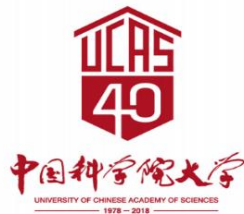


# Weakly Supervised Object Detection, Localization, and instance segmentation

Fang Wan, Yi Zhu, Yanzhao Zhou, **Qixiang Ye**



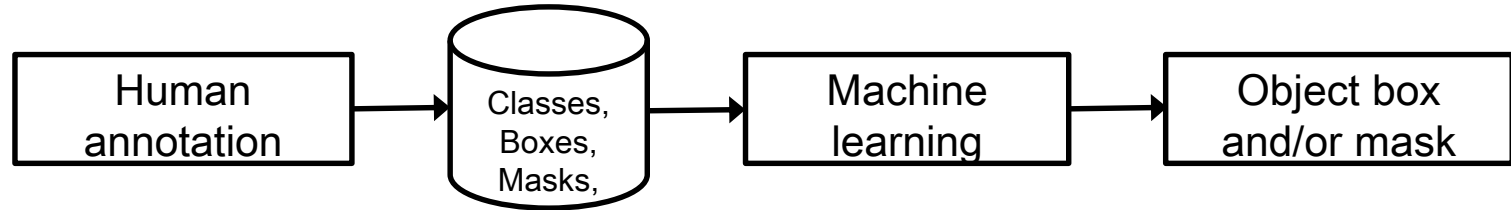
[www.ucasdl.cn](http://www.ucasdl.cn)

[qxye@ucas.ac.cn](mailto:qxye@ucas.ac.cn)

[people.ucas.ac.cn/~qxye](http://people.ucas.ac.cn/~qxye)

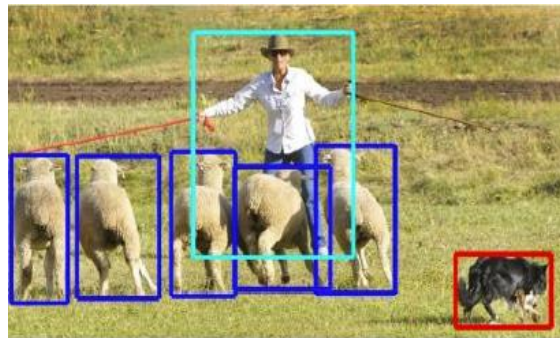
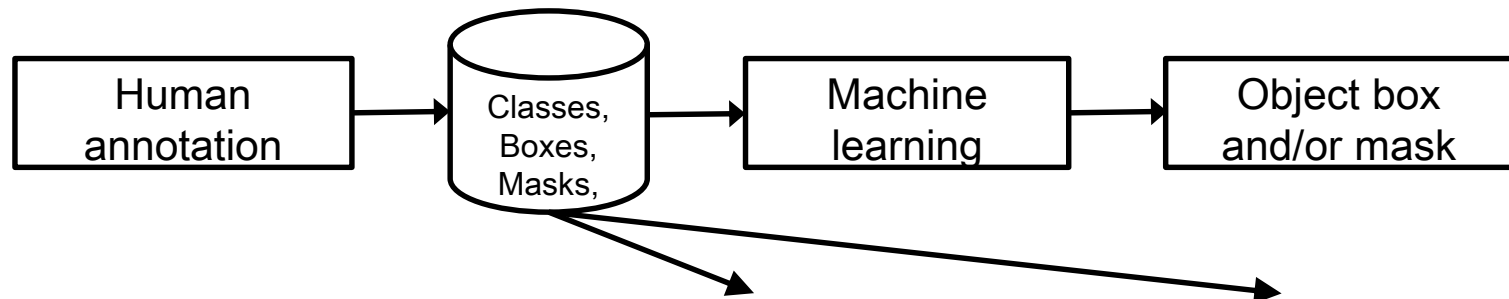
# Problem

Supervised **object detection** and **instance segmentation** pipeline



# Problem

Supervised **object detection** and **instance segmentation** pipeline



Bounding box annotation



Mask annotation

# Problem



20k+  
class

100/c  
→

2M  
instances

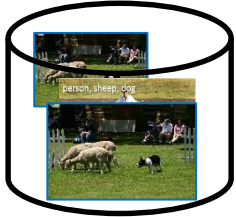
5min/l  
→

**19 years  
/person**

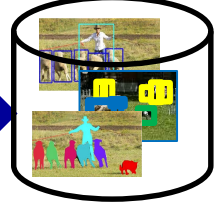
Copy from UIUC Yunchao's Slides

# Problem

Imagery  
databases

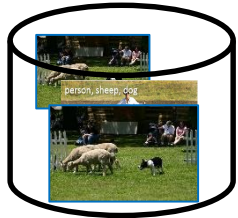


Training  
sets

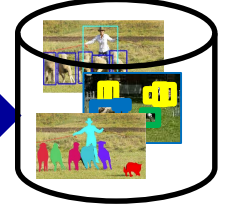


# Solutions

Imagery  
databases

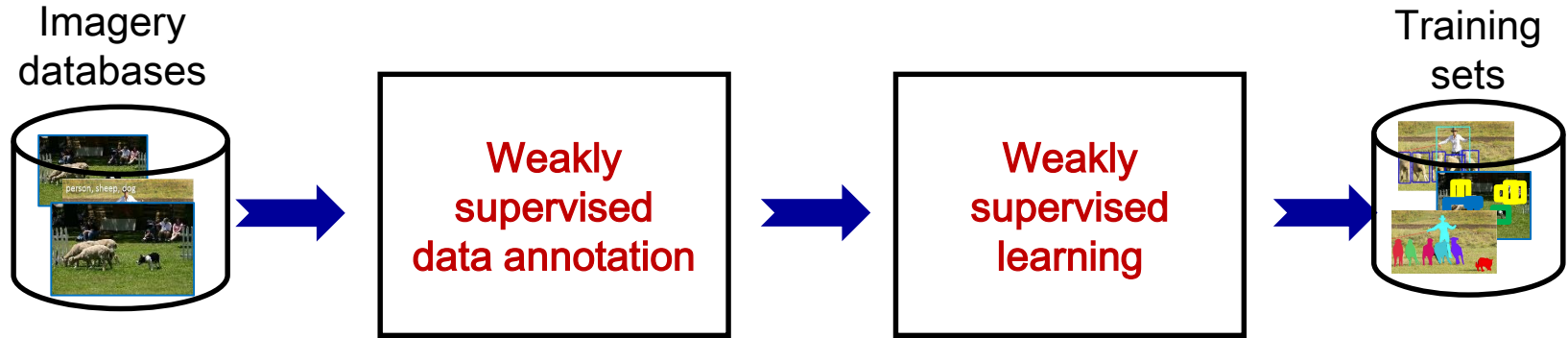


Training  
sets



Data annotation is **expensive**

# Solutions



Data annotation is **efficient and low-cost**

# Solutions

## Weakly Supervised Annotations



Scribes

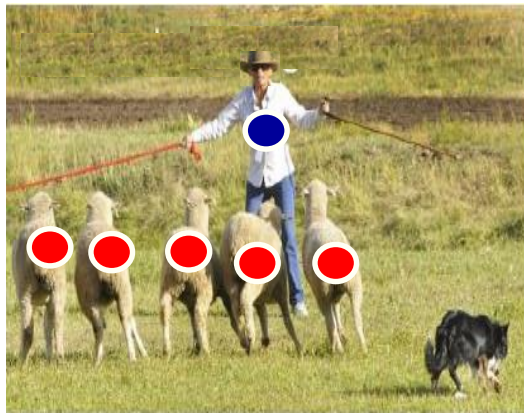
Copy from UIUC Yunchao's Slides

# Solutions

## Weakly Supervised Annotations



Scribes



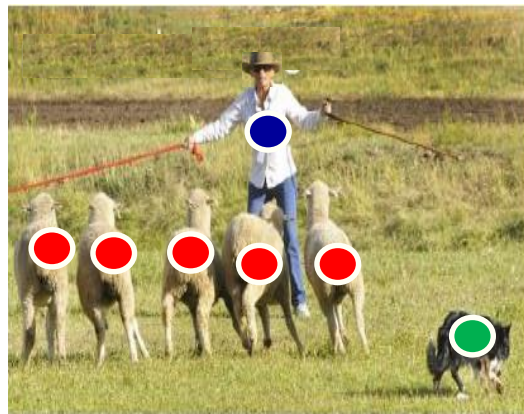
Point

# Solutions

## Weakly Supervised Annotations



Scribes



Point



Image-level labels

**The most efficient one**

# Solutions

Weakly labeled imagery is widely available on the Web


国内版 国际版

草原羊群

登录 人 三

网页 图片 视频 学术 词典 地图

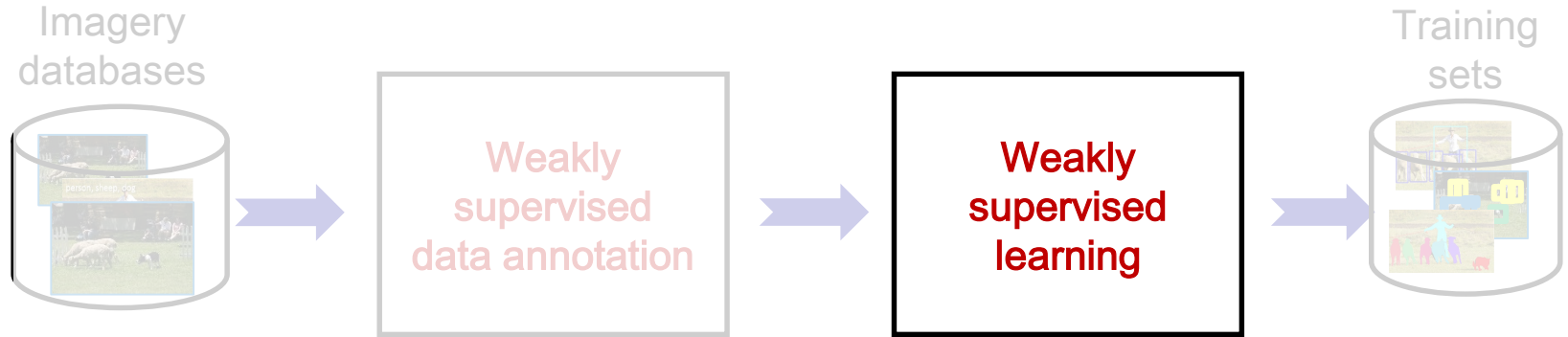
筛选器



The image displays a search results page for '草原羊群' (Grassland Sheep) on a Chinese search engine. The page features a grid of 24 images showing various scenes of sheep grazing in green pastures. The images are arranged in three rows and eight columns. The top row shows a large flock of sheep in a field, a sheep grazing on a green hillside, a sheep grazing near a lake, a large flock of sheep in a field, a sheep grazing in a field, and a large flock of sheep in a field. The middle row shows a sheep grazing near a river, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, and a sheep grazing in a field. The bottom row shows a sheep grazing in a field, a sheep grazing in a field, a solid green image, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, and a sheep grazing in a field. The search engine interface includes a search bar with the text '草原羊群', a search icon, and navigation tabs for '国内版' (Domestic Version) and '国际版' (International Version). The search results are categorized by '网页' (Web), '图片' (Images), '视频' (Videos), '学术' (Academic), '词典' (Dictionary), and '地图' (Maps). The '图片' (Images) tab is selected. The search results are displayed in a grid format, with each image showing a different scene of sheep grazing in a green pasture. The images are arranged in three rows and eight columns. The top row shows a large flock of sheep in a field, a sheep grazing on a green hillside, a sheep grazing near a lake, a large flock of sheep in a field, a sheep grazing in a field, and a large flock of sheep in a field. The middle row shows a sheep grazing near a river, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, and a sheep grazing in a field. The bottom row shows a sheep grazing in a field, a sheep grazing in a field, a solid green image, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, a sheep grazing in a field, and a sheep grazing in a field.

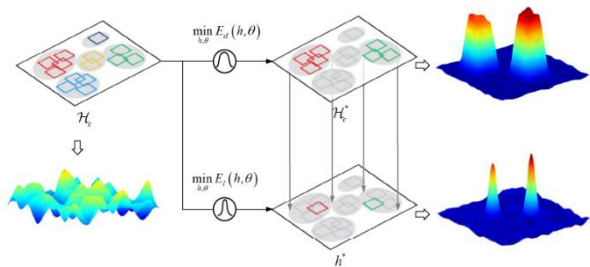


# Solutions



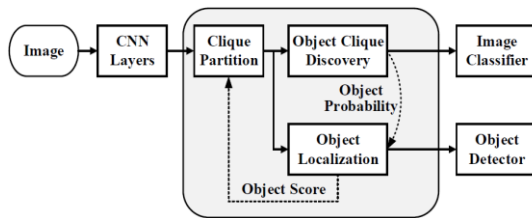
Data annotation is **efficient and low-cost**

# Our works



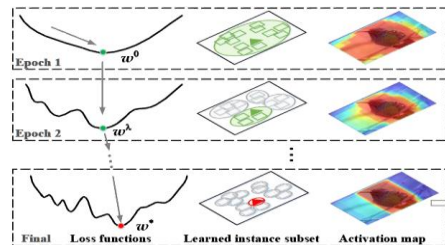
**MELM**

CVPR18: Min-entropy Latent Model (**WSOD**)



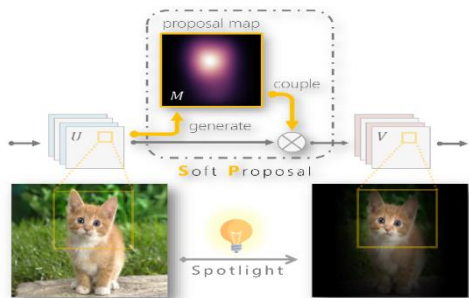
**MELM+Recurrent Learning**

PAMI2019: Recurrent Learning (**WSOD**)



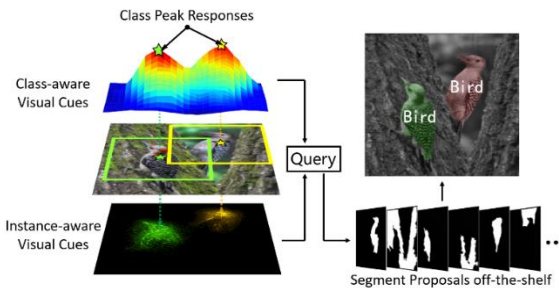
**CMIL: Continuation Multiple Instance Learning**

CVPR19 (**WSOD**)



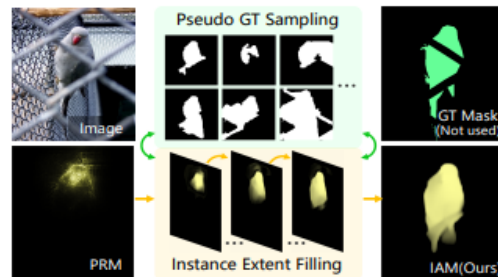
**SPN**

ICCV17: Soft Proposal Network (**WSOL**)



**PRM**

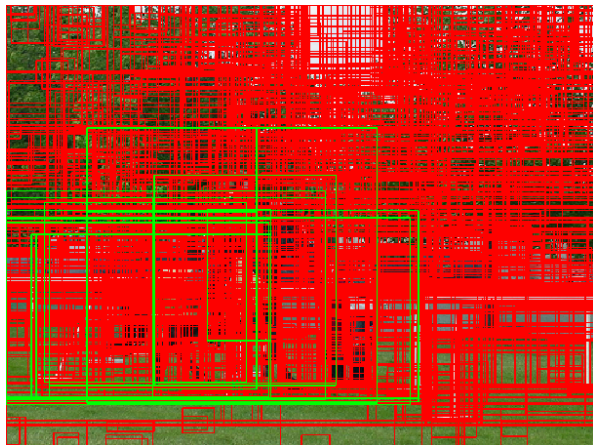
CVPR18: Peak Response Mapping (**WSIS**)



**IAM**

CVPR19: Instance Activation Map (**WSIS**)

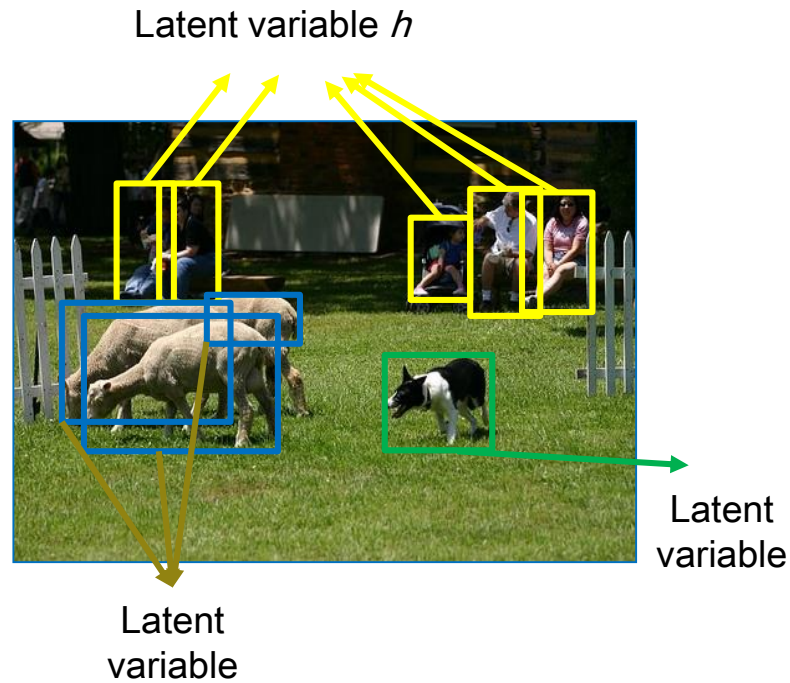
# Our works-**Challenge analysis**



Latent variable  
learning



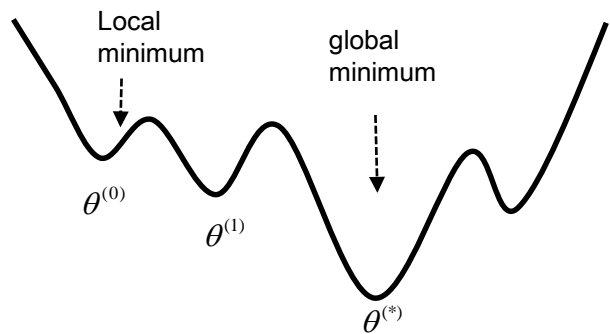
Multiple Instance  
learning



$$L(\theta) = \frac{1}{2} \|\theta\|^2 + \lambda \sum_i \max(0, 1 - y_i f(x_i, h_i))$$

$$f(x, h) = \max_h \theta \cdot \Phi(x, h)$$

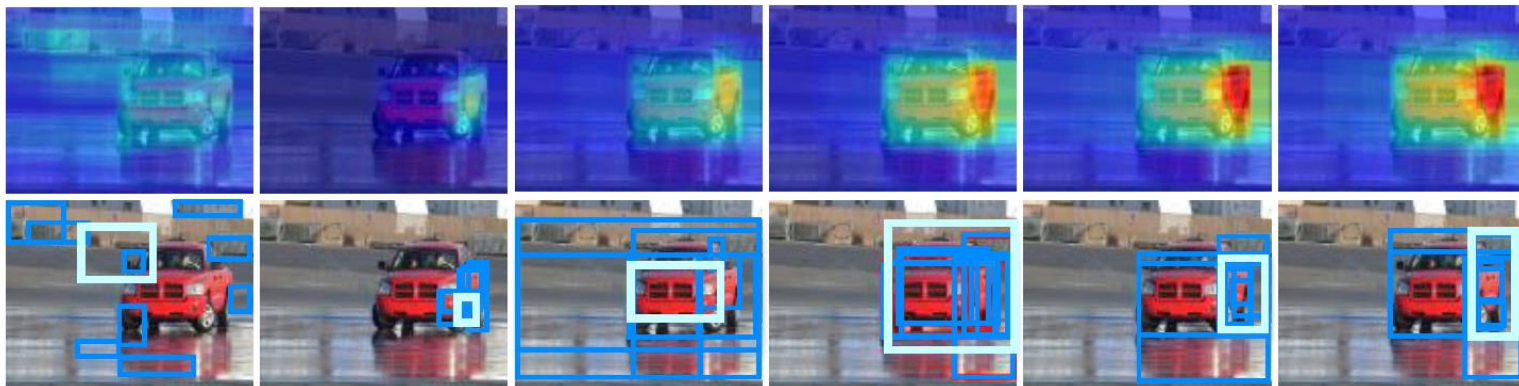
# Our works-**Challenge** analysis



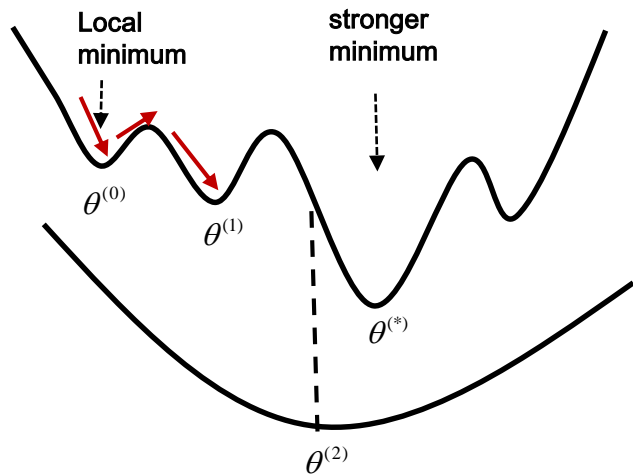
$$L(\theta) = \frac{1}{2} \|\theta\|^2 + \lambda \sum_i \max(0, 1 - y_i f(x_i, h_i))$$

$$f(x, h) = \max_h \theta \cdot \Phi(x, h)$$

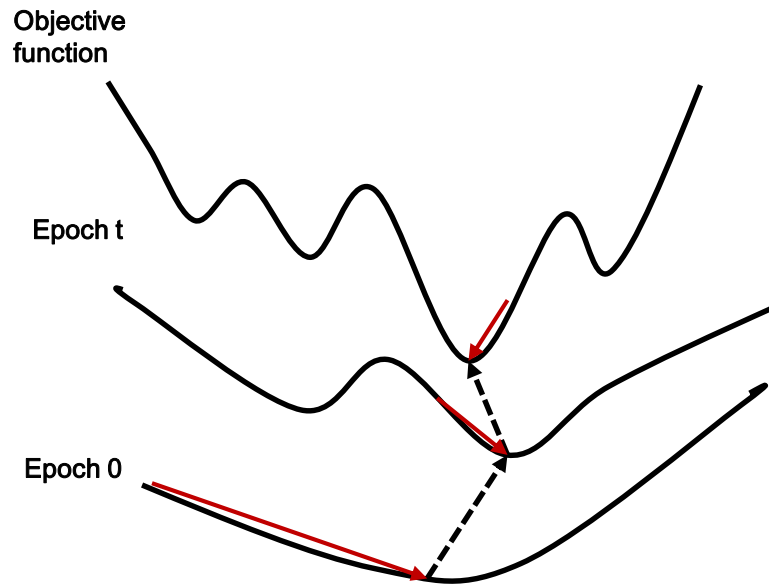
WSDDN



# Our works-Methodology

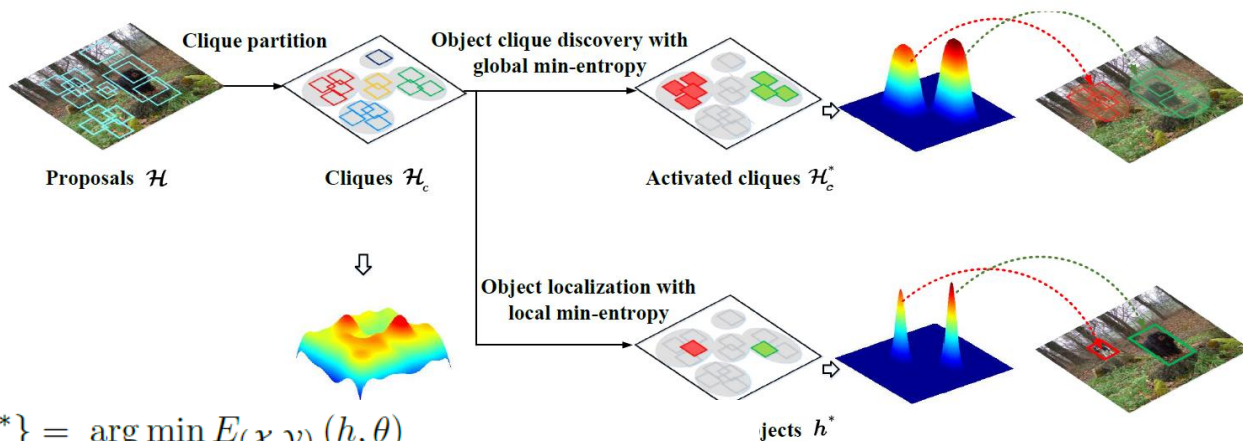


Convex Regularization



Continuation Optimization

# Our works-**Min-entropy** latent model



$$\begin{aligned} \{h^*, \theta^*\} &= \arg \min_{h, \theta} E_{(\mathcal{X}, \mathcal{Y})}(h, \theta) \\ &= \arg \min_{h, \theta} E_{(\mathcal{X}, \mathcal{Y})}(\mathcal{H}_c, \theta) + \lambda E_{(\mathcal{X}, \mathcal{Y}, \mathcal{H}_c)}(h, \theta) \end{aligned}$$

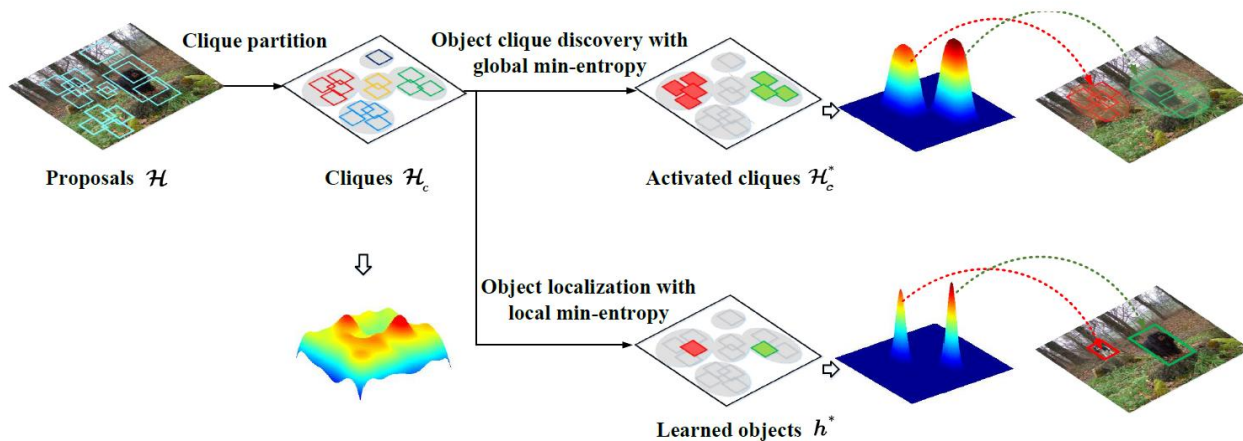
**Object  
discovery**

$$\begin{aligned} E_{(\mathcal{X}, \mathcal{Y})}(\mathcal{H}_c, \theta) &= -\log \sum w_{\mathcal{H}_c} p(y, \mathcal{H}_c; \theta) \\ L_{(\mathcal{X}, \mathcal{Y})}(\mathcal{H}_c, \theta) &= y E_{(\mathcal{X}, \mathcal{Y})}(\mathcal{H}_c, \theta) \\ &\quad - (1 - y) \sum_h \log(1 - p(y, h; \theta)) \end{aligned}$$

**Object  
localization**

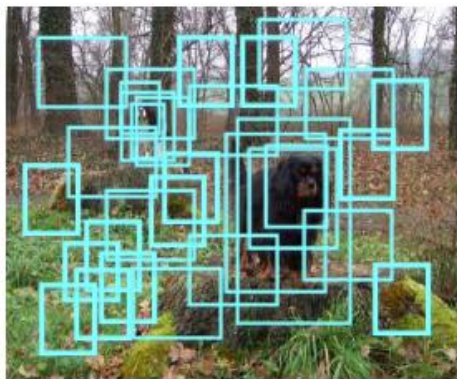
$$\begin{aligned} h^* &= \arg \min_{h \in \mathcal{H}_c^*} E_{(\mathcal{X}, \mathcal{Y}, \mathcal{H}_c^*)}(h, \theta) \\ E_{(\mathcal{X}, \mathcal{Y}, \mathcal{H}_c)}(h, \theta) &= -\sum_{h \in \Omega_{h^*}} w_h p(y, h; \theta) \log p(y, h; \theta) \\ L_{(\mathcal{X}, \mathcal{Y}, \mathcal{H}_c)}(h, \theta) &= E_{(\mathcal{X}, \mathcal{Y}, \mathcal{H}_c^*)}(h, \theta). \end{aligned}$$

# Our works-**Min-entropy** latent model

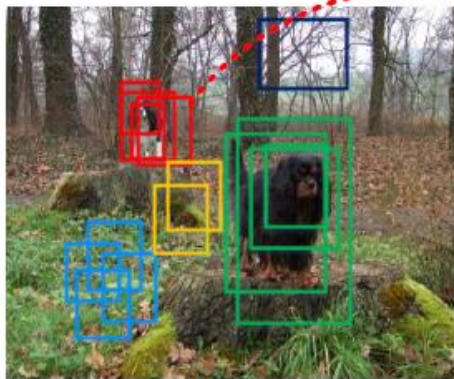


- (1) Instance (object and object part) are collected with a **clique partition module**;
- (2) Object **clique discovery** with a global min-entropy model;
- (3) **Object localization** with a local min-entropy model

# Our works-**Min-entropy** latent model



Proposals



Cliques

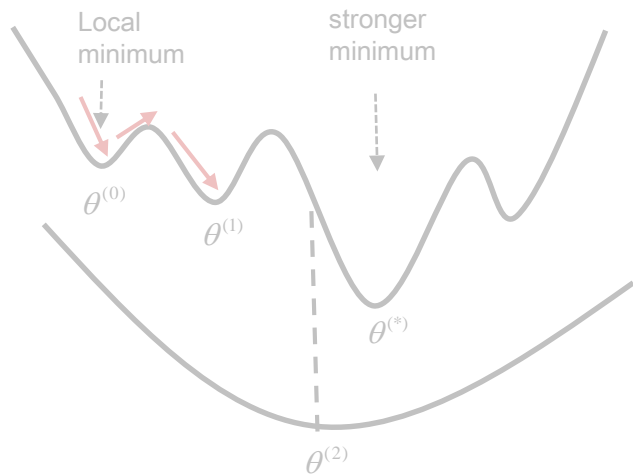


Object Cliques

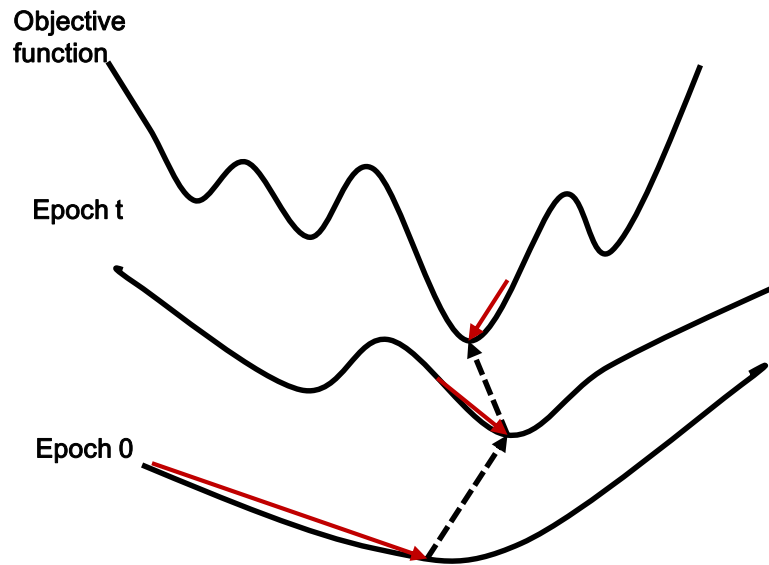
Clique partition:

$$\begin{cases} \bigcup_{c=1}^C \mathcal{H}_c = \tilde{\mathcal{H}} \\ \forall c \neq c', \mathcal{H}_c \cup \mathcal{H}_{c'} = \emptyset \end{cases}$$

# Our works-Methodology

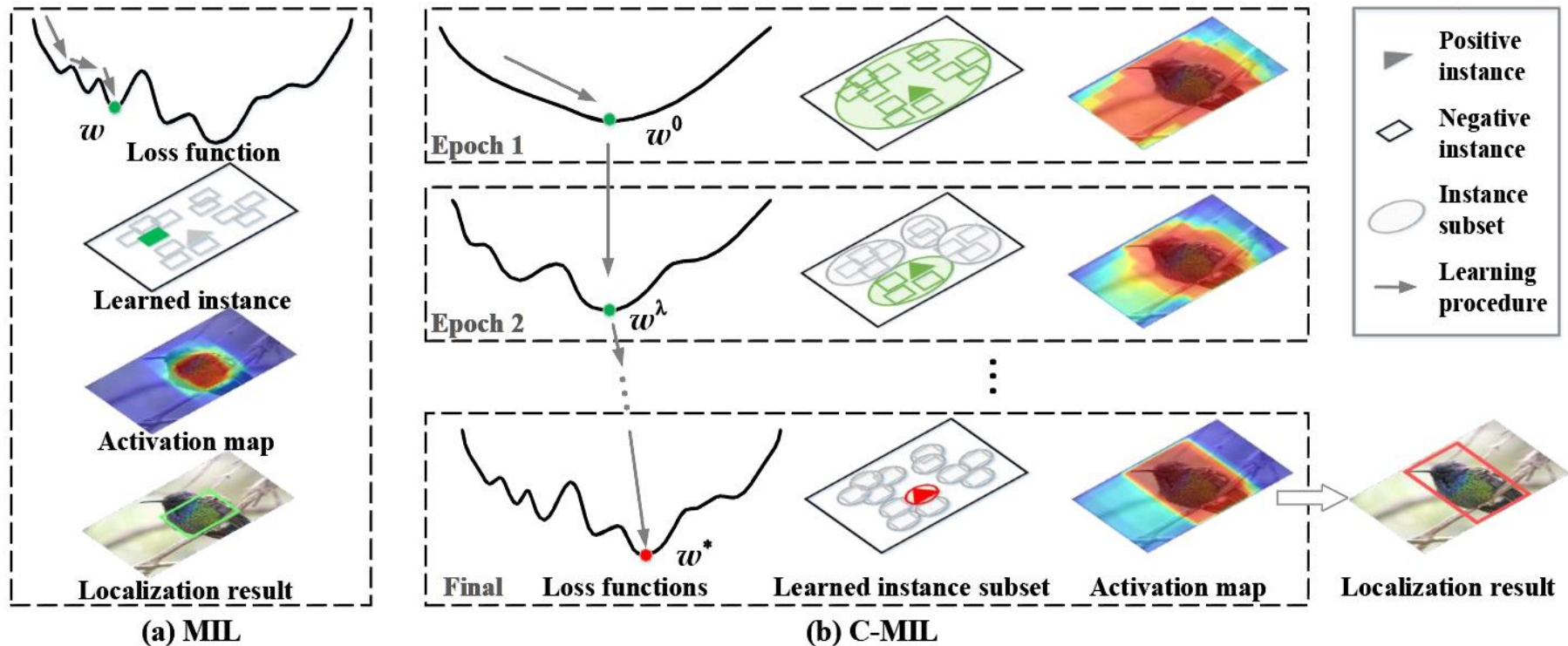


Convex Regularization

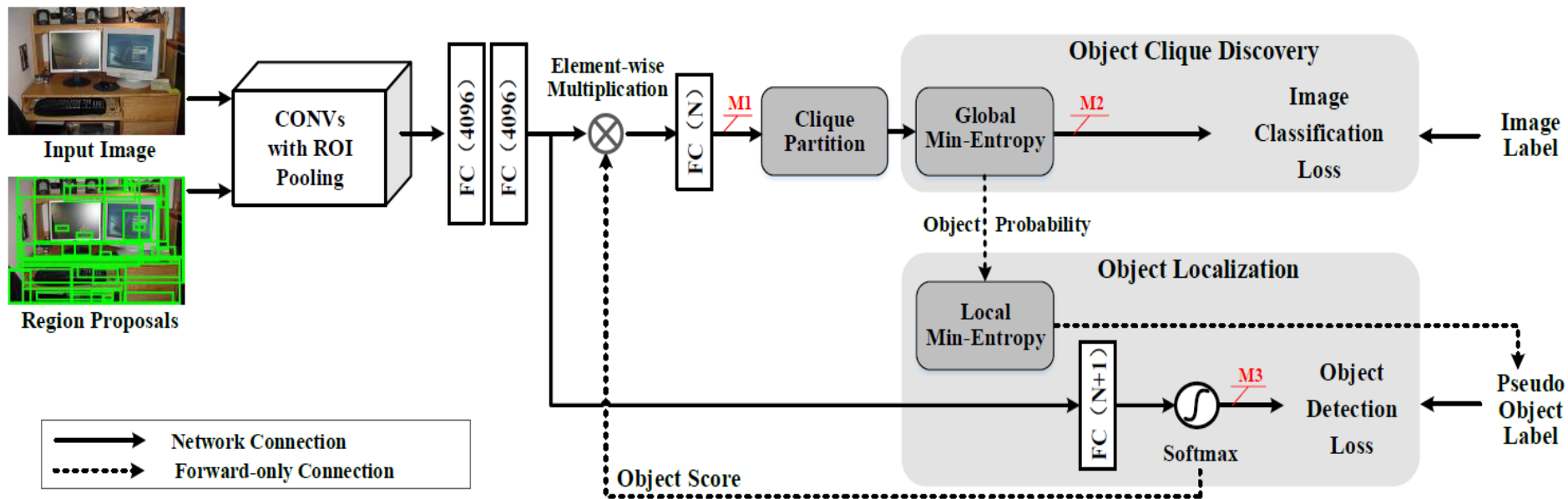


Continuation Optimization

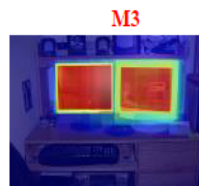
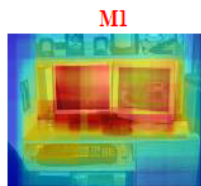
# Our works-Continuation Multiple Instance Learning



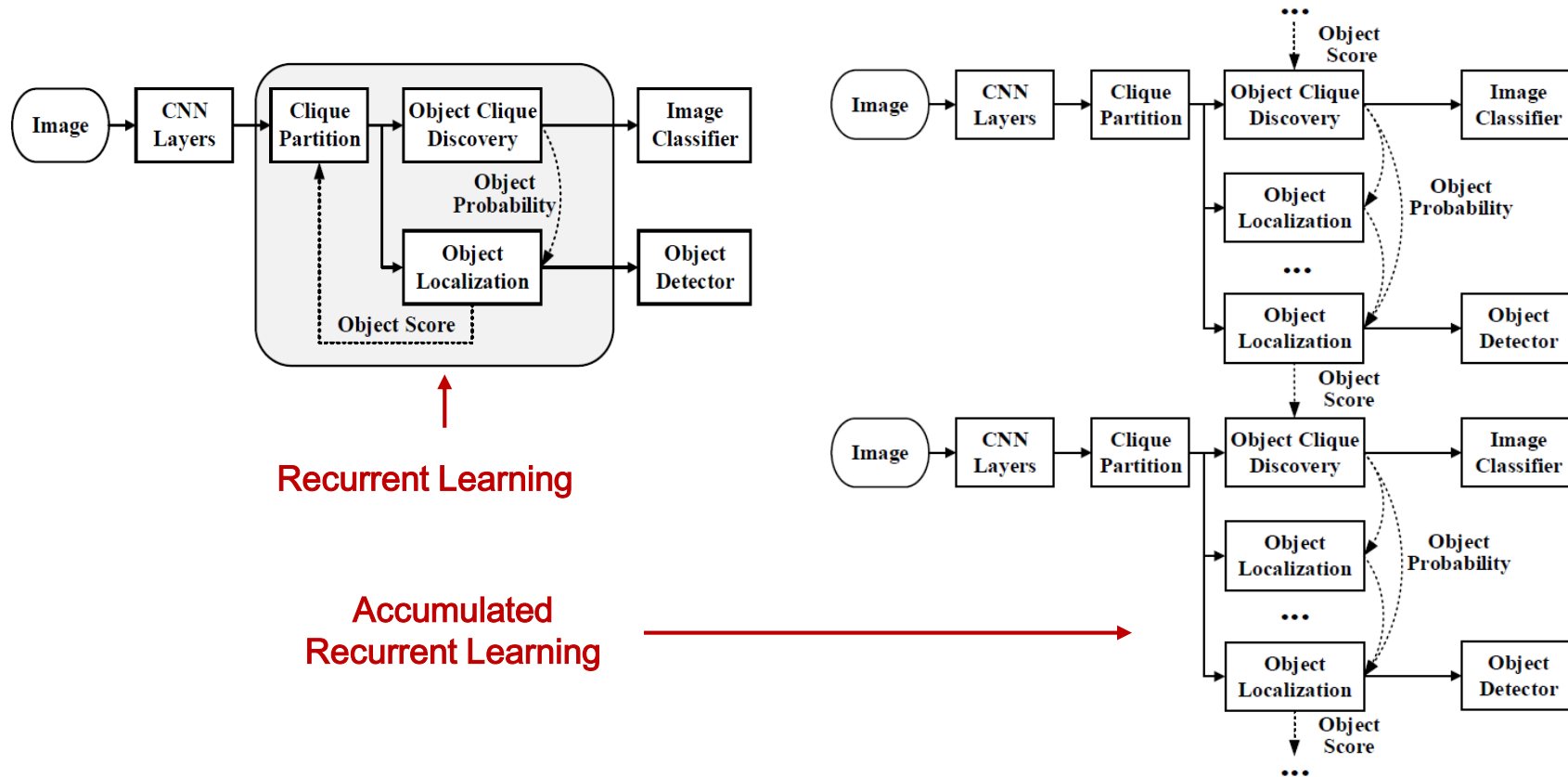
# Our works-**Min-entropy** latent model



Object Score Heatmap

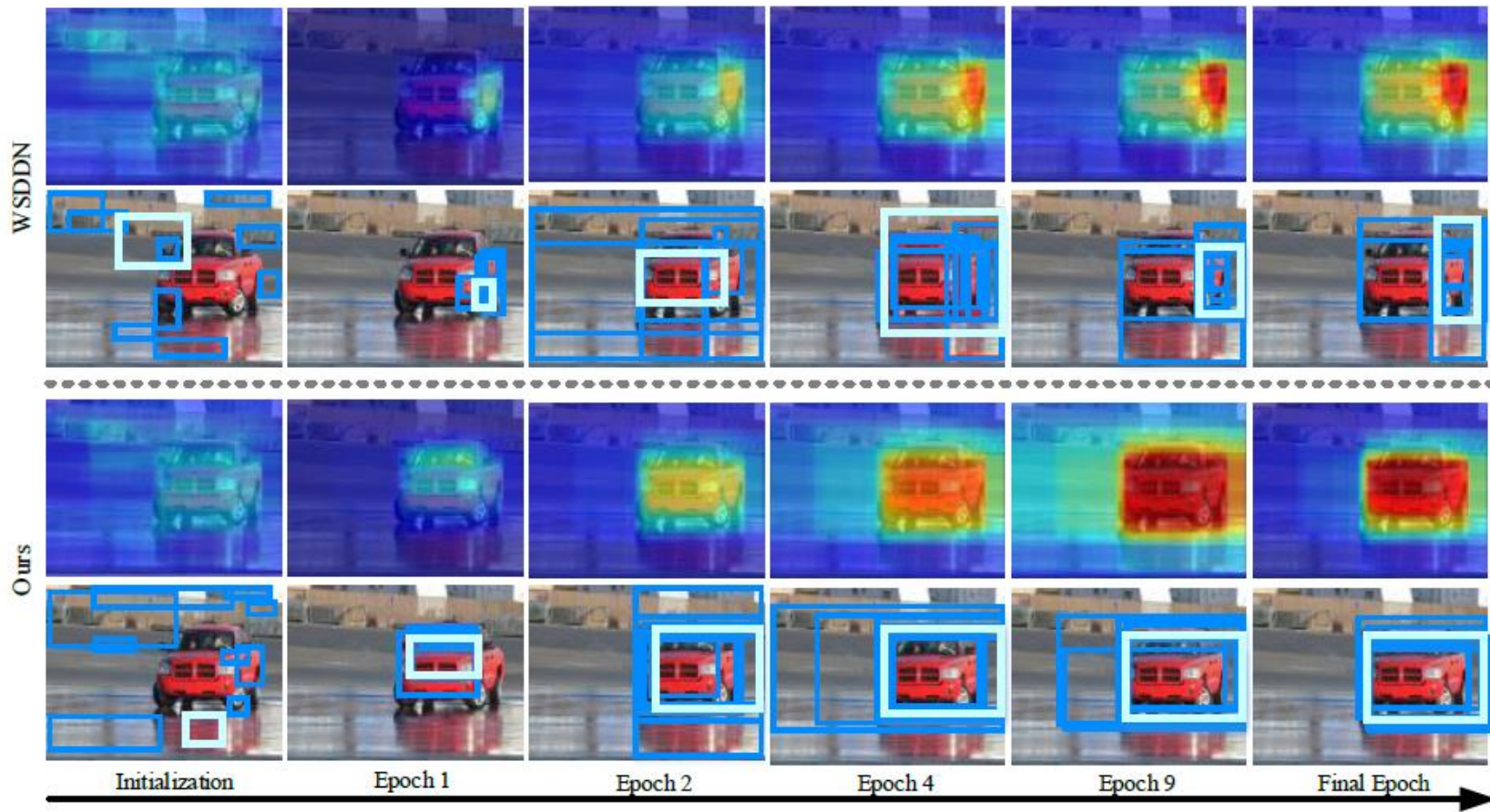


# Our works-**Recurrent Learning**

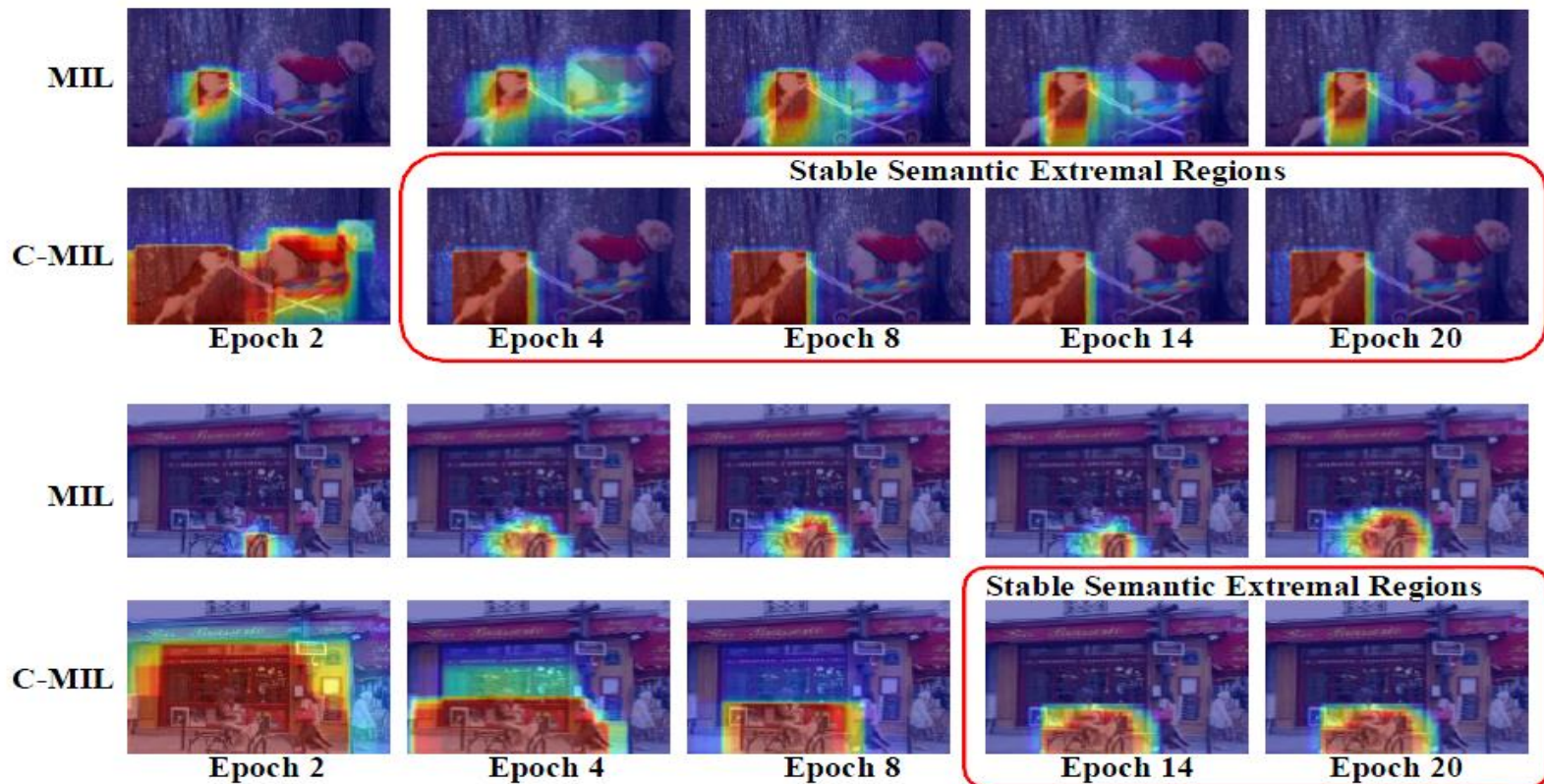


F. Wan, P. Wei, Z. Han, J. Jiao, Q. Ye, "Min-entropy Latent Model for Weakly Supervised object Detection," IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), DOI:10.1109/TPAMI.2019.2898858.

# Our works-Results

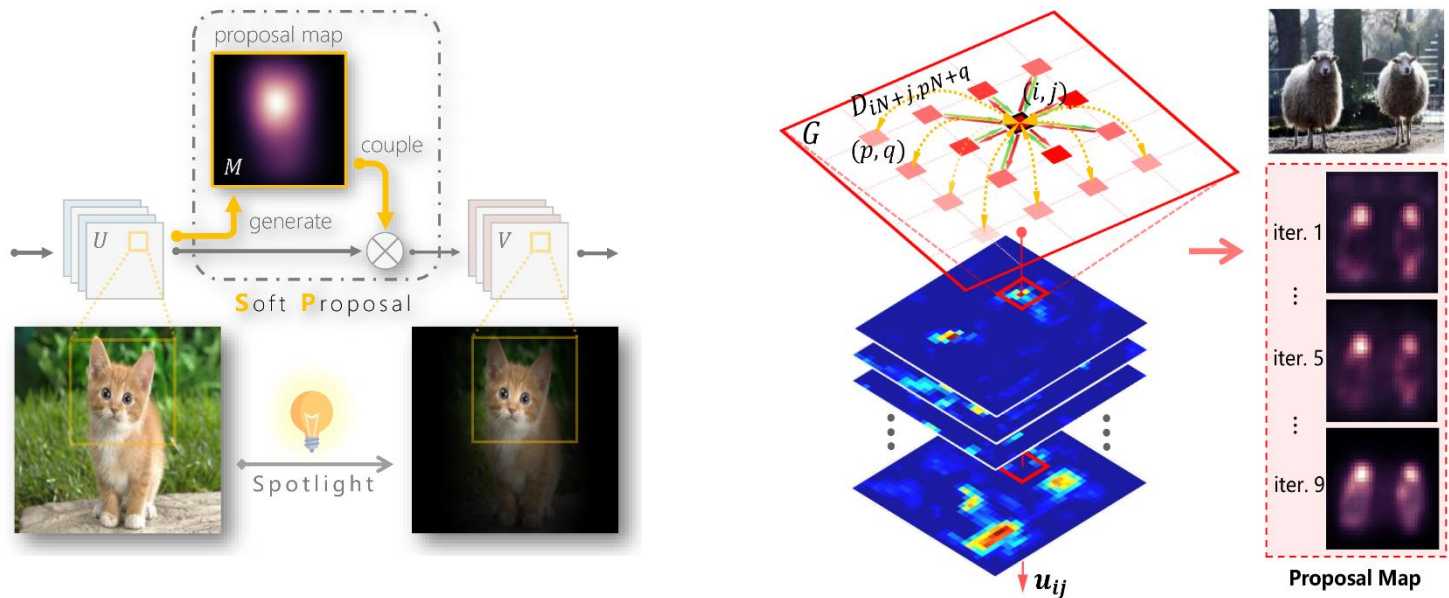


# Our works-Results



**SSER:** Semantic Stable Extremal Region

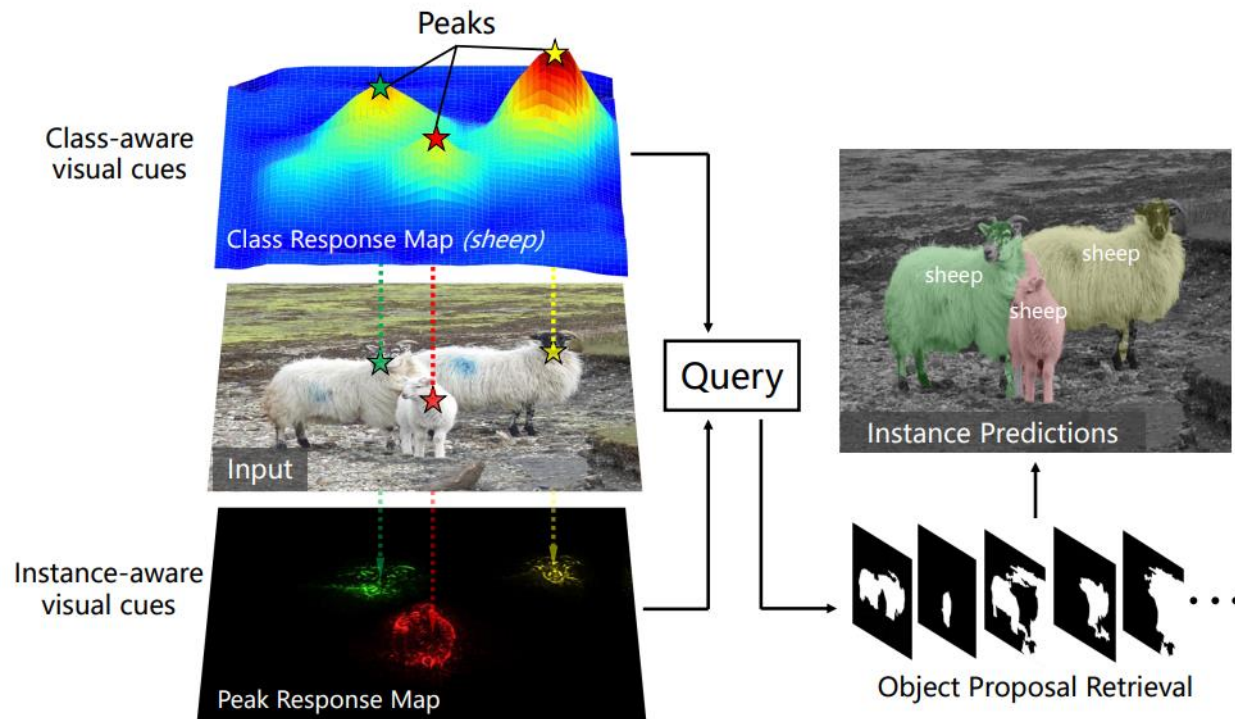
# Our works-Soft Proposal Network



$$M \leftarrow D \times M. \quad M \leftarrow D(U^l(W^l)) \times M. \quad W^l = W^l + \Delta W(M)$$

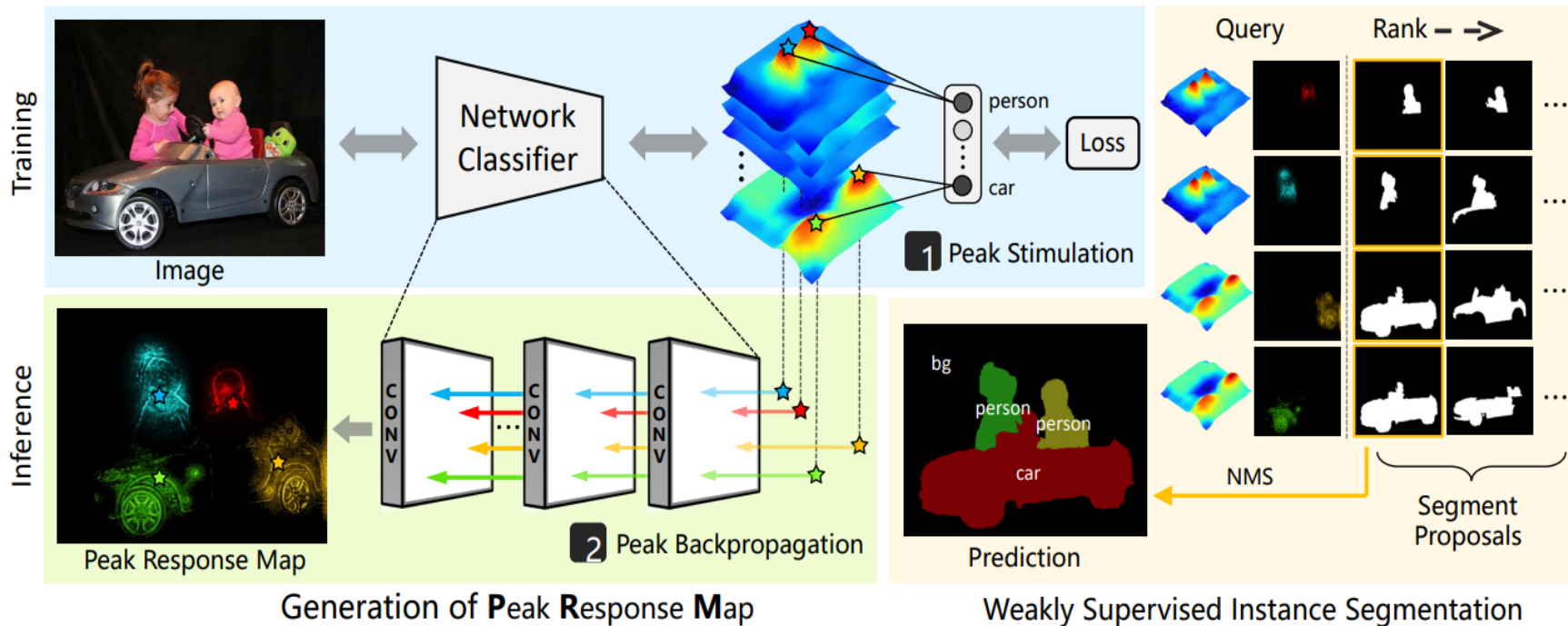


# Our works-**Peak Response Mapping**

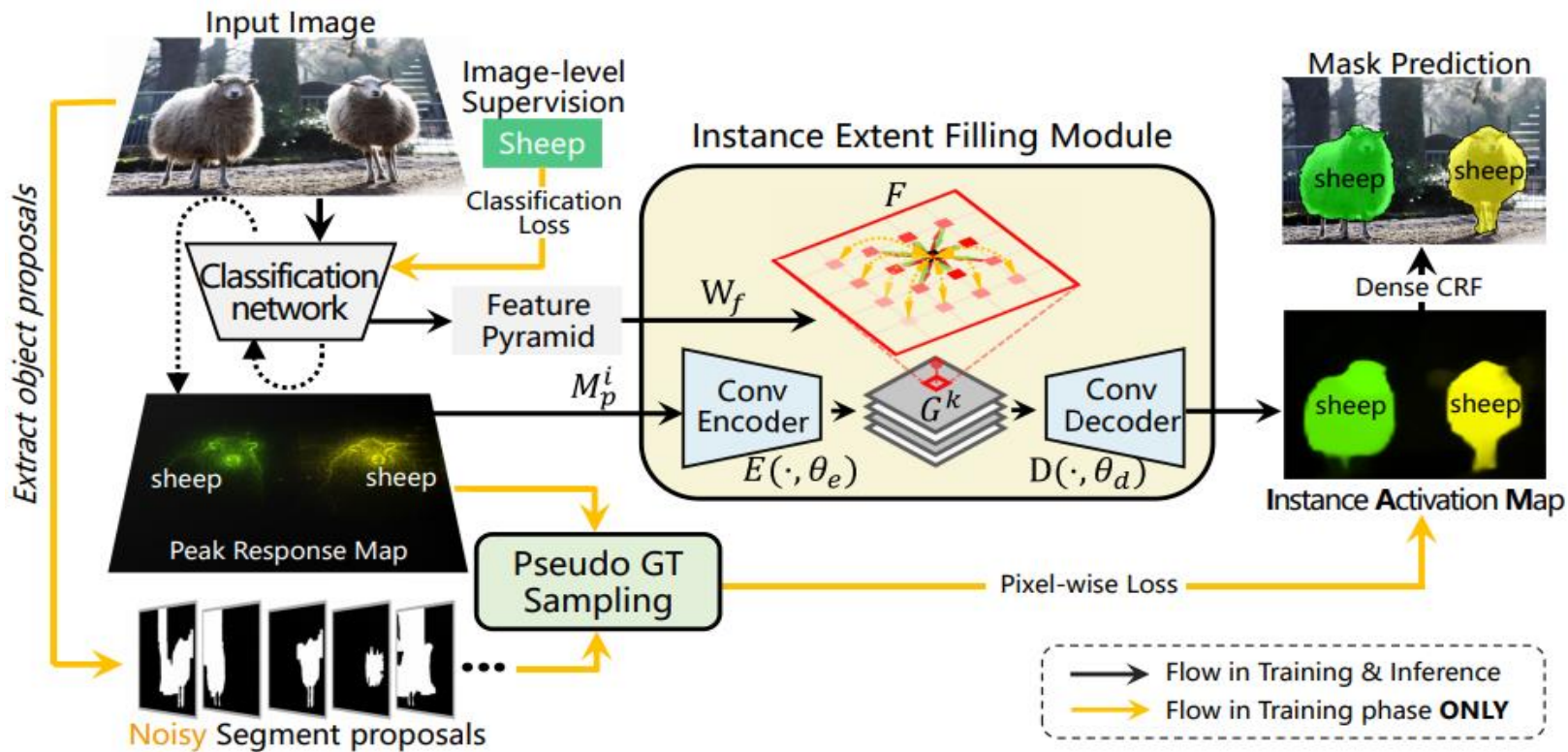


Y. Zhou, Y. Zhu, Q. Ye, Q. Qiu, J. Jiao, "Weakly Supervised Instance Segmentation using Class Peak Response, IEEE CVPR 2018 (Spotlight).

# Our works-**Peak Response Mapping**

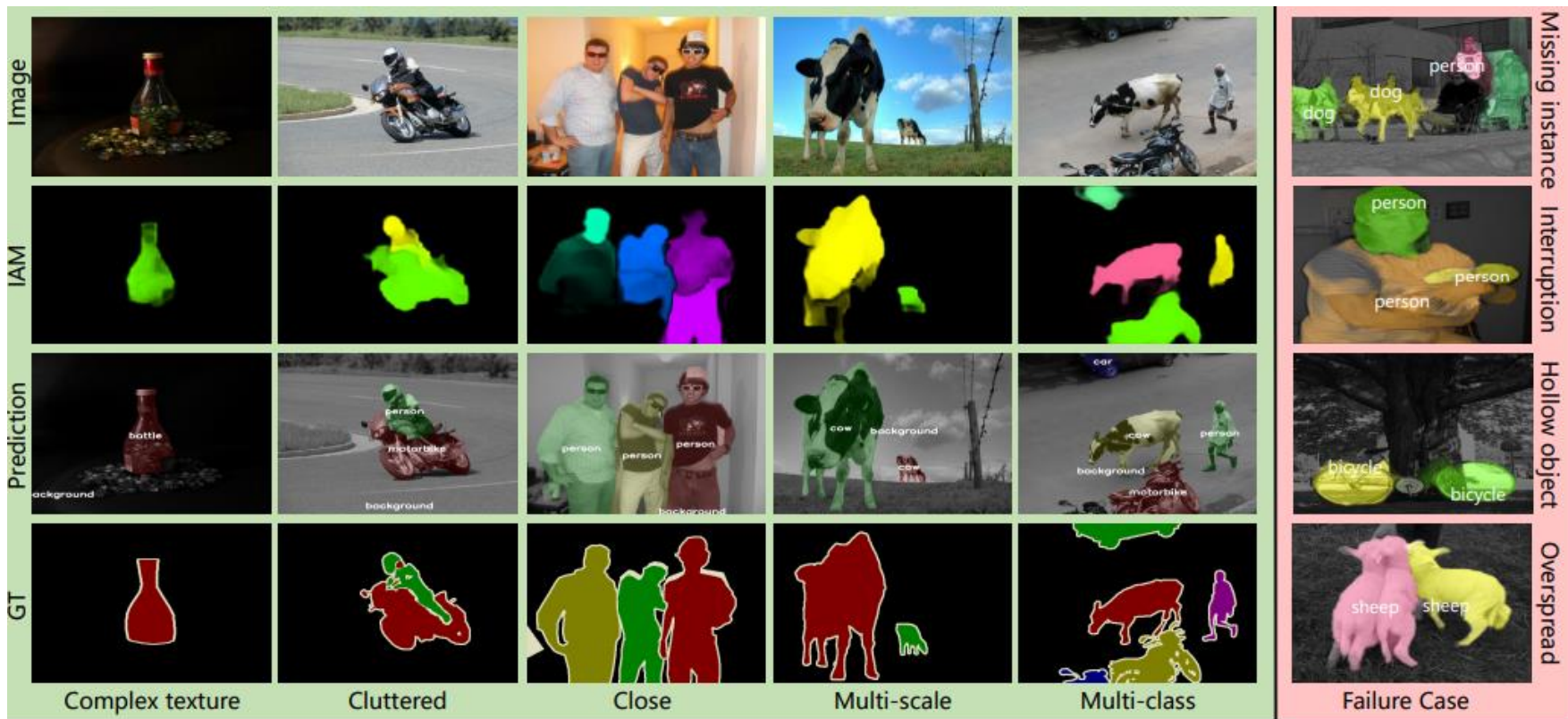


# Our works-learning Instance Activation Maps



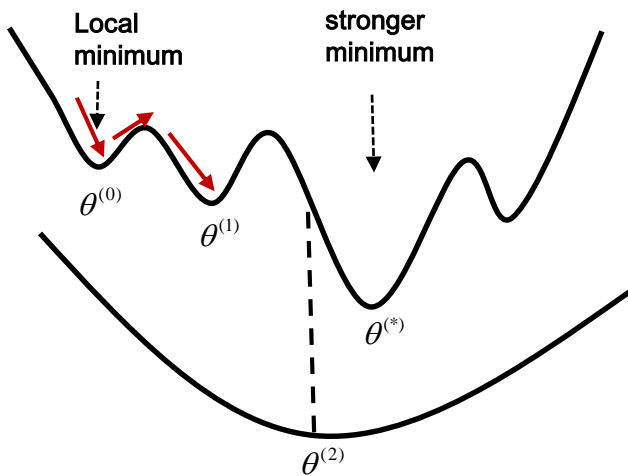
Y. Zhu, Y. Zhou, H. Xu, Q. Ye., D. Doermann, J. Jiao, "Learning Instance Activation Maps for Weakly Supervised Instance Segmentation," IEEE CVPR 2019.

# Our works-learning Instance Activation Maps

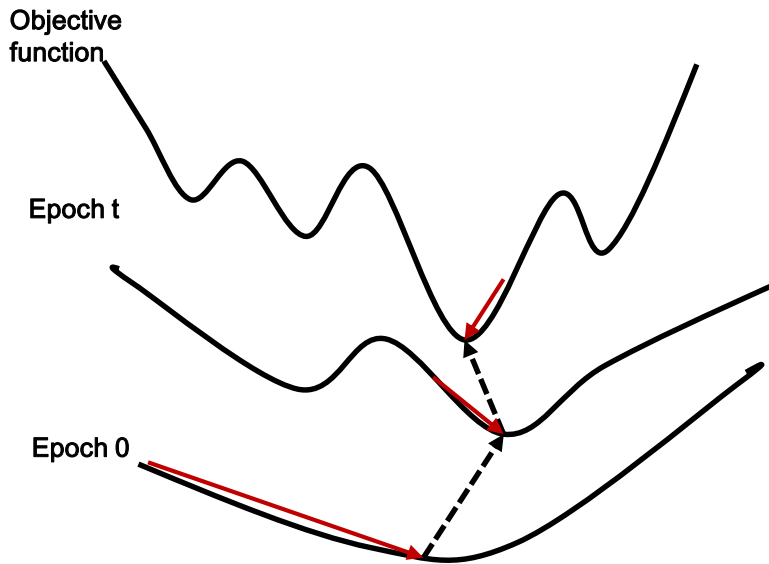


# The future

## Beyond regularization and continuation optimization



Convex Regularization



Continuation Optimization

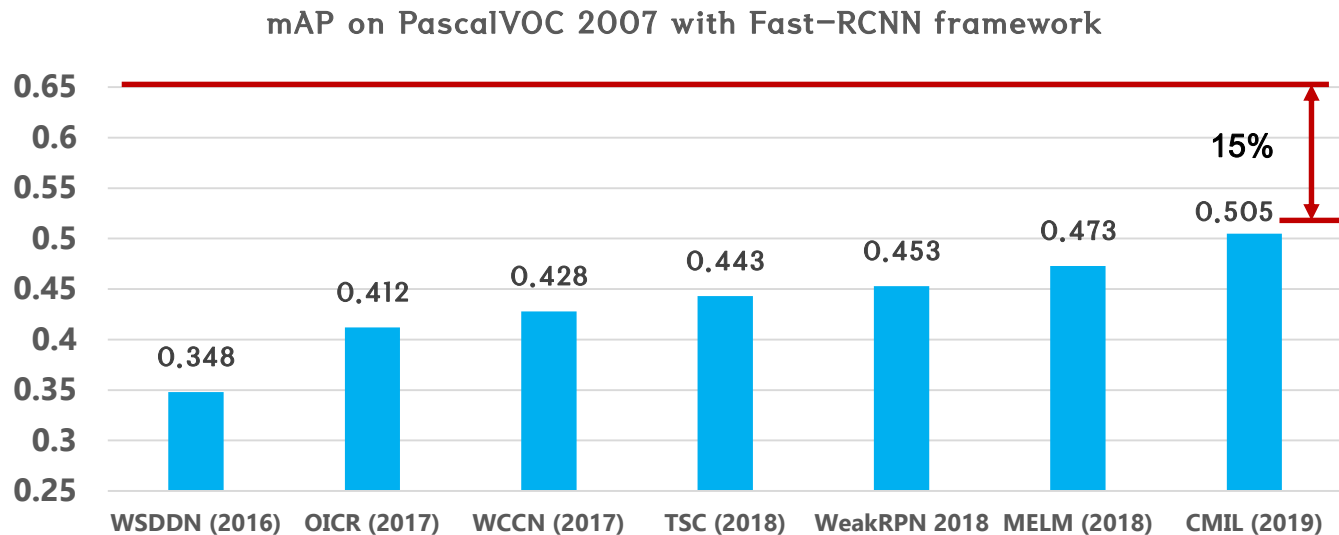
# The future

Beyond weakly supervised detection and segmentation



# The future

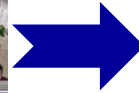
Fill the gap of supervised and weakly supervised methods



# The future

Weakly supervised detection **meets X**

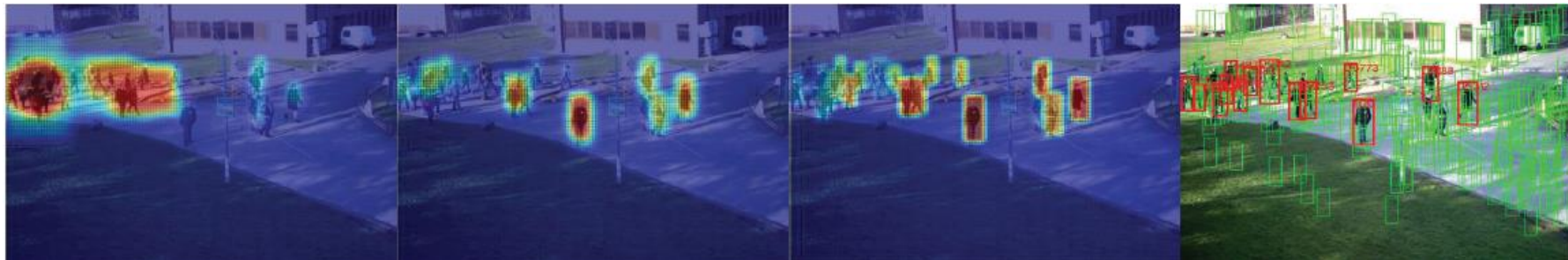
X= Few-shot Active Learning | Online Feedback | Temporal



# The future

X= Few-shot Active Learning | Online Feedback | **Temporal**

Pets2009 (crowd)



Towncenter (moving distractors)

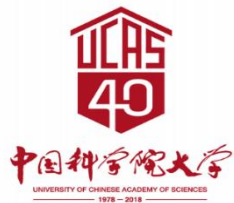


Q. Ye, Z. Zhang, Q. Qiu, B. Zhang, J. Chen, and G. Sapiro, "Self-learning Scene-specific Pedestrian Detectors using a Progressive Latent Model," IEEE CVPR, 2017

# Ref.

- [1] F. Wan, P. Wei, Z. Han, J. Jiao, Q. Ye, “Min-entropy Latent Model for Weakly Supervised object Detection,” IEEE Trans. PAMI, DOI:10.1109/TPAMI.2019.2898858. **(MELM+Recurrent Learning)**
- [2] F. Wan, C. Liu, J. Jiao, Q. Ye, “CMIL: Continuation Multiple Instance Learning for Weakly Supervised object Detection (CVPR2019) **(C-MIL)**
- [3] Y. Zhu, Y. Zhou, H. Xu, Q. Ye., D. Doermann, J. Jiao, “Learning Instance Activation Maps for Weakly Supervised Instance Segmentation,” IEEE CVPR 2019. **(IAM)**
- [4] P. Tang, X. Wang, S. Bai, W. Shen, X. Bai, W. Liu, and A. L. Yuille, “Pcl: Proposal cluster learning for weakly supervised object detection,” IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2018. **(PCL)**
- [5] Y. Zhou, Y. Zhu, Q. Ye, Q. Qiu, J. Jiao, “Weakly Supervised Instance Segmentation using Class Peak Response,” in Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2018 (Spotlight). **(PRM)**
- [6] F. Wan, P. Wei, Z. Han, J. Jiao, Q. Ye, “Min-entropy Latent Model for Weakly Supervised object Detection,” in Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR), 2018: 1297-1306. **(MELM)**
- [7] Y. Zhu, Y. Zhou, Q. Ye, Q. Qiu, and J. Jiao, "Soft Proposal Network for Weakly Supervised Object Localization," in Proc. of IEEE Int. Conf. on Computer Vision (ICCV), 2017. **(SPN)**
- [8] Q. Ye, Z. Zhang, Q. Qiu, B. Zhang, J. Chen, and G. Sapiro, "Self-learning Scene-specific Pedestrian Detectors using a Progressive Latent Model," IEEE CVPR, 2017 **(Self-Learning)**
- [9] B. Hakan and V. Andrea, “Weakly supervised deep detection networks,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp. 2846–2854. **(WSDDN)**
- [10] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba, “Learning deep features for discriminative localization,” in Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit. (CVPR), 2016, pp.2921–2929. **(CAM)**

# Thank!



[www.ucassdl.cn](http://www.ucassdl.cn)

[qxye@ucas.ac.cn](mailto:qxye@ucas.ac.cn)

[people.ucas.ac.cn/~qxye](http://people.ucas.ac.cn/~qxye)