Object-centric spatial pooling for image classification

Olga Russakovsky, Yuanqing Lin,
Kai Yu, Li Fei-Fei

ECCV 2012

Russakovsky et al. ECCV 2012

Image classification

Testing: Does this image contain a car?

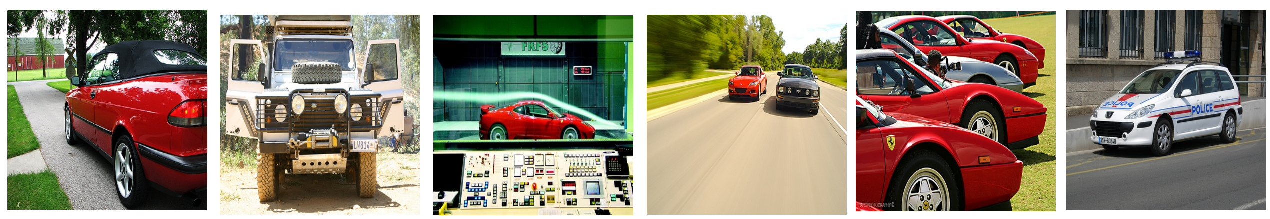
IMAGENET

PASCAL2
Pattern Analysis, Statistical Modelling
Computational Learning

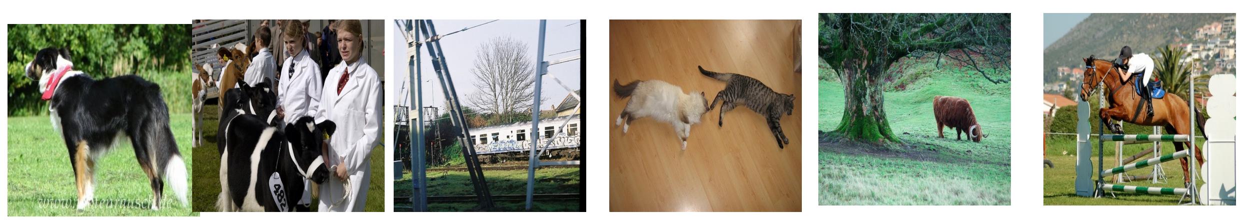
LabelMe SUN database

Training:

cars



not cars



Russakovsky et al. ECCV 2012

Proof of concept experiment

Testing: Does this image contain a car?

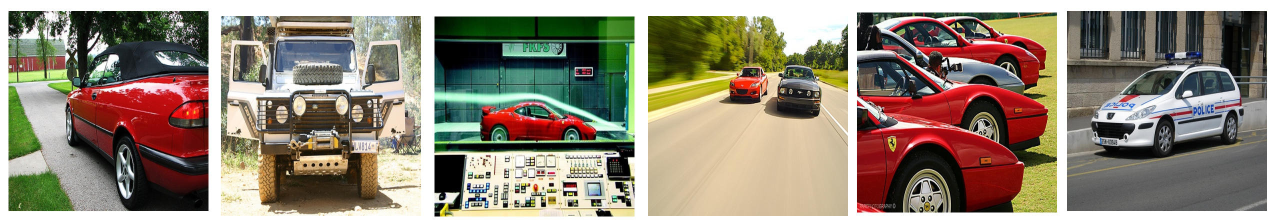
IMAGENET

PASCAL2 Pattern Analysis, Statistical Modelling Computational Learning

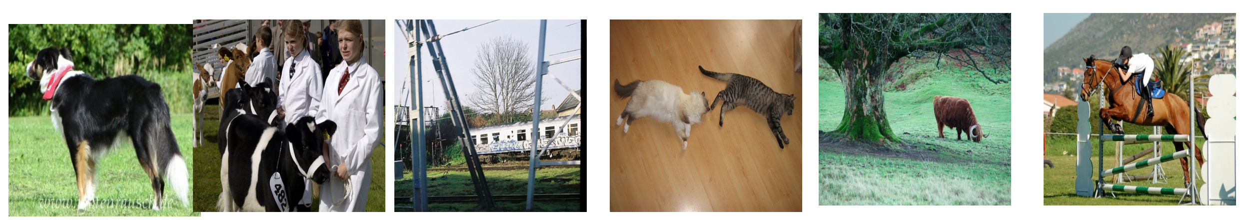
LabelMe SUN database

Training:

cars



not cars



Proof of concept experiment

Testing:

Does this image contain a car?

Build an image classification system

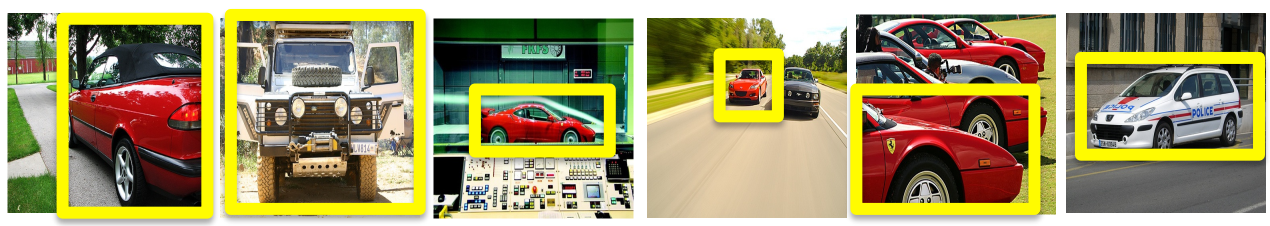
PASCAL2
Pattern Analysis, Statistical Modelling Computational Learning

PASCAL07 val, 20 classes,
DHOG features, LLC coding 8K codebook,
1x1,3x3 SPM, linear SVM

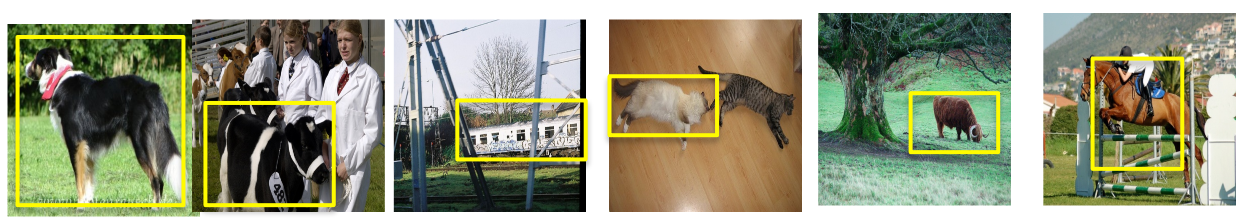
Full images	Cropped objects
52.0 mAP	**69.7 mAP**

Training:

cars



not cars



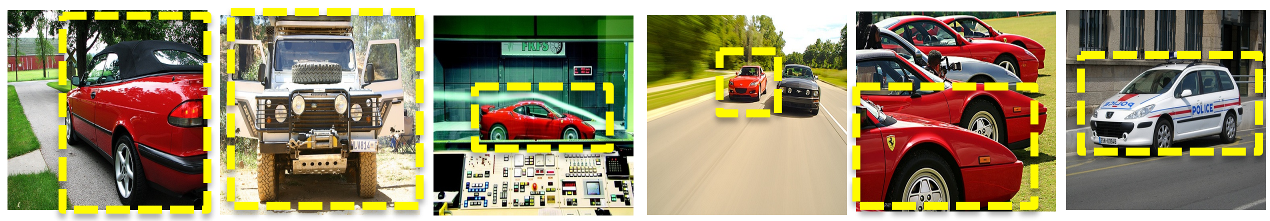
Inferring object locations for classification

Testing: Does this image contain a car?

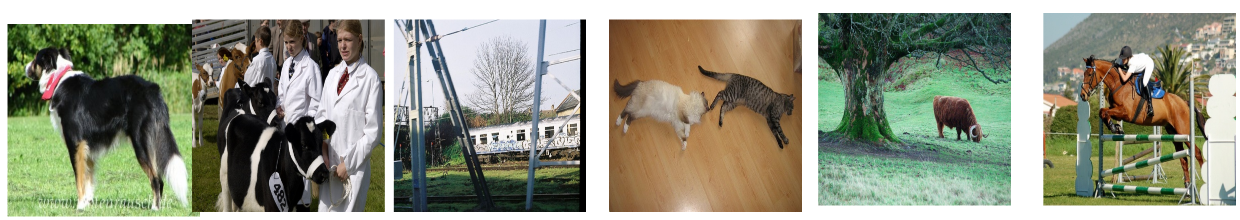
Challenges:
1. *Weakly supervised localization* during training
2. Inferring inaccurate localization will make classification impossible

Training:

cars



not cars



Outline

Object-centric spatial pooling (OCP) image representation

Training the OCP model as a joint image classification and object localization model

Results

- Improved image classification accuracy
- Competitive weakly supervised localization accuracy

Russakovsky et al. ECCV 2012

Image classification system

Image

Low-level
visual features

DHOG features,
LLC coding 8K codebook

.3
1
.2
-.5
...

Image-level
representation

Classifier

Model

Linear SVM

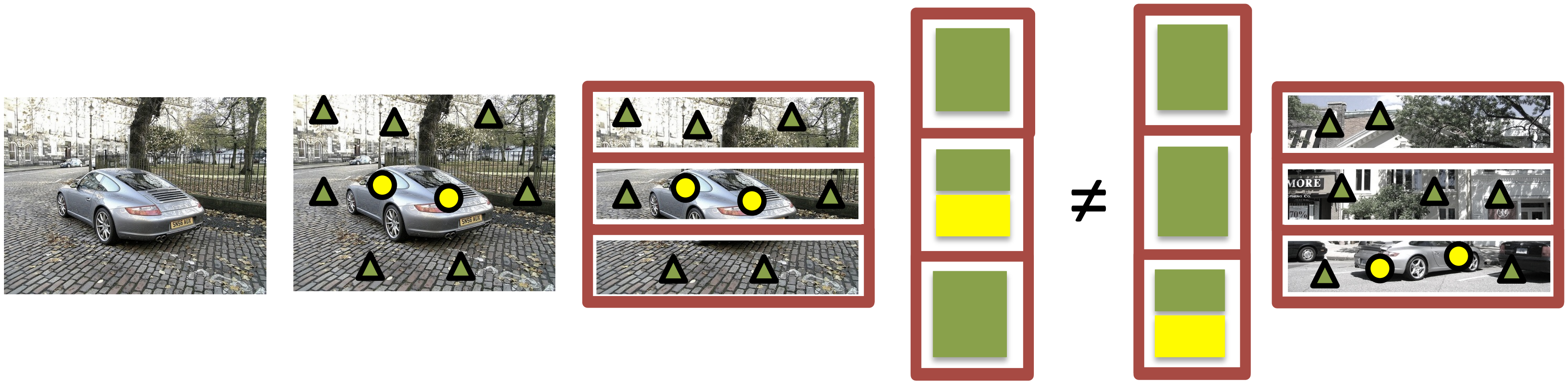
Yes

Result

Russakovsky et al. ECCV 2012

Standard representation: SPM pooling

The Spatial Pyramid Matching (SPM) approach forms the image representation by pooling visual features over pre-defined coarse spatial bins.



SPM-based pooling results in inconsistent image representations when the object of interest appears in different locations within the image.

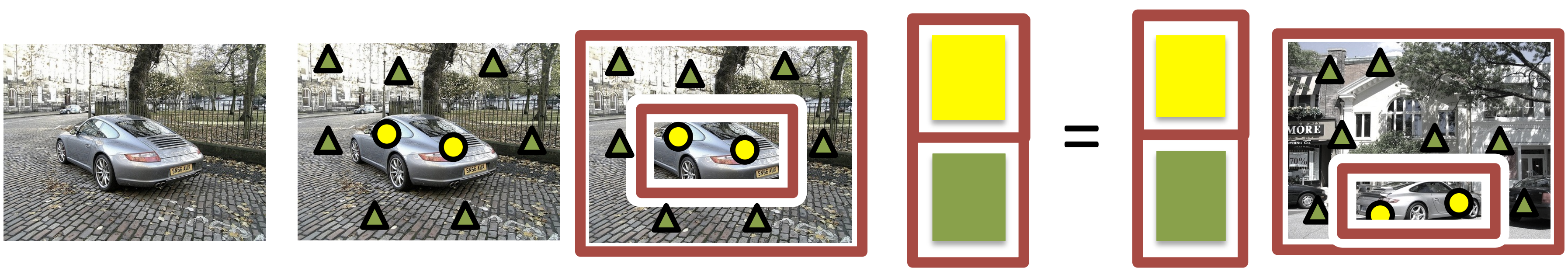
Russakovsky et al. ECCV 2012

Object-centric spatial pooling

We propose an object-centric spatial pooling (OCP) approach which

(1) localizes the object of interest, and then

(2) pools foreground visual features separately from the background features.

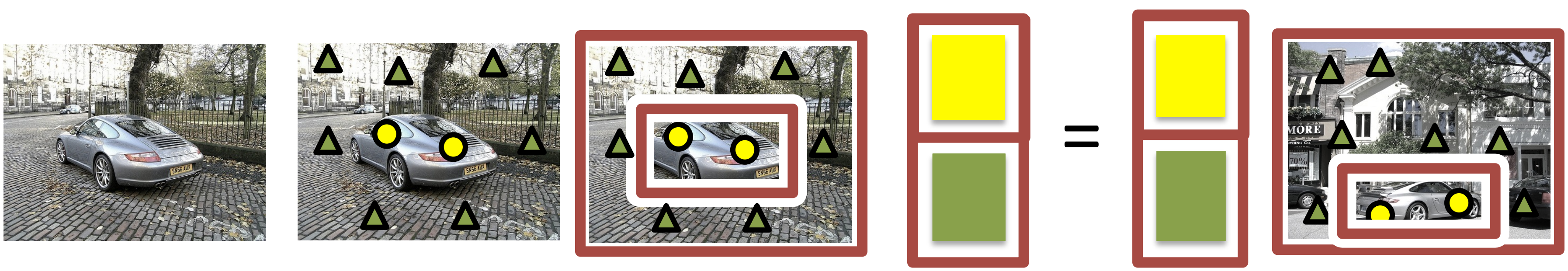


Object-centric spatial pooling

We propose an object-centric spatial pooling (OCP) approach which

(1) localizes the object of interest, and then

(2) pools foreground visual features separately from the background features.



Russakovsky et al. ECCV 2012

OCP training formulation

Given: N images with labels $y_1 \ldots y_N \in \{-1,+1\}$ and no object location information

Know:

Positive images contain at least one instance of the object

Negative images contain no object instances

Positive examples

Negative examples

OCP training formulation

Given: N images with labels $y_1 \ldots y_N \in \{-1,+1\}$ and no object location information

Know:

Positive images contain at least one instance of the object

Negative images contain no object instances

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \text{slack}_i$$

$$\text{s.t. } y_i \max_{\substack{\text{regions} \\ \text{of Image}_\text{i}}} [\mathbf{w}^T F_{\text{region}} + b] \geq 1 - \text{slack}_i \quad \forall i$$

Nguyen et al. ICCV09

OCP training formulation

Given: N images with labels $y_1 \ldots y_N \in \{-1,+1\}$ and no object location information

Know:

Positive images contain at least one instance of the object

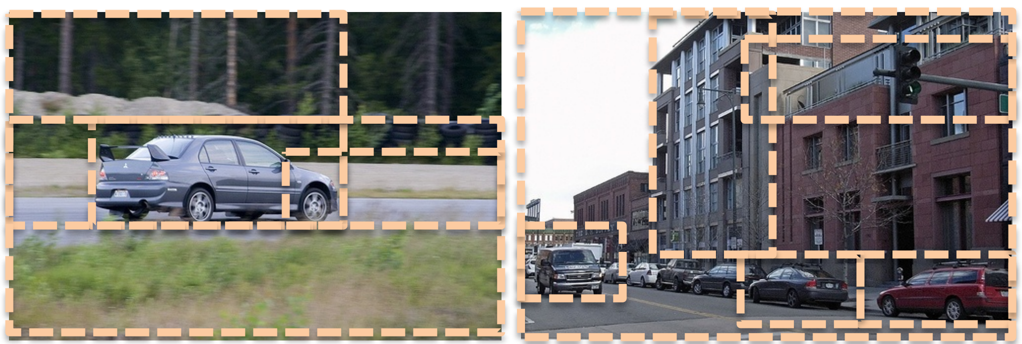
Negative images contain no object instances

Goal: a joint model for accurate image classification and accurate object localization

Russakovsky et al. ECCV 2012

OCP key #1: limiting the search space

Positive examples



Negative examples

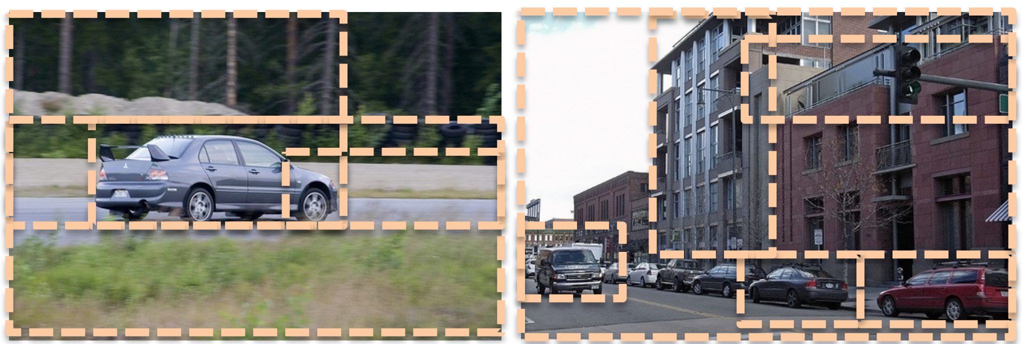
Use an unsupervised algorithm to propose regions likely to contain an object

- e.g., van de Sande et al. ICCV 2011, Alexe et al. TPAMI 2012
- Recall: > 97%, ~1500 regions per image
- Helps with accurate object localization

Russakovsky et al. ECCV 2012

OCP key #2: using all negative data

Positive examples



Negative examples

Dataset: PASCAL07, 20 object classes

~200 examples from positive images +

~5000 negative images x ~1500 regions per image

=> more than 7M examples

Training: stochastic gradient descend with averaging (Lin CVPR'11)

Russakovsky et al. ECCV 2012

OCP training algorithm

Positive examples

Negative examples

- Predict object location is the full image

Russakovsky et al. ECCV 2012

OCP training algorithm

- Predict object location is the full image
- Learn appearance model

Russakovsky et al. ECCV 2012

OCP training algorithm

- Predict object location is the full image
- Learn appearance model
- Update location estimate

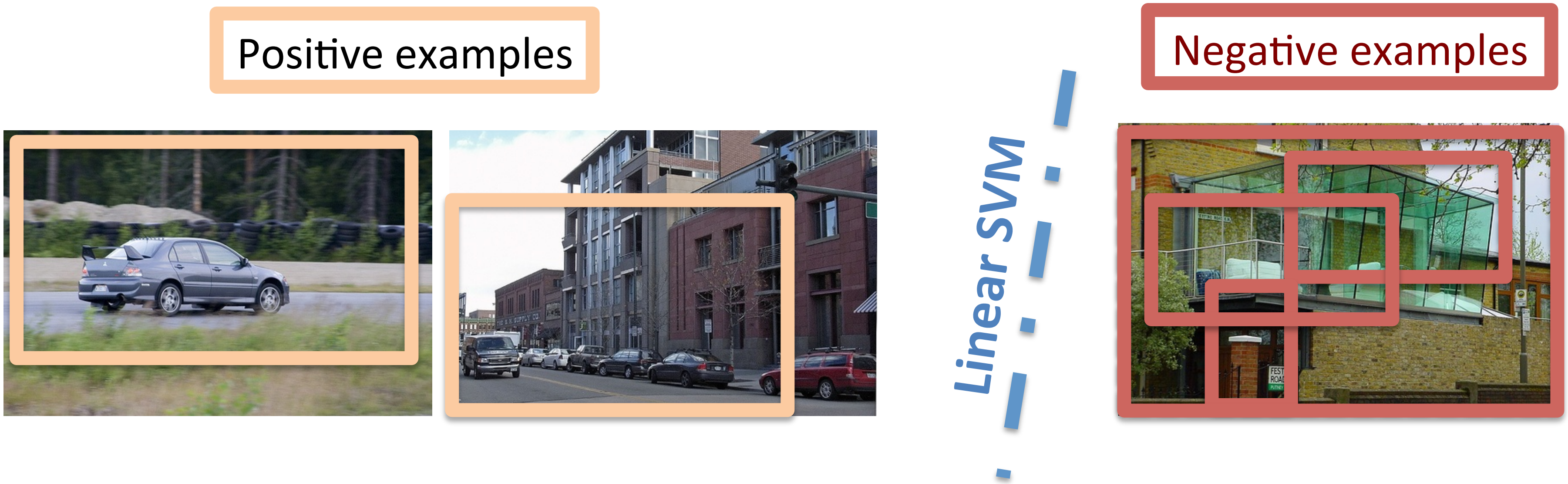
Russakovsky et al. ECCV 2012

OCP training algorithm

- Predict object location is the full image
- Learn appearance model
- Update location estimate
- Re-learn appearance model

Russakovsky et al. ECCV 2012

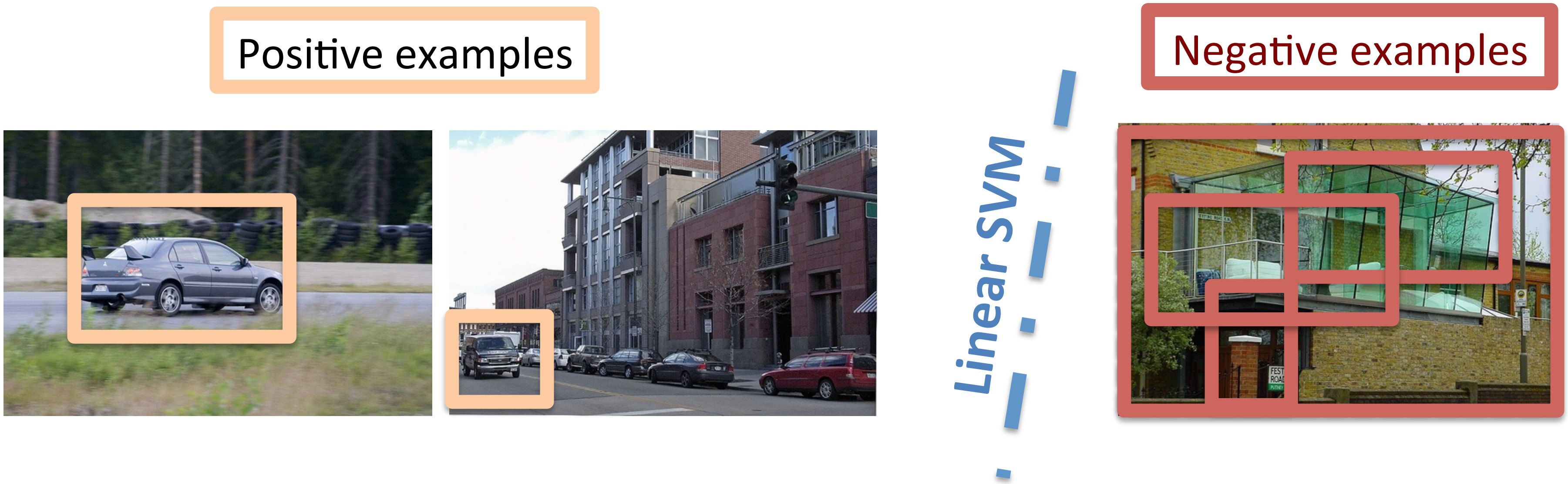
OCP training algorithm



- Predict object location is the full image
- Learn appearance model
- Update location estimate
- Re-learn appearance model

Russakovsky et al. ECCV 2012

OCP training algorithm



- Predict object location is the full image
- Learn appearance model
- Update location estimate
- Re-learn appearance model

OCP training algorithm

- Predict object location is the full image
- Learn appearance model
- Update location estimate
- Re-learn appearance model

Joint model for
image classification and
object localization

Russakovsky et al. ECCV 2012

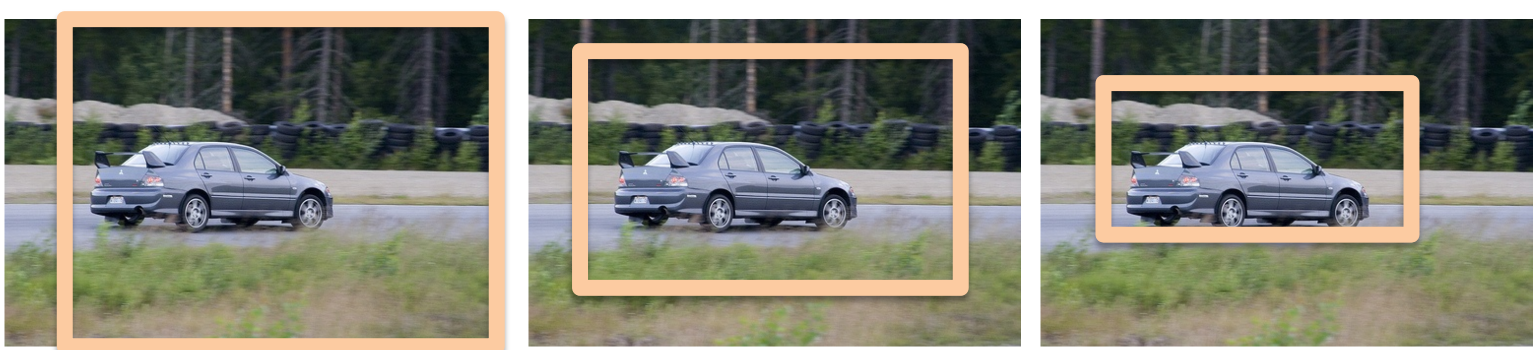
OCP key #3: avoiding local minima

Positive examples

BAD

Negative examples

- Desired training progression:

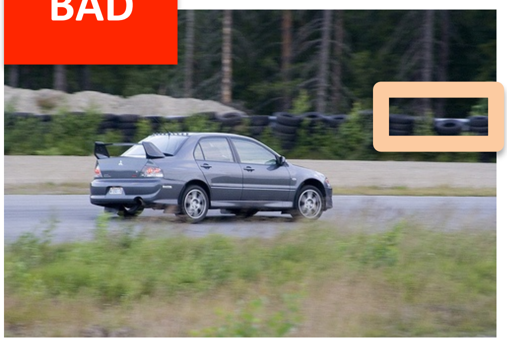


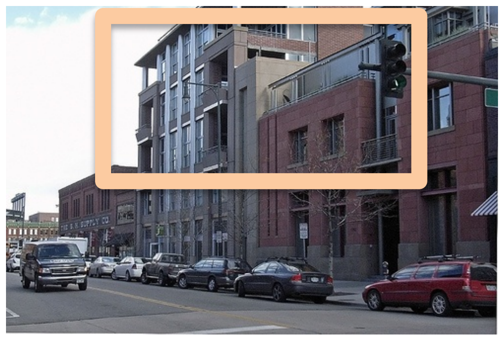
...

OCP key #3: avoiding local minima

Positive examples

BAD





Negative examples

- On each iteration, slowly shrink the minimum allowed size
  - Iteration 0: use full image
  - Iteration 1: use only regions with area > 75% image area
  - Iteration 2: use only regions with area > 70% image area
  - …

Recall OCP training formulation

Given: N images with labels $y_1 \ldots y_N \in \{-1,+1\}$ and no object location information

Know:

Positive images contain at least one instance of the object

Negative images contain no object instances

$$\min_{\mathbf{w},b} \frac{1}{2}||\mathbf{w}||^2 + C\sum_i \text{slack}_i$$

$$\text{s.t. } \; y_i \max_{\substack{\text{regions} \\ \text{of Image}_\text{i}}} [\mathbf{w}^T \boxed{F_{\text{region}}} + b] \geq 1 - \text{slack}_i \quad \forall i$$

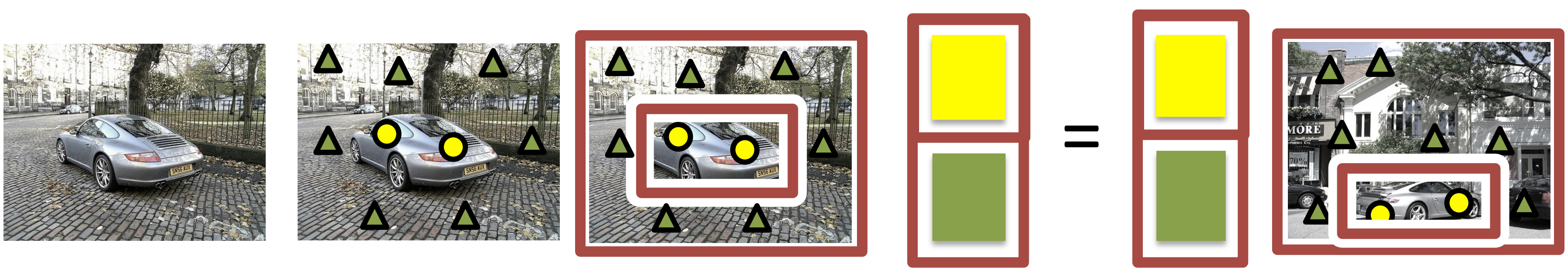
Russakovsky et al. ECCV 2012

Object-centric spatial pooling

We propose an object-centric spatial pooling (OCP) approach which

(1) localizes the object of interest, and then

(2) pools foreground visual features separately from the background features.

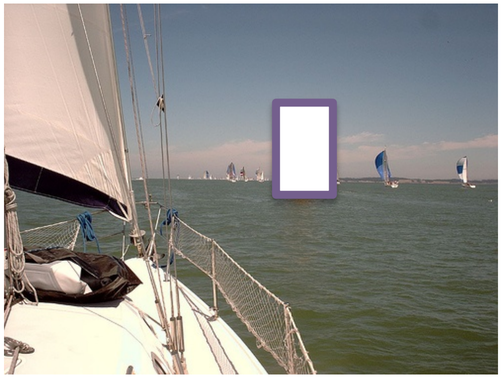


OCP key #4: Foreground-background

- Background provides context to improve classification

Foreground

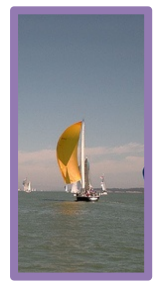
Background



OCP key #4: Foreground-background

- Background provides context to improve classification
- Using a foreground-only model leads to inaccurate localization

Accurate: 

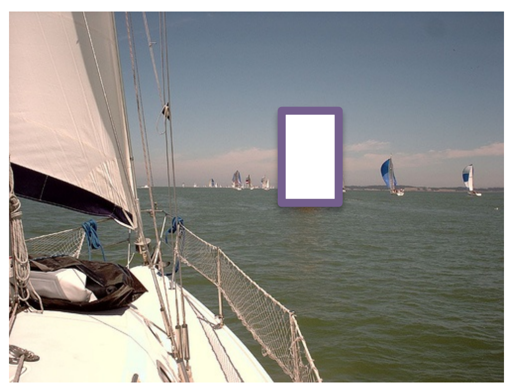
Too big: 

OCP key #4: Foreground-background

- Background provides context to improve classification
- Using a foreground-only model leads to inaccurate localization
- The foreground-background representation is both
  - a bounding box representation (for detection), and
  - an image-level representation (for classification)

Foreground

Background



Outline

Object-centric spatial pooling (OCP) image representation

Training the OCP model as a joint image classification and object localization model:

1. Limit the search space
2. Train with lots of negative data
3. Localize slowly to avoid local minima
4. Use foreground-background representation

Results

- Improved image classification accuracy
- Competitive weakly supervised localization accuracy

Russakovsky et al. ECCV 2012

Results

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Results: image classification

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Baseline SPM on full image: 54.3% classification mAP
Object-centric pooling (OCP): **57.2%** classification mAP

Method	aero	bicycle	bird	boat	bottle	bus	car	cat	chair	cow
SPM	72.5	56.3	**49.5**	63.5	22.4	60.1	76.4	57.5	**51.9**	42.2
OCP	**74.2**	**63.1**	45.1	**65.9**	**29.5**	**64.7**	**79.2**	**61.4**	51.0	**45.0**

Method	dining	dog	horse	mot	person	plant	sheep	sofa	train	tv
SPM	48.9	38.1	75.1	62.8	82.9	20.5	38.1	46.0	**71.7**	50.5
OCP	**54.8**	**45.4**	**76.3**	**67.1**	**84.4**	**21.8**	**44.3**	**48.8**	70.7	**51.7**

Results: image classification

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Baseline SPM on full image:	54.3% classification mAP
Object-centric pooling (OCP):	**57.2%** classification mAP
Baseline with 4-level SPM:	54.8% classification mAP
OCP foreground-only:	55.7% classification mAP

Russakovsky et al. ECCV 2012

Results: image classification

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Baseline SPM on full image:	54.3% classification mAP
Object-centric pooling (OCP):	**57.2%** classification mAP
Baseline with 4-level SPM:	54.8% classification mAP
OCP foreground-only:	55.7% classification mAP

Foreground-only (green) vs. foreground-background (yellow)

Results: image classification

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Baseline SPM on full image:	54.3% classification mAP
Object-centric pooling (OCP):	**57.2%** classification mAP
Baseline with 4-level SPM:	54.8% classification mAP
OCP foreground-only:	55.7% classification mAP

OCP with state-of-the-art
strongly supervised detector
(Felzenszwalb et al.):

Results: image classification

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

Baseline SPM on full image:	54.3% classification mAP
Object-centric pooling (OCP):	**57.2%** classification mAP
Baseline with 4-level SPM:	54.8% classification mAP
OCP foreground-only:	55.7% classification mAP
OCP with state-of-the-art strongly supervised detector (Felzenszwalb et al.):	56.9% classification mAP

Results: weakly supervised localization

PASCAL VOC 2007 **train set**, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin

27.4% localization accuracy

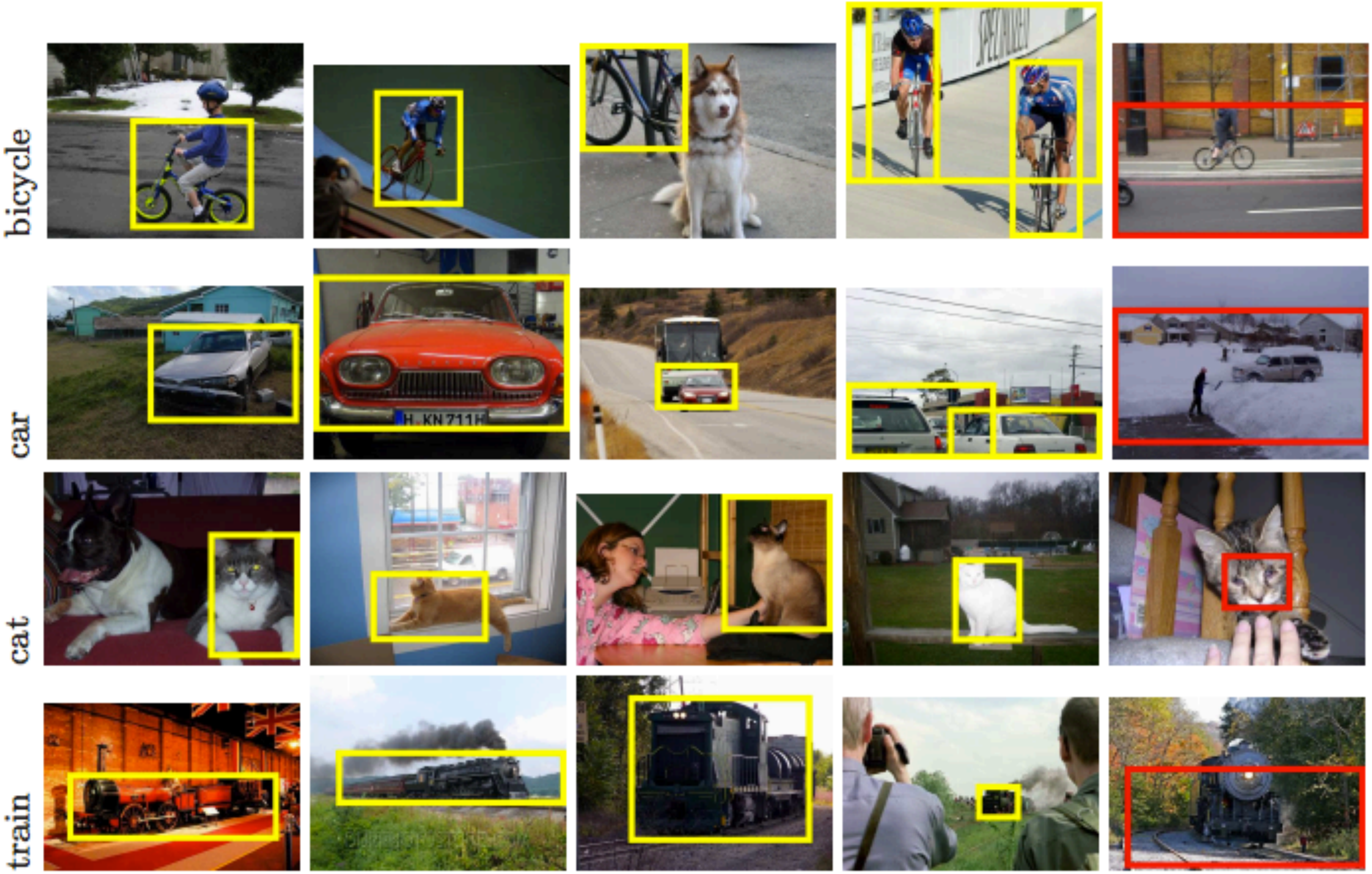
(compare to 28% of Deselaers IJCV12 and 30% of Pandey ICCV11)

PASCAL VOC 2007 **test set**, 6 classes

Method	aeroplane		bicycle		boat		bus		horse		motorbike		average detection mAP
	left	right	left	right	left	right	left	right	left	right	left	right	
Pandey 2011	7.5	21.1	38.5	44.8	0.3	0.5	0	0.3	45.9	17.3	43.8	27.2	20.8
Deselaers 2012	5	18	49	62	0	0	0	16	29	14	48	16	21.4
OCP	**30.8**		25.0		**3.6**		**26.0**		21.3		29.9		**22.8**

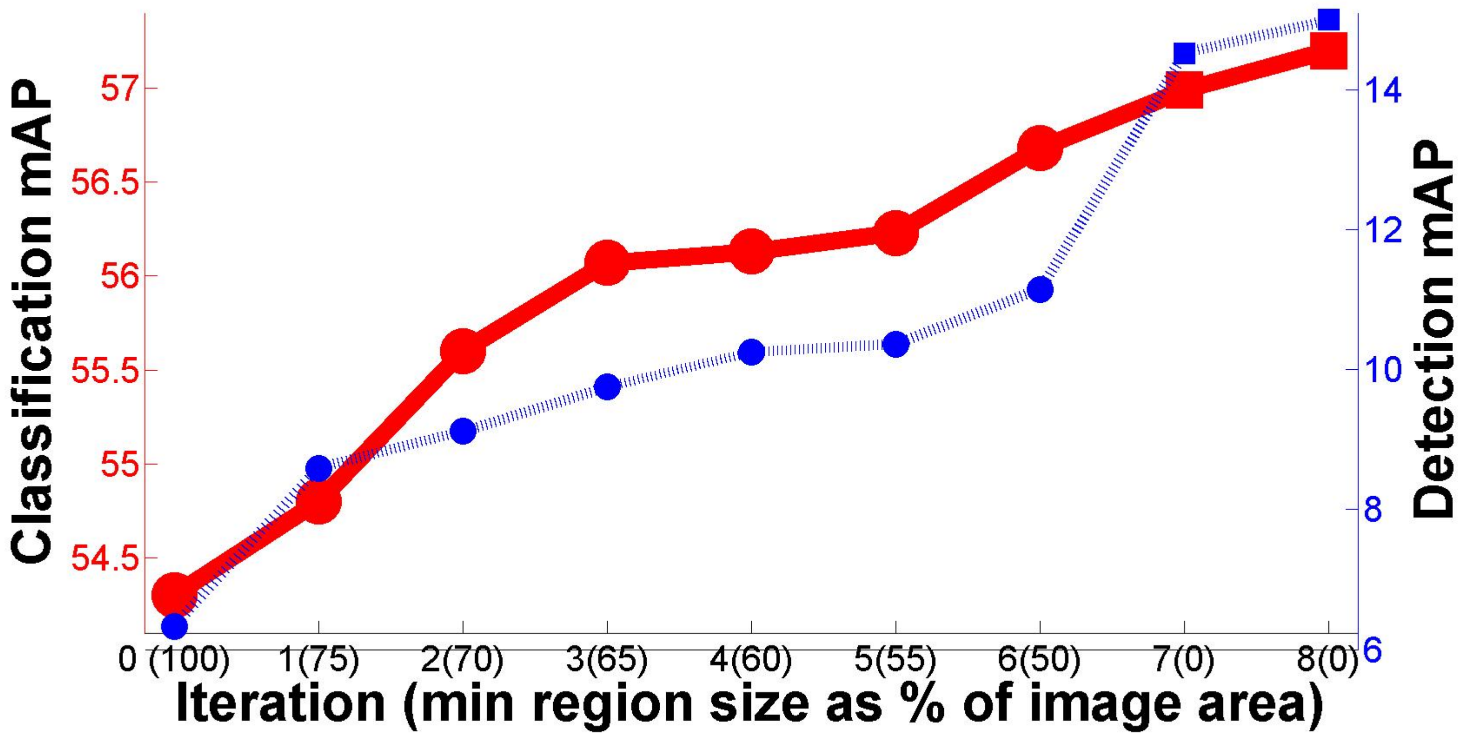
Russakovsky et al. ECCV 2012

Results: weakly supervised localization



Results: classification + detection

PASCAL VOC 2007 test set, 20 classes
DHOG features with LLC coding (codebook size 8192, k=5) and max pooling
1x1,3x3 SPM pooling on foreground + 1 background bin



Conclusions

Object-centric spatial pooling (OCP) framework:

Joint model for image classification and object localization

Foreground-background representation

Competitive results

Image classification

Weakly supervised object localization

Important step towards better image understanding

Without the need for additional costly image annotation

Olga Russakovsky, Yuanqing Lin, Kai Yu, Li Fei-Fei.
Object-centric spatial pooling for image classification. ECCV 2012
http://ai.stanford.edu/~olga **olga@cs.stanford.edu**

NEC Laboratories America
Relentless passion for innovation

NEC Laboratories America

Relentless passion for innovation

Object-centric spatial pooling for image classification

Olga Russakovsky, Yuanqing Lin,

Kai Yu, Li Fei-Fei

ECCV 2012