Microsoft Research 微软亚洲研究院

Unsupervised Object Discovery via Saliency-Guided Multiple Class Learning

Jun-Yan Zhu[1,2], Jiajun Wu[1,2], Yichen Wei[1], Eric Chang[1], Zhuowen Tu[1,3]

1 Microsoft Research Asia, 2 Tsinghua University, 3 UCLA

Ambiguity of Unsupervised Object Discovery

(a) (b)

- Easy to cluster images in (a), but hard to cluster those in (b)
- intrinsic ambiguity of the complex object appearances and the background clutter

Problems with the Existing MIC and Clustering Methods

- **MIC**: treat all the images as positive bags without considering the uses of negative bags
- **Unsupervised Clustering**: strict constraints include large occupation of the foreground objects, clean background, …

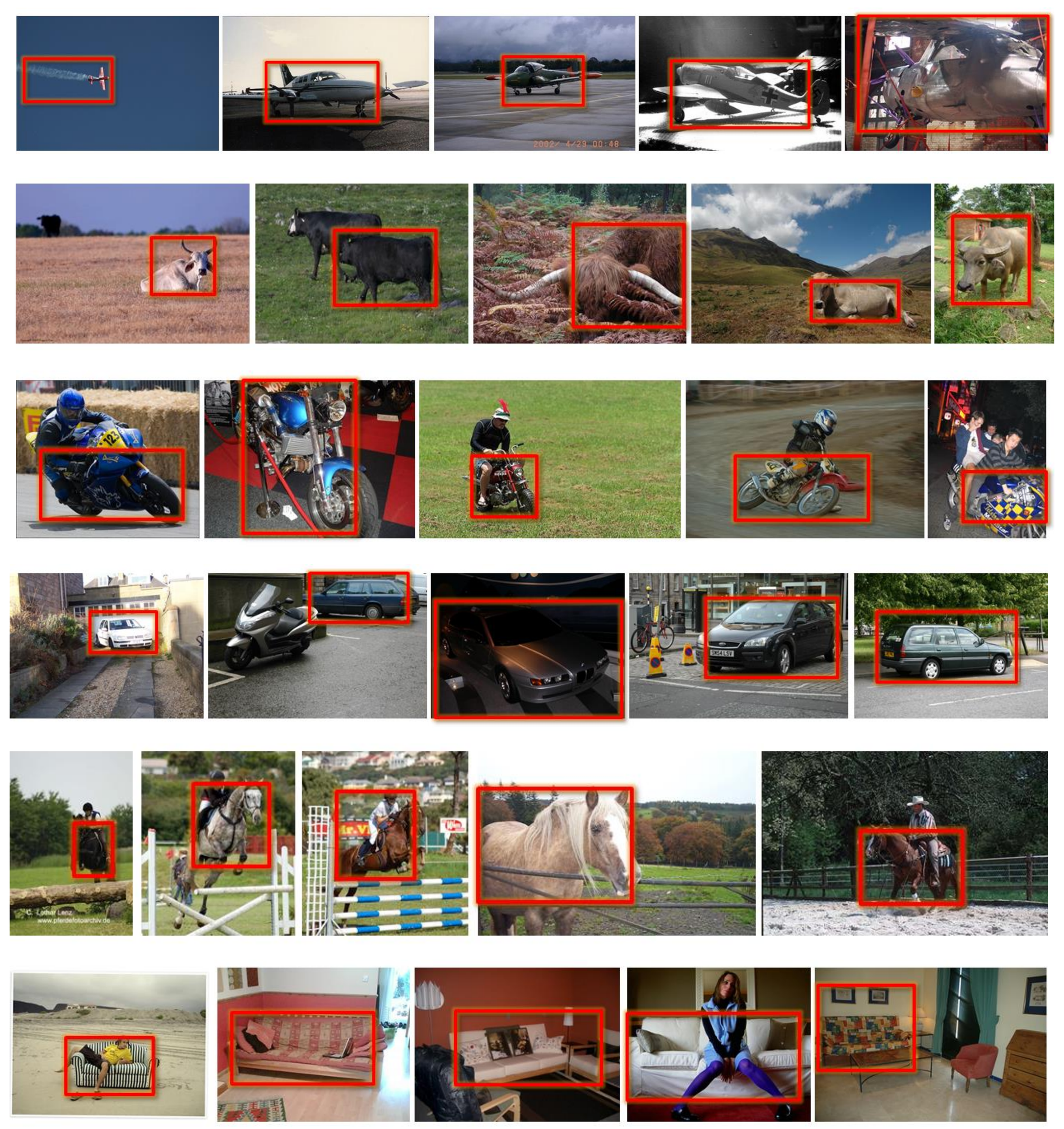
- Cannot perform localization
- Cannot train object classifiers

- Cannot distinguish between objects and background

Weakly Supervised Single Class Learning

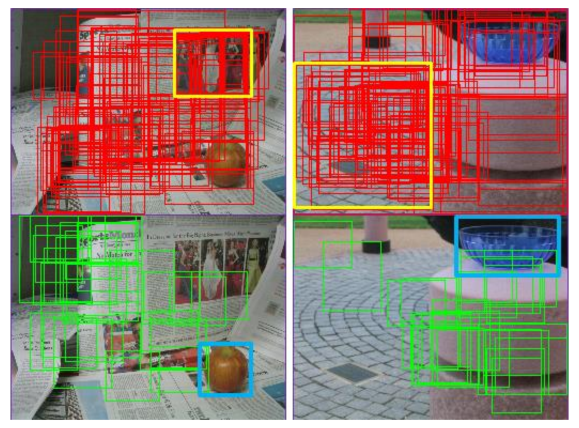
	bMCL	[11]	[9]	[37]	[33]
PASCAL 06-all	45	**49**	34	27	N/A
PASCAL 07-all	**31**	28	19	14	30

comparison with previous weakly supervised learning methods, measured in CorLoc



object localization results of bMCL with a single object class on the challenging PASCAL VOC 07

Saliency-Guided Notion



Blue rectangle: the desired object window.
Yellow rectangle: the most salient window.
Red rectangles: the positive bag.
Green rectangles: the negative bag.

Idea: Although the complex background may create some false detections, objects are mostly covered in the top ranked windows.

Unsupervised ⇒ Weakly Supervised

For each image

- a positive (object) bag consists of the most salient windows, and
- a negative (background) bag consists of the least salient window from a large set of randomly sampled windows

Bottom-up Multiple Class Learning

Maximum Margin Clustering ⇒ Multiple Instance Learning

Goal:

- Discriminate the positive (object) instances from negative (background) instances;
- Maximize the differences between different object classes in the positive bags.

Algorithm 1 Bottom-up Multiple Class Learning

Input: Bags $\{x_1,\ldots,x_n\},\{y_1,\ldots,y_n\},T,K,H_K^0$.
Output: K discriminative classifiers: $h^1,\ldots,h^K$.
$r \leftarrow 0$.
repeat
  $r \leftarrow r+1$.
  for $k=1\rightarrow K$ **do** {**M** Step}
    Given class variables H_K^{r-1}, group terms $\mathcal{L}^k(h_r^k;Y,X,H_K^{r-1})$ by class indices.
    Train a strong MIL classifier h_r^k to minimize $\mathcal{L}^k(h_r^k;Y,X,H_K^{r-1})$ via MIL-Boost. T is the number of weak classifiers in MIL-Boost.
  end for
  for $i=1\rightarrow n$ **do** {**E** Step}
    Compute $\Pr(y_i=1,k_i=k|x_i;\theta_r)$ using estimated model $\theta_r=\{h_r^1,\ldots,h_r^K\}$. Sample k_i via $\Pr(k_i=k|y_i=1,x_i;\theta_r)\sim\Pr(y_i=1,k_i=k|x_i;\theta_r)$.
  end for
until $H_K^r=H_K^{r-1}$

Discriminative EM

Goal: optimize the training set log likelihood $\mathcal{L}(\theta;Y,X)$ *w.r.t.* model parameters in the presence of hidden variable H.

$$\frac{d}{d\theta}\mathcal{L}(\theta;Y,X)=\mathbb{E}_{H\sim\Pr(H|Y,X;\theta)}\left[\frac{d}{d\theta}\mathcal{L}(\theta;Y,X,H)\right]$$

where $\mathcal{L}(\theta;Y,X,H)=-\log\Pr(Y,H|X;\theta)$

$$\Pr(H|Y,X;\theta)=\frac{\Pr(Y,H|X;\theta)}{\Pr(Y|X;\theta)}$$

Approach: iteratively update an initial estimate θ_0 with successively better estimates $\theta_1,\theta_2,\ldots$, until convergence.

E step: Compute $\Pr(H\mid Y,X;\theta)$ via previous estimate r.
M step: Update θ_{r+1} by minimizing $\mathcal{L}(\theta;Y,X)$.

DiscEM for bMCL

Definition: $H=(H_K,H_I)$ are hidden variables where

- $H_K=\{k_i\mid i=1,\ldots,n\}$ are class latent labels.
- $H_I=\{y_{ij}\mid i=1,\ldots,n,j=1,\ldots,m\}$ are instance labels.

$$\frac{d}{d\theta}\mathcal{L}(\theta;Y,X)=\mathbb{E}_{H_K\sim\Pr(H_K|Y,X,\theta)}\left[\frac{d}{d\theta}\mathcal{L}(\theta;Y,X,H_K)\right]$$

where $\mathcal{L}(\theta;Y,X,H_K)=\sum_{k=1}^{K}\mathcal{L}^k(h^k;Y,X,H_K)$

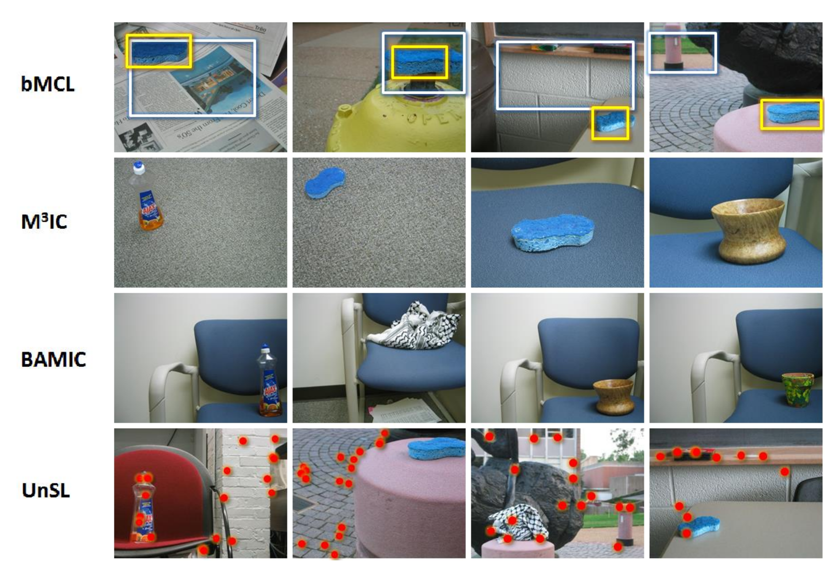
For each $\mathcal{L}^k(h^k;Y,X,H_K)$, hidden instance variables H_I could be further integrated out as

$$\frac{d}{d\theta}\mathcal{L}^k(h^k;Y,X,H_K)=$$
$$\mathbb{E}_{H_I\sim\Pr(H_I|Y,H_K,X;\theta)}\left[\frac{d}{d\theta}\mathcal{L}^k(h^k;Y,X,H)\right]$$

Unsupervised Object Discovery

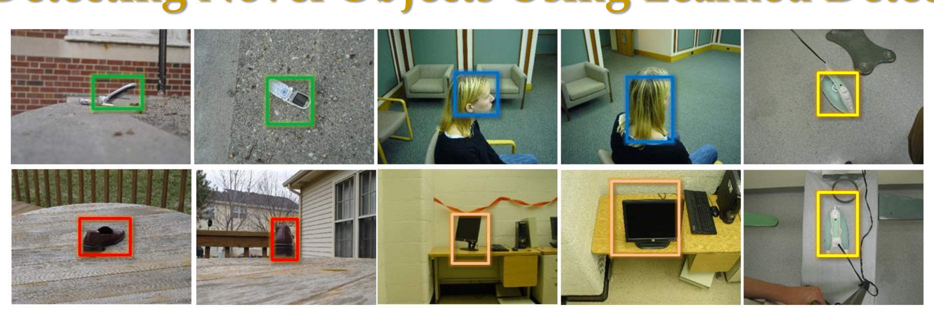
	bMCL	SD	M³IC	BAMIC	UnSL
SIVAL1	**95.3**	80.4	39.3	38.0	27.0
SIVAL2	**84.0**	71.7	40.0	33.3	35.3
SIVAL3	**74.7**	62.7	37.3	38.7	26.7
SIVAL4	**94.0**	86.0	33.0	37.7	27.3
SIVAL5	**75.3**	70.3	35.3	37.7	25.0
CC	**80.0**	73.9	46.1	47.8	60.0
3D1	**81.1**	64.0	46.9	43.2	37.3
3D2	**85.6**	82.9	52.3	51.4	37.5

measured in terms of purity



SIVAL dataset

Detecting Novel Objects Using Learned Detectors

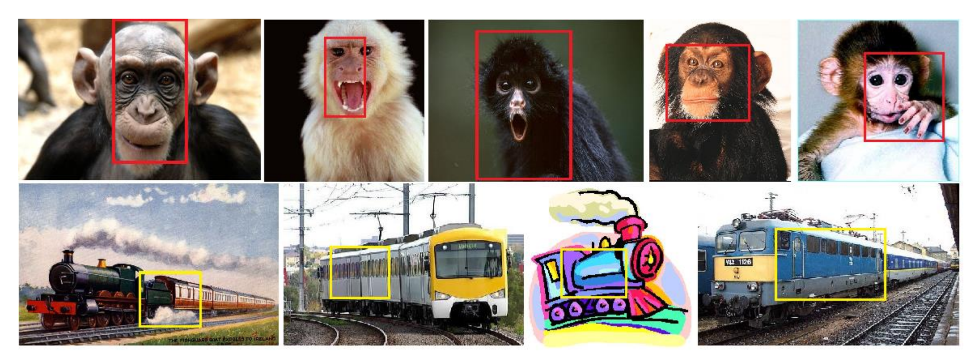


3D object category dataset

Internet Images

Internet	bMCL	SD	M³IC	BAMIC	UnSL
Accuracy	**96.25**	88.75	66.25	75.00	56.25
Purity	**96.25**	88.75	66.25	75.00	56.25
NMI	**77.22**	49.99	8.12	25.75	11.28

clustering results of images returned by Google and Bing image search, measured in terms of three different metrics



Unsupervised object discovery

Detecting novel objects using learned detectors

Conclusions

Bottom-up Multiple Class Learning performs

- object class discovery,
- object localization, and
- object detector training

in an integrated framework.

Advantages:

- Adopt saliency detection to convert unsupervised learning into multiple instance learning;
- Develop Discriminative EM (DiscEM) to solve bMCL;
- Perform localization, object class discovery and object detector training in an integrated framework;
- Observe significant improvements over the existing methods for multi-class object discovery;
- Show single class localization as a special case in our bMCL framework;
- Demonstrate advantages of bMCL over purely data-driven saliency methods;
- Apply bMCL on internet images to verify its generality.

Co-saliency

	apple	book	candle	note	scrunge
bMCL	0.65	**0.75**	**0.66**	**0.65**	**0.68**
[12]	**0.69**	0.74	0.62	0.54	0.61
[4]	0.49	0.71	0.43	0.62	0.52

evaluation of co-saliency in F-measure on SIVAL dataset