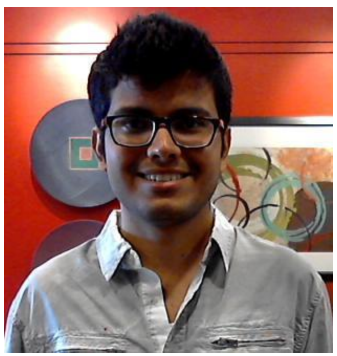
ECCV '14
Zürich, September 6-12th, 2014
European Conference on Computer Vision

Knowing a Good HOG Filter When You See It:

Efficient Selection of Filters for Detection



Ejaz Ahmed[1], Gregory Shakhnarovich[2], and Subhransu Maji[3]

[1] University of Maryland, College Park

[2] Toyota Technological Institute at Chicago

[3] University of Massachusetts, Amherst

Visual Category as Collection of filters

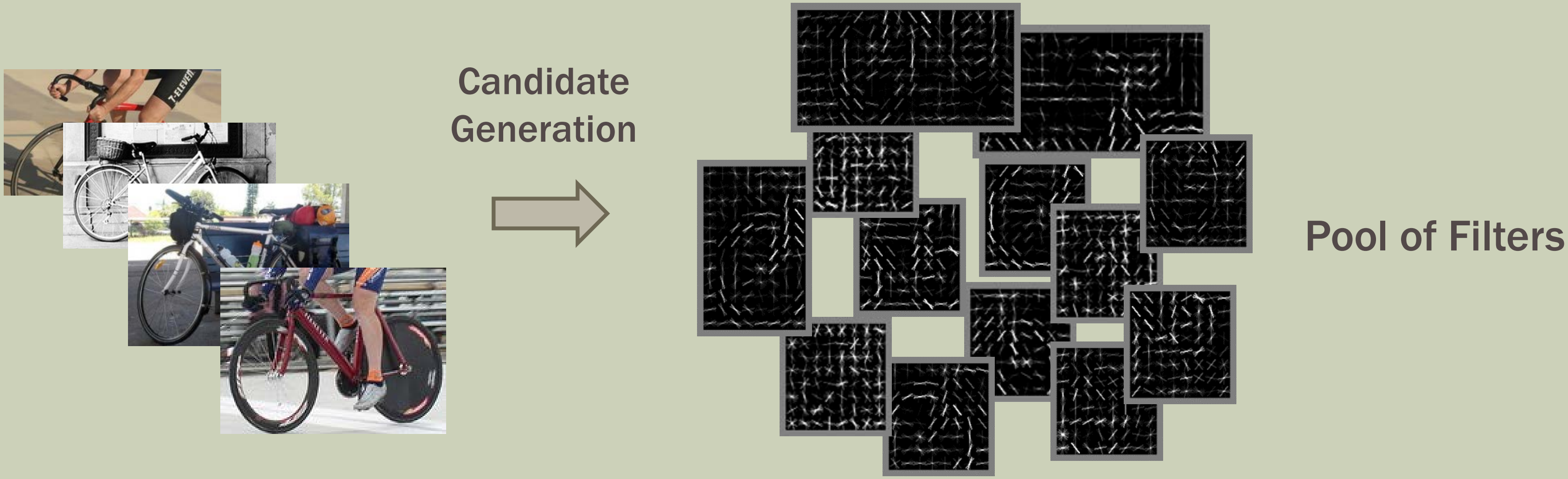
Poselets

Mid Level Discriminative Patches

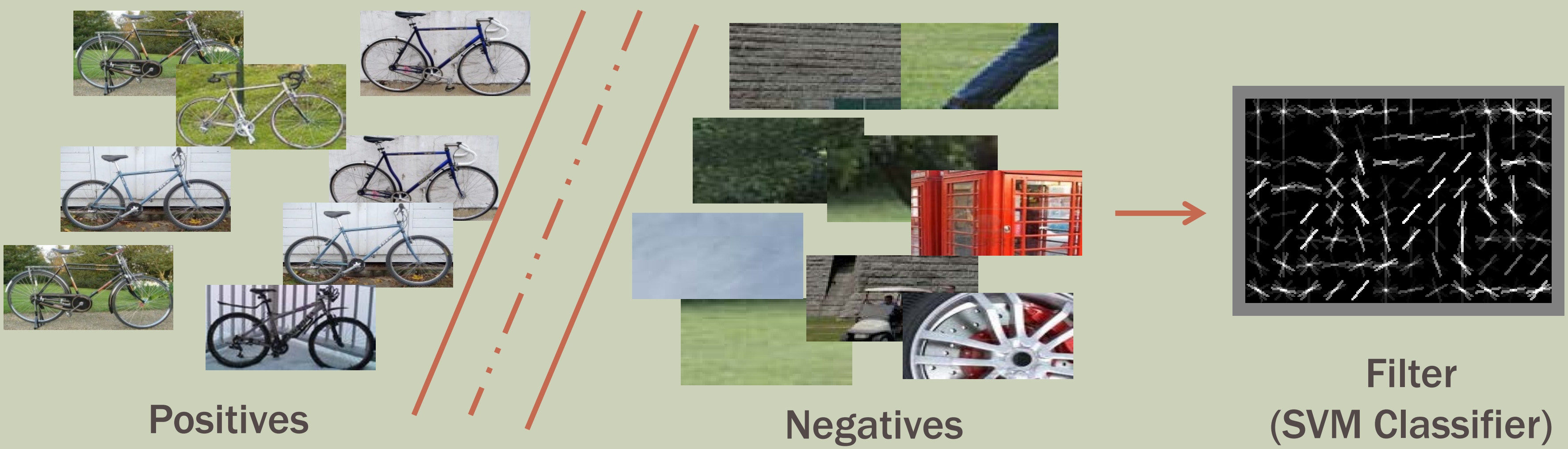
Exemplar SVMs

Candidate Generation

- Generation of a large pool of filters.

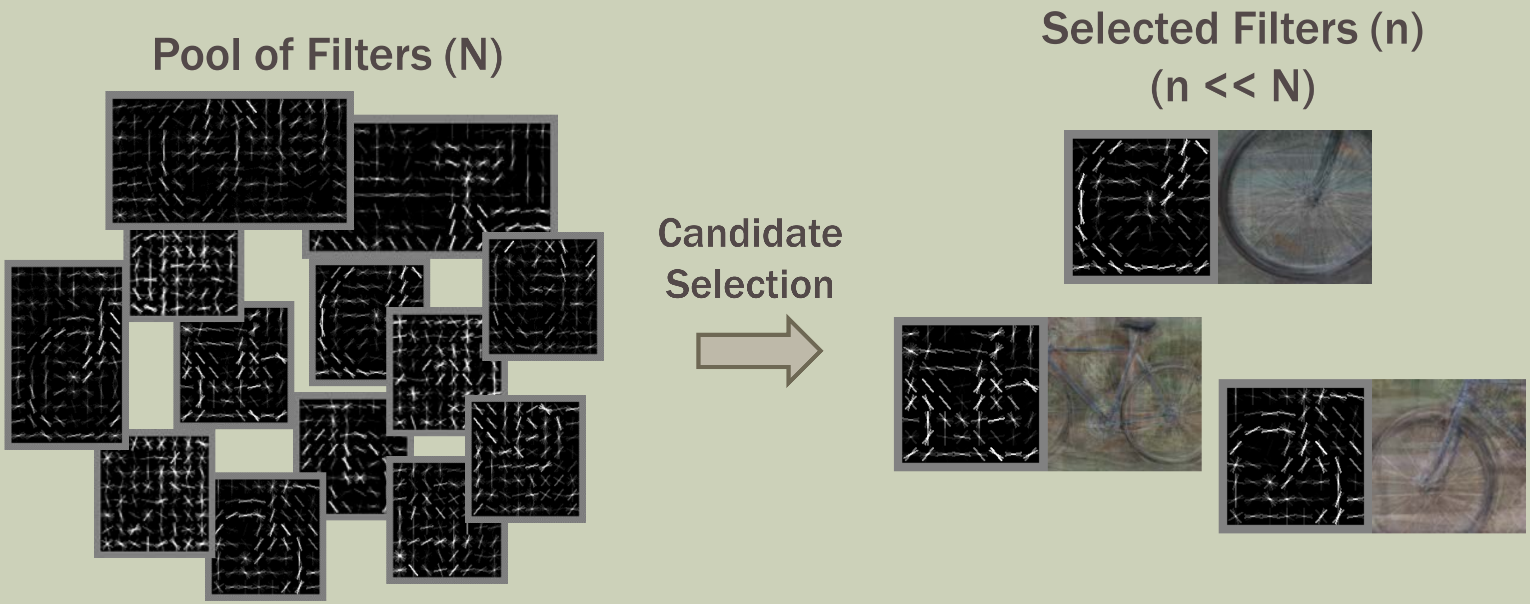


- Filters are generated using positives and negatives examples.

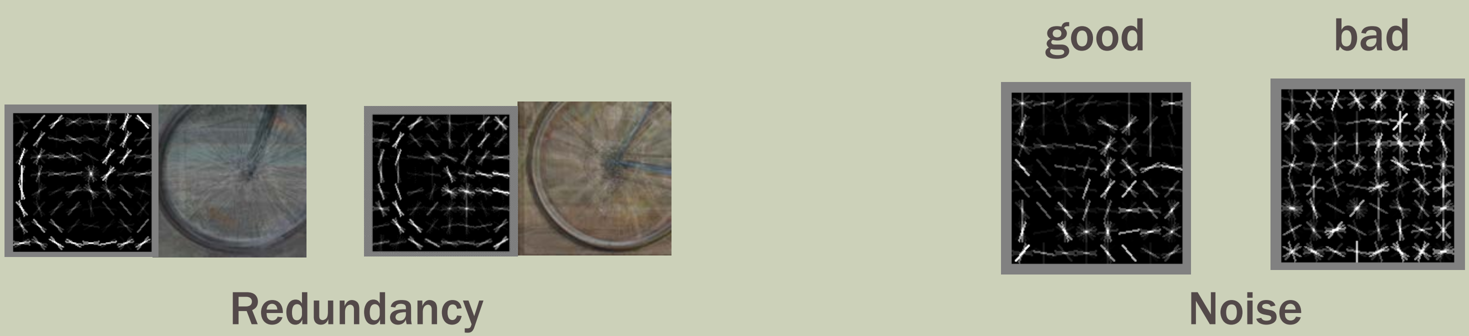


Candidate Selection

- Impractical to use all generated filters.



- Two sources of inefficiency



Candidate Selection Cont…

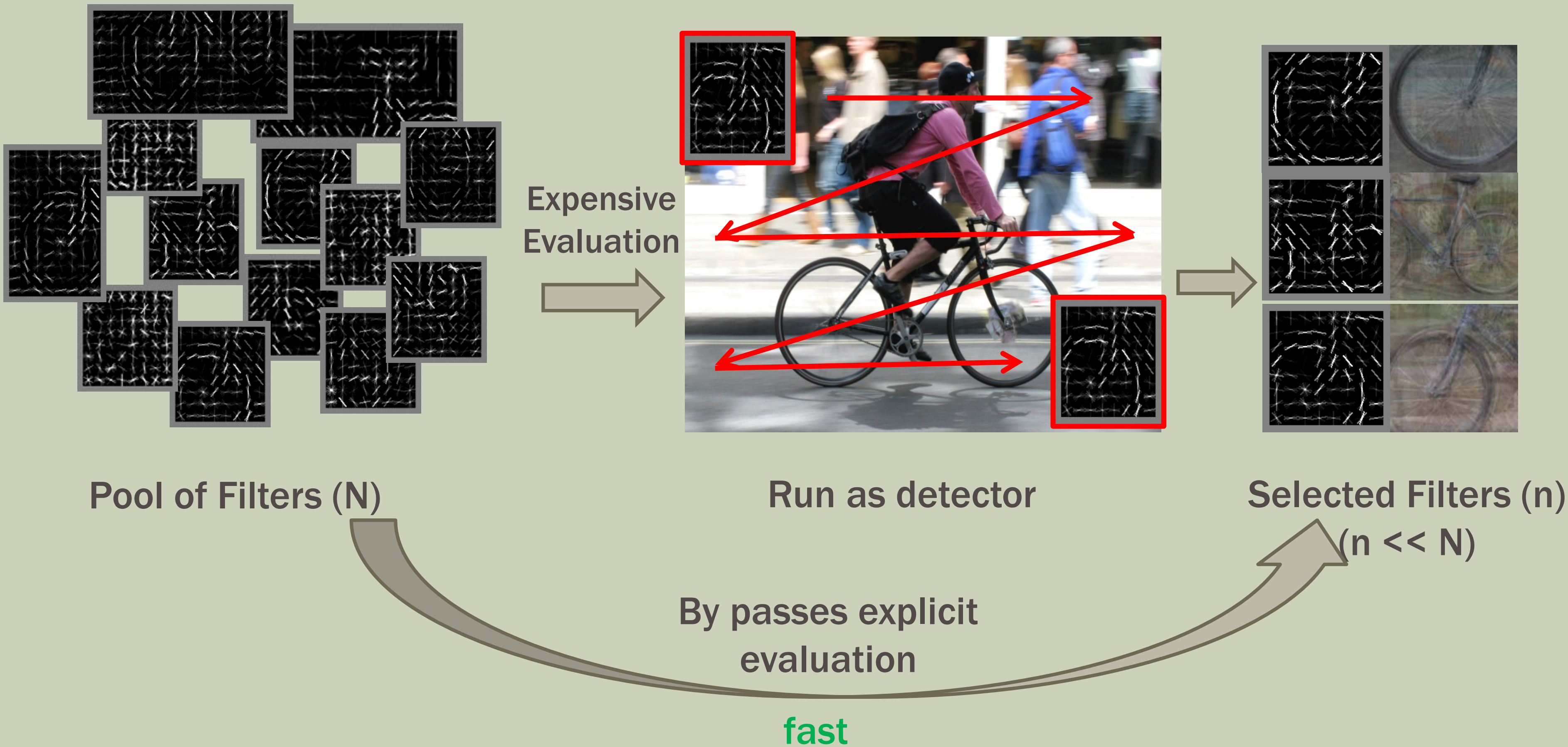
Pool of Filters (N)

Run as detector

Selected Filters (n)
(n << N)

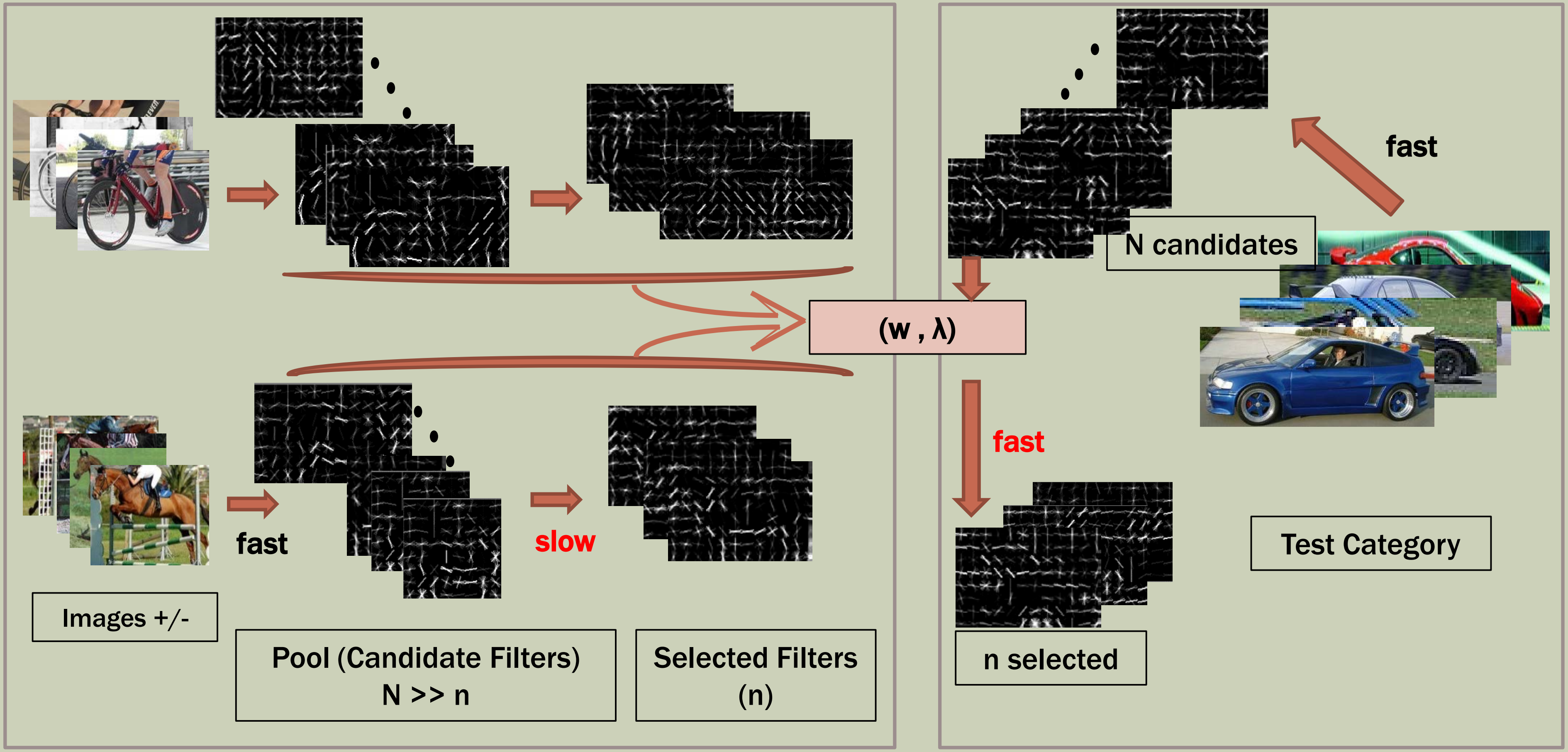
Bottleneck

What we Propose



Our Contribution : fast automatic selection of a subset of discriminative and non redundant filters given a collection of filters

Category Independent Model



- Can rank filters as accurately as a direct evaluation on thousands of examples.

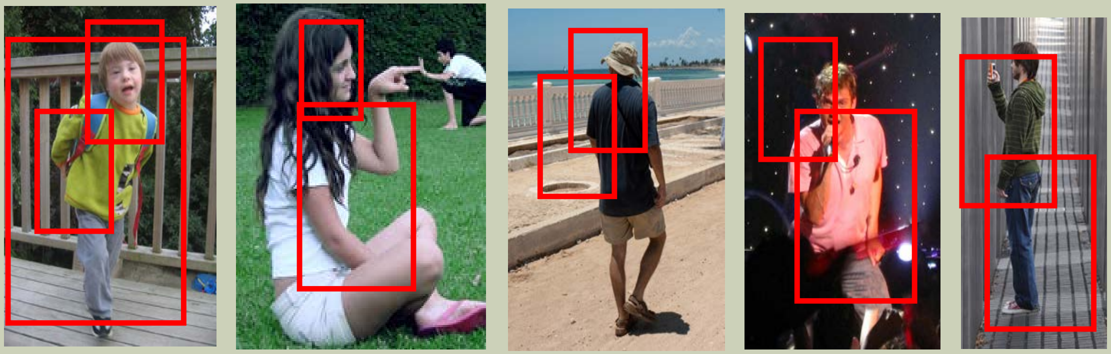
Poselets

- Poselets are semantically aligned discriminative patterns that capture parts of object.

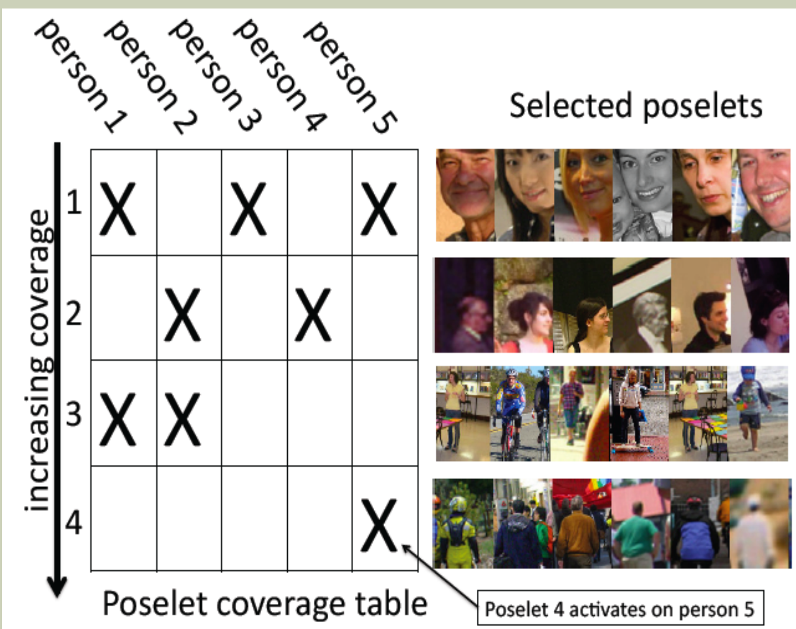
- Patches are often far visually, but they are close semantically

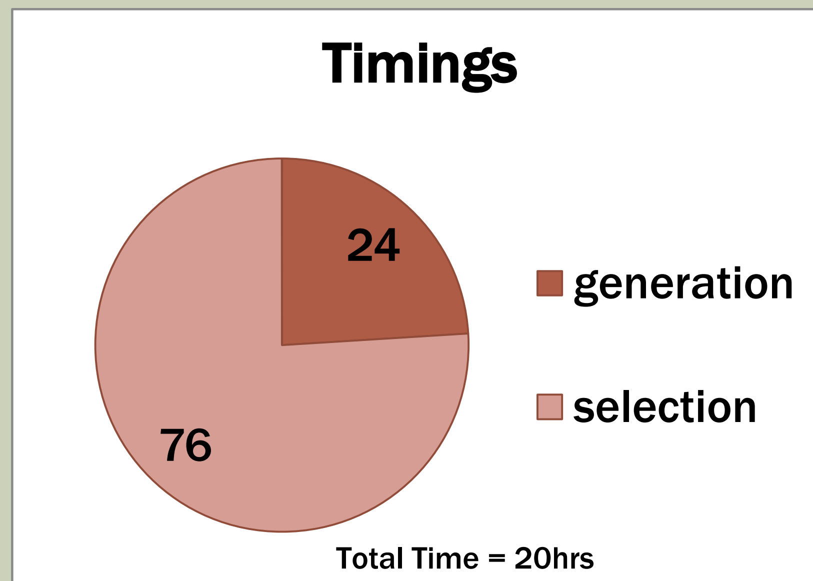
Poselet Architecture

- Candidate Generation :



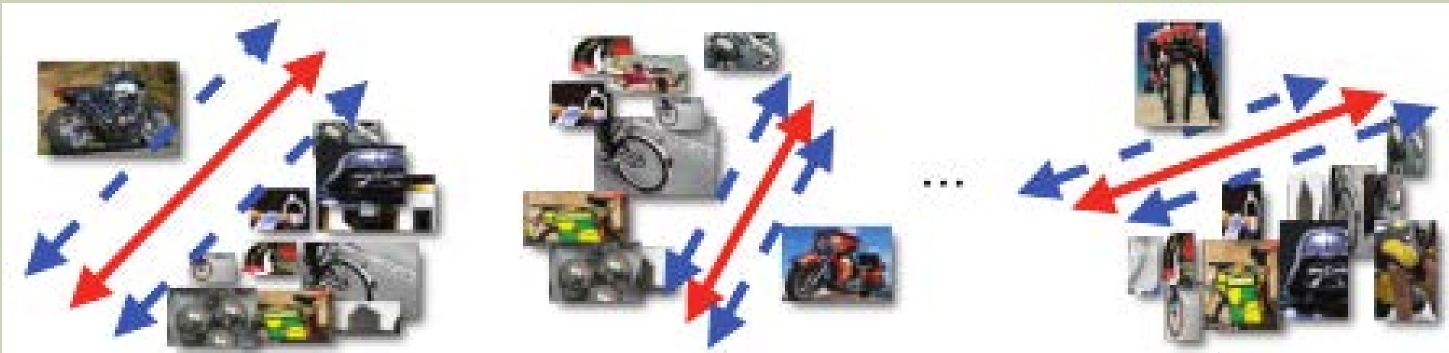
- Candidate Selection :



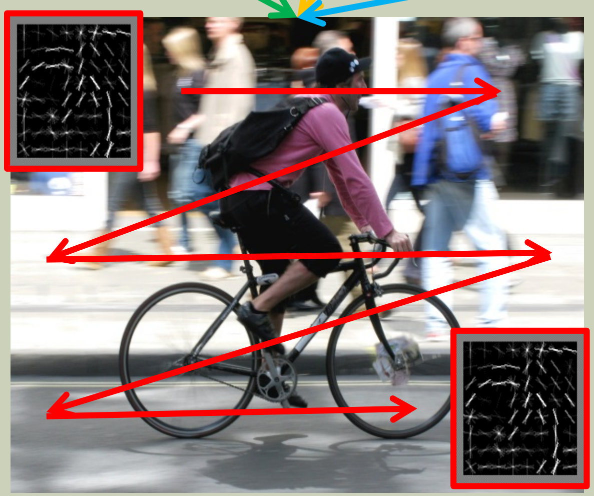


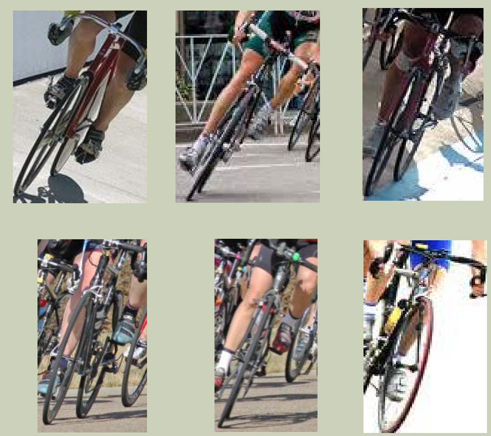
ESVM

SVM for each positive example



Test time

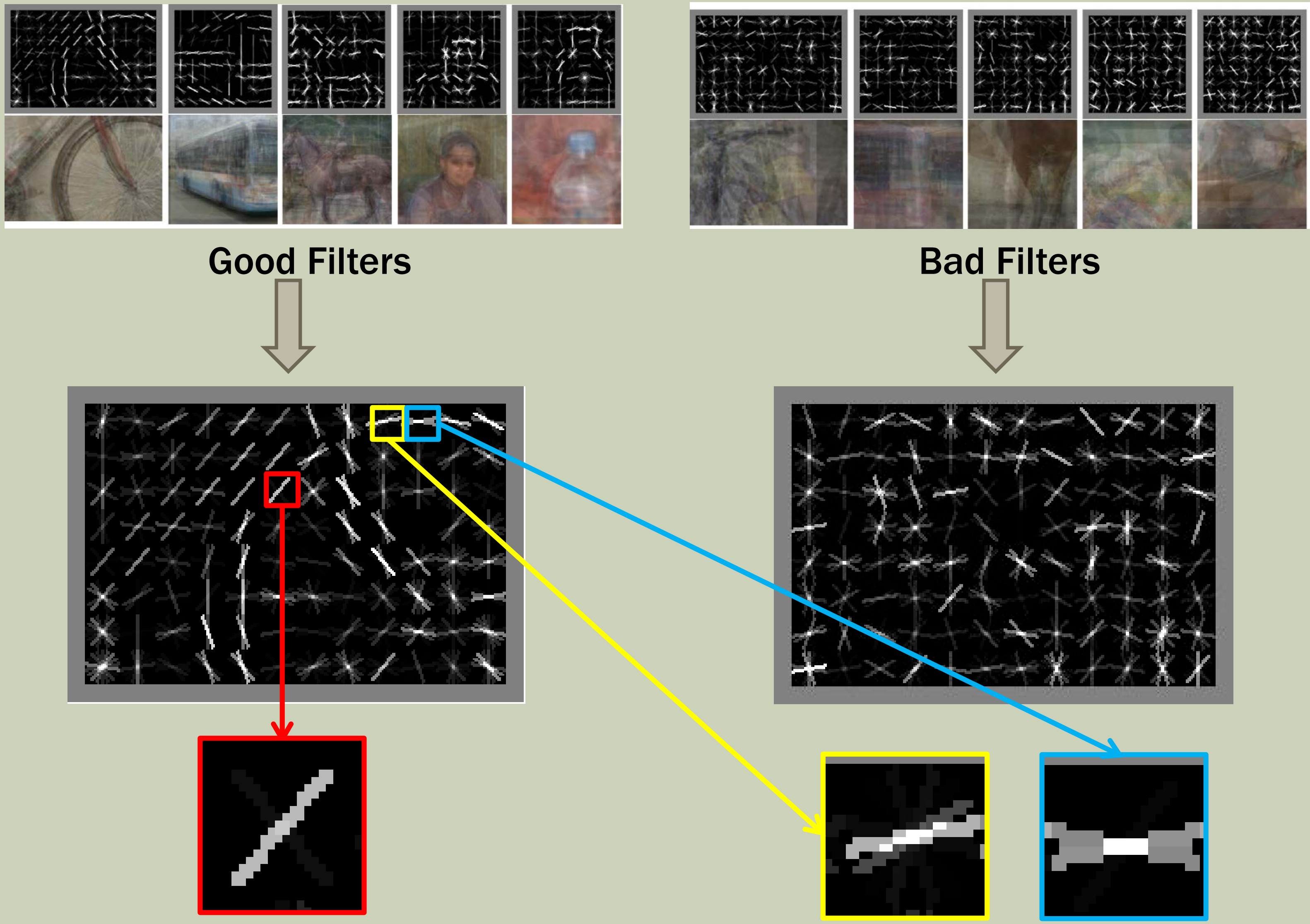




Redundant Exemplars

- Save significantly in training time if we can quickly select small set of relevant exemplars.

Good / Bad Filters

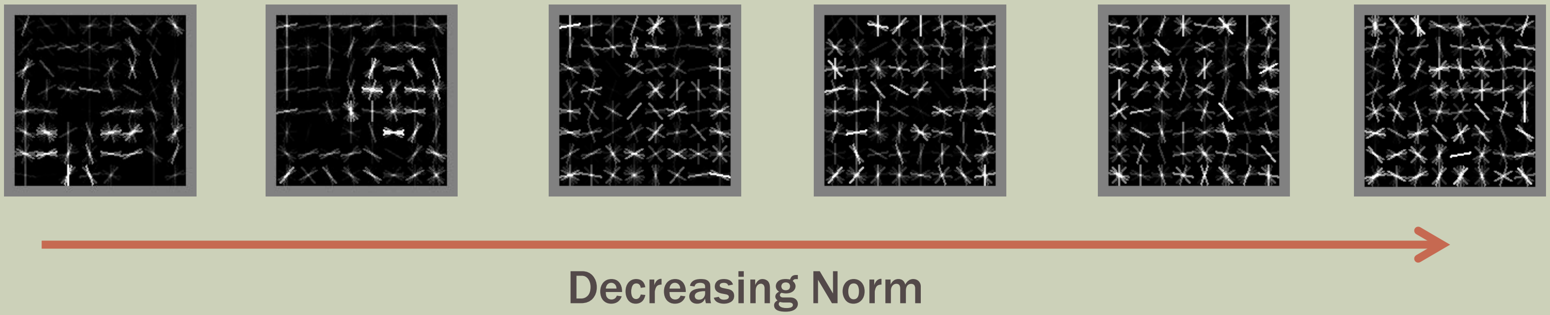


Gradient orientation within a cell
(active simultaneously)

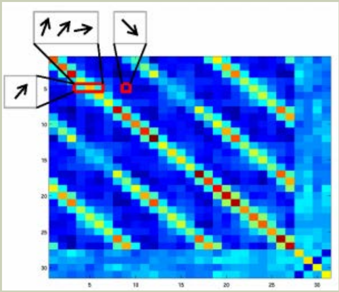
Gradient orientation of neighboring cells
(lines, curves)

Features for filter Ranking

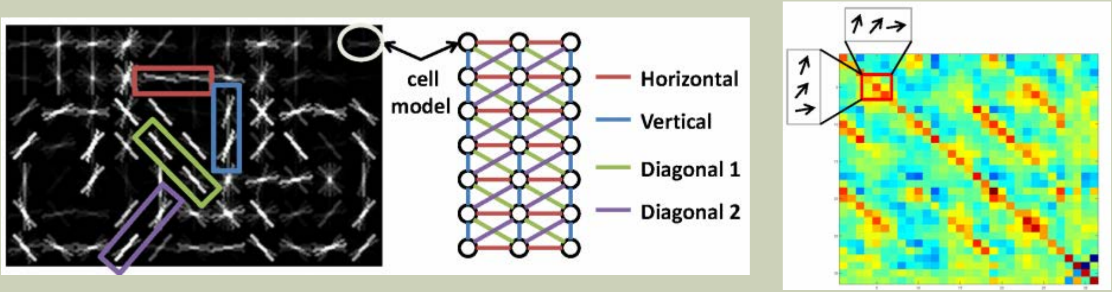
- **Norm:** consistent with high degree of alignment.
- **Normalized Norm:** Makes norm invariant to filter dimension.



- **Cell Covariance:** Different orientation bins within a cell are highly structured. Gao et al. ECCV 2012
- **Cell Cross Covariance:** Strong correlation between filter weights in nearby spatial locations.



Cell Covariance



Cell Cross Covariance

Learning to Rank Filters

- $\Phi(\boldsymbol{f})$ – representation of filter $\boldsymbol{f}$
- Goal : model ranking score of $\boldsymbol{f}$ by a linear function $< \mathrm{w}, \Phi(\boldsymbol{f}) >$
- Training data : $\{\boldsymbol{f}_{\mathrm{g,i}}\}$, $\mathrm{y}_{\mathrm{g},i}$
  - $\mathrm{g} = 1, \dots, G$ where G is number of training categories.
  - $i = 1, \dots, N$ where N is number of filters per category.
  - $\mathrm{y}_{\mathrm{g},i}$ is estimated quality, obtained by expensive method.
- $\boldsymbol{f}_{\mathrm{g},i}$ is ordered in descending value of $\mathrm{y}_{\mathrm{g},i}$
- $\Delta_{\mathrm{g,i,j}} = \mathrm{y}_{\mathrm{g},i} - \mathrm{y}_{\mathrm{g,j}}$, for $i > j$ measures how much better $\boldsymbol{f}_{\mathrm{g},i}$ is from $\boldsymbol{f}_{\mathrm{g,j}}$
- $\delta\Phi_{\mathrm{g,i,j}} = \Phi(\boldsymbol{f}_{\mathrm{g,i}}) - \Phi(\boldsymbol{f}_{\mathrm{g},j})$

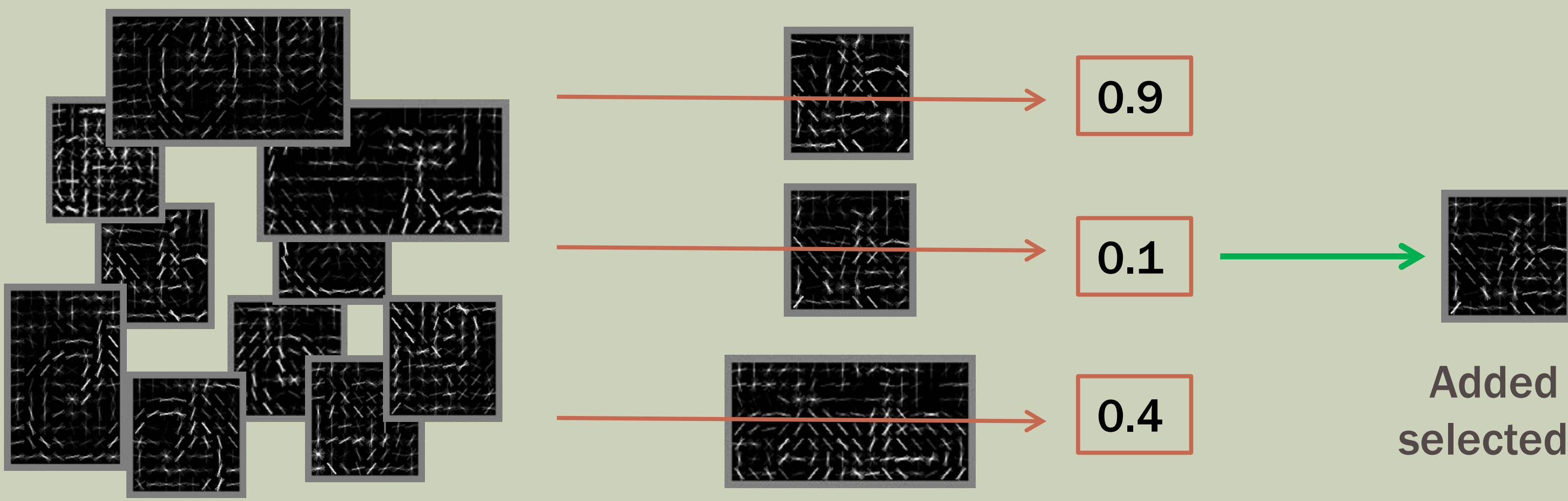
$$\min_{\mathbf{w}} \frac{1}{2}\|\mathbf{w}\|^2 + C \sum_{g=1}^{G} \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \left[1 - \langle \mathbf{w}, \delta\phi_{g,i,j} \rangle\right]_+ \Delta_{g,i,j}$$

- Slack rescaled hinge loss

Greedy approximation for Diversity

- Selected parts should be individually good and complimentary.
- First filter - $argmax_i \, \hat{y}_i$
- t filters selected so far
- Select next filter using following

$$\underset{i}{\operatorname{argmax}} \left\{ \widehat{y_i} - \lambda \max_{j=1,\ldots,t} A_{i,j} \right\}$$

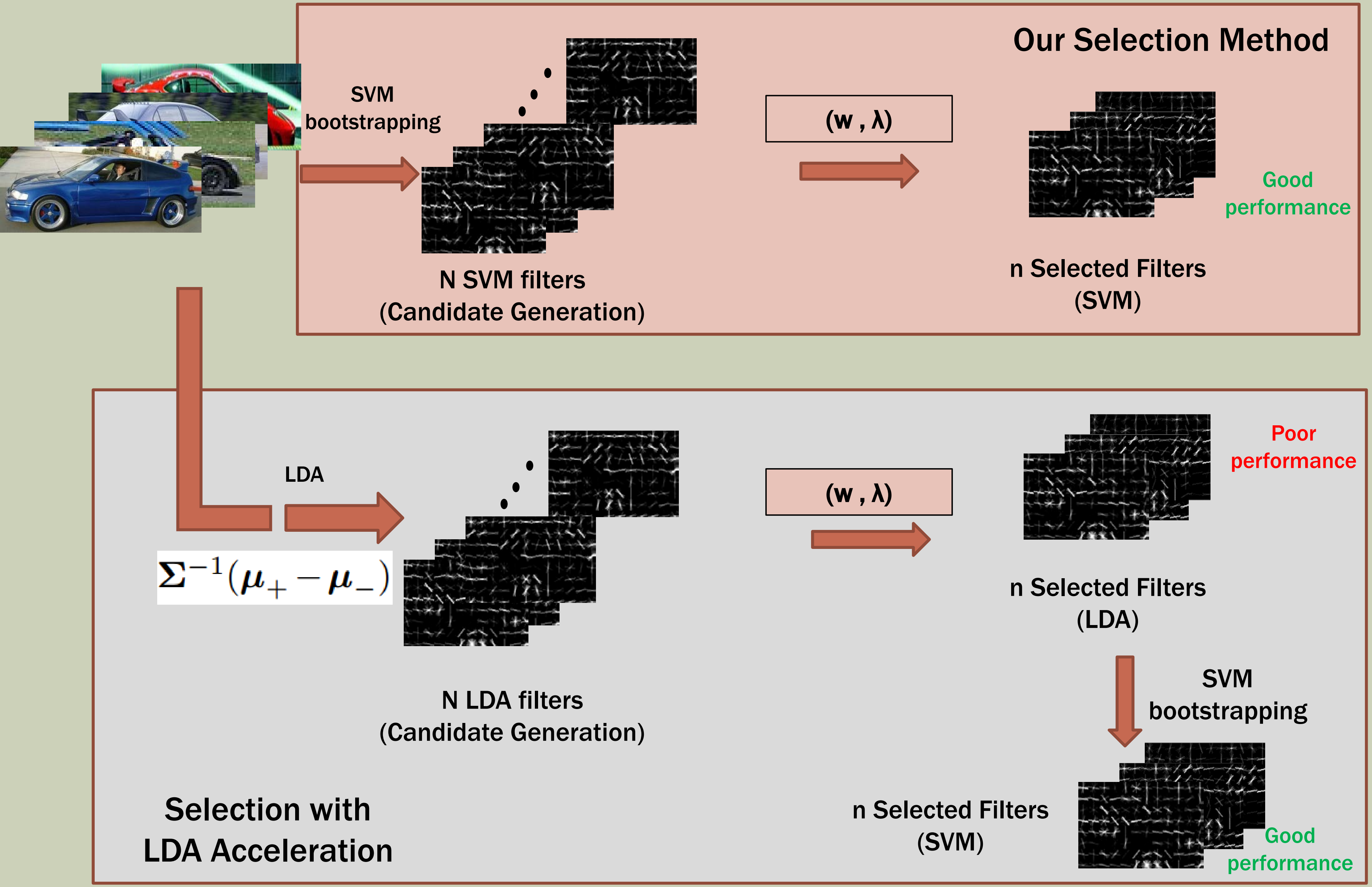


Selected Filters

Not yet Selected

Added to selected set

LDA Acceleration



Experiments with Poselets

Test category

Filters used for training from remaining categories

- 800 poselet filters for each category
- Goal : given a category select 100 out of 800 filters
- Ranking task
- Detection task

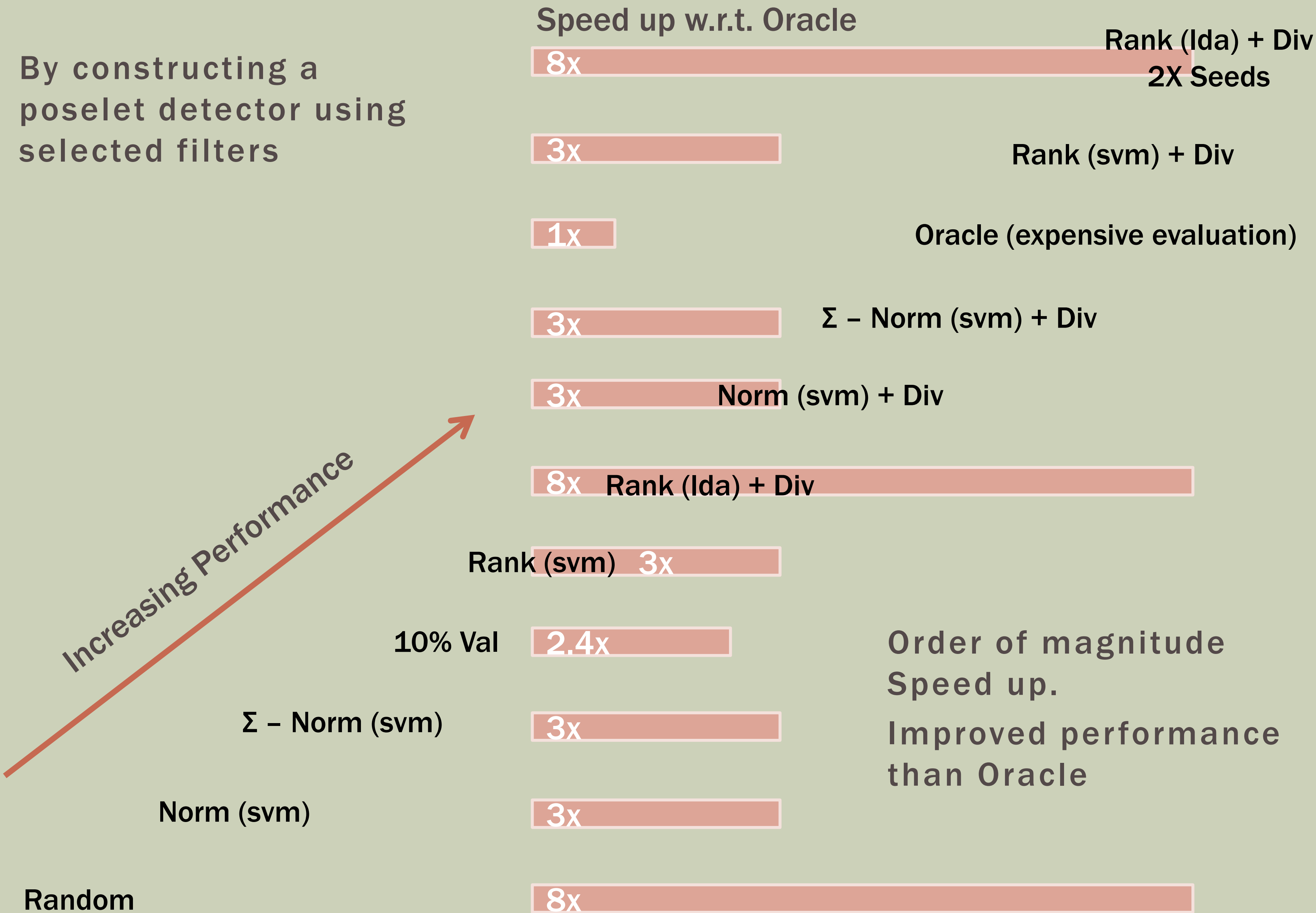
Performance of Ranker

- Predicted ranking vs true ranking as per AP scores.

Norm (svm)	<	Σ – Norm (svm)	<	Rank (lda)	<	Rank (svm)
		Gao et al. ECCV 2012				

Detection Results

By constructing a poselet detector using selected filters



Order of magnitude Speed up.

Improved performance than Oracle

Experiments with exemplar SVMs

- Each category has 630 exemplars on average.
- Goal select 100 exemplars such that they reproduce result for optimal set of 100 exemplars.
- Optimal set – weights of each exemplar in the final scoring model. (Oracle)
- Frequency of exemplars

Frequent Exemplar

Rare Exemplar

- We have presented an automatic mechanism for selecting diverse set of discriminative filters.
- Order of magnitude improvement in training time.
- Our approach is applicable to any discriminative architecture that uses a collection of filters.
- Insight into what makes a good filter for object detection.
- Can be used as an attention mechanism during test time
  - Reduce number of convolutions / hashing lookups.

Bottom line: One can tell whether a filter is useful for a category without knowing what that category is, just by “looking” at the filter.