R-CNN Student Presentations

All The Students

Notes

- Project 3 report due tomorrow at 5 pm
- No Fei-Fei office hours today, out of town
- Last TA office hours today, will continue to answer Piazza

R-CNN Analysis

Dylan Rhodes

Experimental Setup

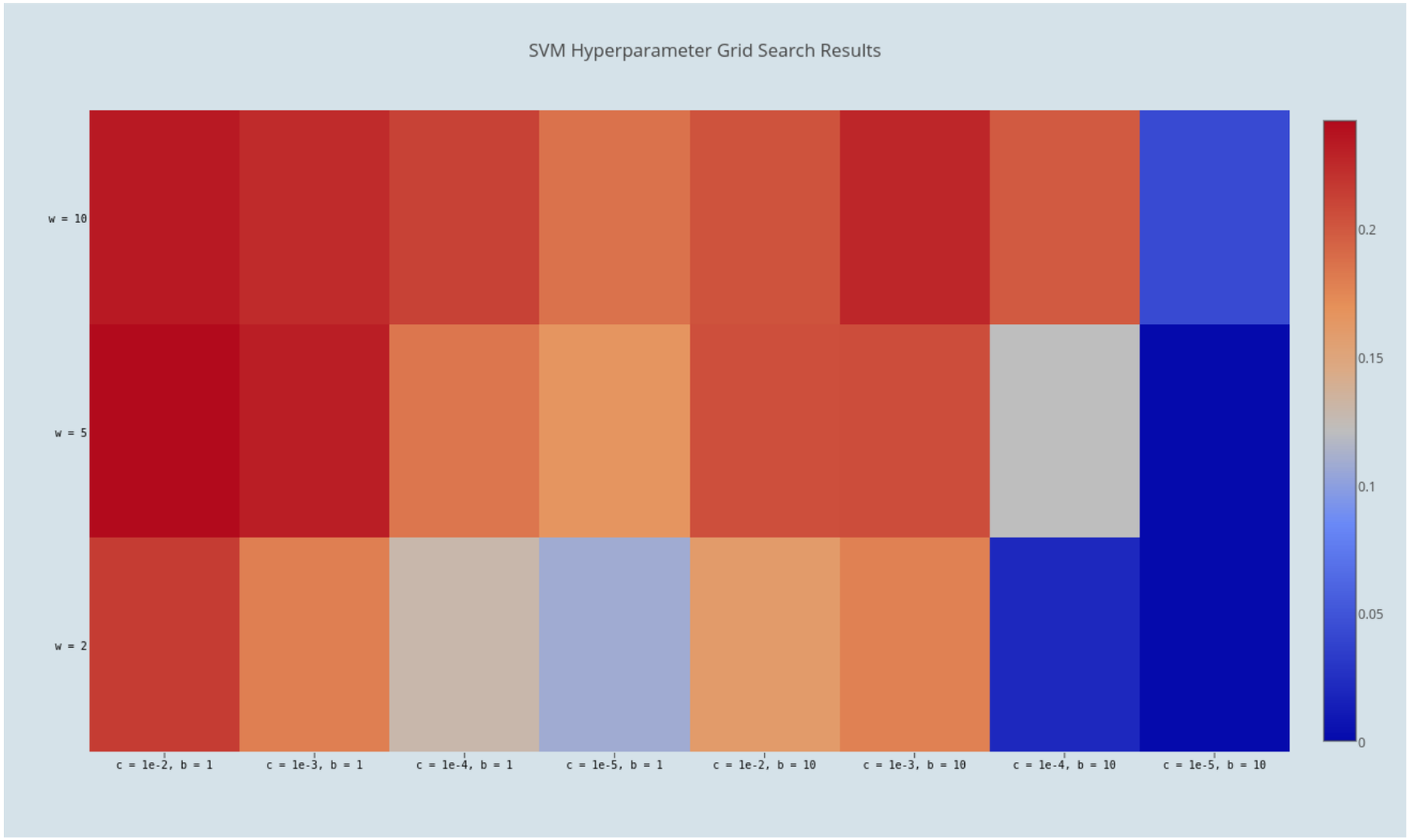
- Training data split into train/val set
  - train = first 500 images
  - val = last 250 images
- Liblinear SVM implementation used
  - L2 regularization, L1 loss

Baseline Performance

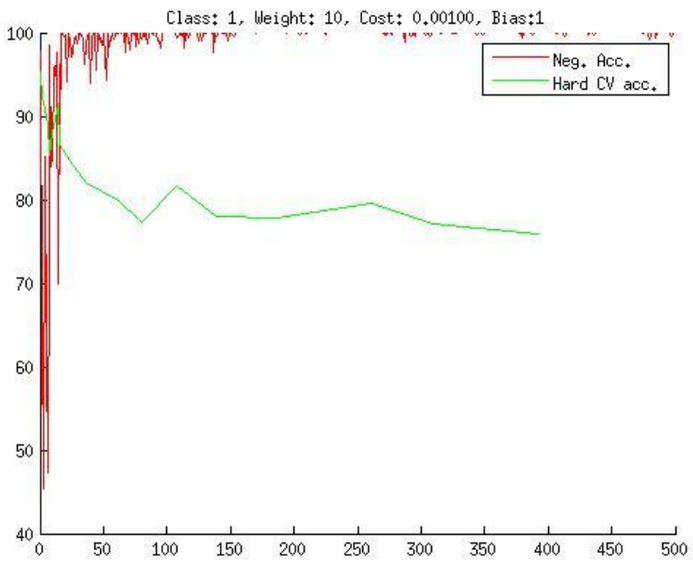
- Hyperparameters taken from paper
  - where not supplied, set to a ‘reasonable’ value
- SVM pipeline evaluated without regression

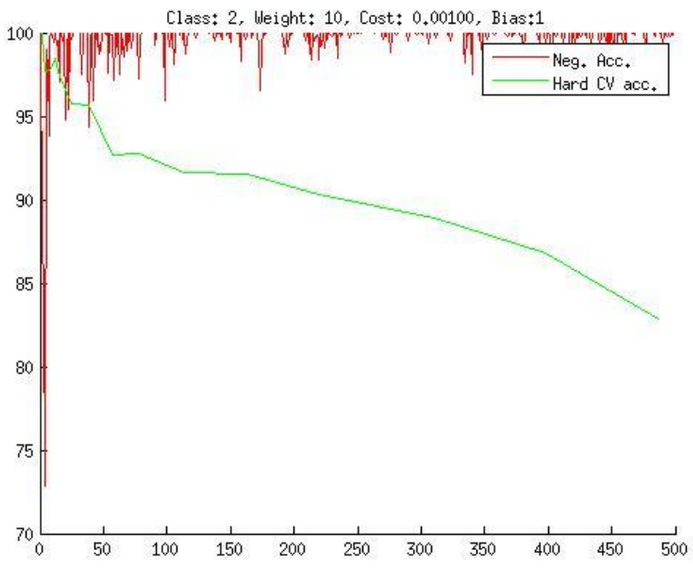
Baseline	Car	Cat	Person	Mean
Avg. Precision	0.1870	0.2515	0.0792	0.1726

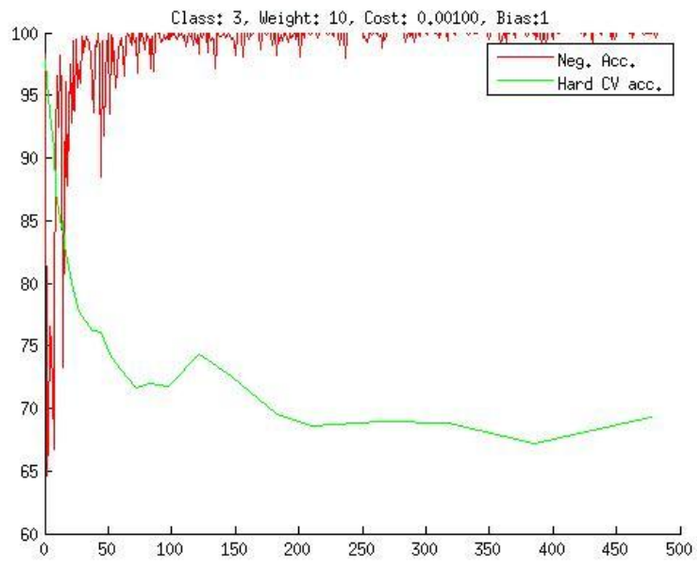
Hyperparameter Tuning



SVM Convergence Analysis







Final Results

- Results given on best hyperparameter set

Baseline	Car	Cat	Person	Mean
Avg. Precision	0.1870	0.2515	0.0792	0.1726

Best	Car	Cat	Person	Mean
Avg. Precision	0.2085	0.3457	0.0922	0.2154

Object Localization with R-CNN:
An Integrated Approach for Robust Detection

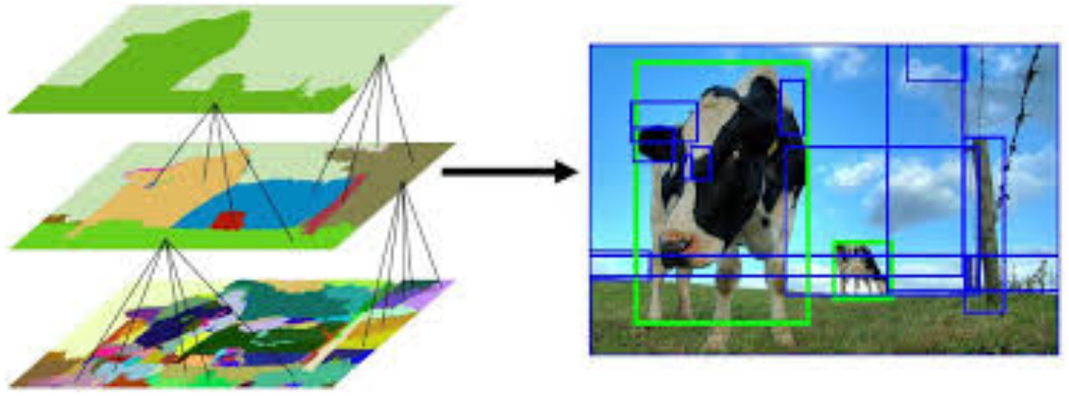
Amani V. Peddada
amanivp@cs.stanford.edu

R-CNN Implementation

- ***Selective Search*** for candidate regions

- ***Convolutional Neural Network (CNN)*** for feature extraction.

- Multiple ***Support Vector Machines (SVMs)*** for region classification.



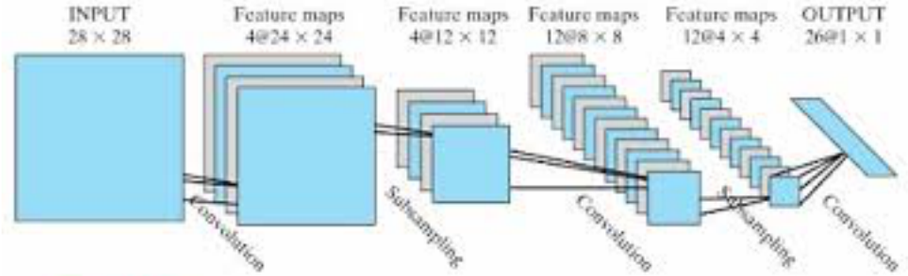
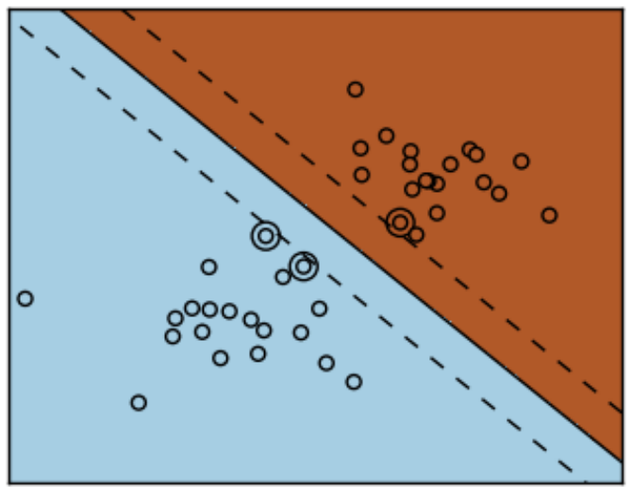


FIGURE 4.23 Convolutional network for image processing such as handwriting recognition. (Reproduced with permission of MIT Press.)



Further Details

- 16 pixel context around regions.
- Ground truth bounding boxes featurized.
- Overlap thresholds of 0.30 and 0.90.
- L2-Normalization (experimented with other types)
- LibLinear implementation
- Weigh positive examples by **ratio**
- 3 epochs, regularization = 50, set bias to one.
- Non-Max Surpression — 0.30 cutoff.

Results

Intermediate Results

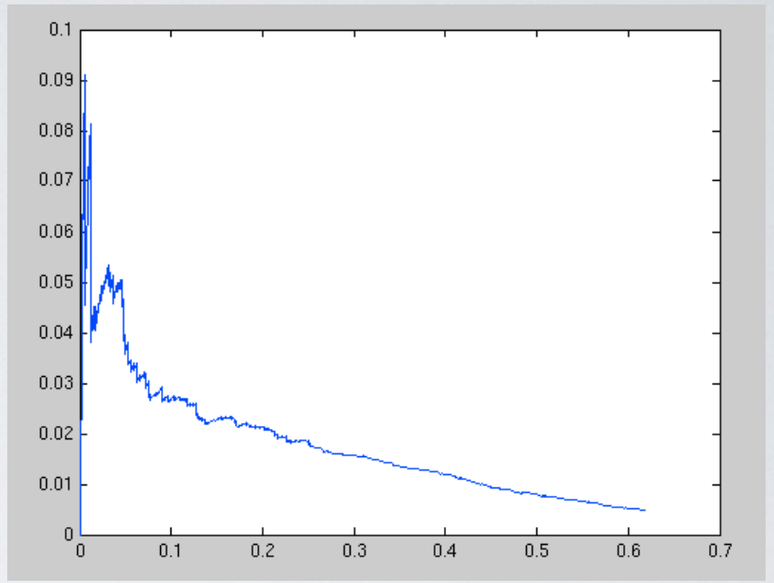
	Car	Cat	Person	mAP
w/o BBR	1.18	1.36	2.91	1.82

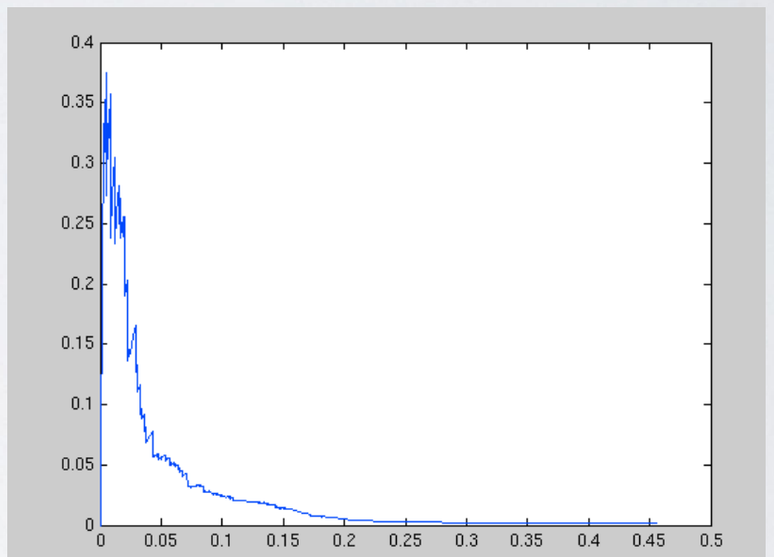
Provided Baseline

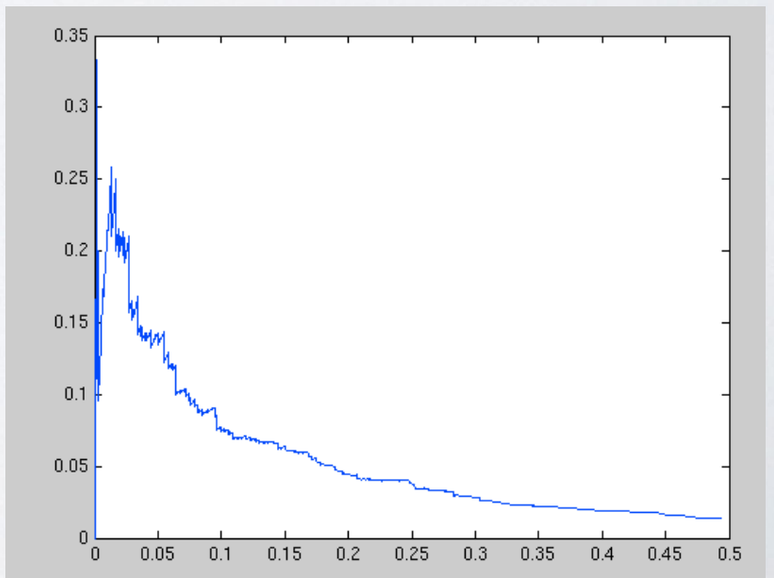
	Car	Cat	Person	mAP
w/o BBR	30.72	35.91	18.83	28.49
w/ BBR	32.97	38.58	20.05	30.53

Analysis

- Several hundreds of bounding boxes are being determined as positive.
- Overfitting training data — 98% accuracy.
- Too much context?
- More clever weighing scheme
- Accept bounding boxes only with higher confidences







Extensions

- Random Forest implementation — ***regress*** jointly on box parameters.

- Other classifiers — MatLab built-in SVM, standard Neural Networks, Softmax

- Other features/descriptors — HOG, SIFT, SURF, other network layers.

Thank you!

Detection with R-CNN

Tugce Tasci
Stanford University
6/3/2015

Extracting features

- Multiple regions at once

- Parallel computing on different cores (2-3 hours in total)

SVM Training

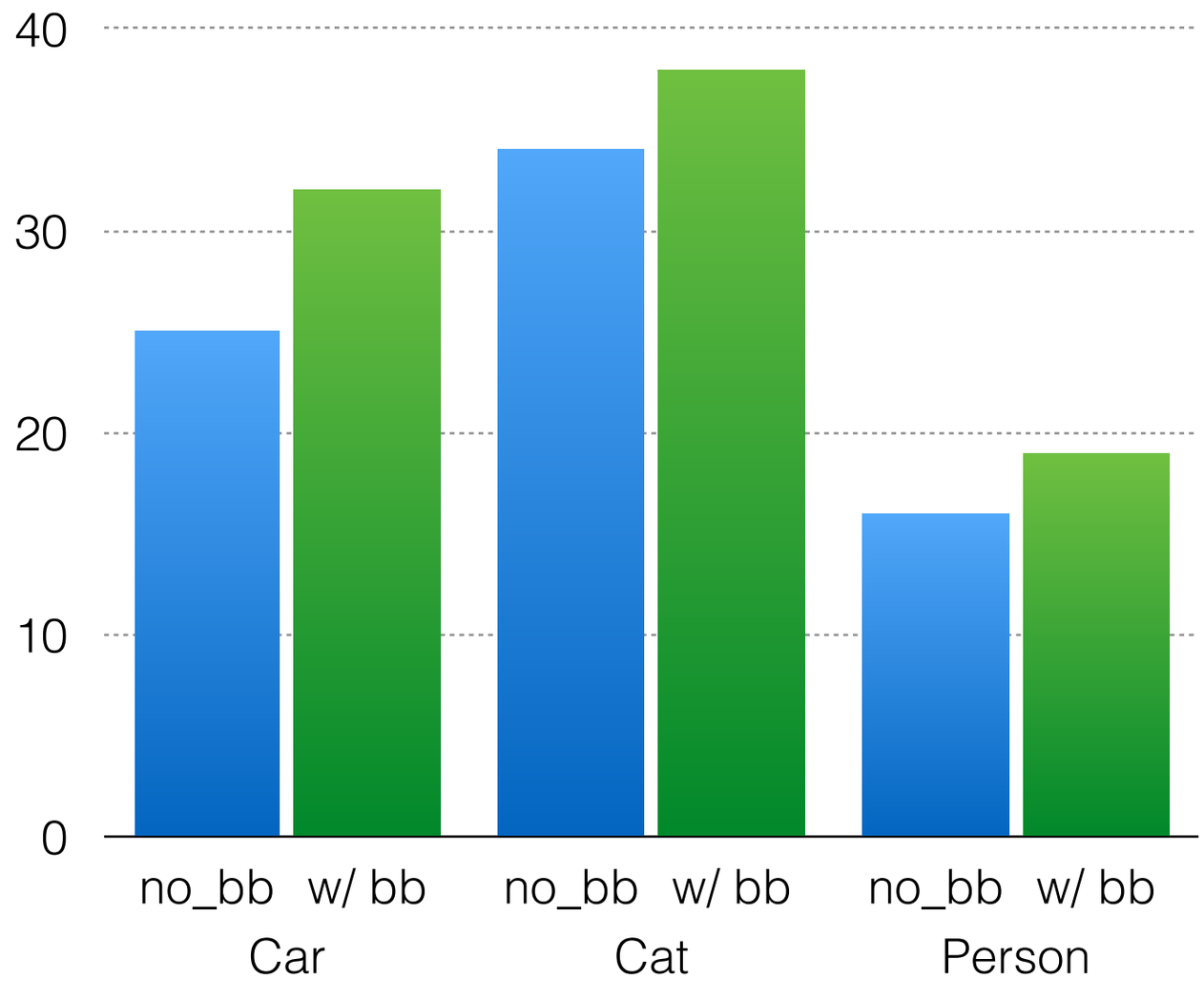
- L2 regularized, L2 loss SVM (LibLinear)
- positive examples: overlap > 0.9
- negative examples: overlap < 0.3
- 10 times more weight on positive examples
- bias added
- Hard negatives determined iteratively

Performance without bounding box regression

- Car: 25.71
- Cat: 34.66
- Person: 16.19
- **mAP: 25.52**

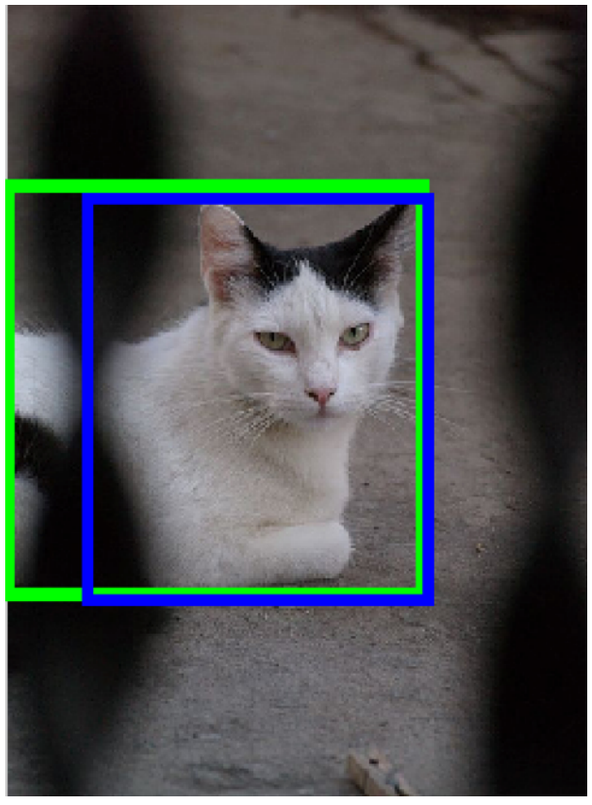
Performance with bounding box regression

- Car: 32.53
- Cat: 38.81
- Person: 19.68
- **mAP: 30.34**



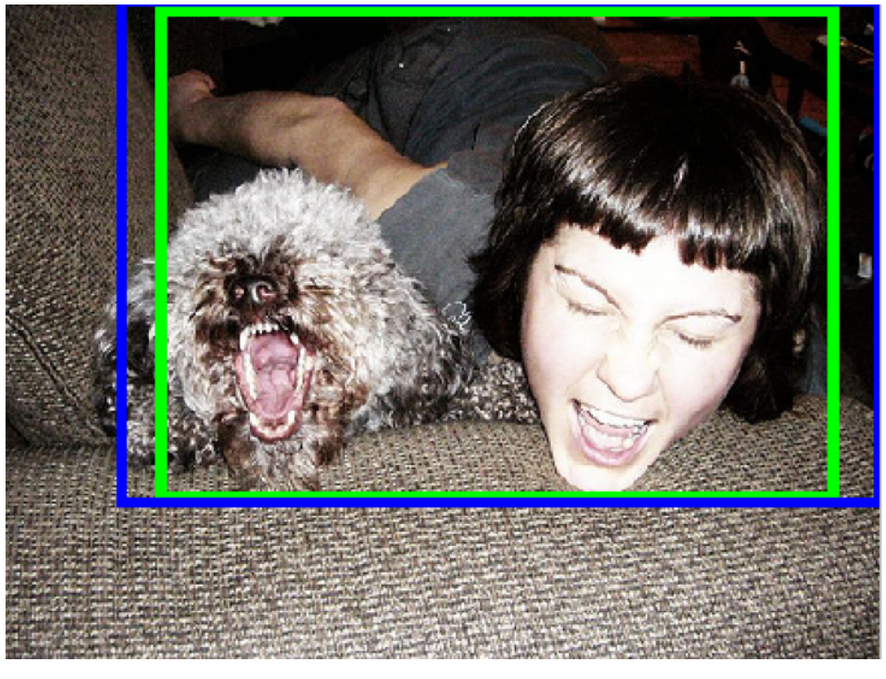
Results

Single class



Prediction

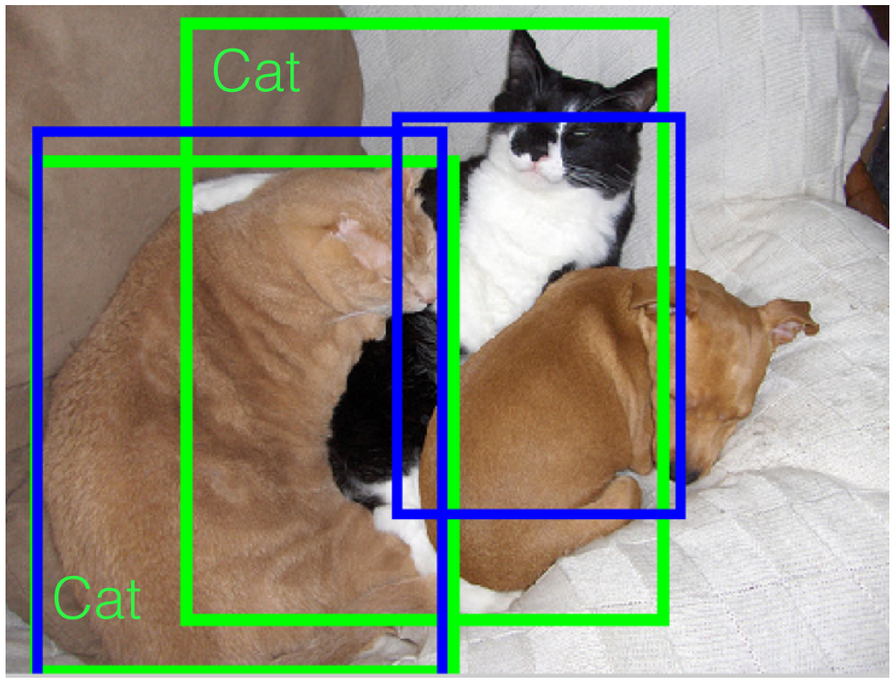
Ground truth

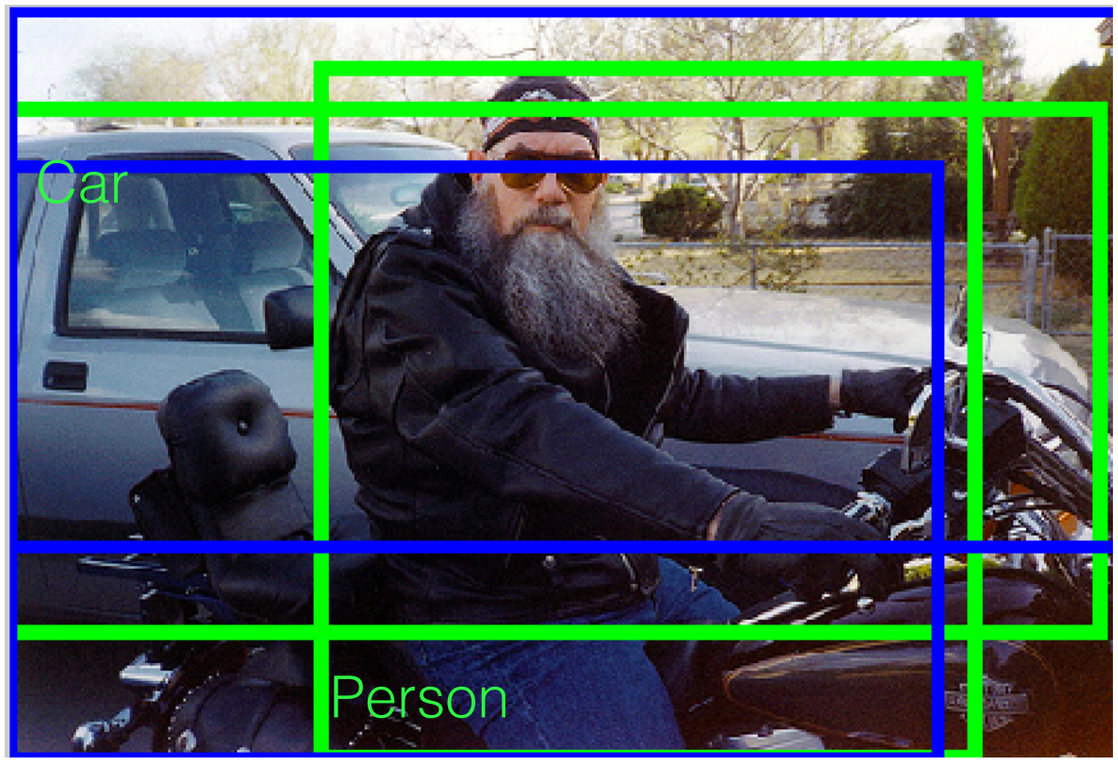


Results

Multiple class

Prediction
Ground truth





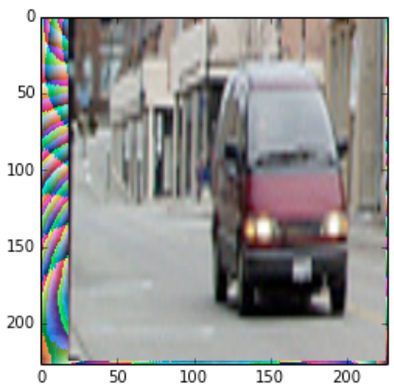
R-CNN

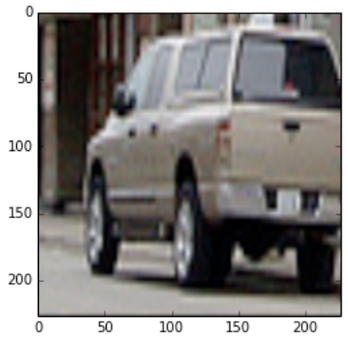
Bryan Anenberg & Michela Meister

3 June 2015

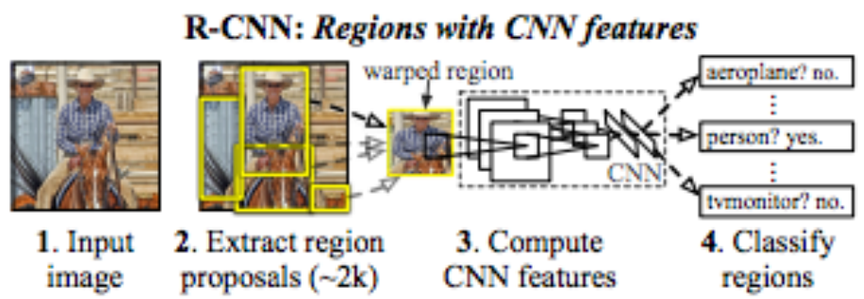
Pre-Processing

Image Cropping



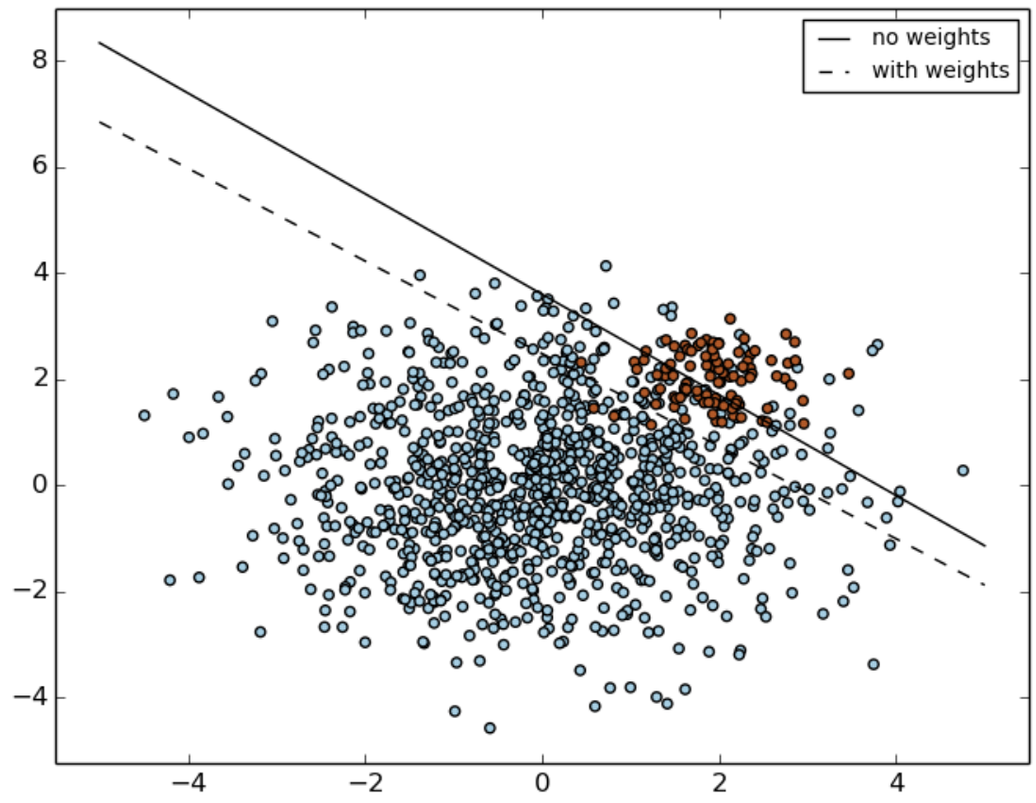


Extract Features



SVM Training

- Large class imbalance.
- Experimented with various options:
  - normalization
  - bias
  - regularization
  - class weights: {1:10}, inverse frequency
  - number of batches
  - number of epochs
- Mixed performance.
- Generates false negatives



	bbox	GT	label	overlap_label	svm_score	svm_1	svm_2	svm_3	iou_1	iou_2	iou_3	image
4	[1, 75, 500, 333]	0	0	1	0.000590	0.000322	-0.000050	0.000590	0.546660	0	0.014734	2008_000189.jpg
5	[1, 95, 500, 333]	0	0	1	0.000593	0.000324	-0.000050	0.000593	0.509377	0	0.010142	2008_000189.jpg
11	[1, 92, 500, 333]	0	0	1	0.000594	0.000325	-0.000051	0.000594	0.515086	0	0.011339	2008_000189.jpg
17	[59, 164, 388, 298]	0	0	1	-0.000008	-0.000508	-0.000008	-0.000869	0.529882	0	0.000000	2008_000189.jpg
18	[1, 60, 500, 333]	0	0	1	0.000592	0.000324	-0.000051	0.000592	0.573472	0	0.013927	2008_000189.jpg

Final Bounding Box Selection

Bounding Box Regression

Find transformation between predicted boxes P and their ground truths G:

$$
\begin{aligned}
t_x &= (G_x - P_x)/P_w \\
t_y &= (G_y - P_y)/P_h \\
t_w &= \log(G_w/P_w) \\
t_h &= \log(G_h/P_h).
\end{aligned}
$$

$$
\mathbf{w}_\star = \underset{\hat{\mathbf{w}}_\star}{\operatorname{argmin}} \sum_i^N (t_\star^i - \hat{\mathbf{w}}_\star^{\mathrm{T}} \boldsymbol{\phi}_5(P^i))^2 + \lambda \|\hat{\mathbf{w}}_\star\|^2 .
$$

$$
\begin{aligned}
\hat{G}_x &= P_w d_x(P) + P_x \\
\hat{G}_y &= P_h d_y(P) + P_y \\
\hat{G}_w &= P_w \exp(d_w(P)) \\
\hat{G}_h &= P_h \exp(d_h(P)).
\end{aligned}
$$

Non-Maximal Suppression (Felzenswalb)

1. Rank boxes by score
2. Iterate from highest to lowest to choose unique boxes.

Results

In progress.

R-CNN

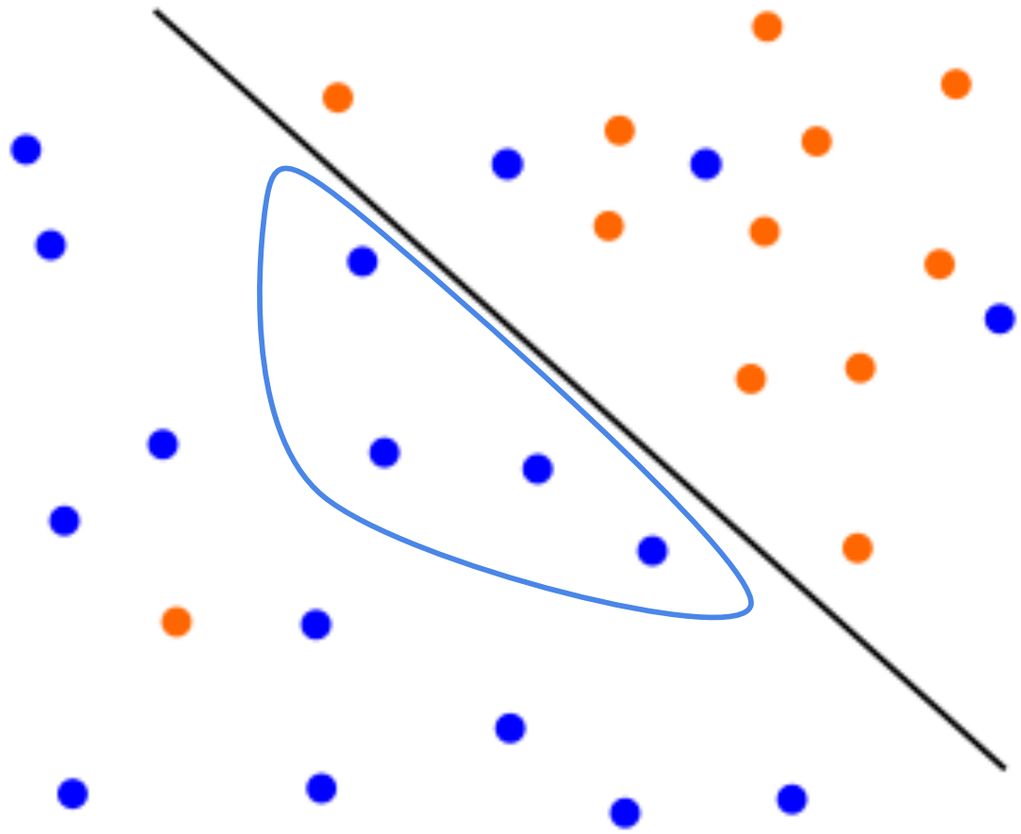
Implementation and Evaluation

Lyne P. Tchapmi
Stanford University/CS231B

Negatives mining

- Batch-wise (100/150 images)
  - Relatively Fast
  - Batch error varies
- Whole set of negatives
  - Slower
  - Training error decreases

Hardest Easy Negatives



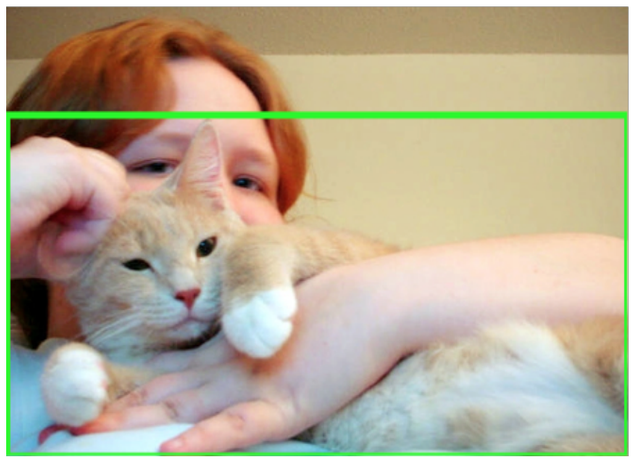
Feature Normalization

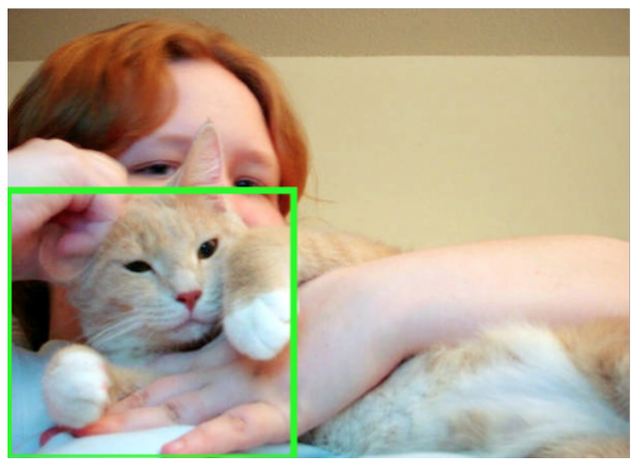
- Standardization (row-wise normalization)
- Zero-mean unit-variance-ZMUV (column-wise)

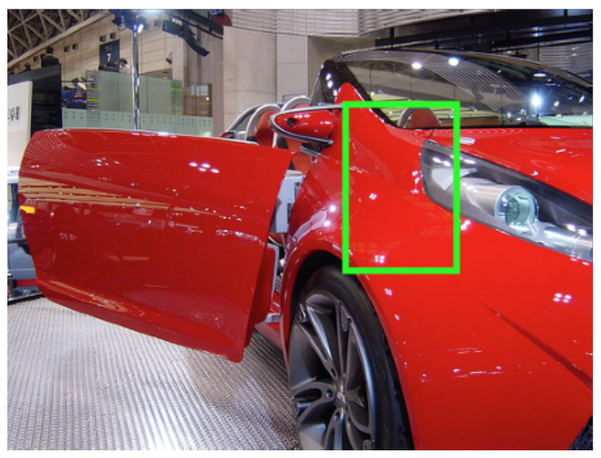
Results

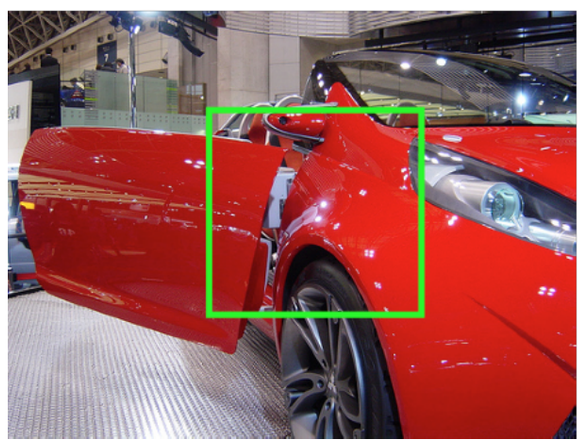
	ZMUV/BatchWise	ZMUV/Whole-set	Standard/Whole-Set
Car MAP	1.28	13.12	**24.36**
Cat MAP	5.23	27.85	**30.14**
Person MAP	2.29	12.79	**13.61**

Standardization vs ZMUV









Project 3: R-CNN

Kelsie Zhao

Content

- Formulation
- Results
- Extensions

Formulation

1. Feature extraction: 16 pixels padding

2. SVM training with varying C, Bias, Weights
   a. Standard hard negative mining: no limits on no. of hard negatives
   b. Set max. number of hard negatives

3. Predicting labels for proposals

4. Bounding box regression

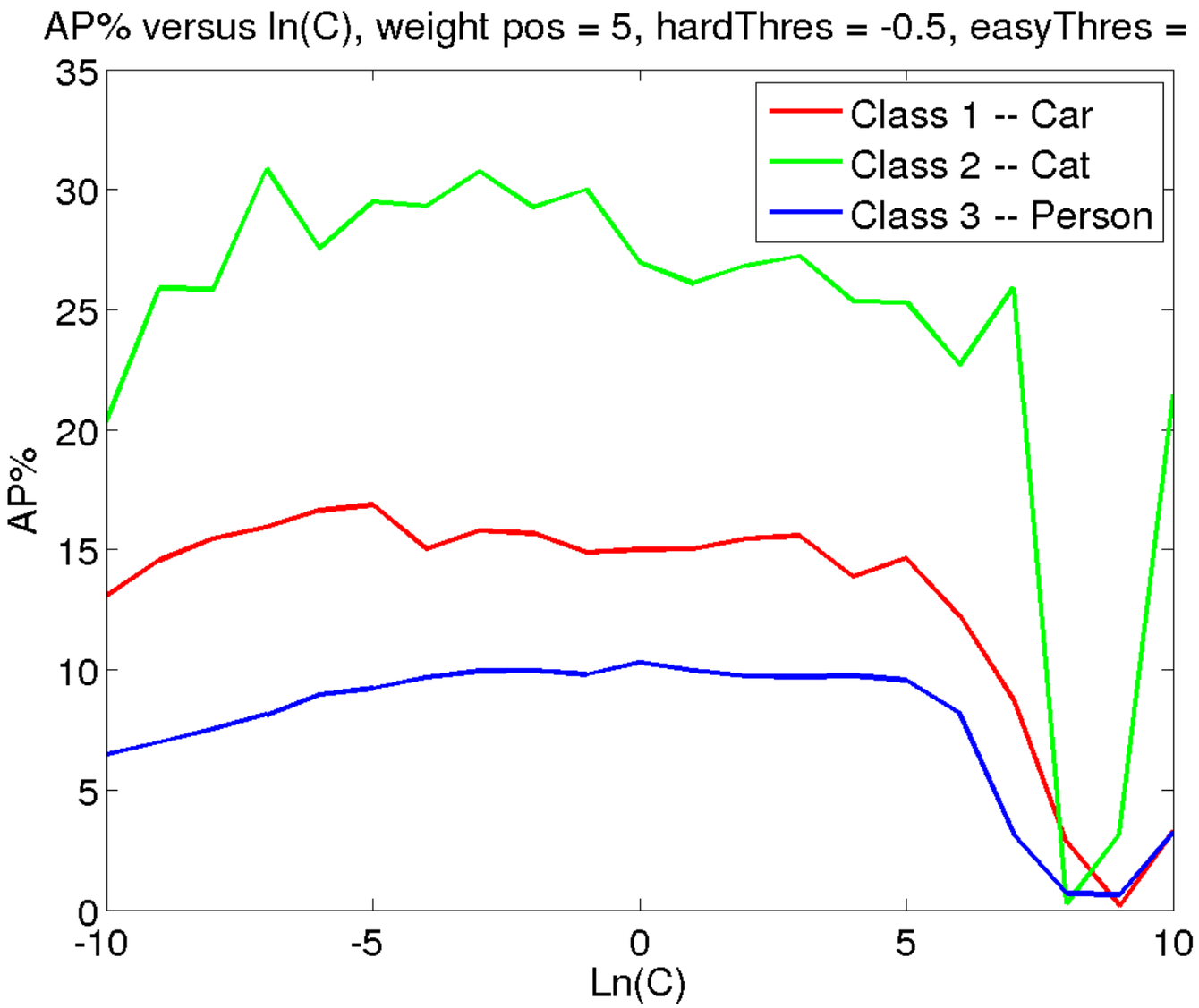
5. Non-maximum suppression with complete enclosure hard suppression

Results

Class	Average Precision %
Car	15.8
Cat	30.8
Person	9.9

1. Varying Weights of Positive class: 1, 2, 3, 5, 10
2. Varying Hard/Easy negative threshold: (-0.5, -1), (-1, -1.2)
3. Varying Bias term: 1, 10
4. Varying C: 2^{-10} to 2^{10}
5. Varying Classification threshold: [0, 0, 0] to [-2, -2, -2]
6. Change SVM Retrain condition

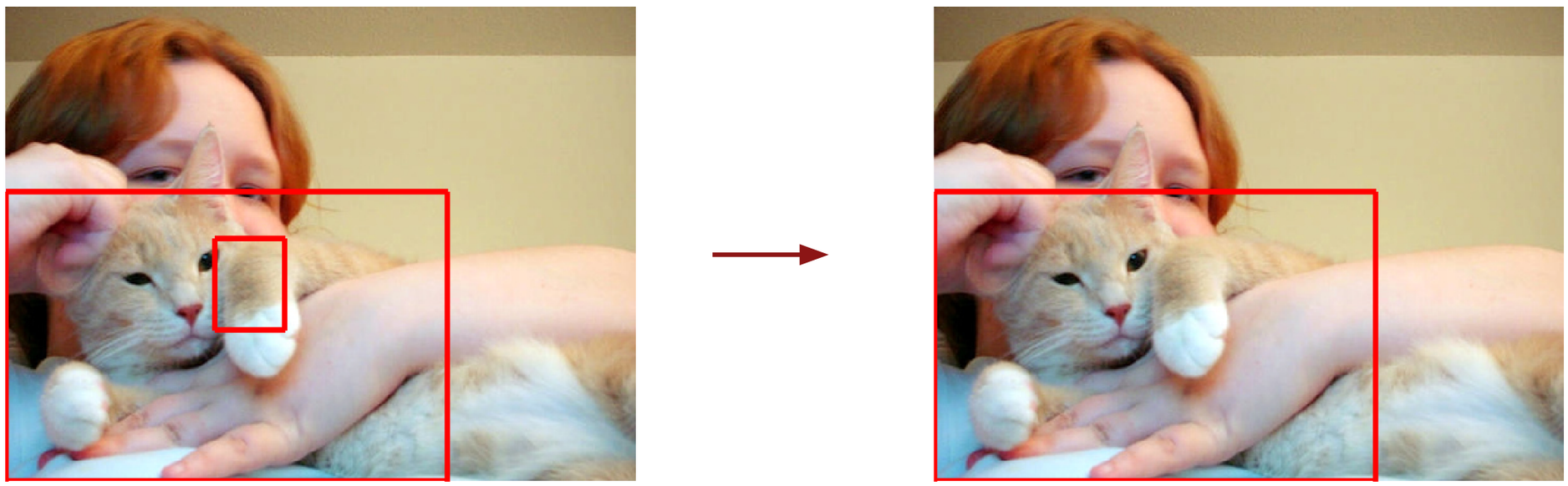
Results


AP% versus ln(C), weight pos = 5, hardThres = -0.5, easyThres = -1
Class 1 -- Car
Class 2 -- Cat
Class 3 -- Person
AP%
Ln(C)
35
30
25
20
15
10
5
0
-10
-5
0
5
10

Extensions

1. HOG features

2. NMS: complete enclosure suppression



Thank You!

Q & A

Object Detection with R-CNN

Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik
University of California, Berkeley

Project Presentation By

Albert Haque and Fahim Dalvi

June 3, 2015

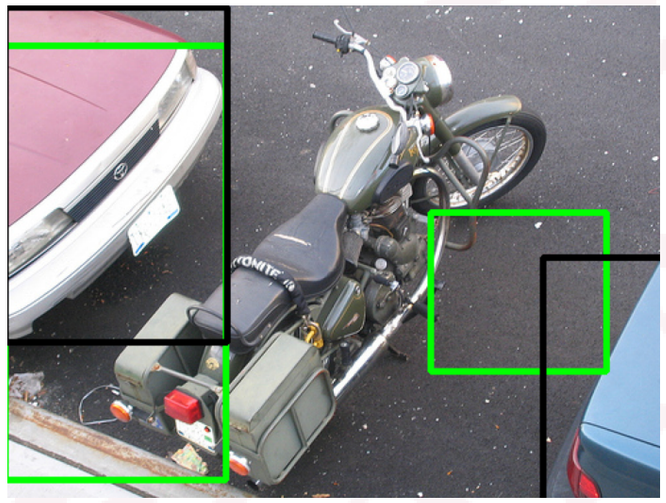
Outline

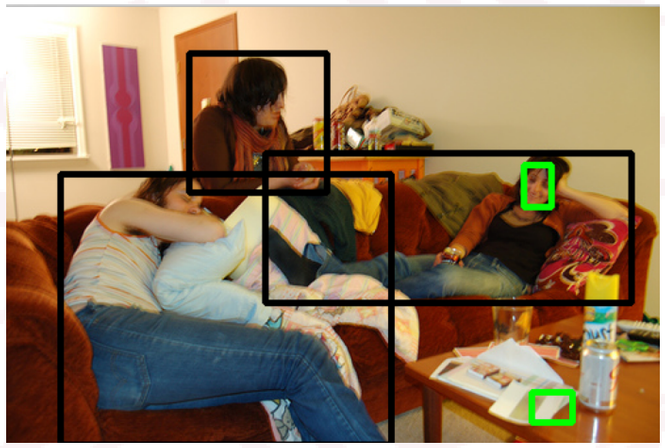
- Baseline Results
  - Iterative SVM
  - Full SVM
- Extensions
  - SGD based implementation
  - VGG Network

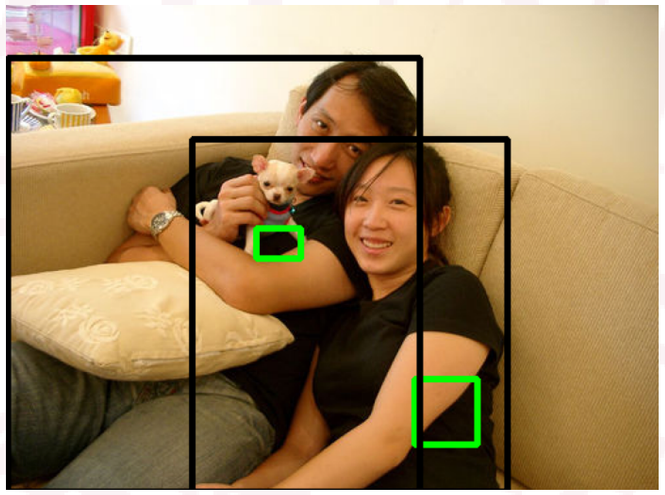
Baseline Results

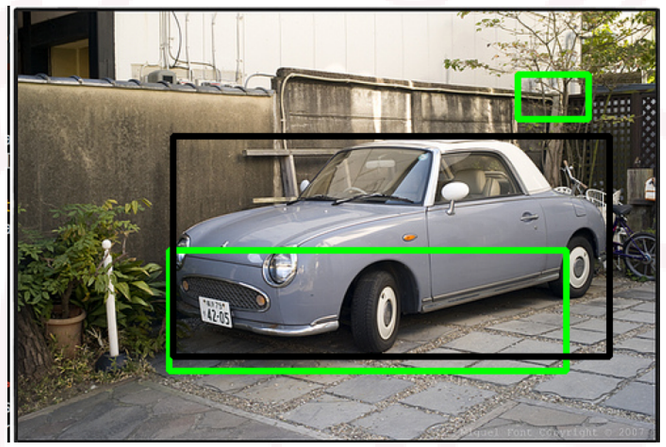
- Setup
  - Validation set (100 images)
  - Zero-mean unit-variance normalization
  - ℓ_2 regularization loss
  - Number of negatives: 30,000 for iterative SVM
  - Positives weighted 10 times more than negatives
- Results
  - Full SVM gives no improvement over Iterative SVM
  - 65% accuracy on the validation set

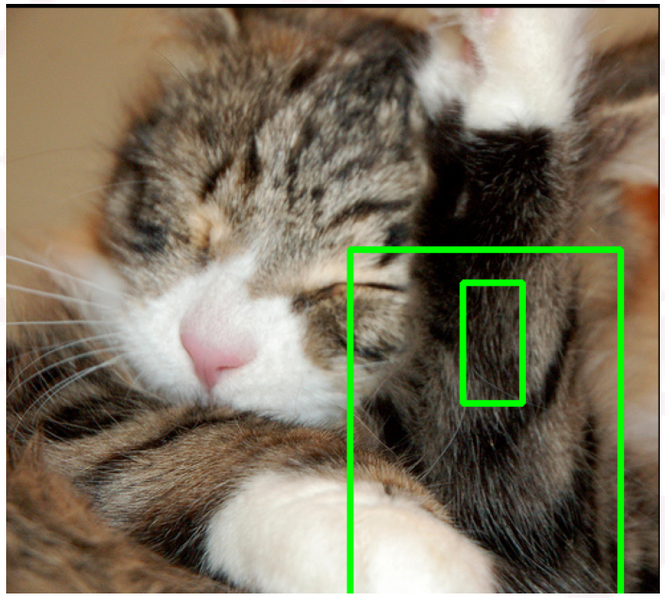
Baseline Results

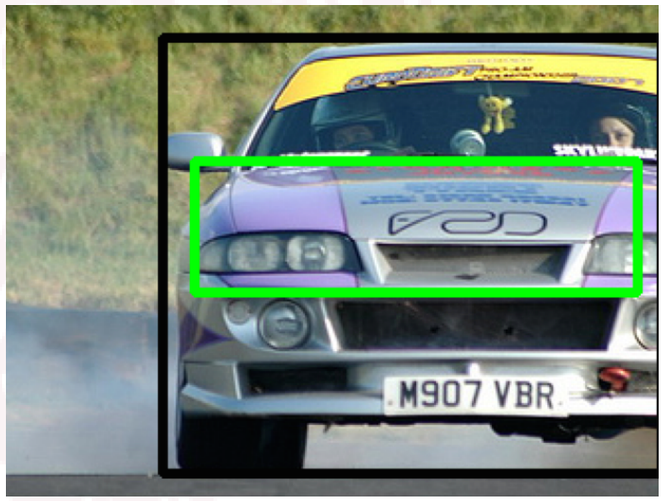










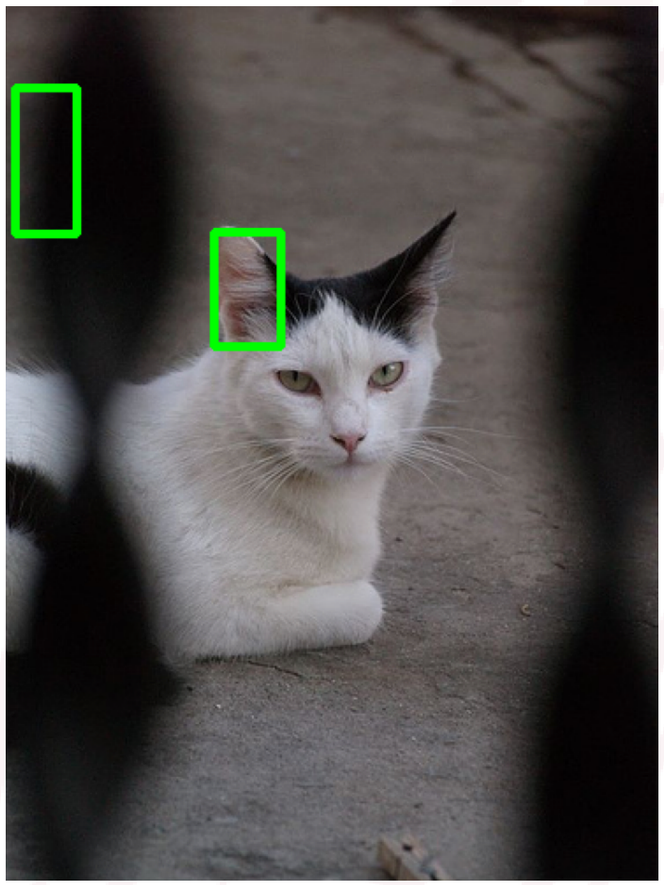


Baseline Results

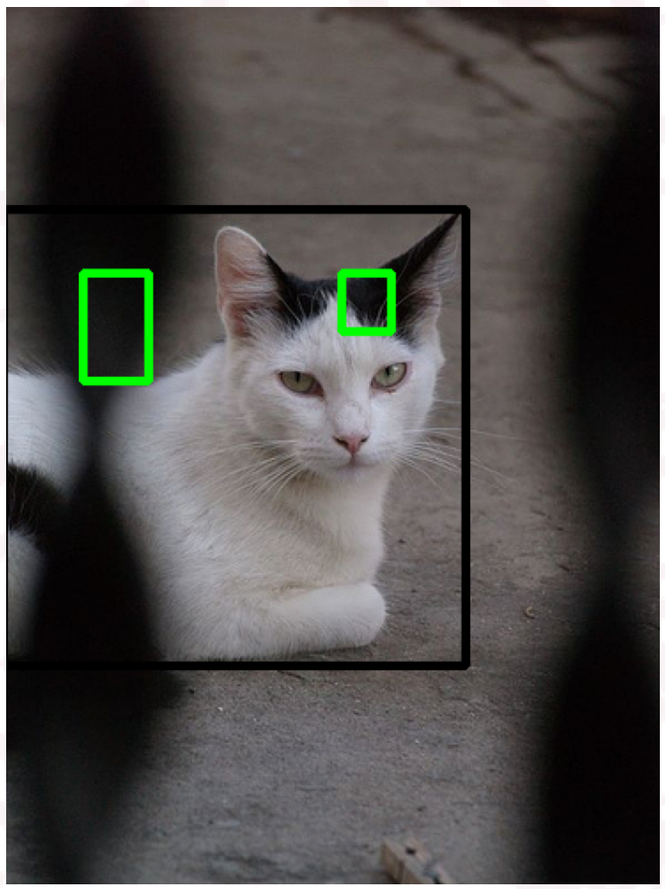
Extensions

- SGD Classifier
  - Trained iteratively using small sets of images
  - Same setup as SVM
- Results
  - SGD classifier does half as worse as the SVM
  - 55% accuracy on the validation set

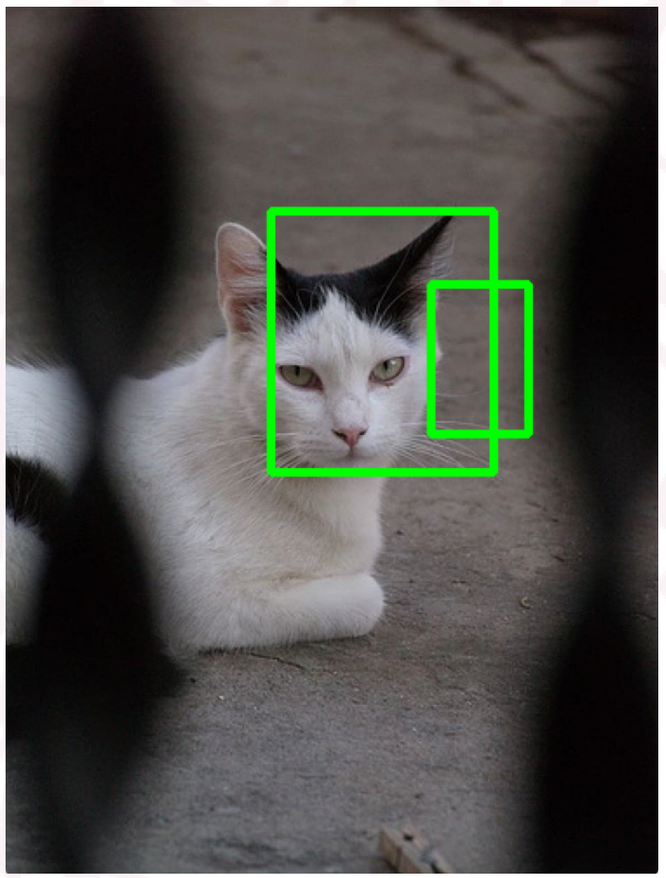
Extensions



Car



Cat



Person

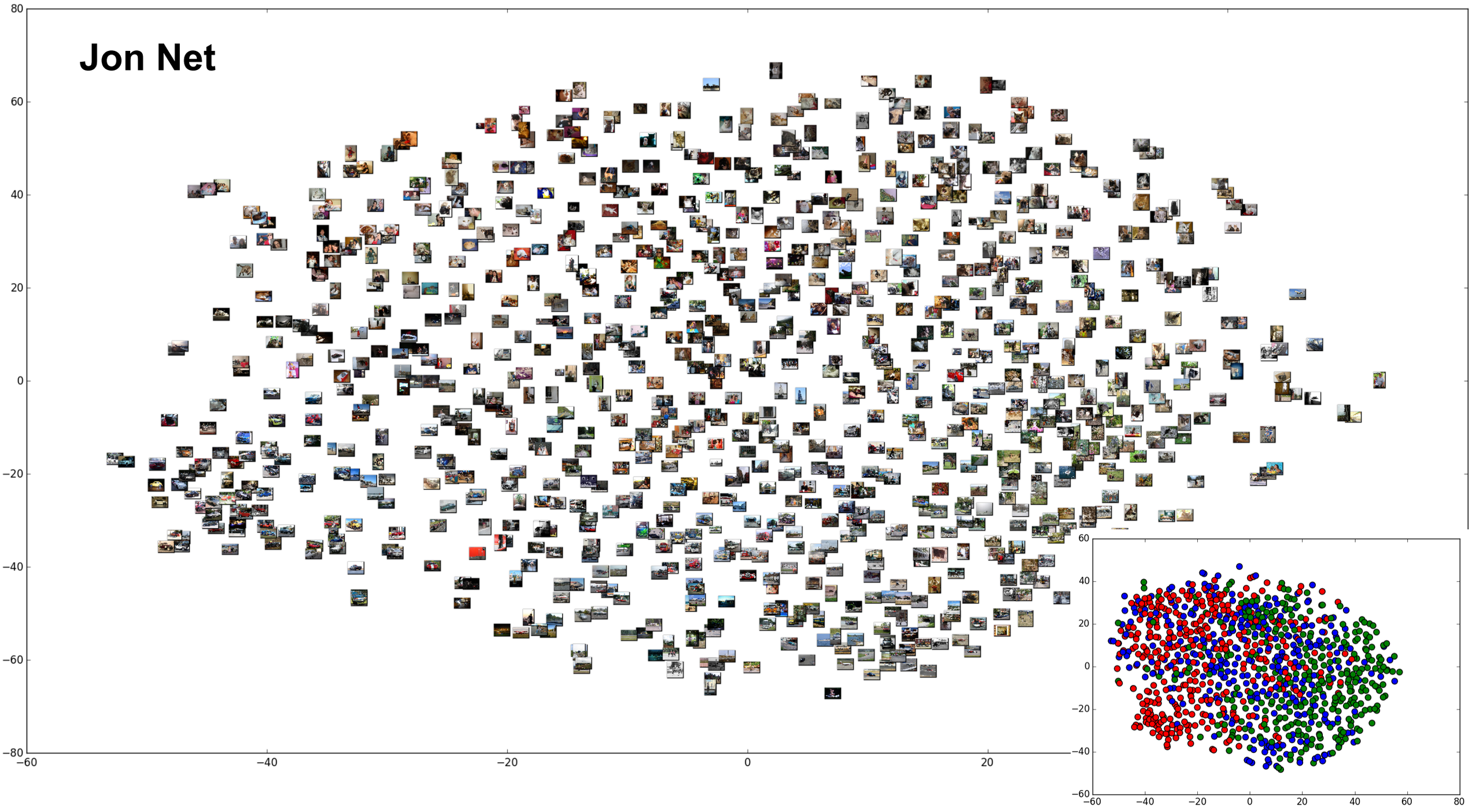
Extensions

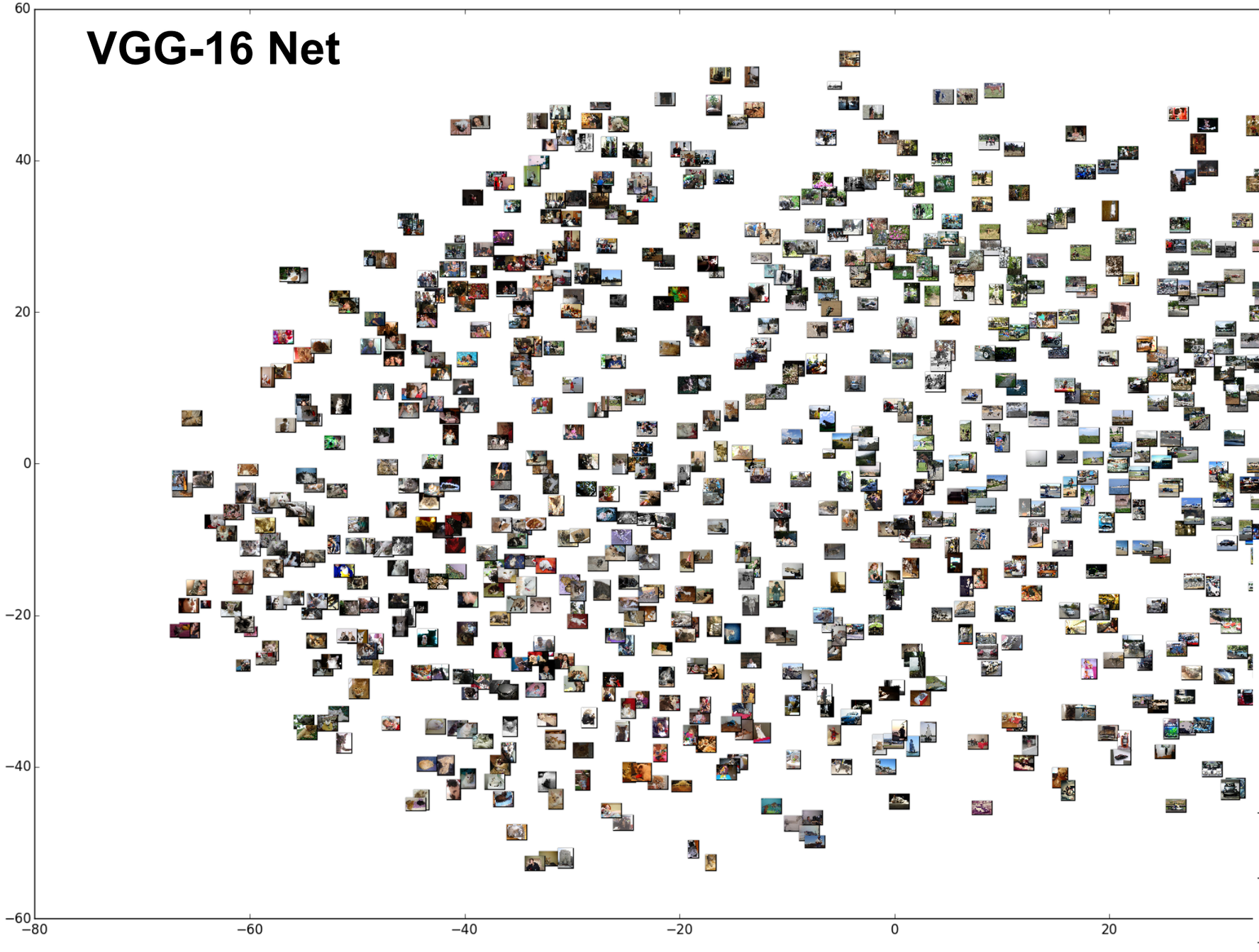
- VGG Network
  - Features extraction is complete (6 hours on 20 GPUs)
  - Currently using fc7 non-rectified features
  - SVM training is in progress

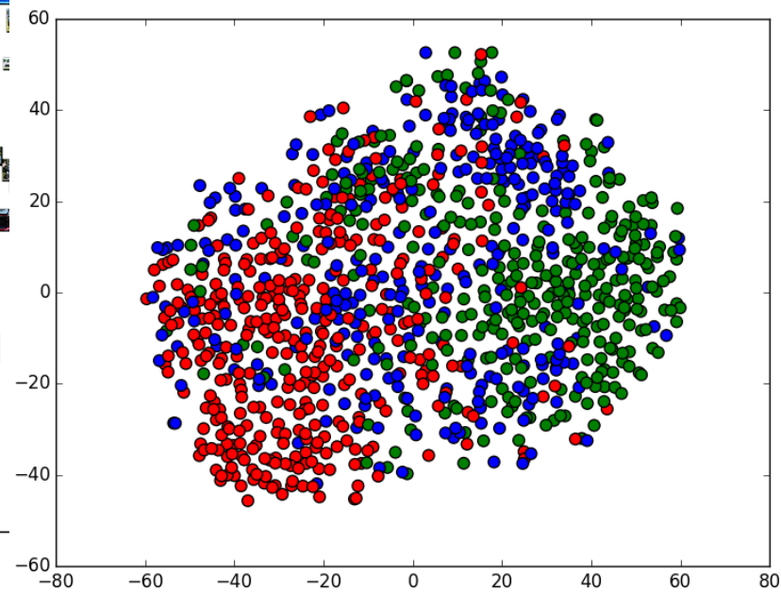
R-CNN

Ranjay Krishna

Jon Net


80
60
40
20
0
−20
−40
−60
−80
−60
−40
−20
0
20
60
40
20
0
−20
−40
−60
−60
−40
−20
0
20
40
60
80


VGG-16 Net
60
40
20
0
−20
−40
−60
−80
−60
−40
−20
0
20


60
40
20
0
−20
−40
−60
−80
−60
−40
−20
0
20
40
60
80

Initial Results

	Car	Cat	Person	mAP
INITIAL	2.90	15.83	3.12	6.28
HOG	6.73	9.86	9.85	8.8133

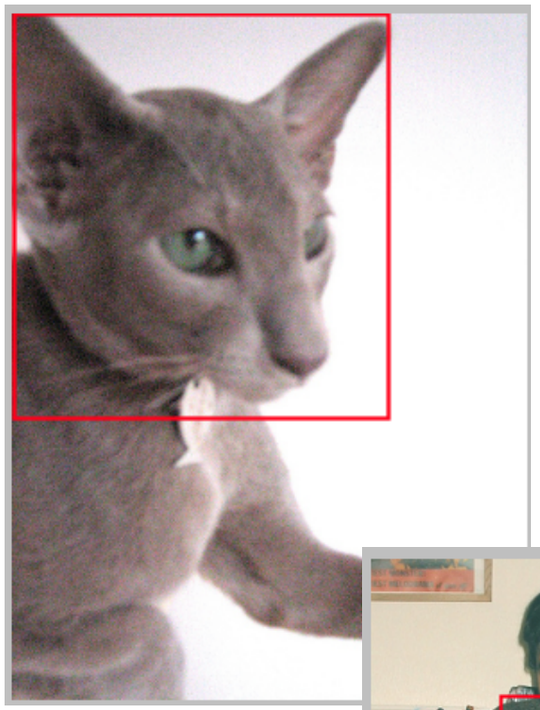
Training

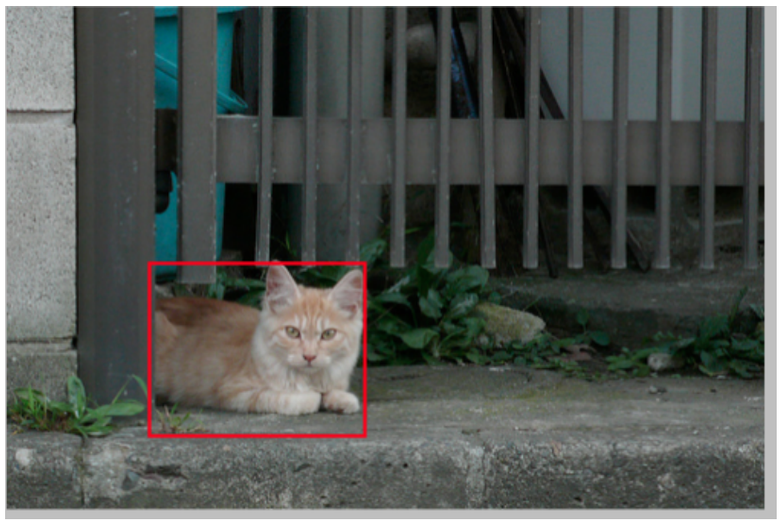
Testing

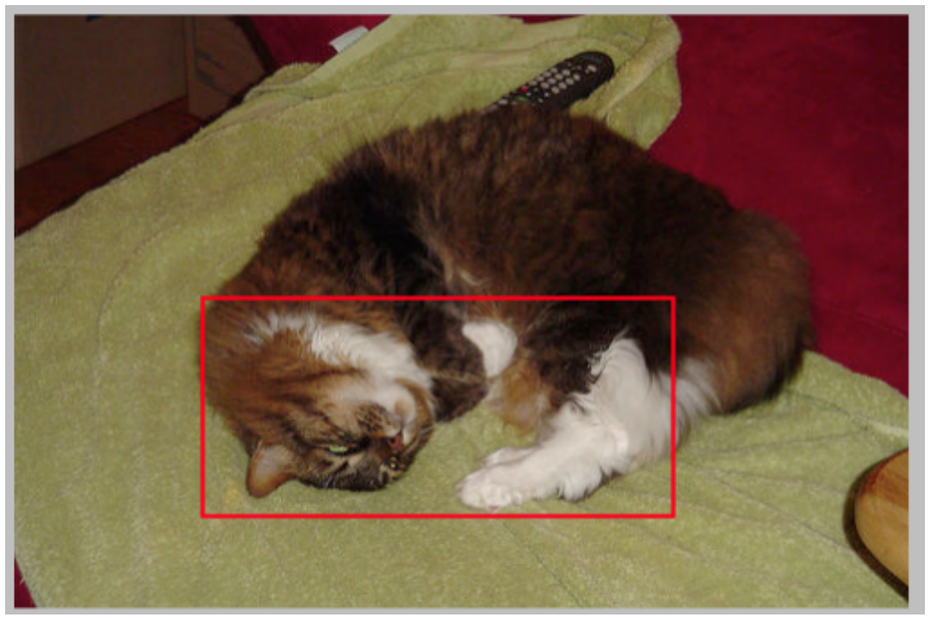


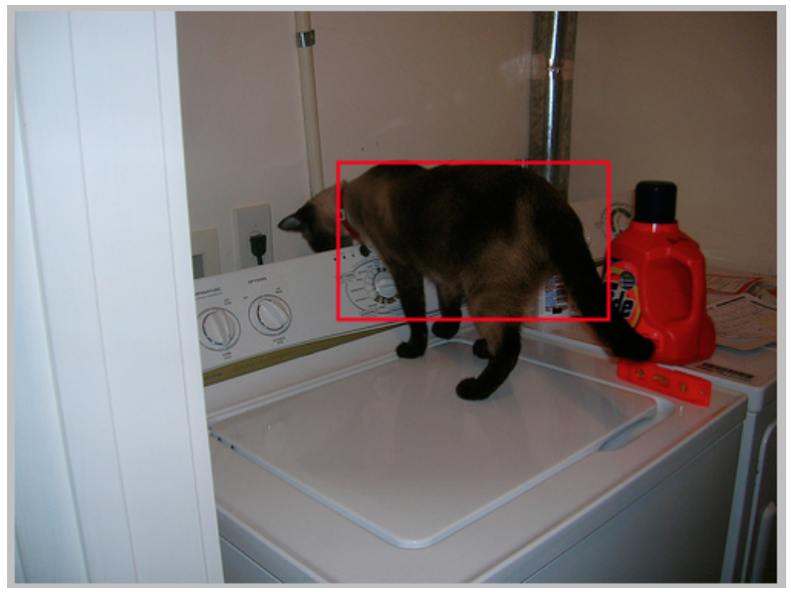



Cats Cats Cats









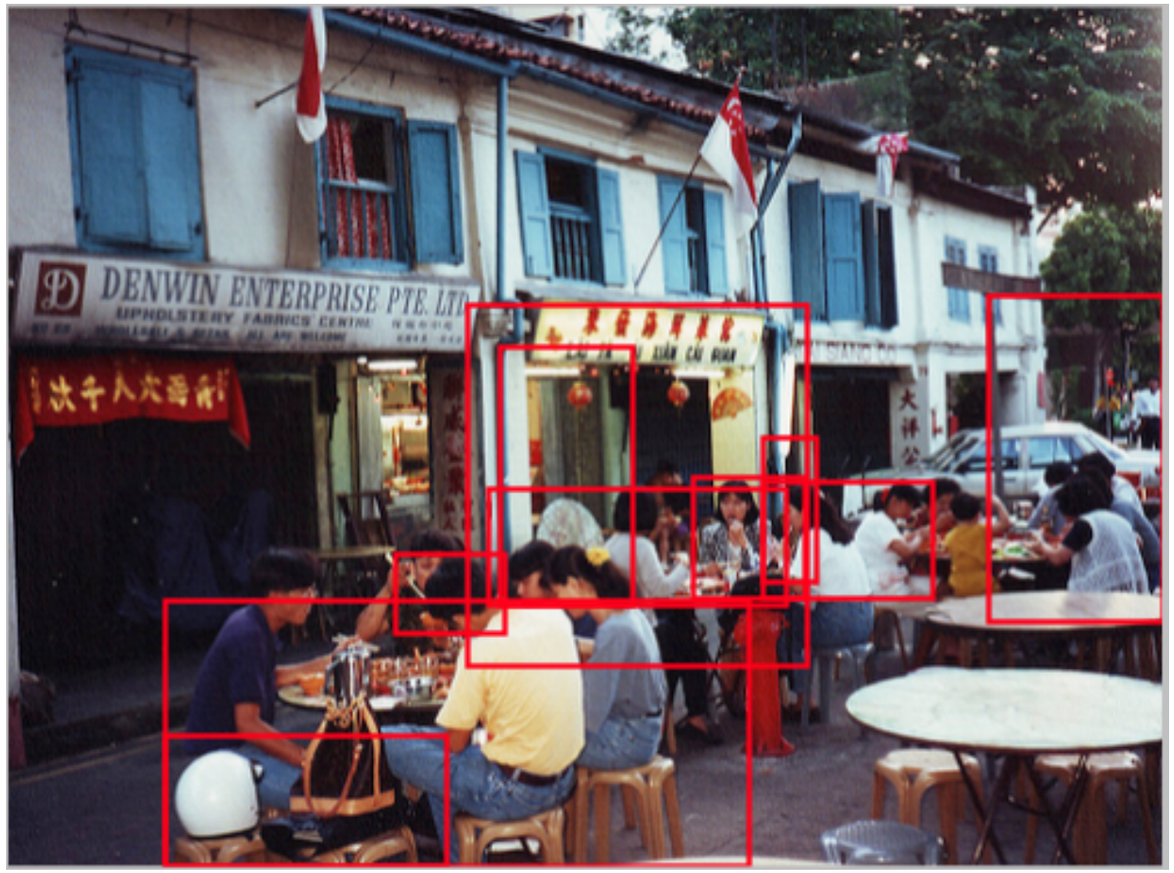
Results with Corrected Features

	Car	Cat	Person	mAP
INITIAL	2.90	15.83	3.12	6.28
HOG	6.73	9.86	9.85	8.8133
CORRECTED	20.23	21.56	22.44	21.41

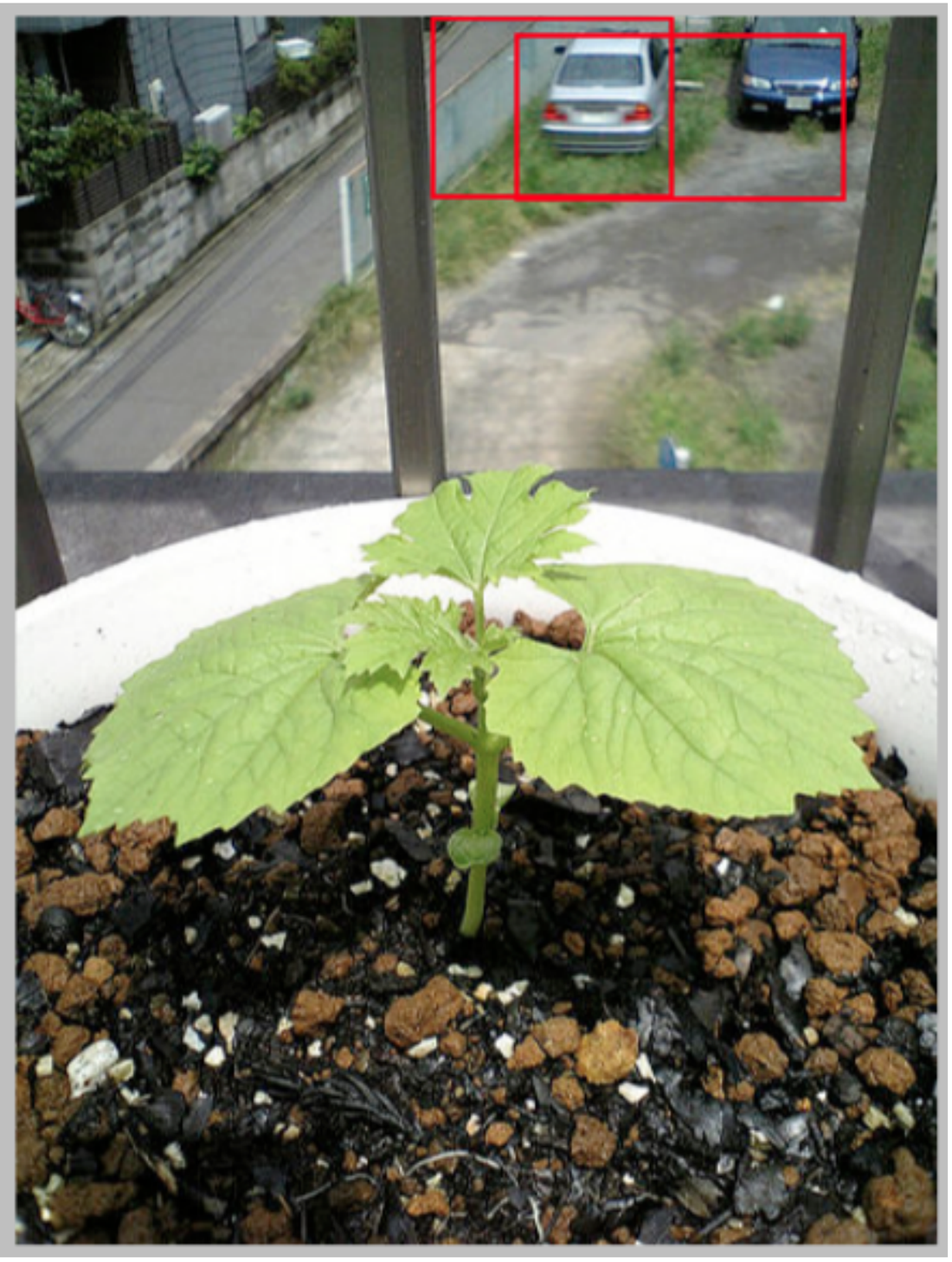
Results with Corrected Features

	Car	Cat	Person	mAP
INITIAL	2.90	15.83	3.12	6.28
HOG	6.73	9.86	9.85	8.8133
CORRECTED	20.23	21.56	22.44	21.41
VGG-16	21.43	22.48	23.12	22.34

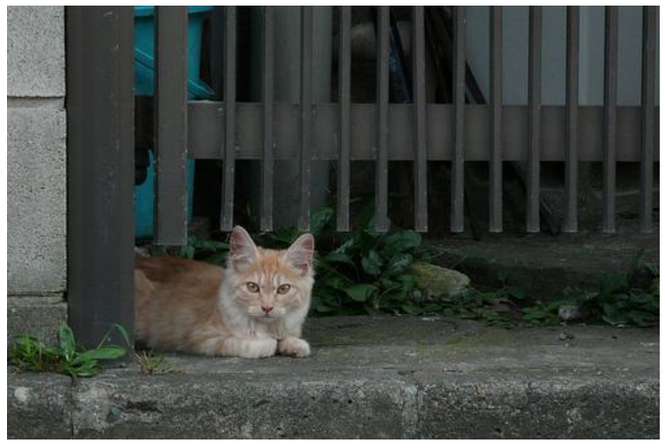
Failure Cases



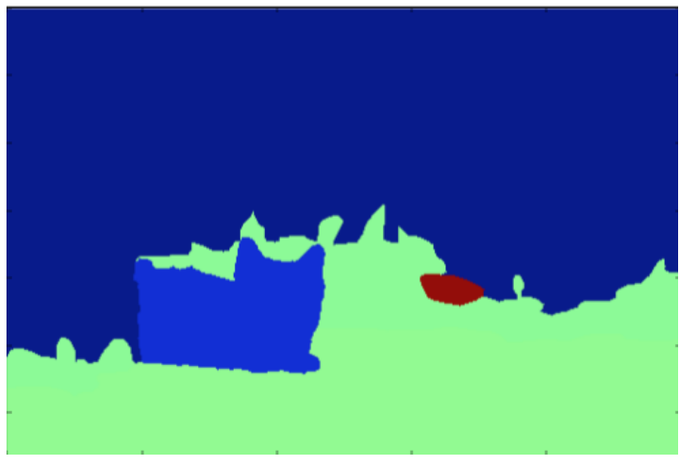
Non Maximum Suppression Errors



Segmentation

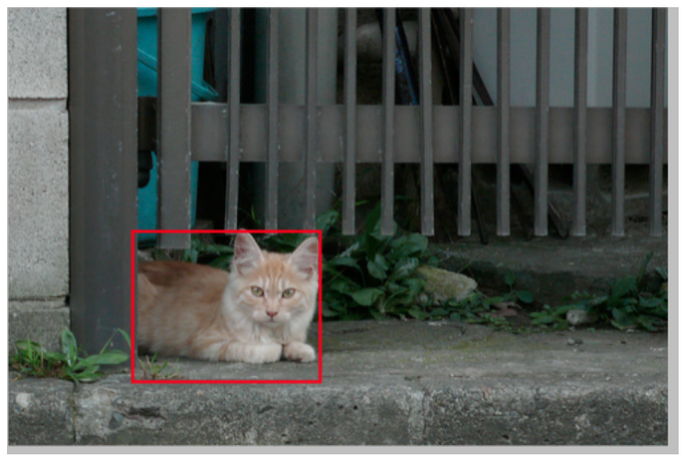


Original



Pascal Segmentation

Segmentation

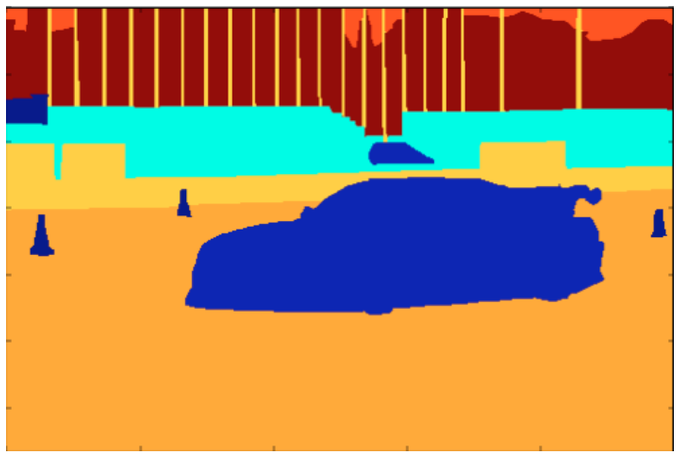


Detection

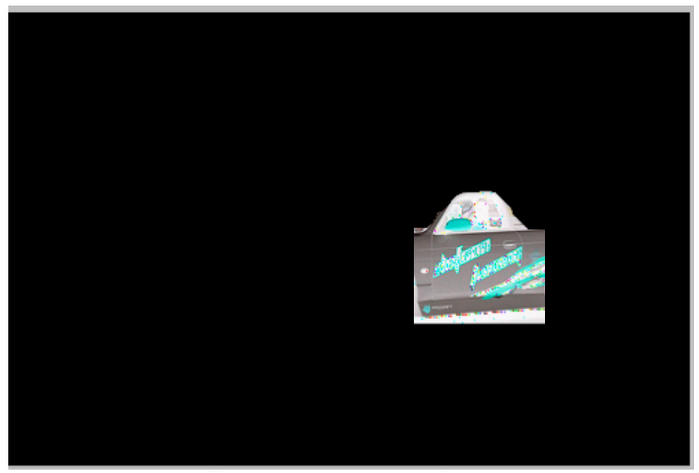
Oversegmentation

Segmentation + Context

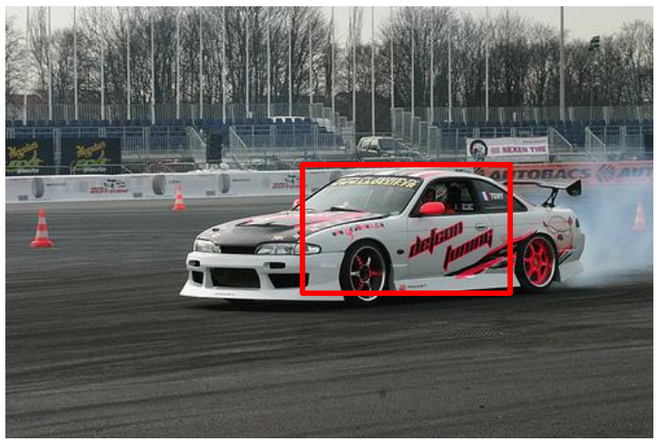
Original



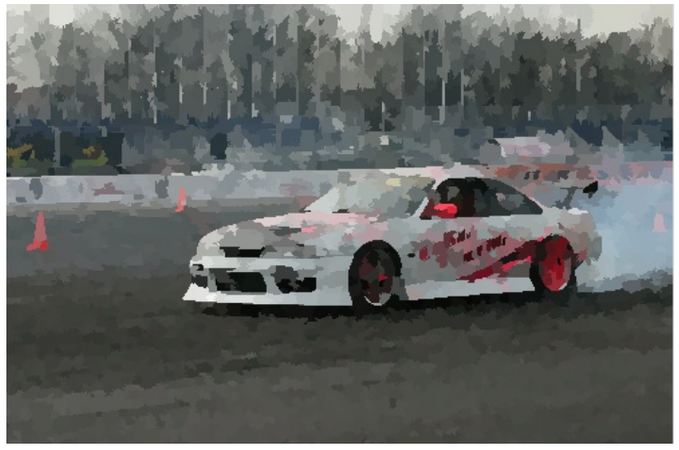
Pascal Segmentation



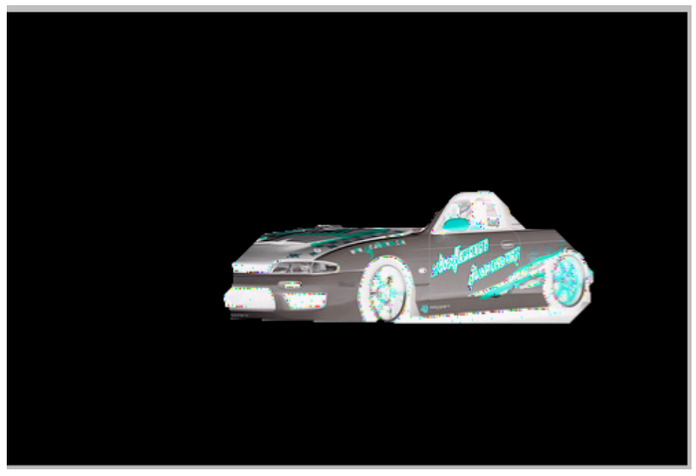
Segmentation



Detection

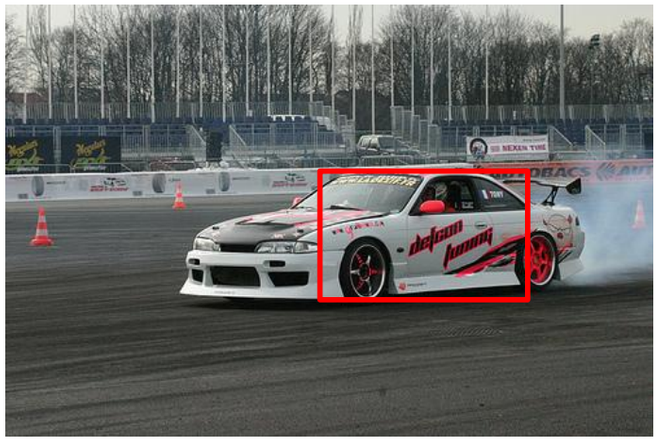


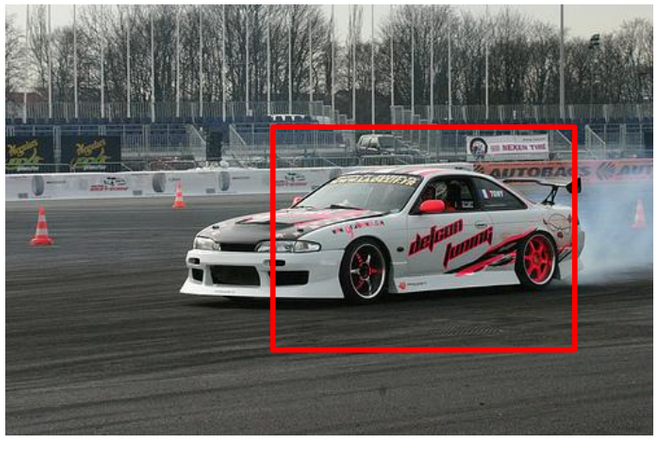
Oversegmentation



Segmentation with Context

Segmentation + Context





	Accuracy	Jaccard
Large Bounding Boxes	88.90	64.96
Perfect Bounding Boxes	95.16	75.80
Default Bounding Boxes	93.23	75.73
+ Superpixels	93.23	81.12
+ Dynamic Center Prior	95.66	83.00
+ RCNN + Context	**64.76**	**43.21**

On randomly selected 50 images from test set

RCNN

CS231b 2015 Project 3

Iretiayo Akinola
Josh Tennefoss

Stanford University

Overview

- Full implementation in Python, used Amazon GPUs
- Computed features using 16 pixel context, no account for CNN scaling
- Trained the SVM using finite set of positive and negative examples
- Grid search for hyperparameters
- Sub-maximal suppression by intersection and clustering
- Bounding box regression did not help
- Best MAP = .33, with no bounding box regression
- Aspect ratio extension

SVM

- Positive Examples
  - 5 with highest IOU with truth
- Negative Examples
  - 10 with IOU closest to .2
  - 10 with IOU closest to 0
- Parameters - extensive grid search
  - C = 1
  - gamma = 1e-3
  - kernel = poly

Visualizations:
white = true bbox
blue = person
green = car
red = cat

Sub-maximal Suppression

- Ignore all boxes whose probability of being no-object is > .85
- Intersection
  - Remove bounding boxes with IOU > .3 and not-max
- Clustering
  - DBSCAN clustering on the center of the bounding boxes
  - Did not improve performance, only helped with Cats

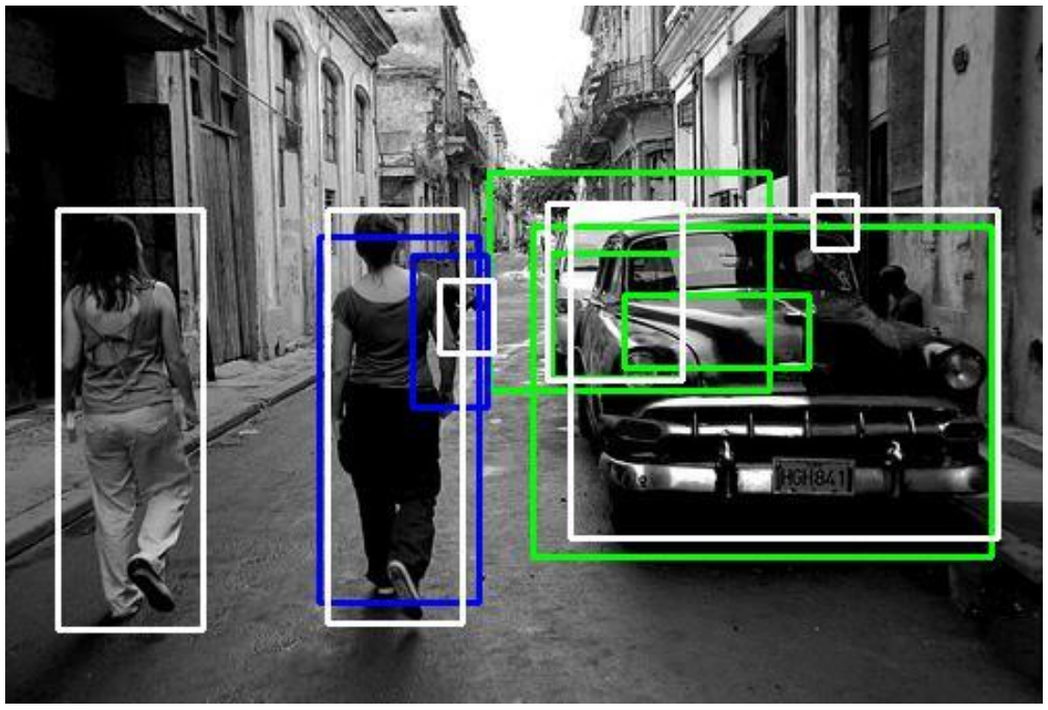
Intersection

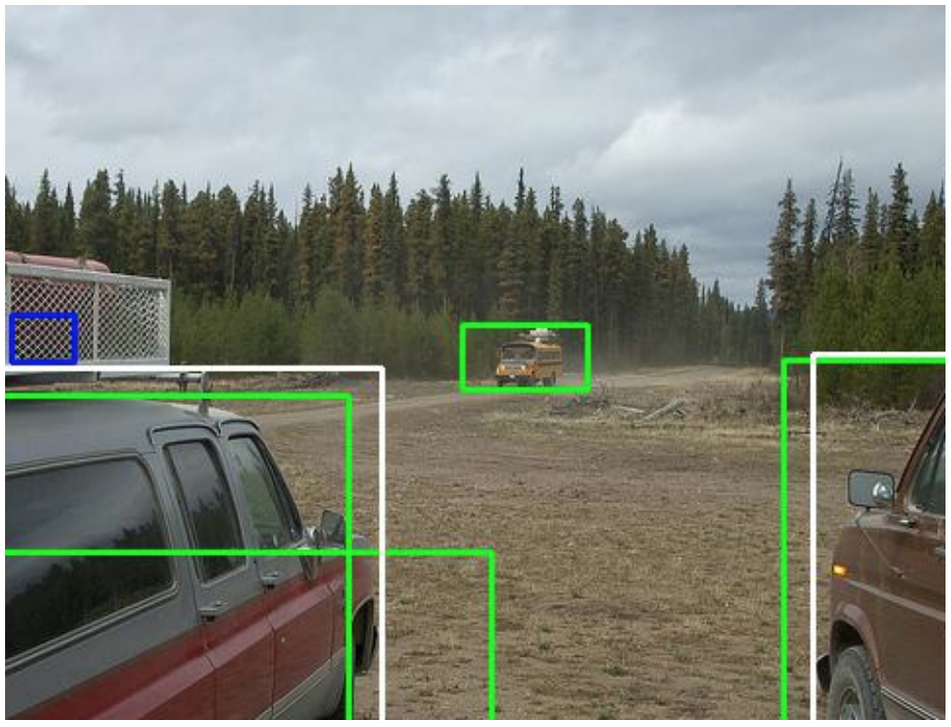
Measure	no Reg	with Reg
Car AP	.28	.29
Cat AP	.45	.46
Person AP	.27	.26
MAP	.3349	.3345

Clustering

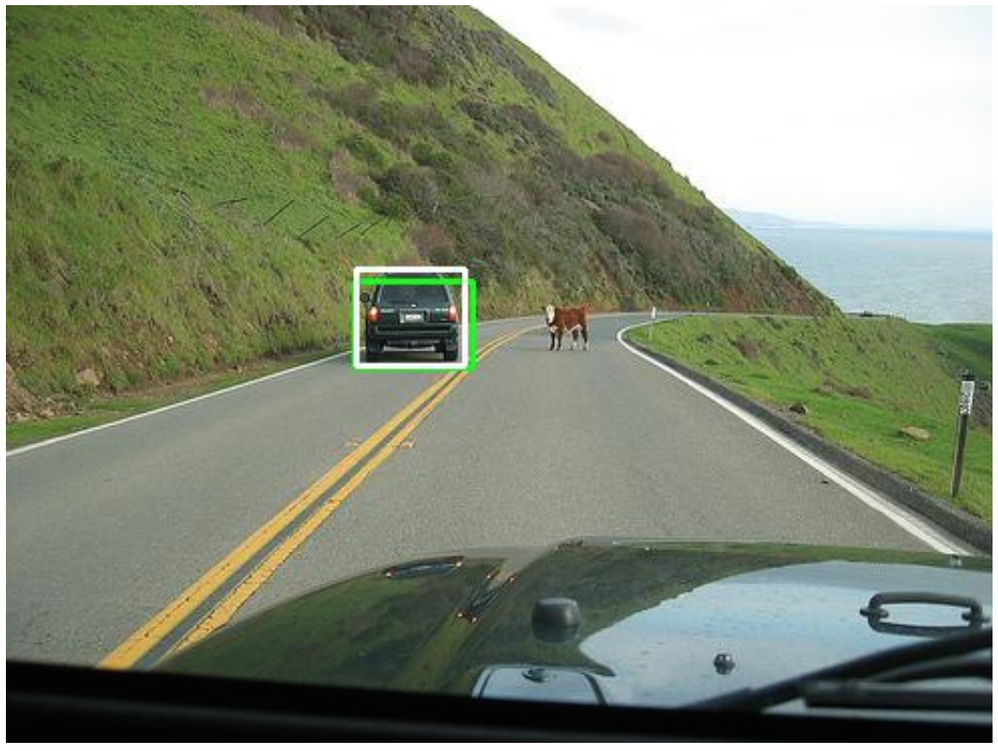
Measure	no Reg	with Reg
Car AP	.23	.23
Cat AP	.49	.49
Person AP	.19	.19
MAP	.304	.303

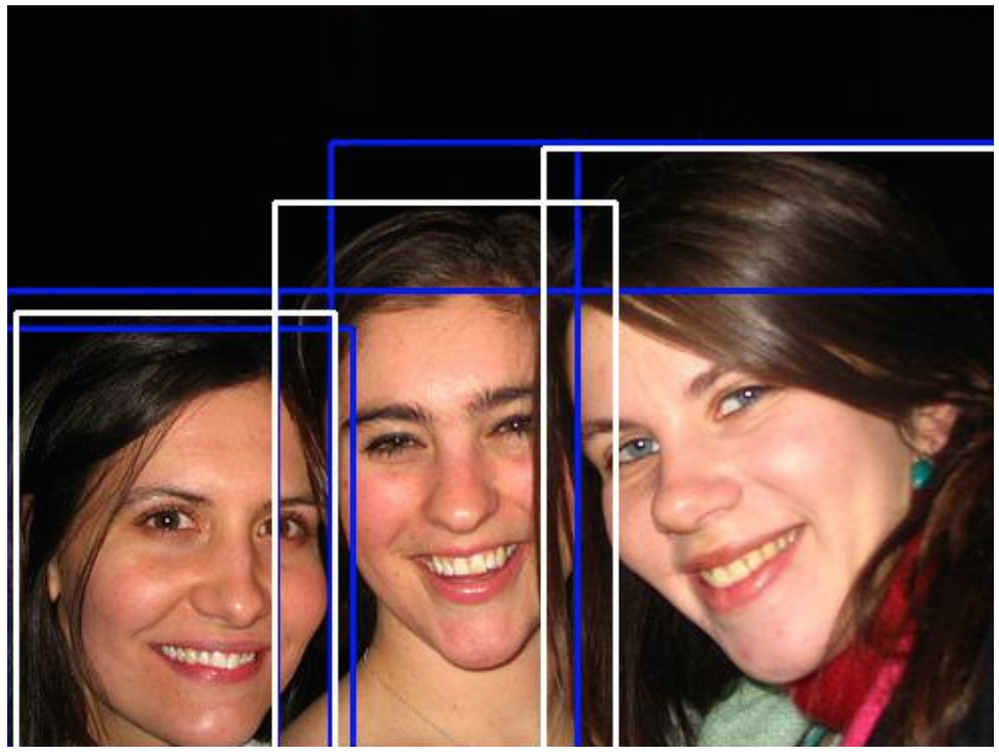
Bad Predictions



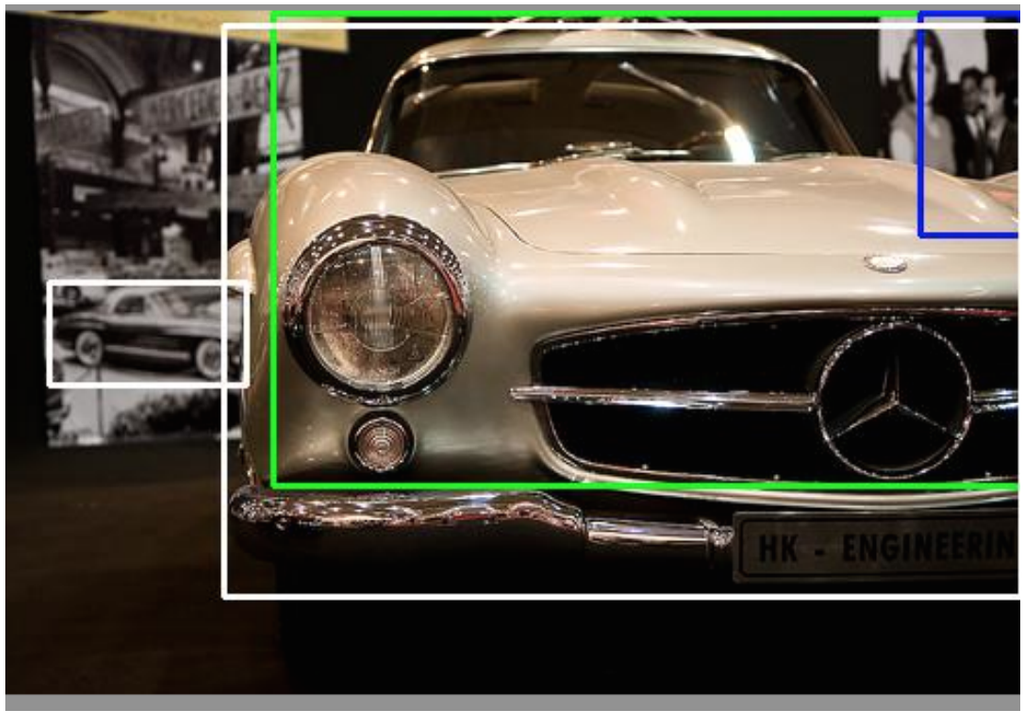


Good Predictions

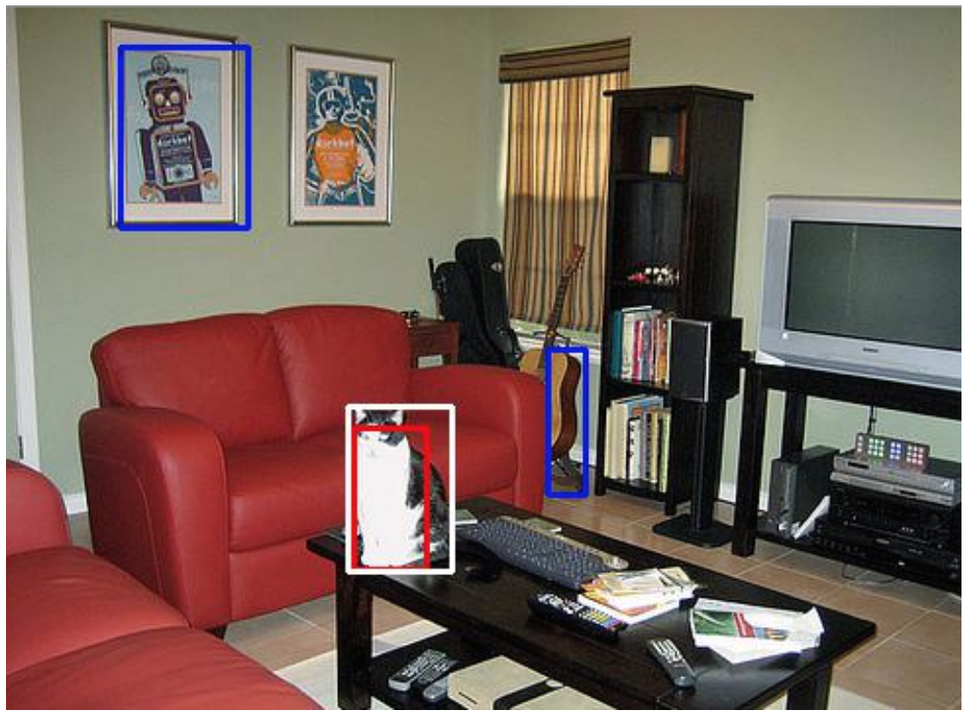




Interesting Predictions



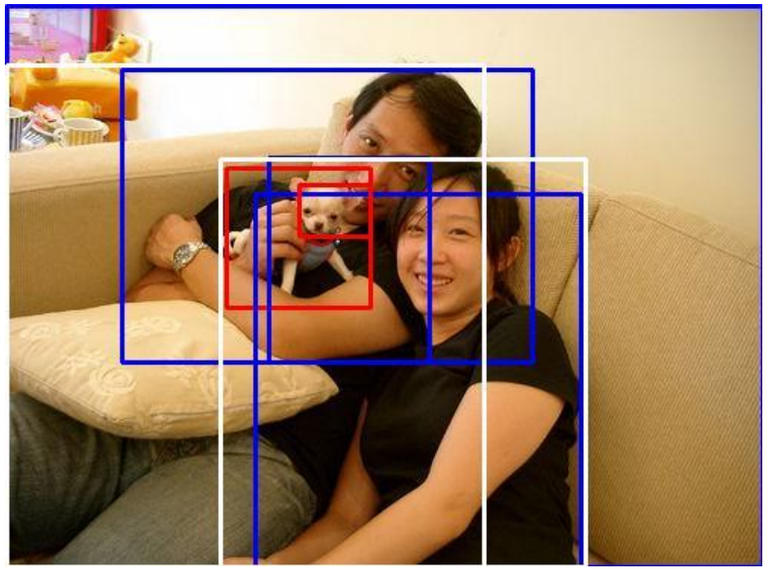
Interesting Predictions



Other Experiments

- Recovering from warping Distortion
  - Integrate aspect ratio into feature vector
  - Result:
    - No Reg - [Car: 0.31, Cat: 0.48, Person: 0.27, mAP: 0.35]

      [Car: 0.28, Cat: 0.45, Person: 0.27, mAP: 0.33]



Other Experiments

Ongoing experiments

- Hard Negative Mining
- BBox merging
- Heuristics for Selecting Training Examples:
  - Positive Examples: IOU_t > 0.9
  - Negative Examples: (IOU_t < 0.3) AND (IOU_n > 0.7)

*IOU_t - IOU with truth box

*IOU_n - IOU with assigned negative example

For The Curious

Precision and Recall By Prediction

Prediction Num	Tag	Avg Precision	Avg Recall
1	car	.57	.39
2	car	.39	.47
3	car	.28	.45
4	car	.22	.41
5	car	.20	.46
1	person	.60	.36
2	person	.43	.43
3	person	.32	.43
4	person	.33	.46
5	person	.28	.43
1	cat	.63	.60
2	cat	.40	.77
3	cat	.28	.75
4	cat	.23	.75
5	cat	.20	.63

Object Detection

Jasper Lin

6/2/15

Code Difficulties

- Finding the correct regularization weight for rcnn_train
- Different regularizations can can cause large amounts of detections or no detections
- Training had the fewest cat features

Training Performance

- Experimented with regularization between 0.001 to 1
- Experimented with different weighting of classes (5x, 10x, 50x)

Qualitative Notes

- Low regularization for negative examples results in a lot of noisy detections

- Too high regularization for positive examples made it difficult to converge

Training Results

1 Epoch

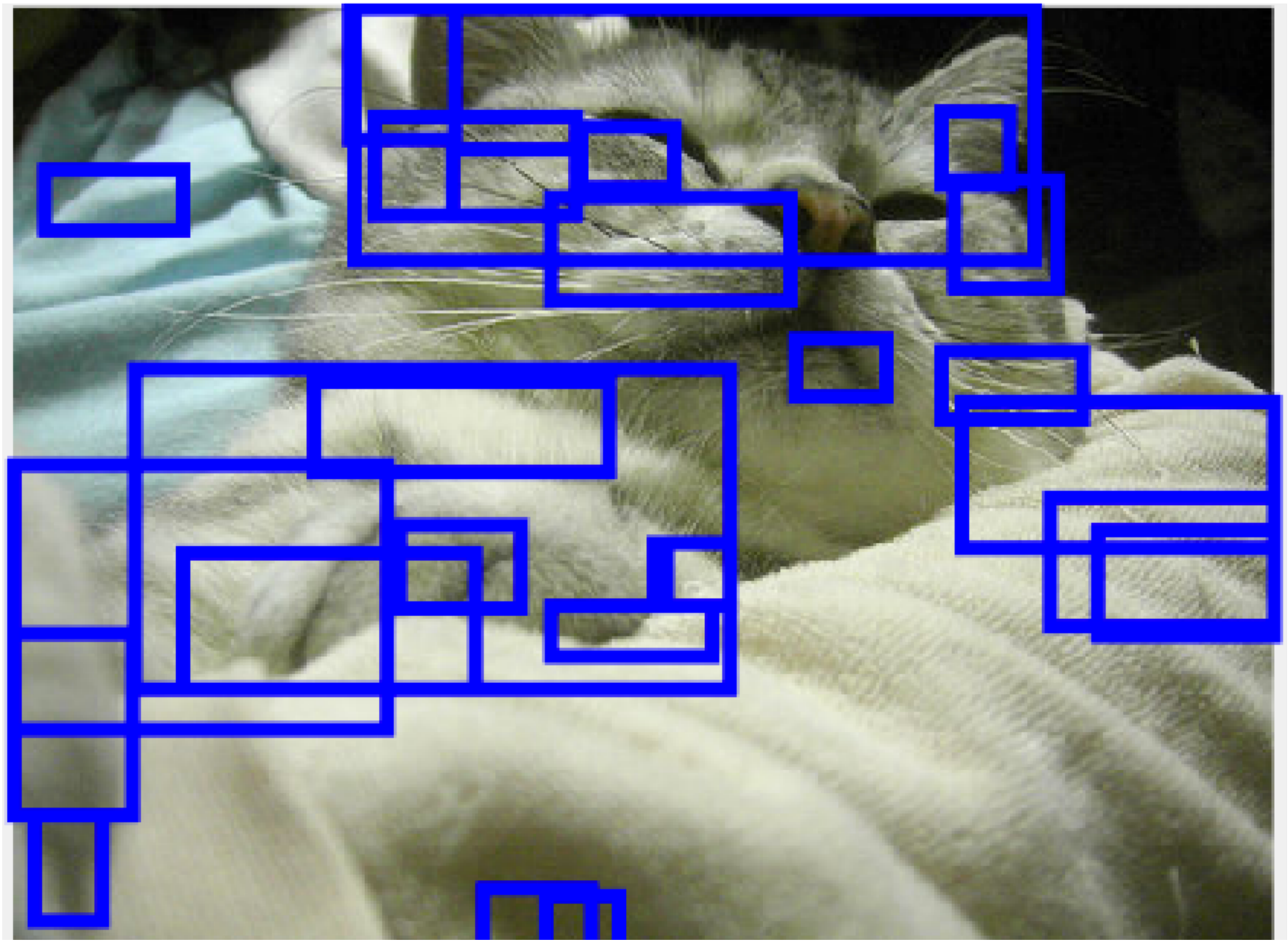
	Number of Features	Accuracy
Car	*54650*	*98.6112%*
Cat	*22671*	*98.3944%*
Human	*78825*	*95.608%*

2 Epochs

	Number of Features	Accuracy
Car	*106010*	*98.619%*
Cat	*42165*	*98.2782%*
Human	*158914*	*95.8651%*

Average Precision

	1 Epoch	2 Epochs
Car	*0.00021*	*0.00043*
Cat	*0.000332*	*0.00002*
Human	*0.00007*	*0.00005*



R-CNN Project

Eric Holmdahl

Implementation Details

- Haven’t gotten a working implementation yet

Extensions

- Dropout
- Testing different overlap thresholds
- Comparison of CNN with Fisher Vectors

Hope to have something soon!

R-CNN

Kyunghee Kim

- Feature extraction

- SVM training <--- working on this

extension

- Fine-tuning on VOC 2011 from VGG pre-trained model

  ----> trained this model

Taking late days

CS231B Project 3 R-CNN

Meng Wu

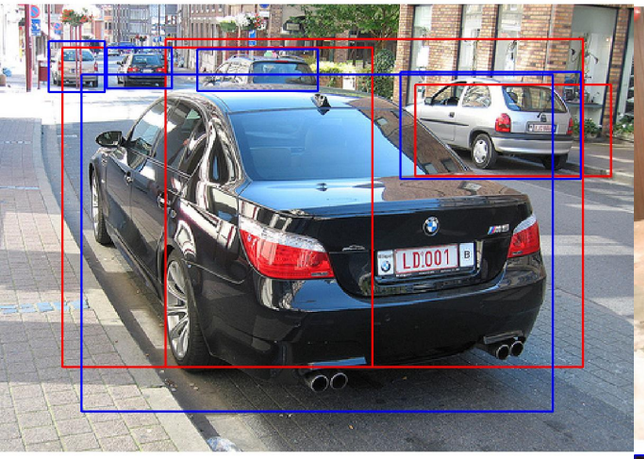
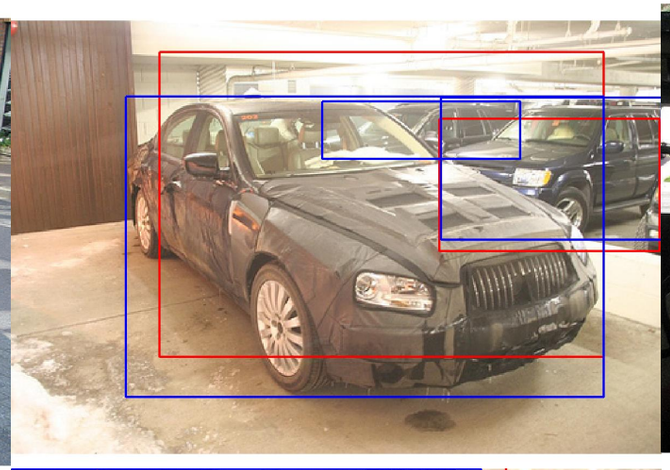
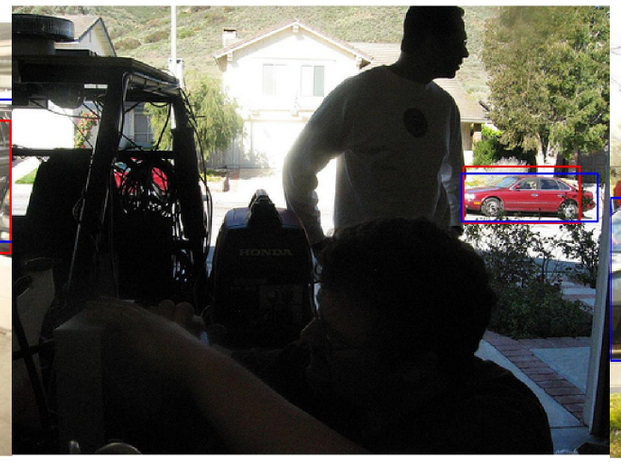
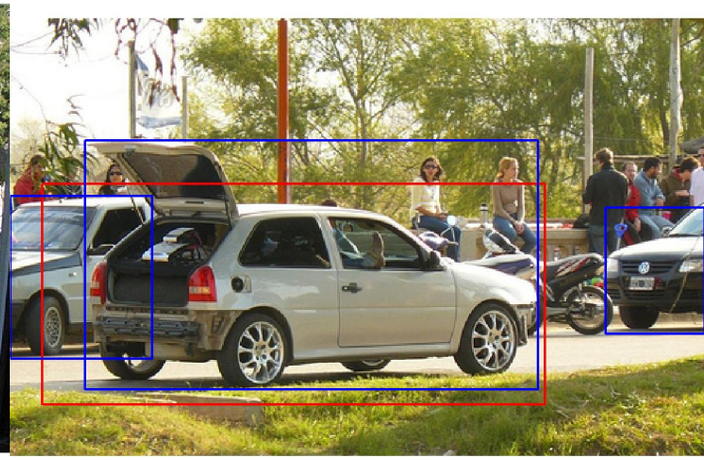
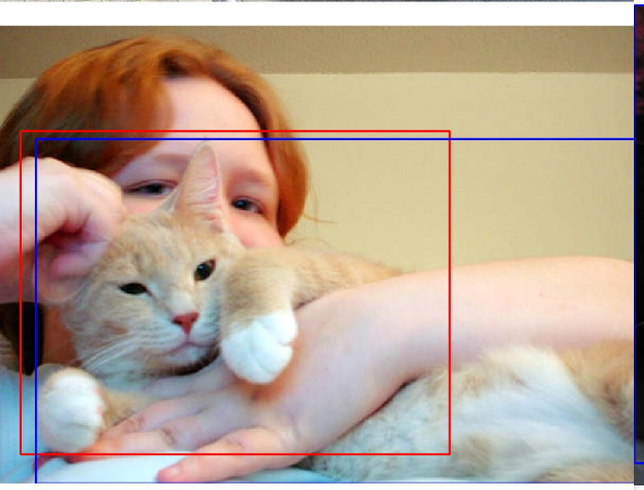
R-CNN

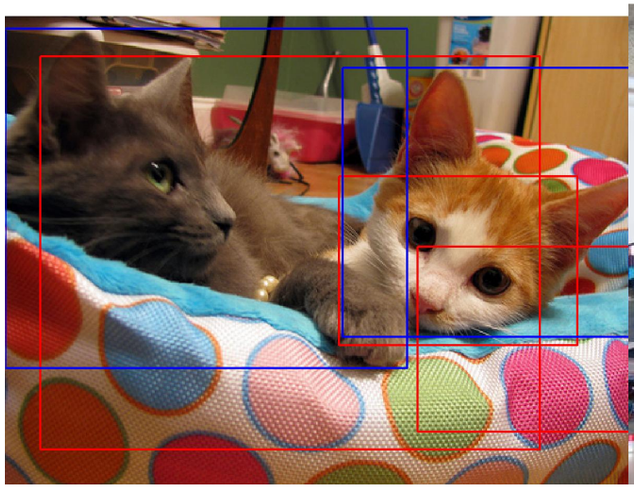
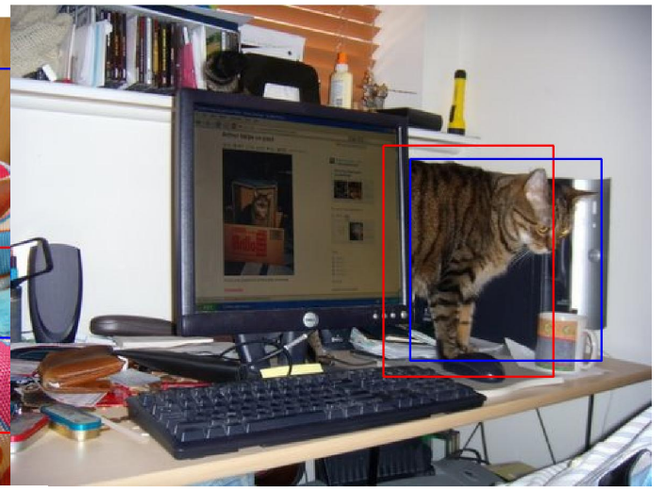
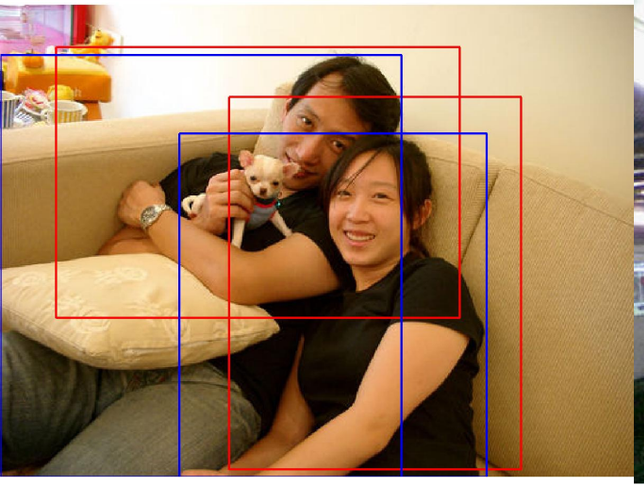
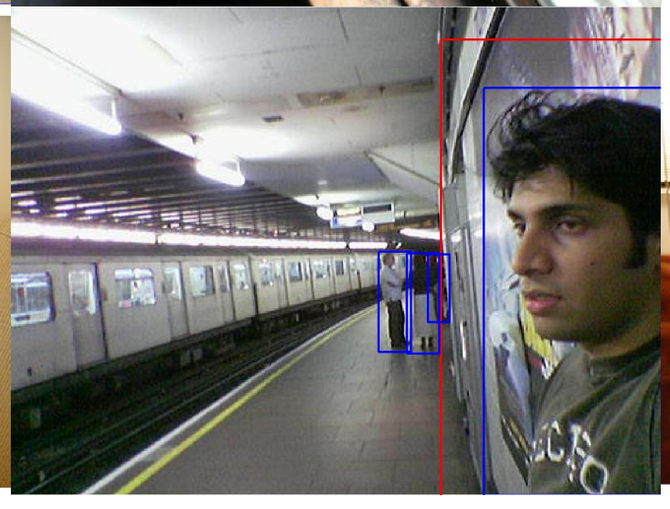
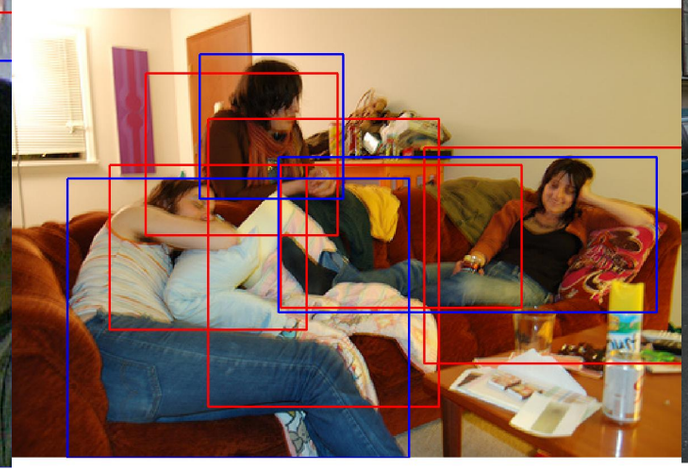
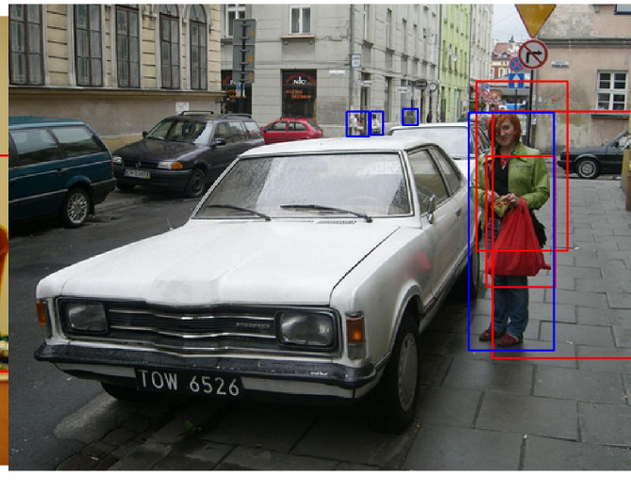
- 512-d image features using pre-trained AlexNet FC layer
- Warp proposed image patch to 277x277 with 8, 16, and 24 pixels around region
- Linear SVM used liblinear
- Non-maximum suppression and bounding box suppression generates final output

Many parameters to be decided

- Bounding box overlapping (IoU) threshold
  - Generate negative and positive example for SVM (0.3P / 0.01N)
  - Bounding box regression examples (0.6)
  - Non-maximum suppression (0.3)
- SVM Parameters
  - Numbers of positive and negative examples (P:20k/N:200k for hard SVM)
  - C parameter (1 or 1e-3)
  - SVM detection threshold (not constant for different cases -0.1 ~ 0.1)
- Bounding box regression
  - Lambda in ridge regression (not constant for different cases 1e4 – 5e4)

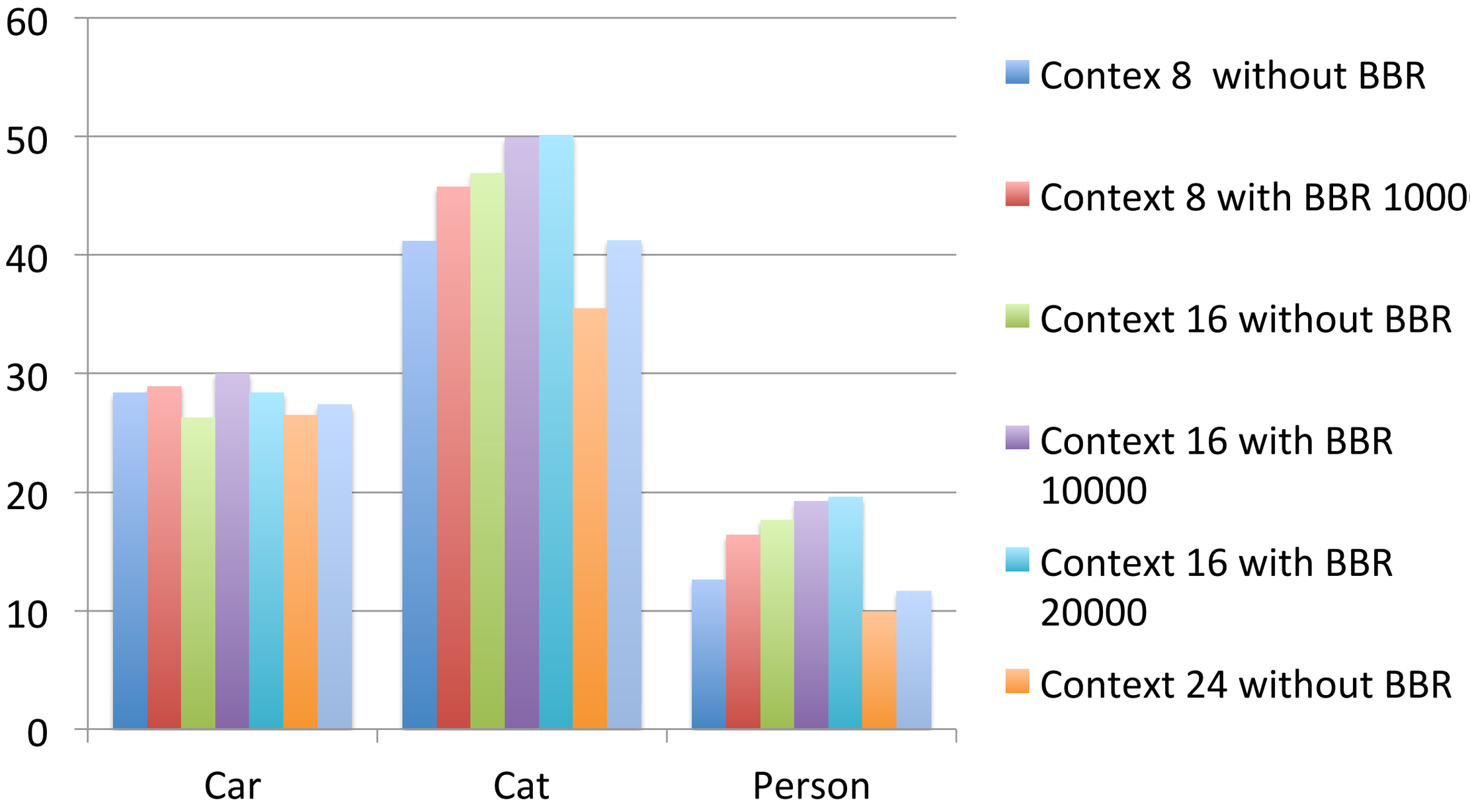
Results

R-CNN: red box, Ground truth: blue box

Quantitative Results



Other Challenges

- Multiple objects are close together
- Block
- Hard to include the entire object (region proposal)
- Confuse with similar objects
  - Car with bus and motorcycle
  - Cat with dog
  - Person with cat?