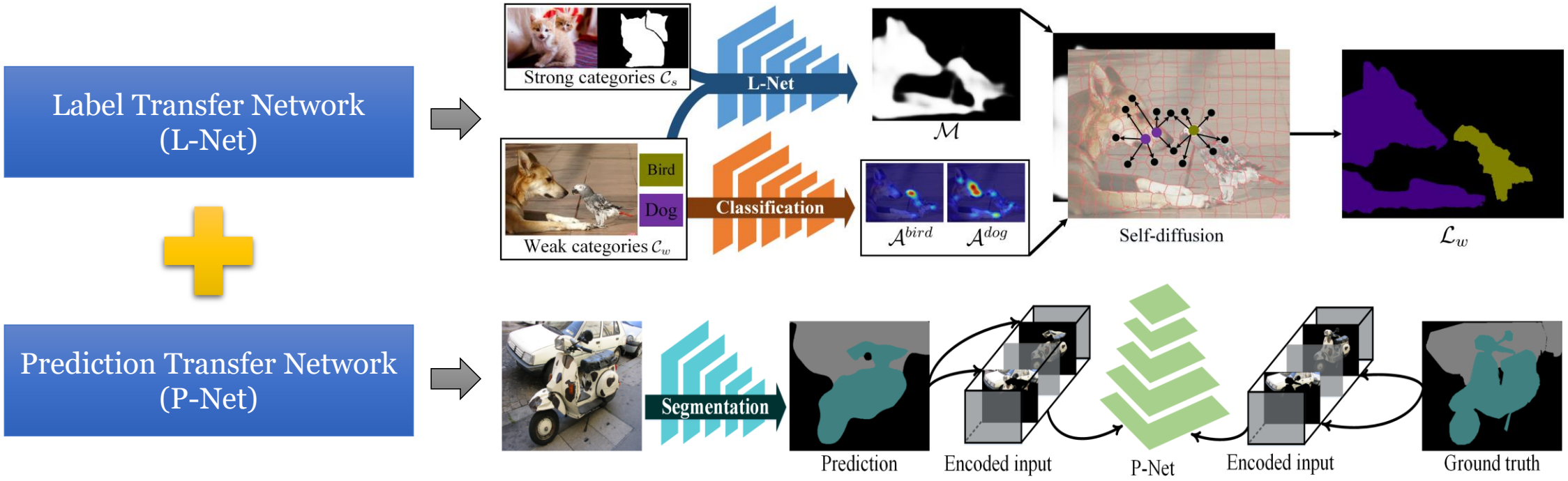
Semi-Supervised Learning



[Xiao AAAI18]

Semi-Supervised Learning

Transferable Semi-supervised Semantic Segmentation



[Xiao AAAI18]

Semi-Supervised Learning

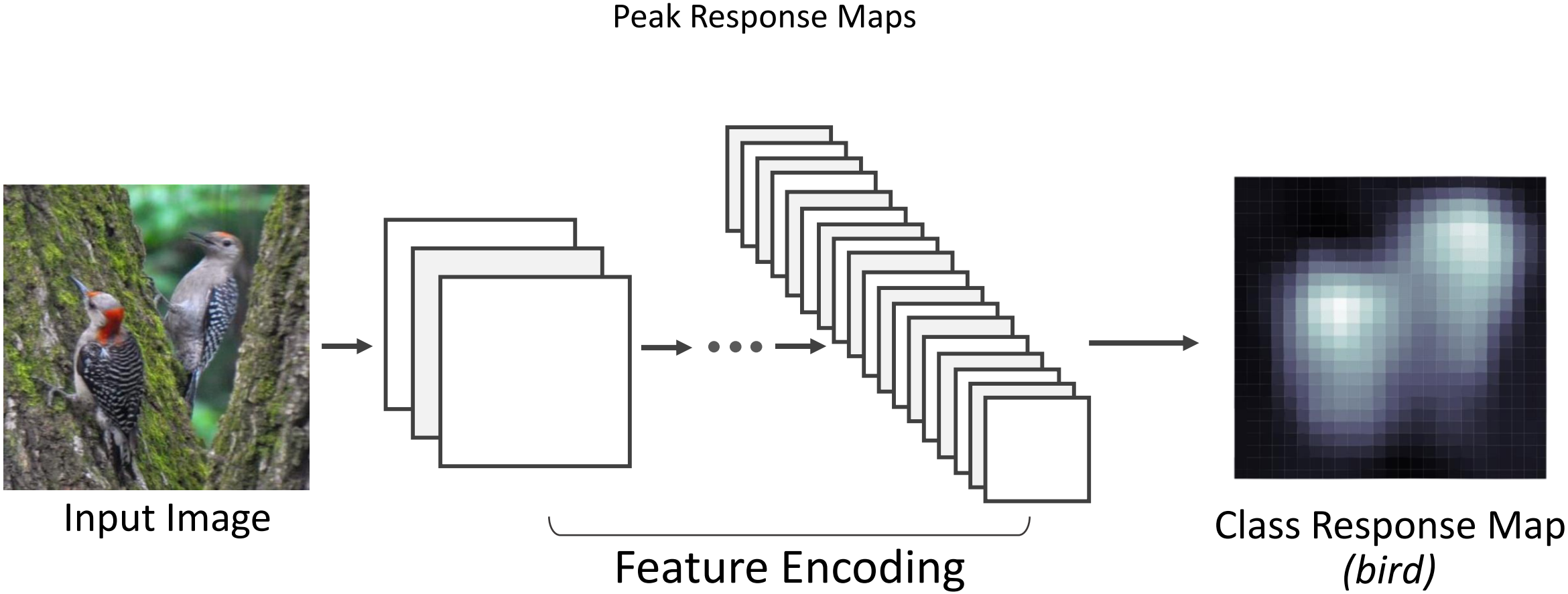
Transferable Semi-supervised Semantic Segmentation

[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

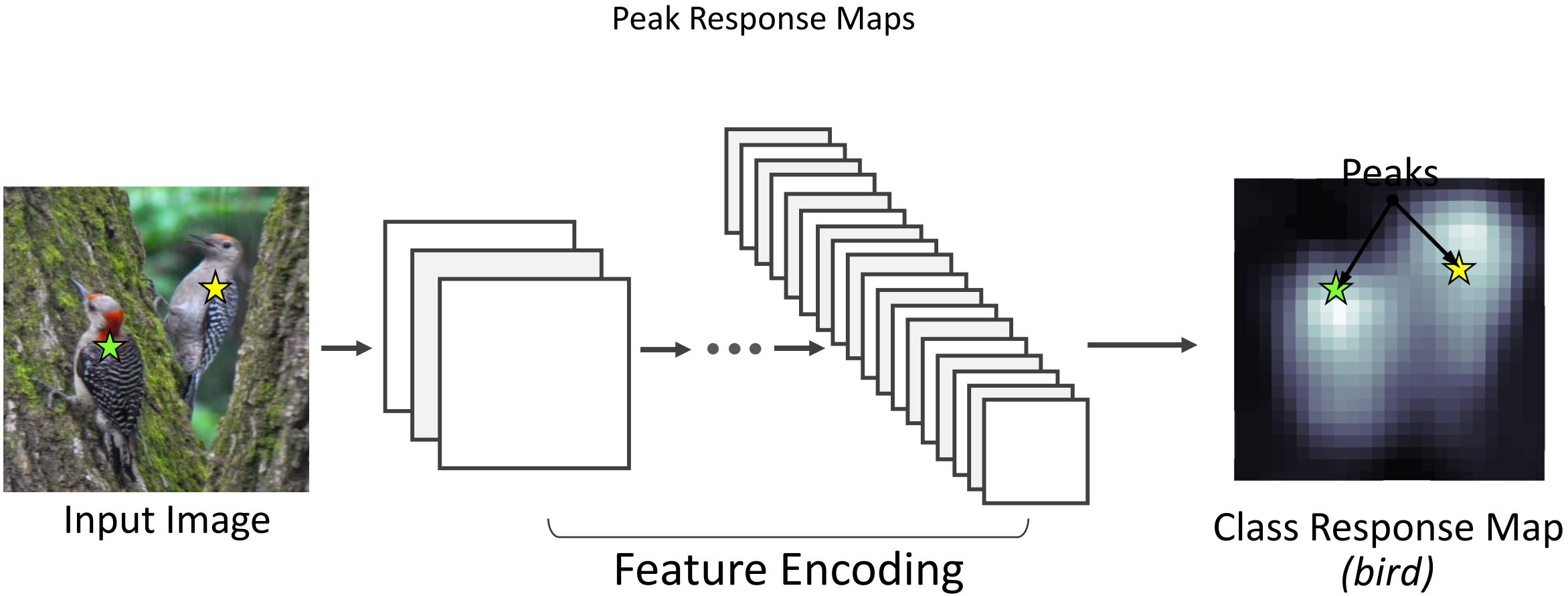
image-level labels

Instance Segmentation



[Zhou CVPR18]

Instance Segmentation

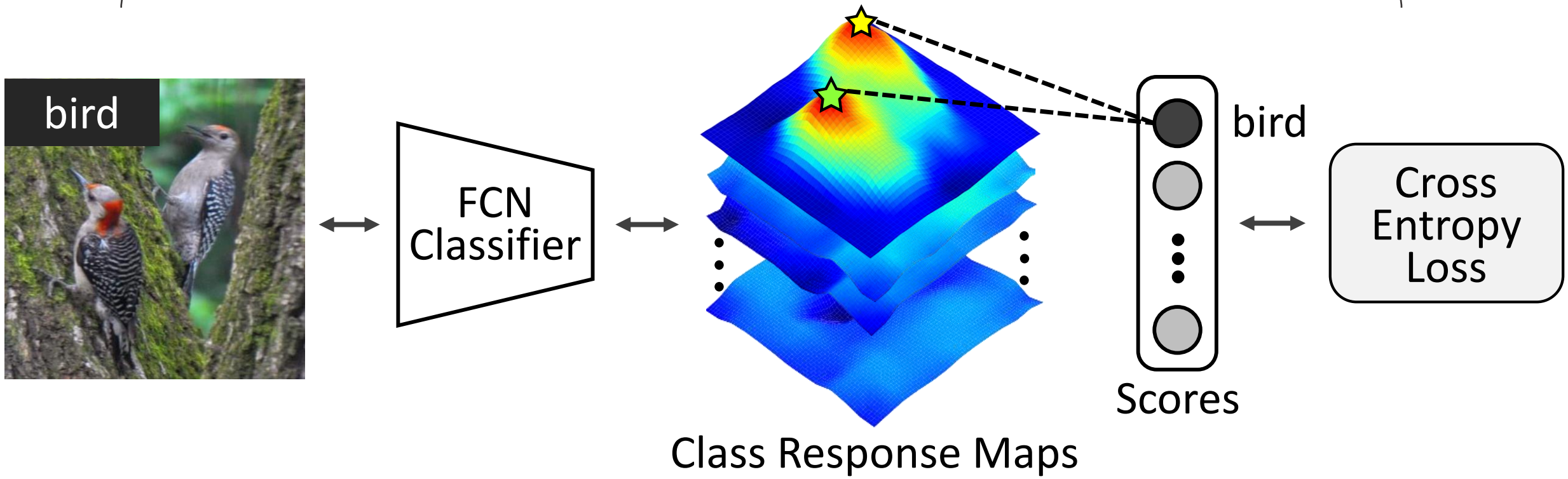


[Zhou CVPR18]

Instance Segmentation

Peak Response Maps

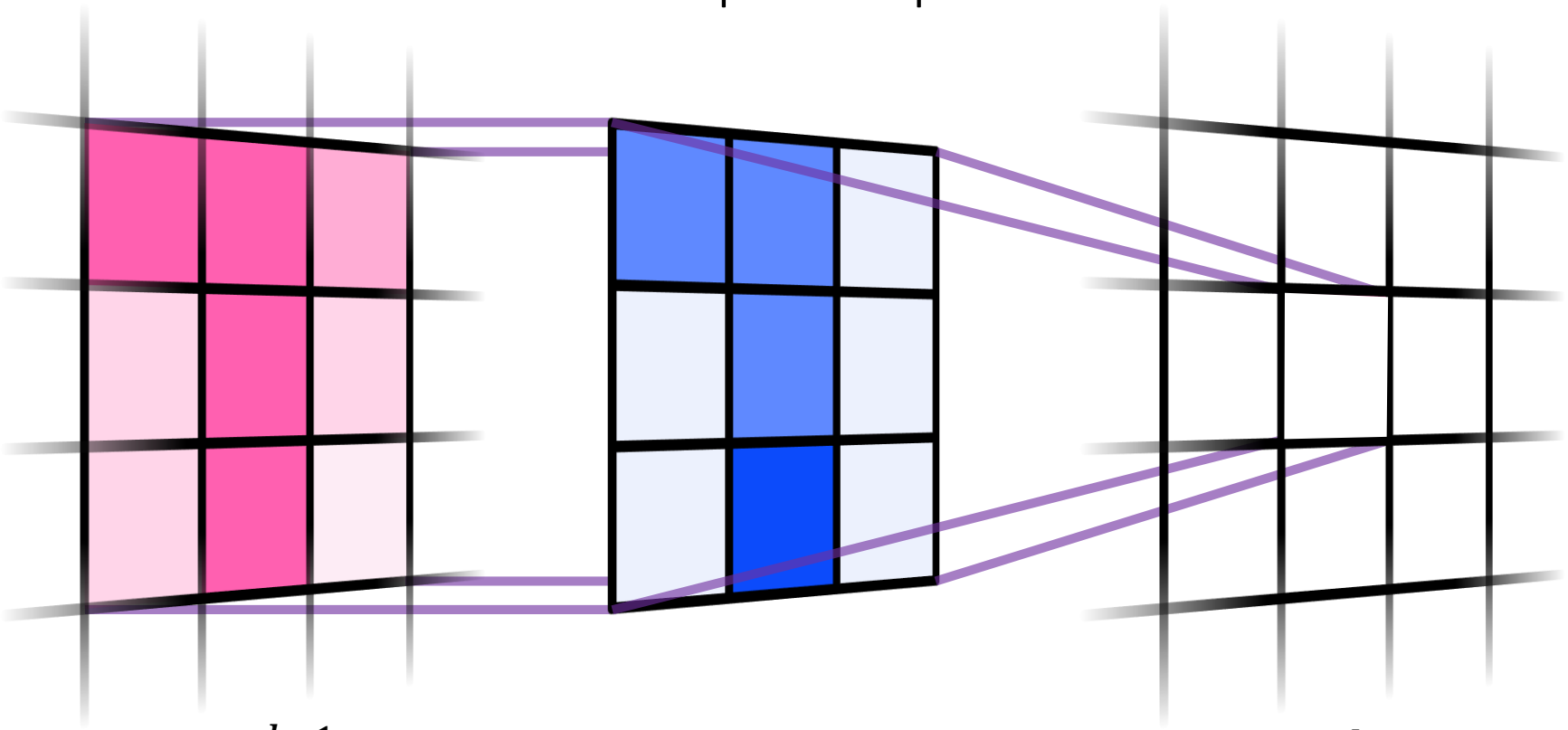
End-to-End training with standard classification settings



[Zhou CVPR18]

Instance Segmentation

Peak Response Maps



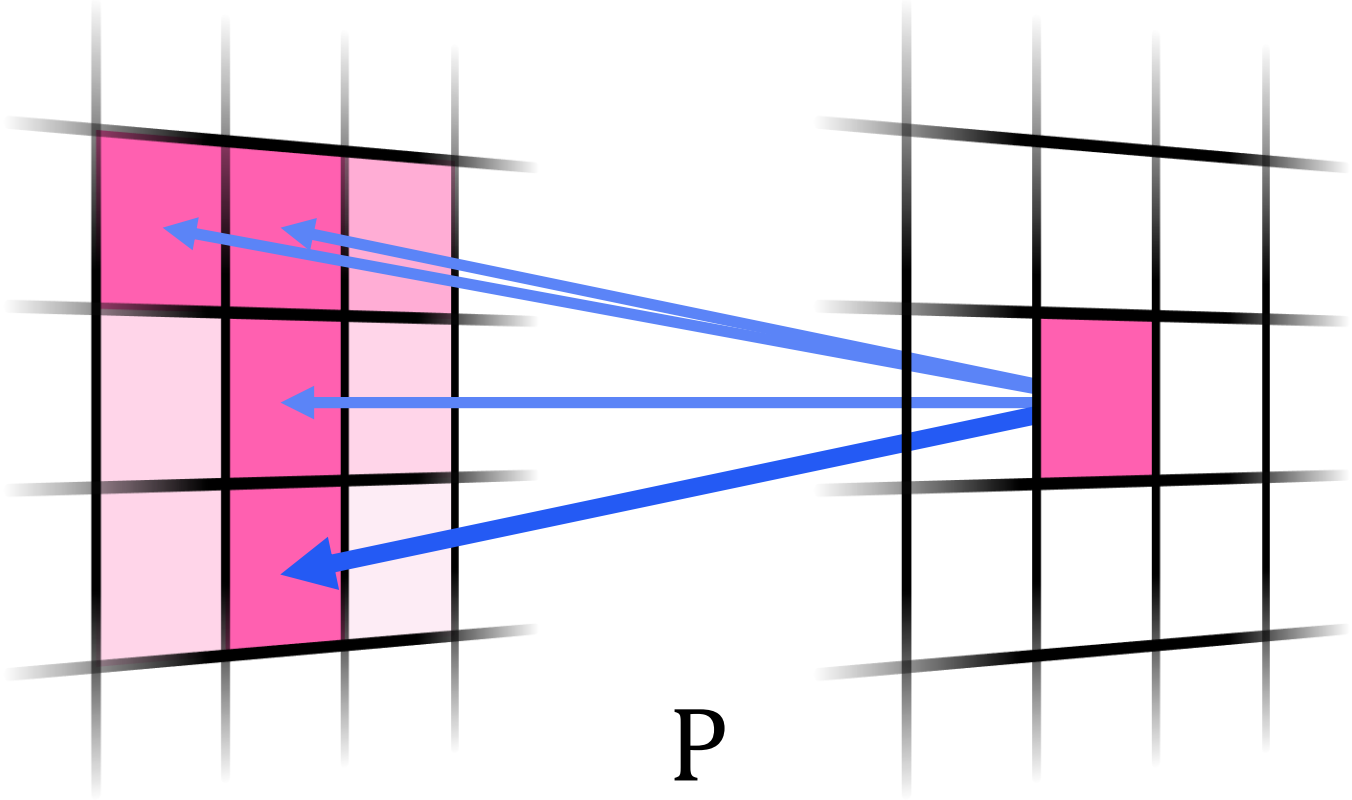
$f(X^{l-1})$ feature map $*$ w conv filter $=$ X^{l} feature map

Model Top-down relevance between locations

[Zhou CVPR18]

Instance Segmentation

Peak Response Maps



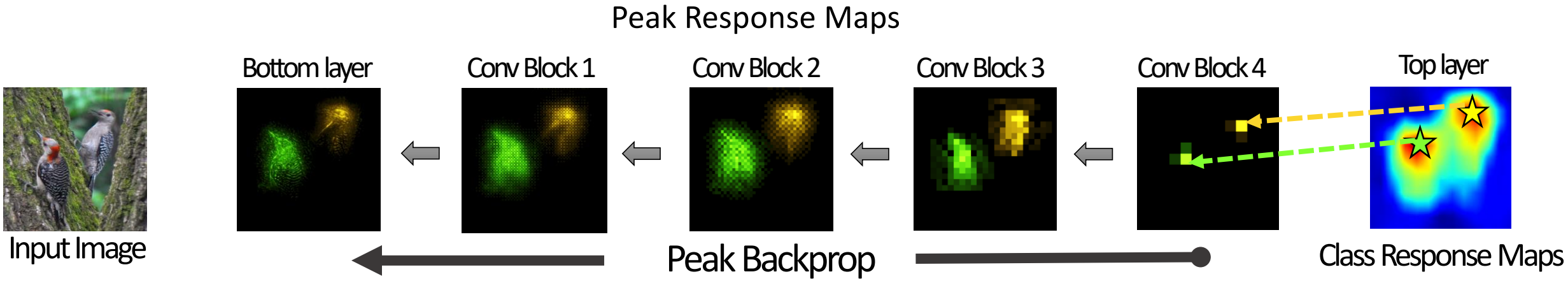
P

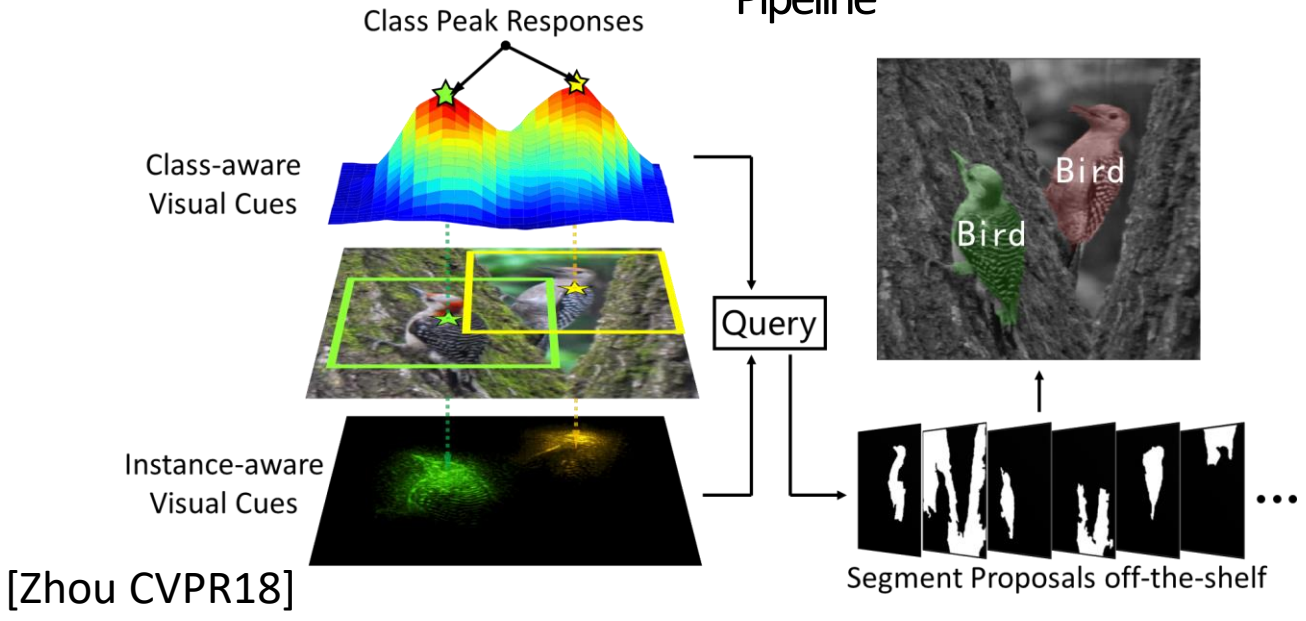
transition probability

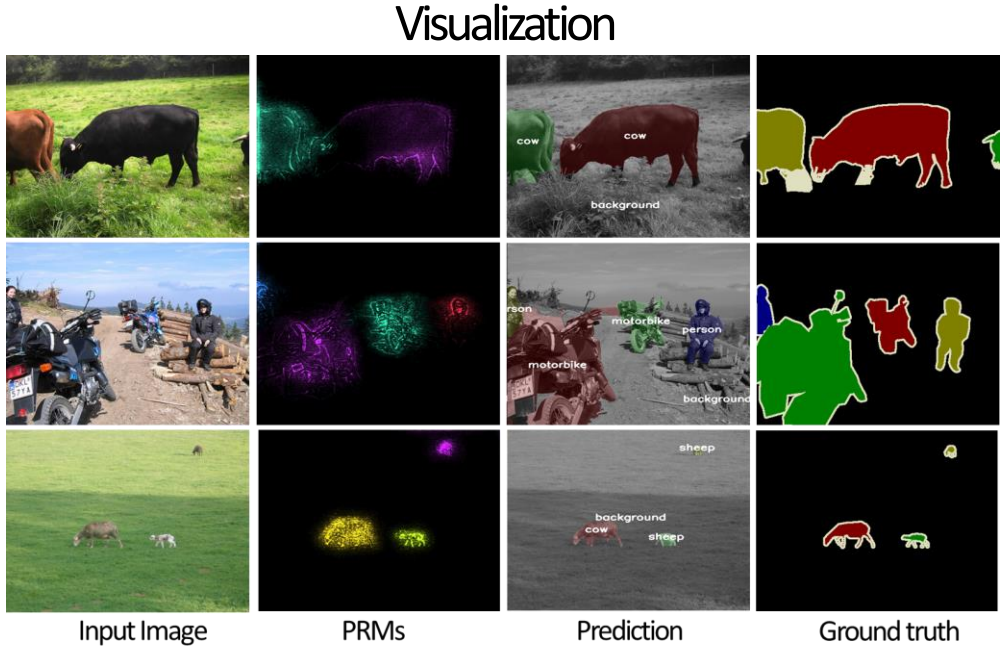
Model Top-down relevance between locations

[Zhou CVPR18]

Instance Segmentation







Outline

image-level labels

points

bounding boxes

scribbles

- PointSup: Object Semantic Segmentation
- PointSup: Scene Parsing

points

- PointSup: Object Semantic Segmentation

- PointSup: Scene Parsing

points

PointSup: Object Semantic Segmentation

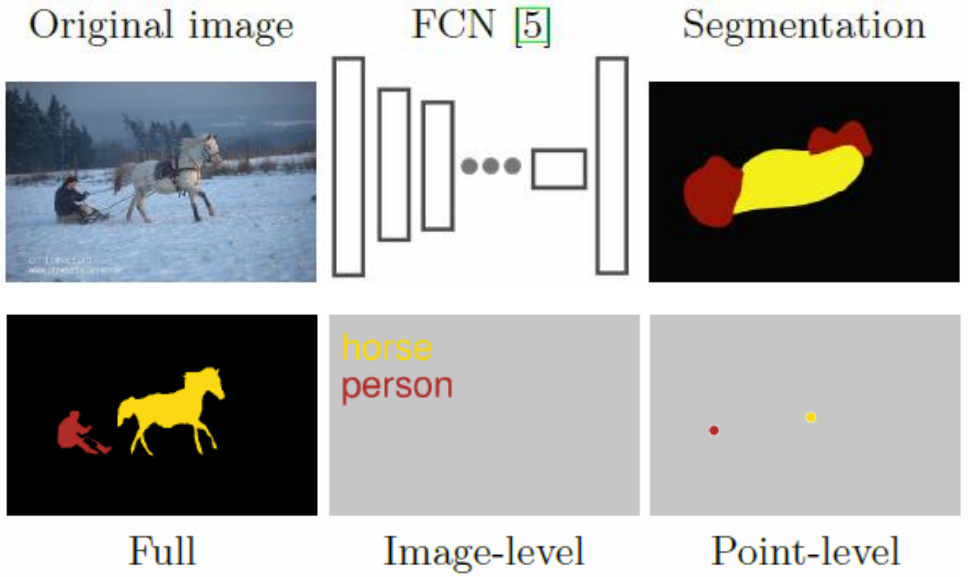
What's the Point

Image-Level Supervision:

$$\mathcal{L}_{img}(S, L, L') = -\frac{1}{|L|}\sum_{c \in L}\log(S_{t_c c}) - \frac{1}{|L'|}\sum_{c \in L'}\log(1 - S_{t_c c})$$

Point-Level Supervision:

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



[Bearman ECCV16]

- PointSup: Object Semantic Segmentation

- PointSup: Scene Parsing

points

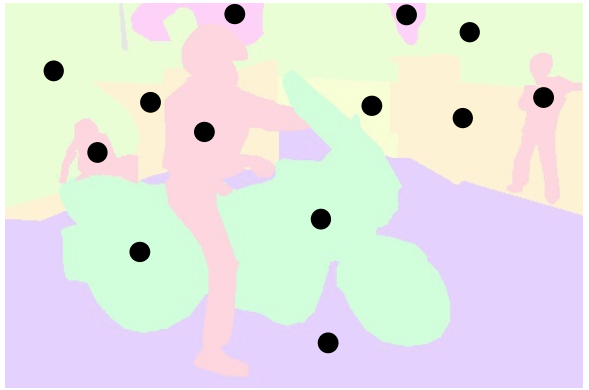
PointSup: Scene Parsing

Point-based Distance Metric Learning

Image

FullSup

PointSup



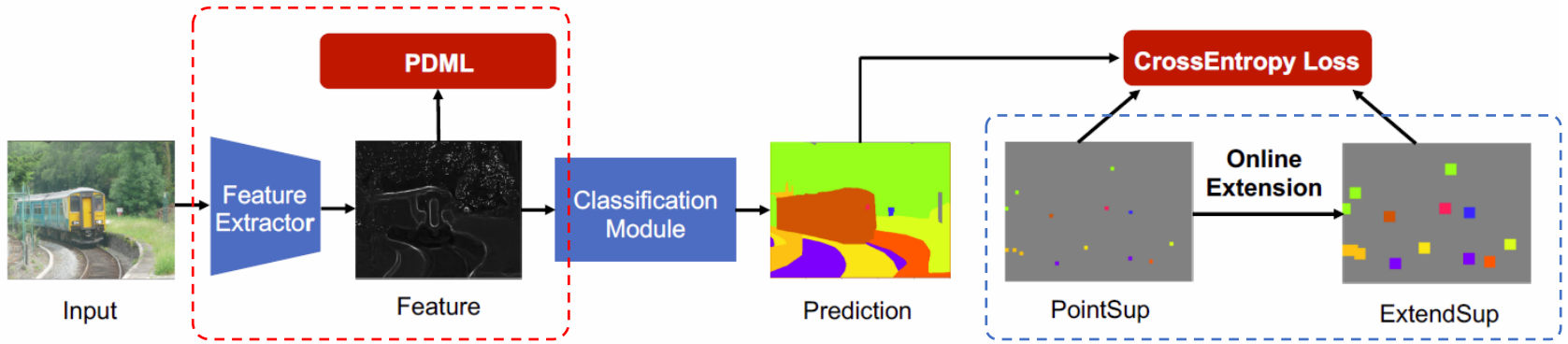
The Number of Annotated Pixels

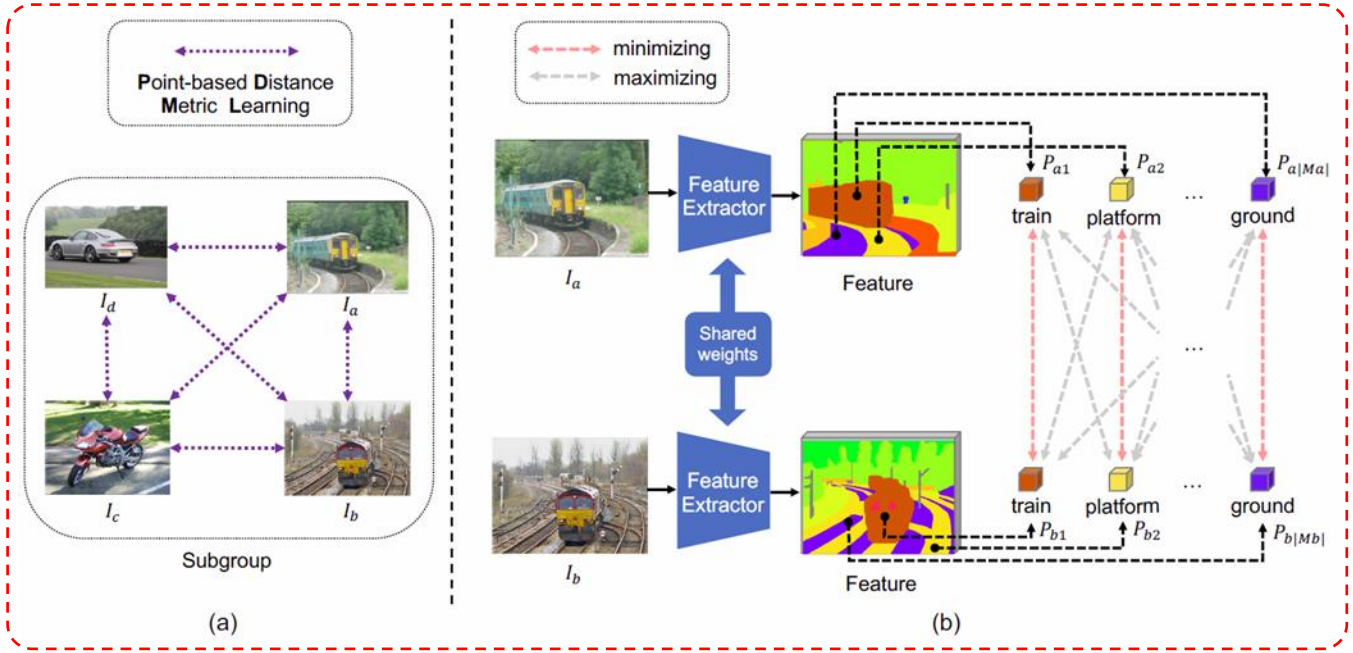
170K → **12.26**

[Qian AAAI19]

PointSup: Scene Parsing

PDML





[Qian AAAI19]

Confident Constraint

$$M_{a\_score} = \bigcup_{u} \{\max(g_{u,c\in C}(f(I_a;\theta_{f1});\theta_{g1})) > thr\}$$

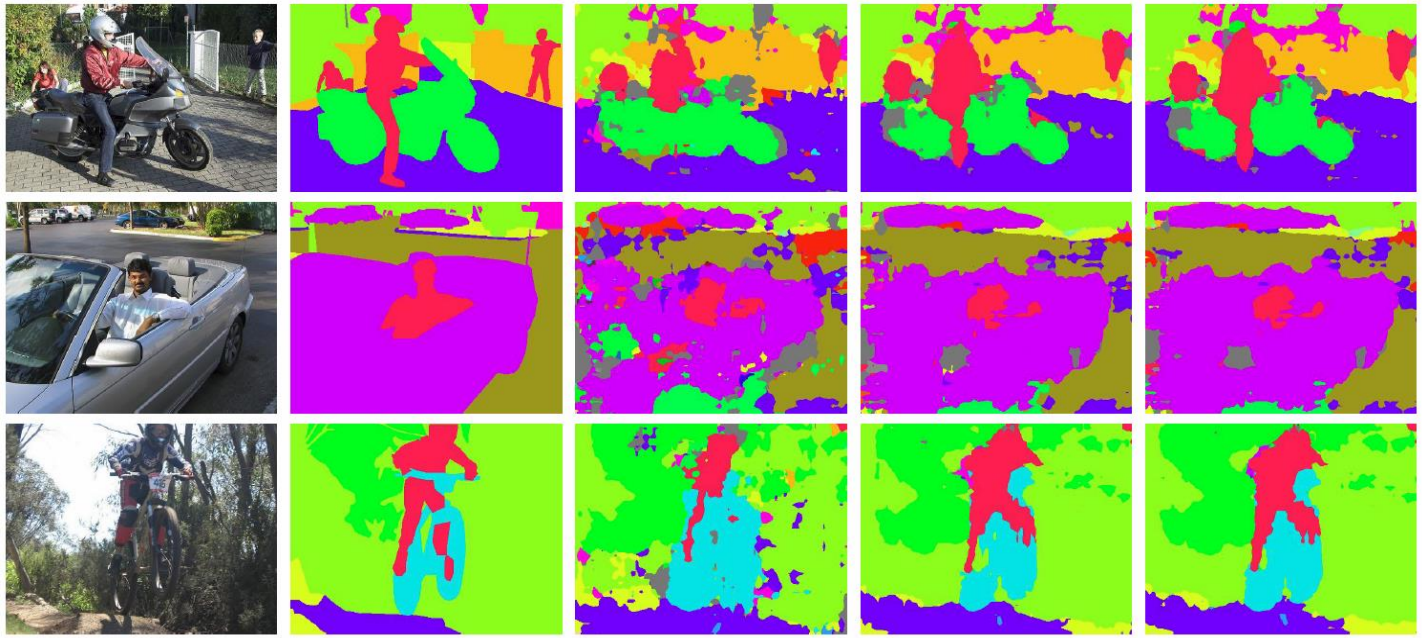
Spatial Constraint

M_{a_region} is a 5X5 square around the ground-truth point

$$M_{a\_extend} = M_{a\_score} \cap M_{a\_region}$$

PointSup: Scene Parsing

Point-based Distance Metric Learning



(a) Image (b) Ground Truth (c) PointSup (d) PointSup+PDML (e) Our method

Qualitative Comparison on PASCAL-Context

Method				Metrics	
FullSup	PointSup	PDML	Online Ext.	mIOU	Pixel Acc
PASCAL-Context validation dataset					
√				39.6	78.6%
	√			27.9	55.3%
	√	√		29.7	57.5%
	√	√	√	**30.0**	**57.6%**
ADE 20K validation dataset					
√				33.9	75.8%
	√			17.7	58.0%
	√	√		19.0	59.0%
	√	√	√	**19.6**	**61.0%**

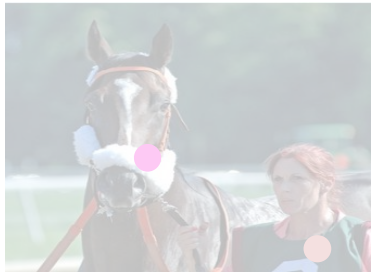
Results on PASCAL-Context and ADE 20K

[Qian AAAI19]

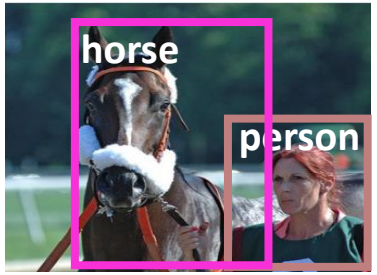
Outline

image-level labels

points



bounding boxes



scribbles

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

bounding boxes

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

bounding boxes

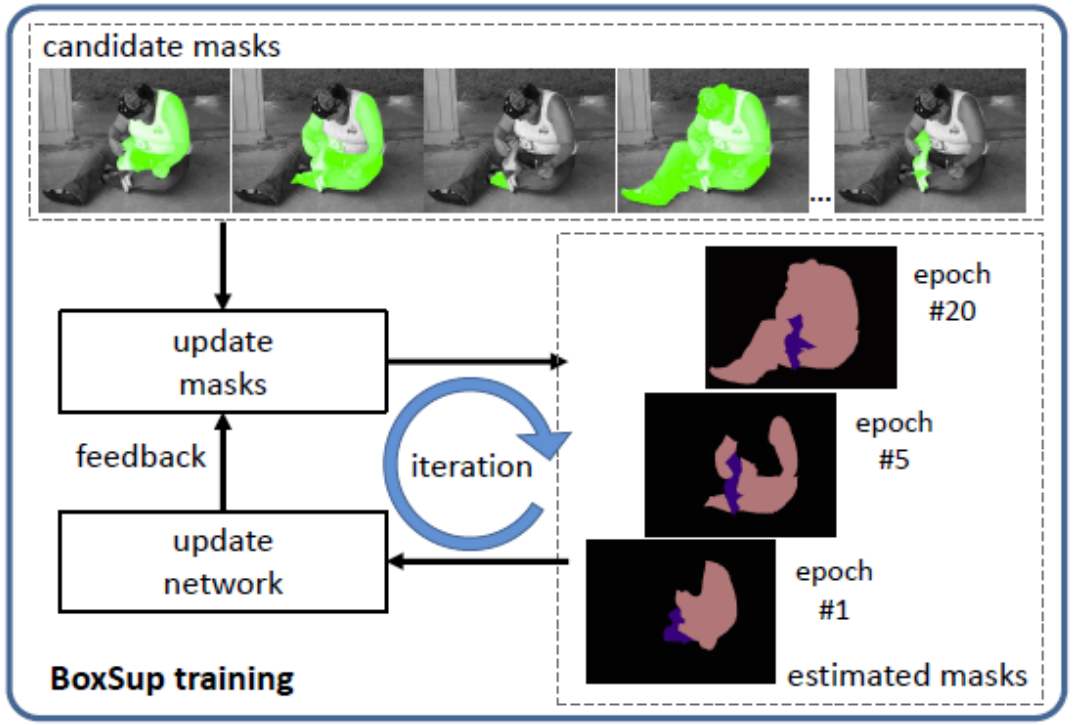
BoxSup: Iterative Feedback

$$\min_{\theta,\{l_S\}} \sum_i (\mathcal{E}_o + \lambda \mathcal{E}_r)$$

$$\mathcal{E}_o = \frac{1}{N} \sum_S (1 - \mathrm{IoU}(B, S)) \delta(l_B, l_S)$$

$$\mathcal{E}_r = \sum_p e(X_\theta(p), l_S(p)).$$

train image
with gt boxes



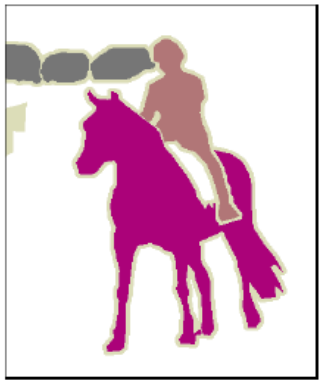
[Dai ICCV15]

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

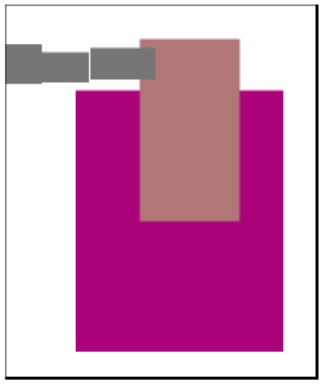
bounding boxes

BoxSup: Mining Strategies for Mask Generation

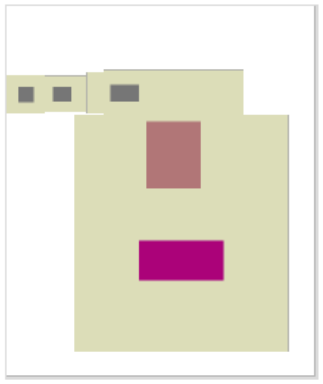
(a) Input image



(b) Ground truth

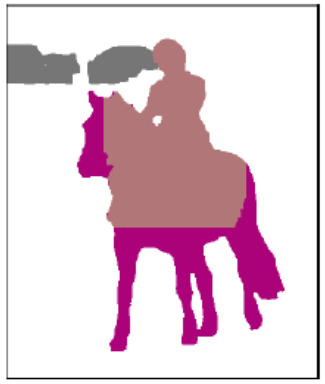


(c) Box

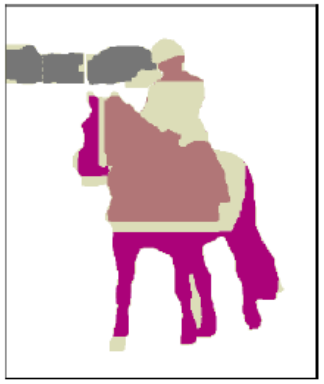


(d) Box^i

(e) GrabCut



(f) GrabCut+



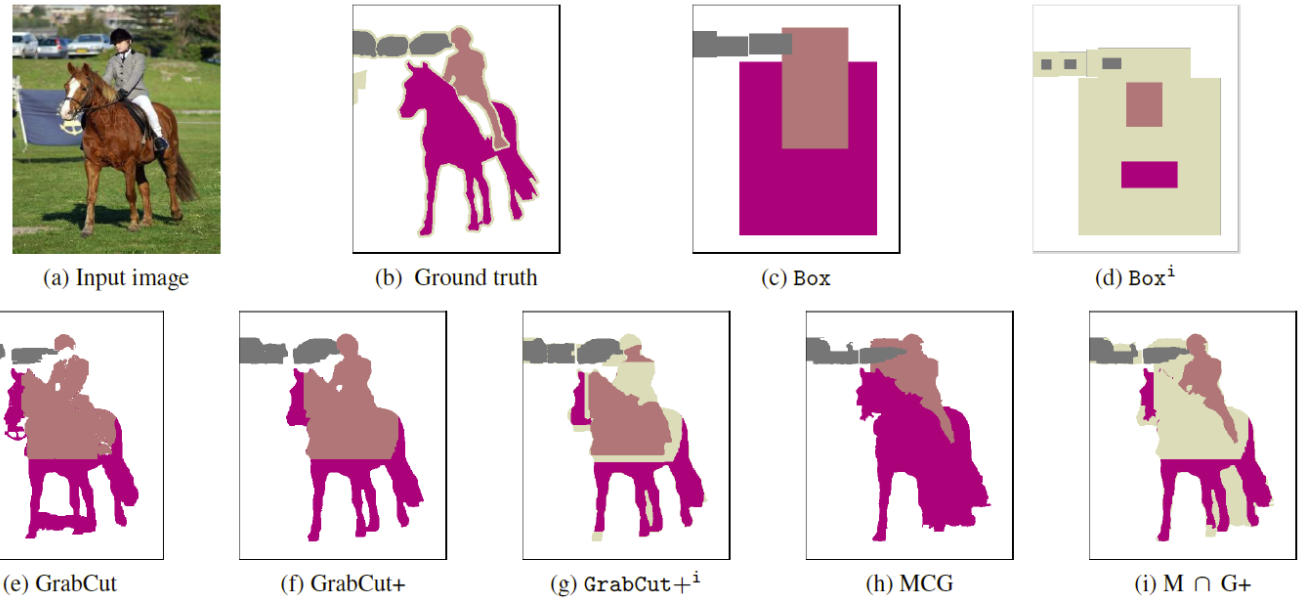
(g) GrabCut+^{i}

(h) MCG

(i) M ∩ G+

[Chen ICCV15, Khoreva CVPR17]

BoxSup: Mining Strategies for Mask Generation



(a) Input image (b) Ground truth (c) Box (d) Box^i

(e) GrabCut (f) GrabCut+ (g) GrabCut+^{i} (h) MCG (i) M ∩ G+

Method		val. mIoU
-	Fast-RCNN	44.3
	GT Boxes	62.2
Weakly supervised	Box	61.2
	Box^i	62.7
	MCG	62.6
	GrabCut+	63.4
	GrabCut+^{i}	64.3
	M ∩ G+	**65.7**
Fully supervised	DeepLab$_{ours}$ [5]	69.1

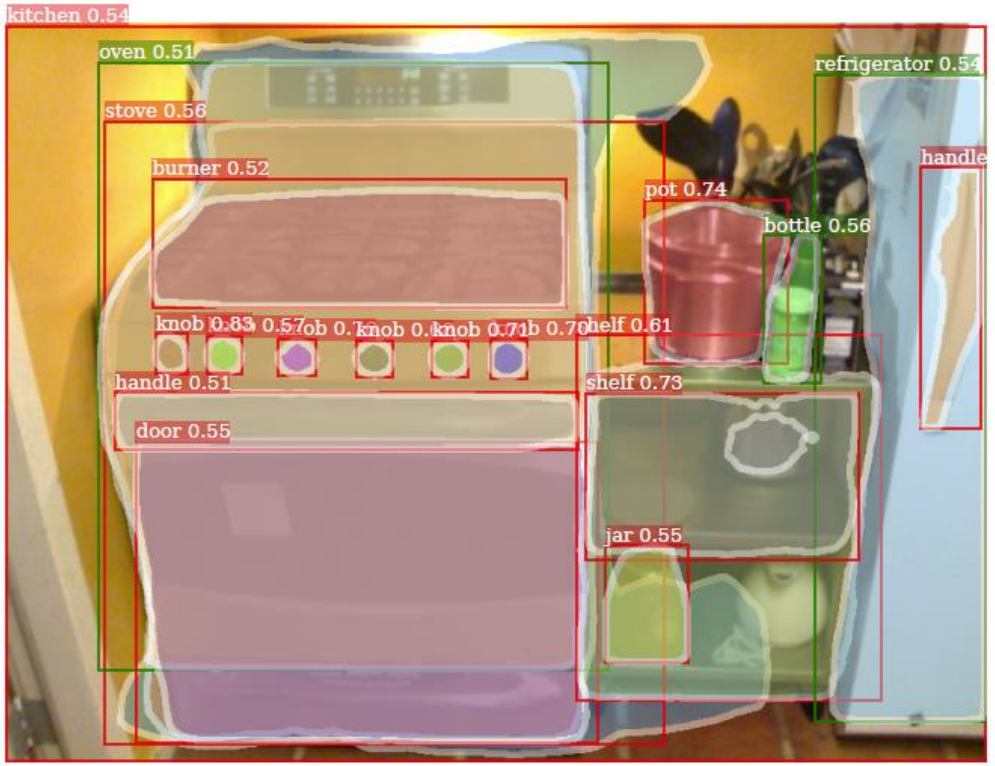
[Chen ICCV15, Khoreva CVPR17]

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

bounding boxes

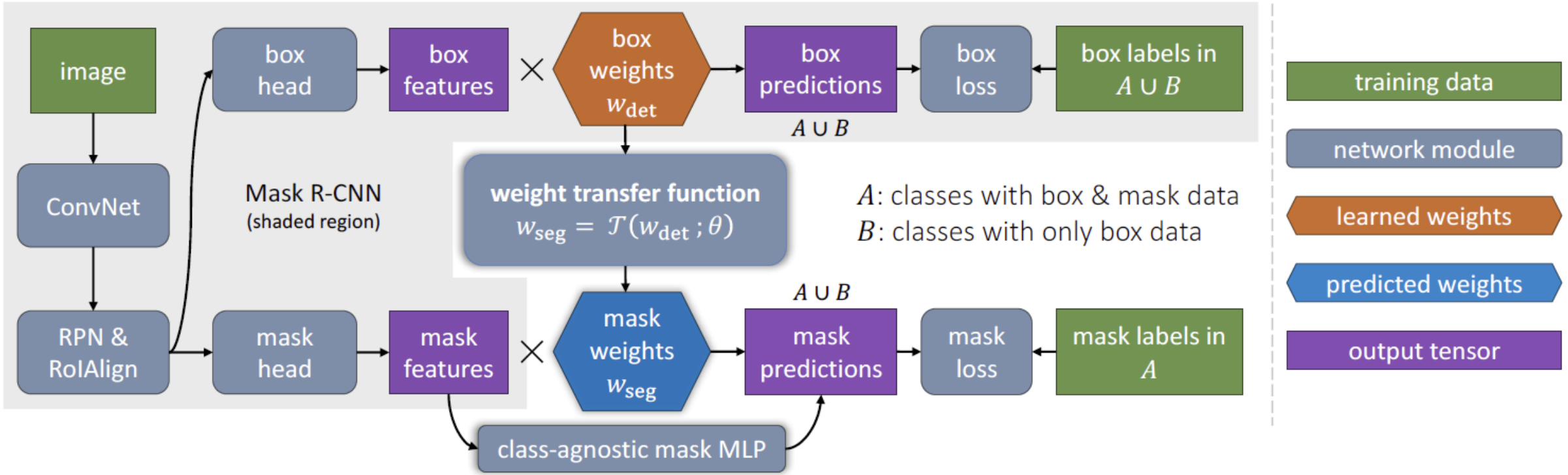
BoxSup: Semi-supervised Learning

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



[Hu CVPR18]

BoxSup: Semi-supervised Learning

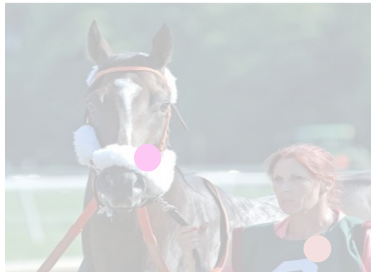


[Hu CVPR18]

Outline

image-level labels

points



bounding boxes

scribbles

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

scribbles

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

scribbles

ScribbleSup: Learning with Alternative Optimization

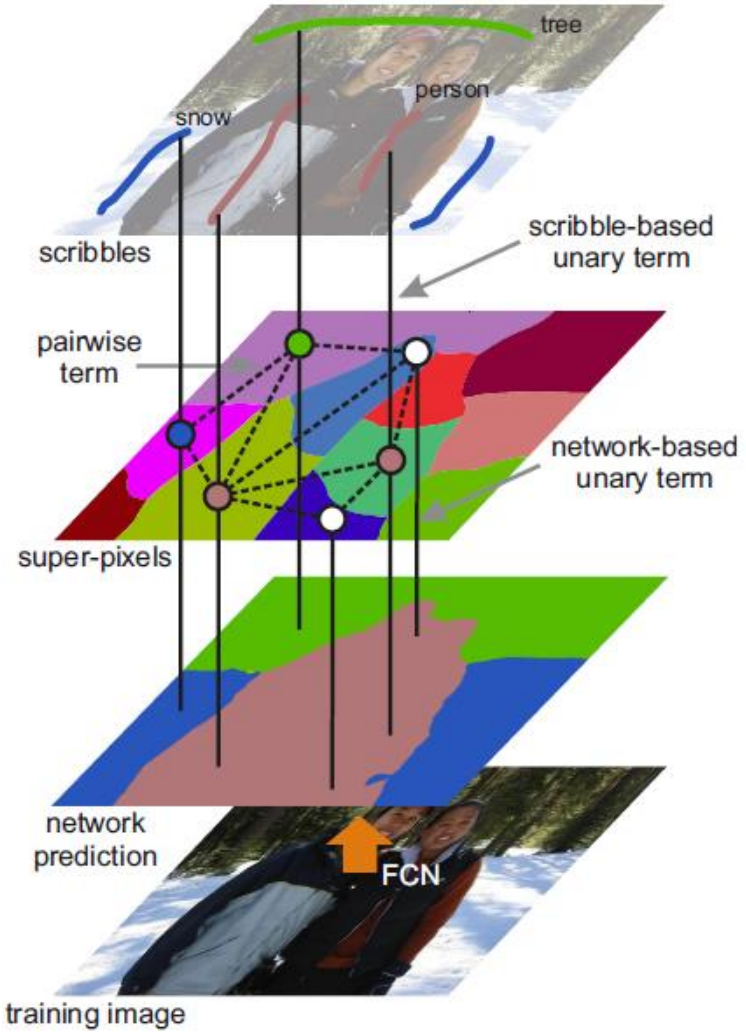
$$\sum_i \psi_i(y_i|X,S) + \sum_{i,j} \psi_{ij}(y_i,y_j|X)$$

$$\psi_i^{scr}(y_i) = \begin{cases} 0 & \text{if } y_i = c_k \text{ and } x_i \bigcap s_k \neq \varnothing \\ -\log(\frac{1}{|\{c_k\}|}) & \text{if } y_i \in \{c_k\} \text{ and } x_i \bigcap S = \varnothing \\ \infty & \text{otherwise} \end{cases}$$

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

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

[Lin CVPR16]

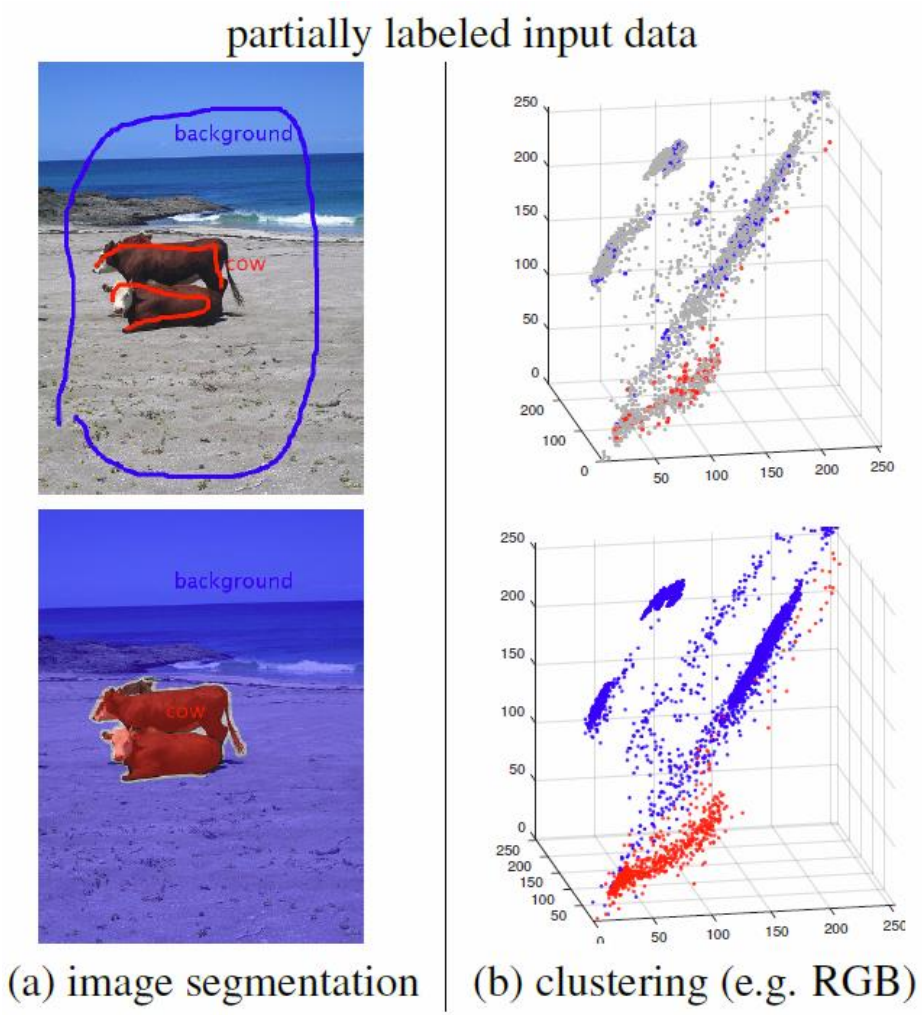


- ScribbleSup: Learning with Alternative Optimization

- ScribbleSup: Learning with Joint Optimization

scribbles

ScribbleSup: Learning with Joint Optimization



[Tang CVPR18, Tang ECCV18]

Semi-supervised Optimization Problem:

$$\min_{\theta} \underbrace{\ell(f_\theta(I), Y)}_{\text{Ground-truth Loss}} + \lambda \cdot \underbrace{R(f_\theta(I))}_{\text{Regularization Loss}}$$

$$\Rightarrow \sum_{p \in \Omega_{\mathcal{L}}} H(Y_p, S_p) + \lambda \cdot R(S)$$

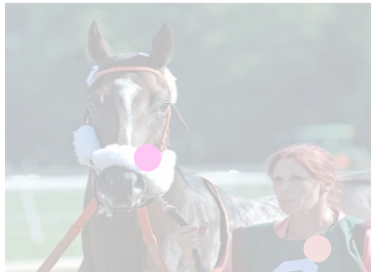
$$R_{NC}(S) = \sum_k \frac{S^{k'} \hat{W} (1 - S^k)}{d' S^k}$$

$$R_{KC}(S) = \sum_k S^{k'} W (\mathbf{1} - S^k) + \gamma \sum_k \frac{S^{k'} \hat{W} (1 - S^k)}{d' S^k}$$

Outline

image-level labels

points



bounding boxes

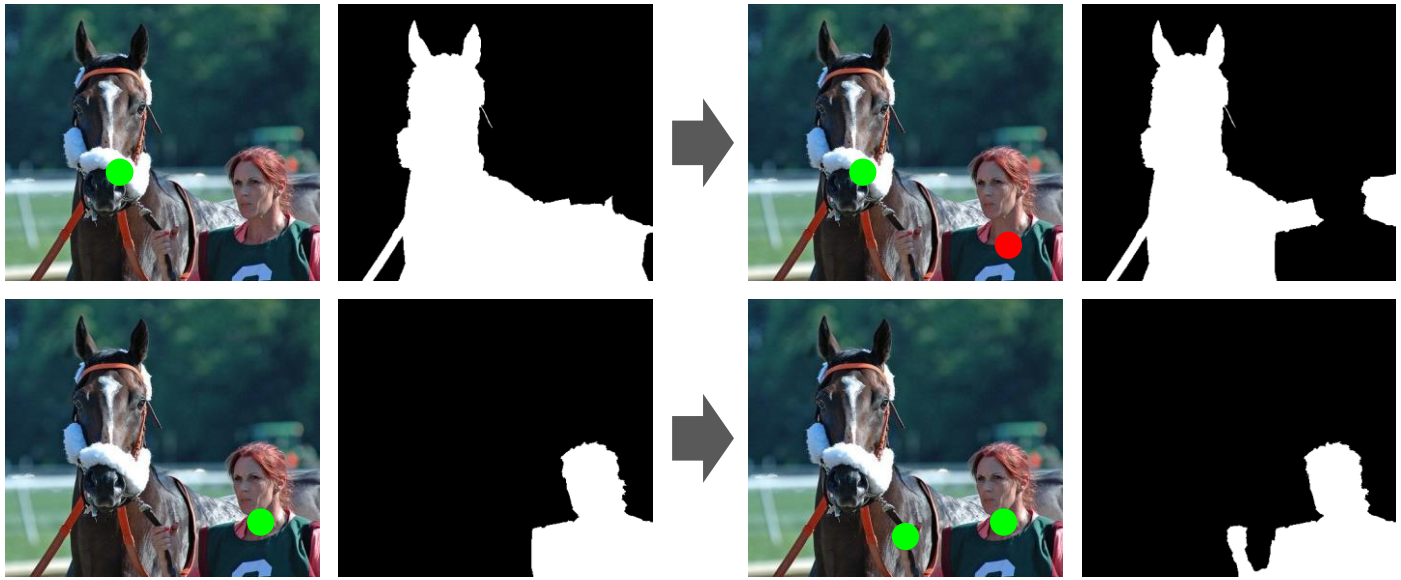
scribbles

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

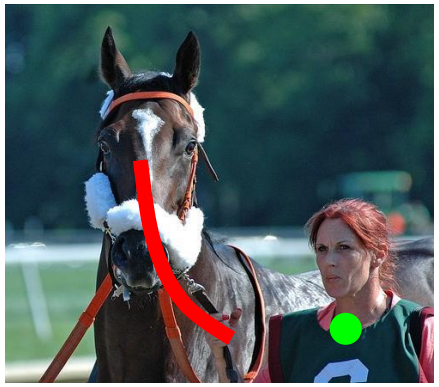
Semantic segmentation



Interactive segmentation

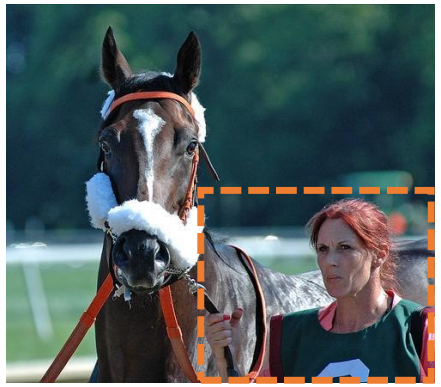
Interactive Image Segmentation

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



(a) Regional points

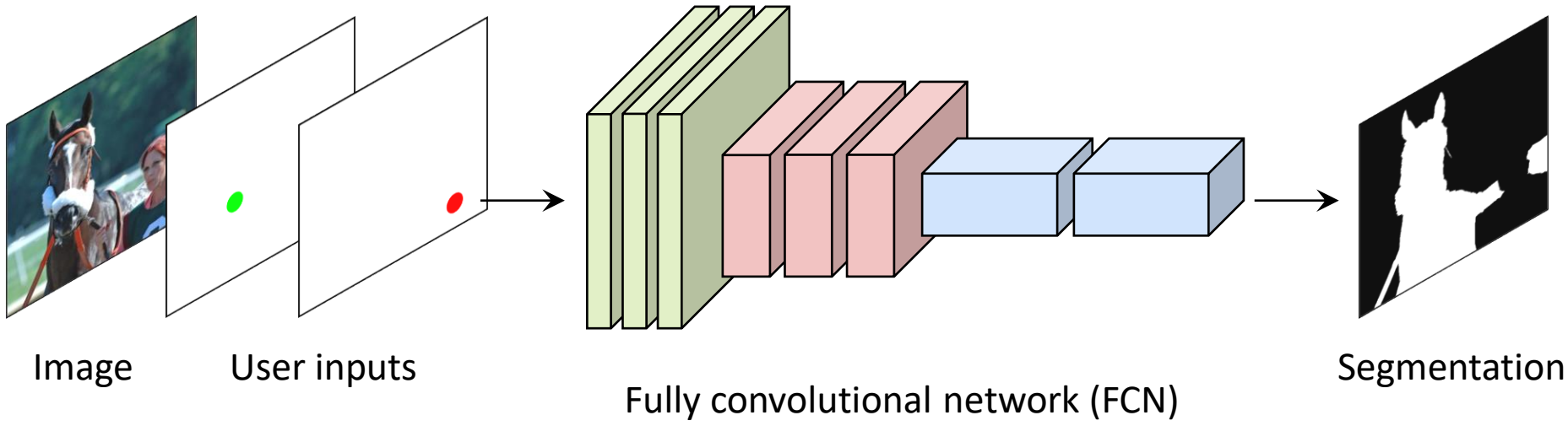
(b) Boundary points



(c) Bounding box

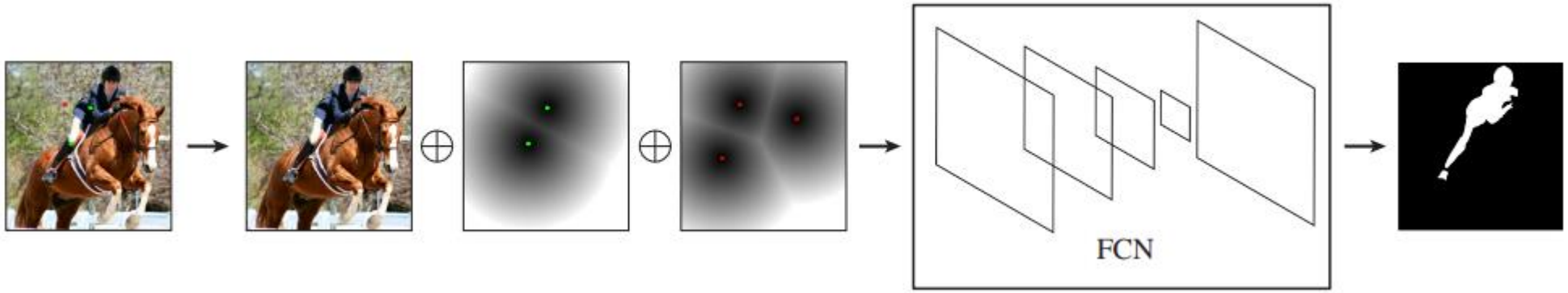
Interactive Image Segmentation

- 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+)



Interactive Image Segmentation

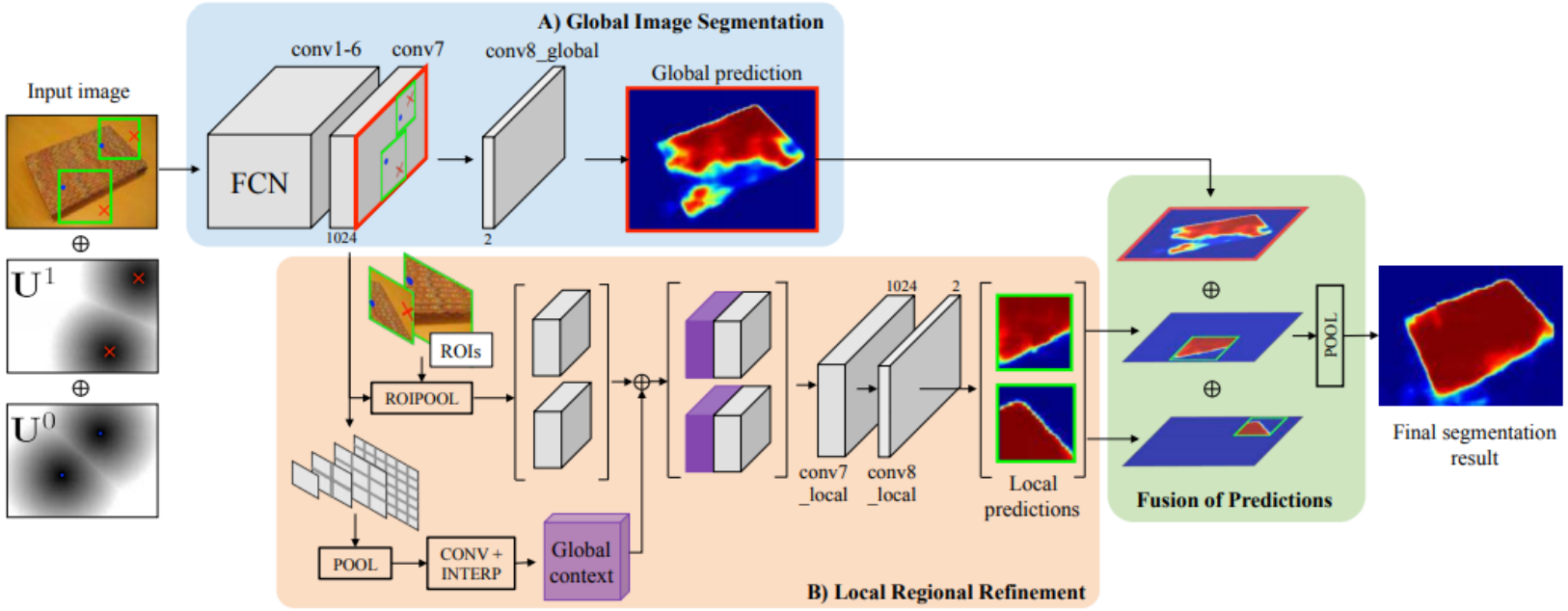
- [Xu CVPR16]
  - First **deep** interactive image segmentation work
  - Simulate user clicks with different clicks sampling strategies
  - Encode user inputs with truncated Euclidean distance maps
  - FCN-8s as the backbone architecture
  - Apply graph cut optimization for better boundary localization



[Xu CVPR16]

Interactive Image Segmentation

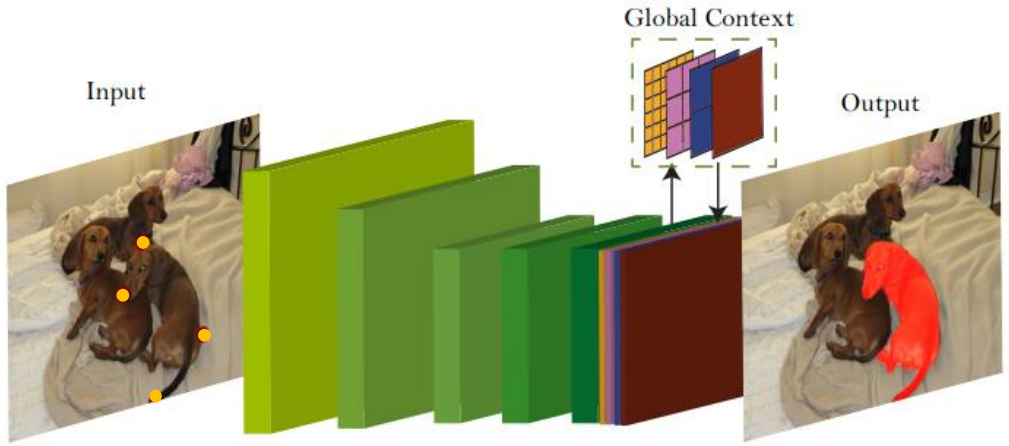
- RIS-Net [Liew ICCV 17]
  - A local branch to focus on the local region around each (+ve, -ve) click pairs
  - Append PSP feature as multiscale global context
  - Click discounting factor to enforce the network to use minimal amount of user inputs for refinement



[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

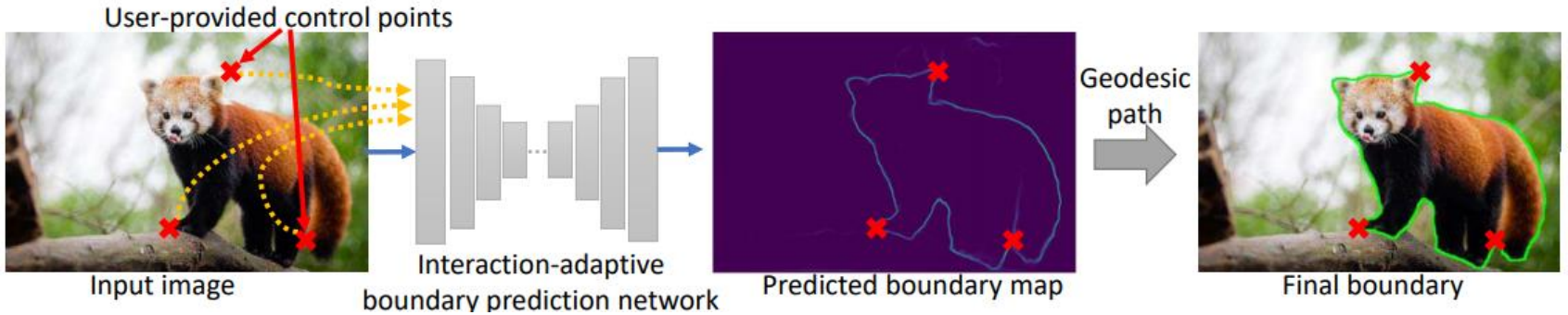


[Maninis CVPR18]

Interactive Image Segmentation

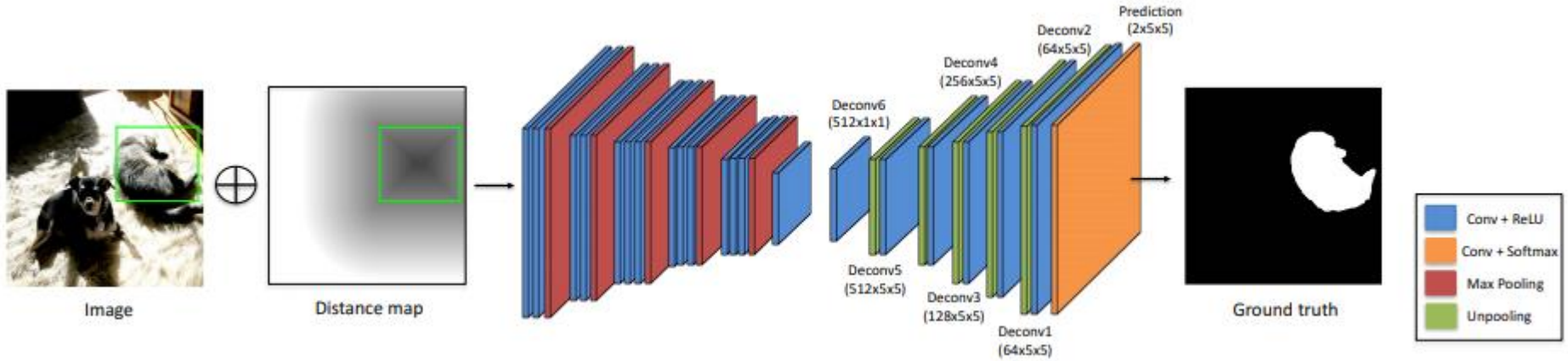
- [Le ECCV 18]
  - Predict object boundaries using boundary clicks (control points)
  - Encode the control points with multiscale "Gaussian" $S_{c_i}^{\sigma}(p) = \exp\left(\frac{-d(p,c_i)^2}{2(\sigma \cdot L)^2}\right)$
  - Encoder-decoder with skip connections as backbone
  - Train with 3 types of 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

[Le ECCV18]

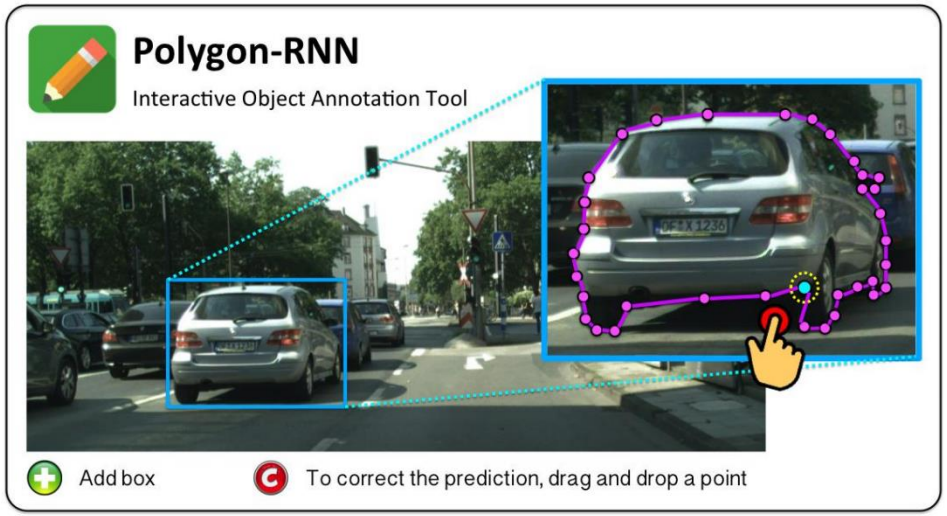


Interactive Image Segmentation

- Deep GrabCut [Xu BMVC 17]



- Polygon-RNN

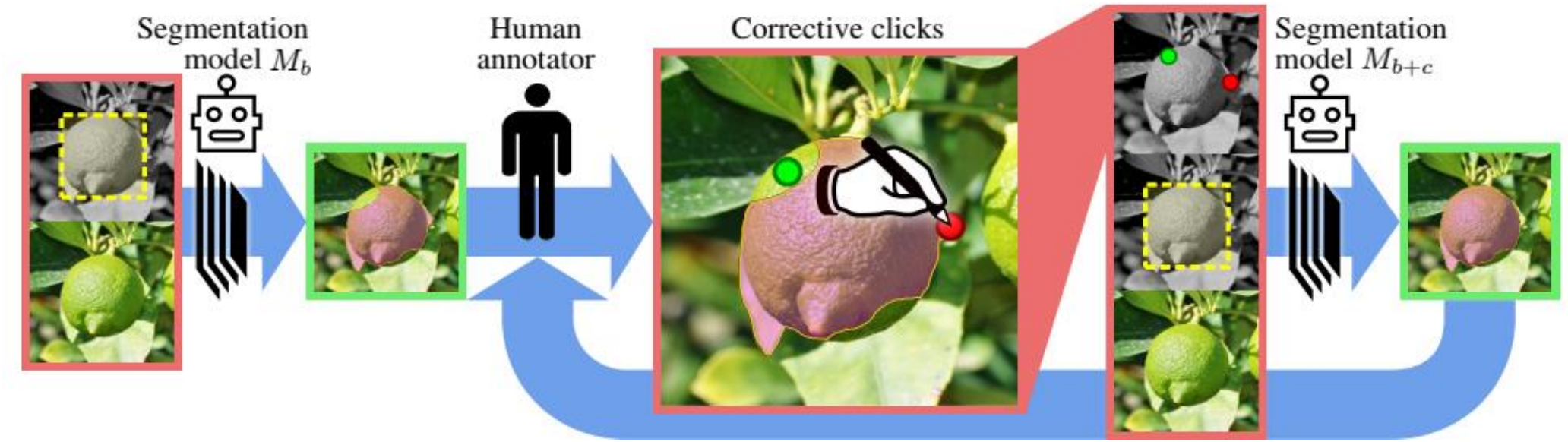


[Xu BMVC17, Castrejon CVPR17]

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

[1] J. Ahn and S. Kwak. Learning pixel-level semantic affinity with image-level supervision for weakly supervised semantic segmentation. In IEEE CVPR, pages 4981–4990, 2018.

[2] A. Bearman, O. Russakovsky, V. Ferrari, and L. Fei-Fei. Whats the point: Semantic segmentation with point supervision. In ECCV, pages 549–565, 2016.

[3] R. Benenson, S. Popov, and V. Ferrari. Large-scale interactive object segmentation with human annotators. arXiv preprint arXiv:1903.10830, 2019.

[4] A. Chaudhry, P. K. Dokania, and P. H. Torr. Discovering class-specific pixels for weakly-supervised semantic segmentation. In BMVC, 2017.

[5] J. Dai, K. He, and J. Sun. Boxsup: Exploiting bounding boxes to supervise convolutional networks for semantic segmentation. In IEEE ICCV, pages 1635–1643, 2015.

[6] R. Fan, Q. Hou, M.-M. Cheng, G. Yu, R. R. Martin, and S.-M. Hu. Associating inter-image salient instances for weakly supervised semantic segmentation. In ECCV, pages 367–383, 2018.

[7] W. Ge, S. Yang, and Y. Yu. Multi-evidence filtering and fusion for multi-label classification, object detection and semantic segmentation based on weakly supervised learning. In IEEE CVPR, pages 1277–1286, 2018.

[8] S. Hong, D. Yeo, S. Kwak, H. Lee, and B. Han. Weakly supervised semantic segmentation using web-crawled videos. In IEEE CVPR, pages 7322–7330, 2017.

[9] Q. Hou, P. Jiang, Y. Wei, and M.-M. Cheng. Self-erasing network for integral object attention. In NIPS, pages 549–559, 2018.

[10] R. Hu, P. Doll´ar, K. He, T. Darrell, and R. Girshick. Learning to segment every thing. In IEEE CVPR, pages 4233–4241, 2018.

[11] Z. Huang, X. Wang, J. Wang, W. Liu, and J. Wang. Weakly-supervised semantic segmentation network with deep seeded region growing. In IEEE CVPR, pages 7014–7023, 2018.

[12] B. Jin, M. V. Ortiz Segovia, and S. Susstrunk. Webly supervised semantic segmentation. In IEEE CVPR, pages 3626–3635, 2017.

[13] A. Khoreva, R. Benenson, J. Hosang, M. Hein, and B. Schiele. Simple does it: Weakly supervised instance and semantic segmentation. In IEEE CVPR, pages 876–885, 2017.

[14] D. Kim, D. Cho, D. Yoo, and I. So Kweon. Two-phase learning for weakly supervised object localization. In IEEE ICCV, pages 3534–3543, 2017.

[15] A. Kolesnikov and C. H. Lampert. Seed, expand and constrain: Three principles for weakly-supervised image segmentation. In ECCV, pages 695–711, 2016.

[16] H. Le, L. Mai, B. Price, S. Cohen, H. Jin, and F. Liu. Interactive boundary prediction for object selection. In ECCV, pages 18–33, 2018.

[17] K. Li, Z. Wu, K.-C. Peng, J. Ernst, and Y. Fu. Tell me where to look: Guided attention inference network. In IEEE CVPR, pages 9215–9223, 2018.

[18] Q. Li, A. Arnab, and P. H. Torr. Weakly-and semi-supervised panoptic segmentation. In ECCV, pages 102–118, 2018.

[19] Z. Li, Q. Chen, and V. Koltun. Interactive image segmentation with latent diversity. In IEEE CVPR, pages 577–585, 2018.

[20] J. Liew, Y. Wei, W. Xiong, S.-H. Ong, and J. Feng. Regional interactive image segmentation networks. In IEEE ICCV, pages 2746–2754, 2017.

References

[21] D. Lin, J. Dai, J. Jia, K. He, and J. Sun. Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In IEEE CVPR, pages 3159–3167, 2016.
[22] S. Mahadevan, P. Voigtlaender, and B. Leibe. Iteratively trained interactive segmentation. In BMVC, 2018.
[23] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool. Deep extreme cut: From extreme points to object segmentation. In IEEE CVPR, pages 616–625, 2018.
[24] S. J. Oh, R. Benenson, A. Khoreva, Z. Akata, M. Fritz, and B. Schiele. Exploiting saliency for object segmentation from image level labels. In IEEE CVPR, pages 5038–5047, 2017.
[25] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille. Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation. In IEEE ICCV, pages 1742–1750, 2015.
[26] D. Pathak, P. Krahenbuhl, and T. Darrell. Constrained convolutional neural networks for weakly supervised segmentation. In IEEE ICCV, pages 1796–1804, 2015.
[27] D. Pathak, E. Shelhamer, J. Long, and T. Darrell. Fully convolutional multi-class multiple instance learning. In ICLR Workshop, 2015.
[28] P. O. Pinheiro and R. Collobert. From image-level to pixel-level labeling with convolutional networks. In IEEE CVPR, pages 1713–1721, 2015.
[29] X. Qi, Z. Liu, J. Shi, H. Zhao, and J. Jia. Augmented feedback in semantic segmentation under image level supervision. In ECCV, pages 90–105, 2016.
[30] R. Qian, Y. Wei, H. Shi, J. Li, J. Liu, and T. Huang. Weakly supervised scene parsing with point-based distance metric learning. In AAAI, 2019.
[31] A. Roy and S. Todorovic. Combining bottom-up, top-down, and smoothness cues for weakly supervised image segmentation. In IEEE CVPR, pages 3529–3538, 2017.
[32] F. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, S. Gould, and J. M. Alvarez. Built-in foreground/background prior for weakly-supervised semantic segmentation. In ECCV, pages 413–432, 2016.
[33] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, and J. M. Alvarez. Bringing background into the foreground: Making all classes equal in weakly-supervised video semantic segmentation. In IEEE ICCV, pages 2125–2135, 2017.
[34] F. S. Saleh, M. S. Aliakbarian, M. Salzmann, L. Petersson, J. M. Alvarez, and S. Gould. Incorporating network built-in priors in weakly-supervised semantic segmentation. IEEE TPAMI, 40(6):1382–1396, 2018.
[35] W. Shimoda and K. Yanai. Distinct class-specific saliency maps for weakly supervised semantic segmentation. In ECCV, pages 218–234, 2016.
[36] K. K. Singh and Y. J. Lee. Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In IEEE ICCV, pages 3544–3553, 2017.
[37] M. Tang, A. Djelouah, F. Perazzi, Y. Boykov, and C. Schroers. Normalized cut loss for weakly-supervised cnn segmentation. In IEEE CVPR, pages 1818–1827, 2018.

References

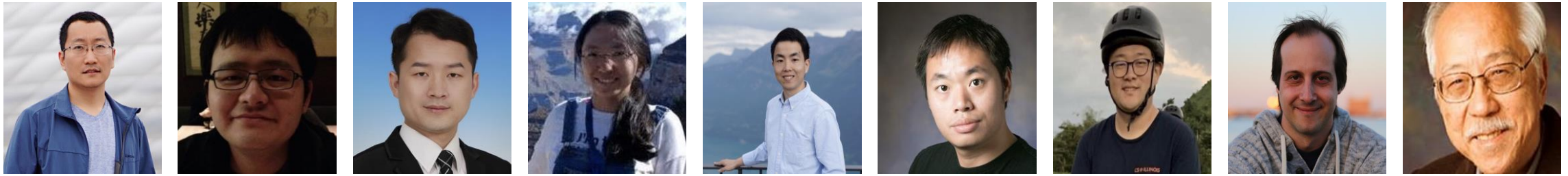
[38] M. Tang, F. Perazzi, A. Djelouah, I. Ben Ayed, C. Schroers, and Y. Boykov. On regularized losses for weakly-supervised cnn segmentation. In ECCV, pages 507–522, 2018.
[39] X. Wang, S. You, X. Li, and H. Ma. Weakly-supervised semantic segmentation by iteratively mining common object features. In IEEE CVPR, pages 1354–1362, 2018.
[40] Y. Wei, J. Feng, X. Liang, M.-M. Cheng, Y. Zhao, and S. Yan. Object region mining with adversarial erasing: A simple classification to semantic segmentation approach. In IEEE CVPR, pages 1568–1576, 2017.
[41] Y. Wei, X. Liang, Y. Chen, Z. Jie, Y. Xiao, Y. Zhao, and S. Yan. Learning to segment with image-level annotations. Pattern Recognition, 59:234–244, 2016.
[42] Y. Wei, X. Liang, Y. Chen, X. Shen, M.-M. Cheng, J. Feng, Y. Zhao, and S. Yan. Stc: A simple to complex framework for weakly-supervised semantic segmentation. IEEE TPAMI, 39(11):2314–2320, 2017.
[43] Y. Wei, H. Xiao, H. Shi, Z. Jie, J. Feng, and T. S. Huang. Revisiting dilated convolution: A simple approach for weakly-and semi-supervised semantic segmentation. In IEEE CVPR, pages 7268–7277, 2018.
[44] H. Xiao, Y. Wei, Y. Liu, M. Zhang, and J. Feng. Transferable semi-supervised semantic segmentation. In AAAI, 2018.
[45] N. Xu, B. Price, S. Cohen, J. Yang, and T. Huang. Deep grabcut for object selection. In BMVC, 2017.
[46] N. Xu, B. Price, S. Cohen, J. Yang, and T. S. Huang. Deep interactive object selection. In IEEE CVPR, pages 373–381, 2016.
[47] J. Zhang, S. A. Bargal, Z. Lin, J. Brandt, X. Shen, and S. Sclaroff. Top-down neural attention by excitation backprop. IJCV, 126(10):1084–1102, 2018.
[48] X. Zhang, Y. Wei, J. Feng, Y. Yang, and T. S. Huang. Adversarial complementary learning for weakly supervised object localization. In IEEE CVPR, pages 1325–1334, 2018.
[49] X. Zhang, Y. Wei, G. Kang, Y. Yang, and T. Huang. Self-produced guidance for weakly-supervised object localization. In ECCV, pages 597–613, 2018.
[50] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In IEEE CVPR, pages 2921–2929, 2016.
[51] Y. Zhou, Y. Zhu, Q. Ye, Q. Qiu, and J. Jiao. Weakly supervised instance segmentation using class peak response. In IEEE CVPR, pages 3791–3800, 2018.
[52] Y. Zhu, Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao. Soft proposal networks for weakly supervised object localization. In IEEE ICCV, pages 1841–1850, 2017.

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