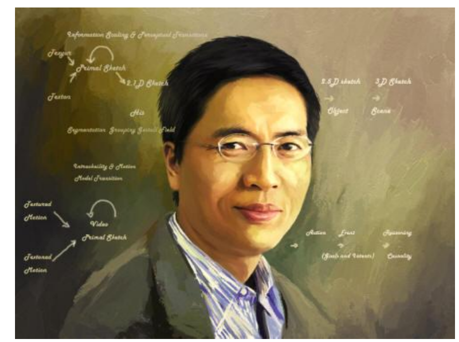
Integrating Context and Occlusion for Car Detection by Hierarchical And-Or Model



Bo Li[1,2] Tianfu Wu[2] Song-Chun Zhu[2]

[1]Beijing Institute of Technology

[2]University of California, Los Angeles (UCLA)

Car Detection

E.g., in PASCAL VOC *-- not much occlusion (D. Hoiem et al., ECCV 2012)*

Occlusion and Car-to-Car Context in Car Detection

KITTI

(A. Geiger et al., CVPR2012)

Street Parking

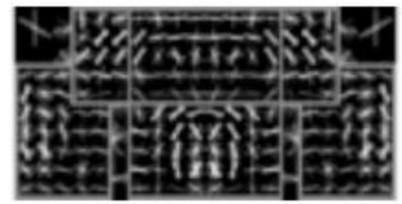
(B. Li et al., ICCV 2013)

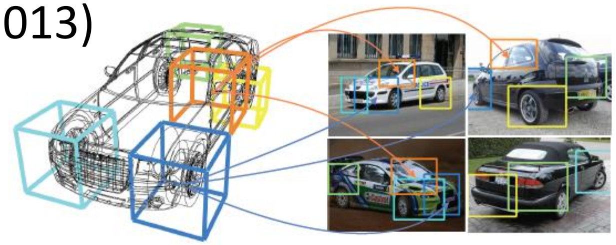
Parking Lot

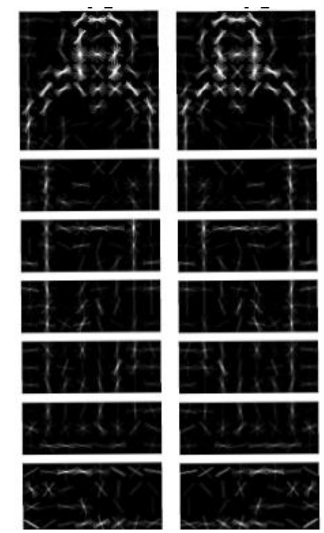
(self-collected in this paper)

Literature Review

- Single Object Model & Occlusion Modeling
  - Felzenszwalb et al., TPAMI (2010)
  - Girshick et al., In: NIPS (2011)
  - Zhu et al., In: CVPR (2010)
  - Pepik et al., In: CVPR (2012) & CVPR (2013)
  - Li et al, In: ICCV (2013)
  - Song et al., In: CVPR (2013)

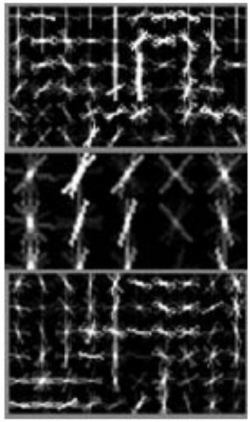


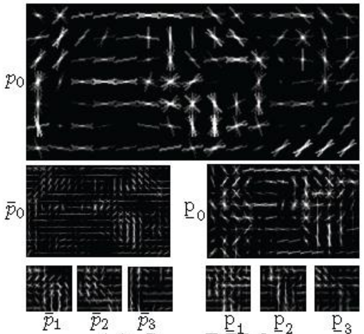




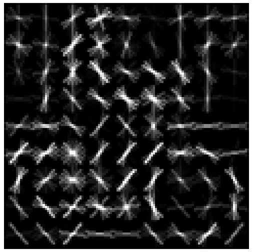
Literature Review

- Single Object Model & Occlusion Modelling
  - Felzenszwalb et al., TPAMI (2010)
  - Girshick et al., In: NIPS (2011)
  - Zhu et al., In: CVPR (2010)
  - Pepik et al., In: CVPR (2012) & CVPR (2013)
  - Li et al, In: ICCV (2013)
  - Song et al., In: CVPR (2013)
- Object-Pair & Visual Phrase Models
  - Li et al., PR (2014)
  - Tang et al., In: BMVC (2012)
  - Ouyang, W. & Wang, X., In: CVPR (2013)
  - Pepik et al., In: CVPR (2013)
  - Sadeghi, M. & Farhadi, A., In: CVPR (2011)

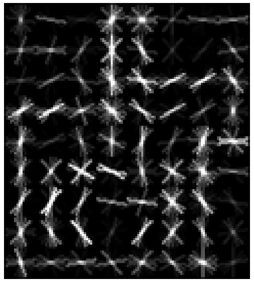




(e) Sym-DPM



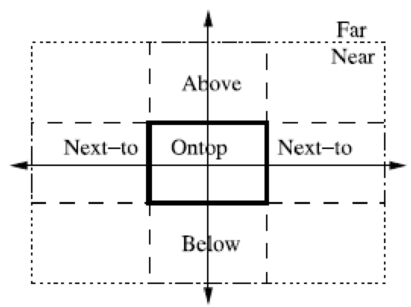
person_riding_bicycle

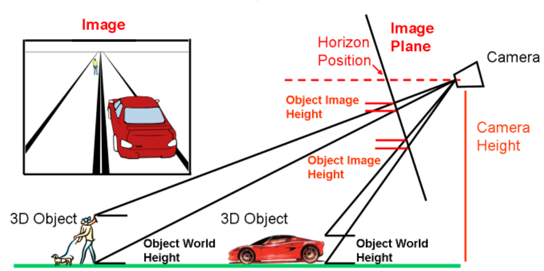


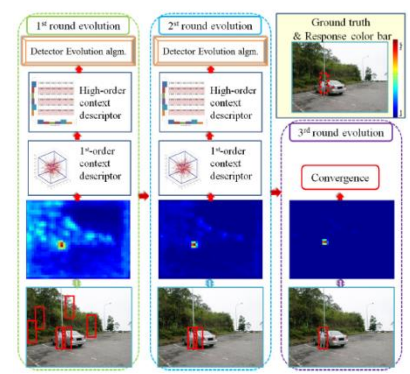
person_riding_horse

Literature Review

- Single Object Model & Occlusion Modelling
  - Felzenszwalb et al., TPAMI (2010)
  - Girshick et al., In: NIPS (2011)
  - Zhu et al., In: CVPR (2010)
  - Pepik et al., In: CVPR (2012) & CVPR (2013)
  - Li et al, In: ICCV (2013)
  - Song et al., In: CVPR (2013)
- Object-Pair & Visual Phrase Models
  - Li et al., PR 47, 3254 - 3264 (2014)
  - Tang et al., In: BMVC (2012)
  - Ouyang, W. & Wang, X., In: CVPR (2013)
  - Pepik et al., In: CVPR (2013)
  - Sadeghi, M. & Farhadi, A., In: CVPR (
- Context Model
  - Hoiem et al., IJCV (2008)
  - Desai et al., IJCV (2011)
  - Tu, Z. & Bai, X., TPAMI (2010)
  - Chen et al., In: CVPR (2013)
  - Yang et al., In: CVPR (2012)

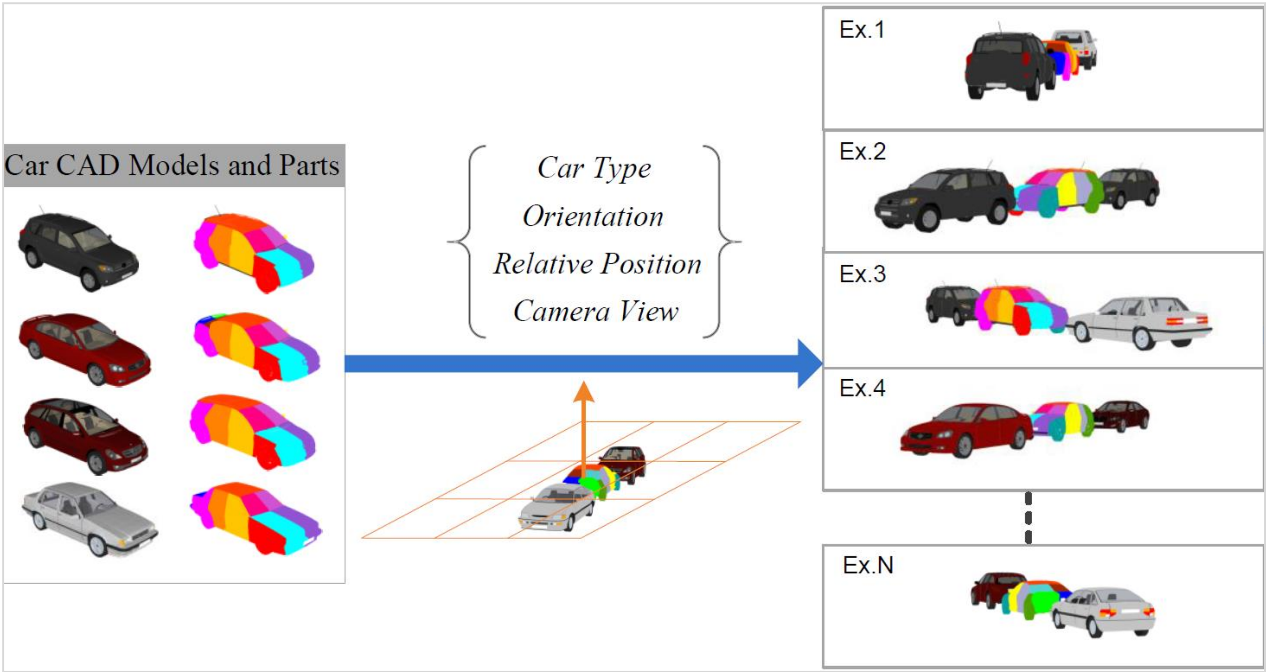


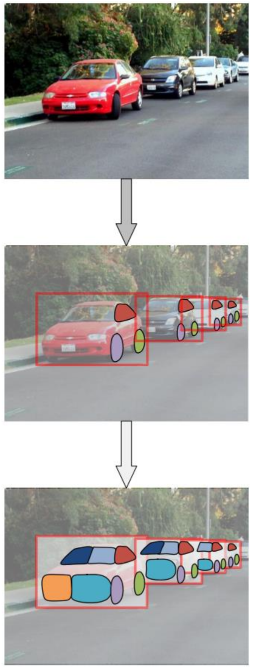




Regularity and Reconfigurability

from the object detection grammar perspective





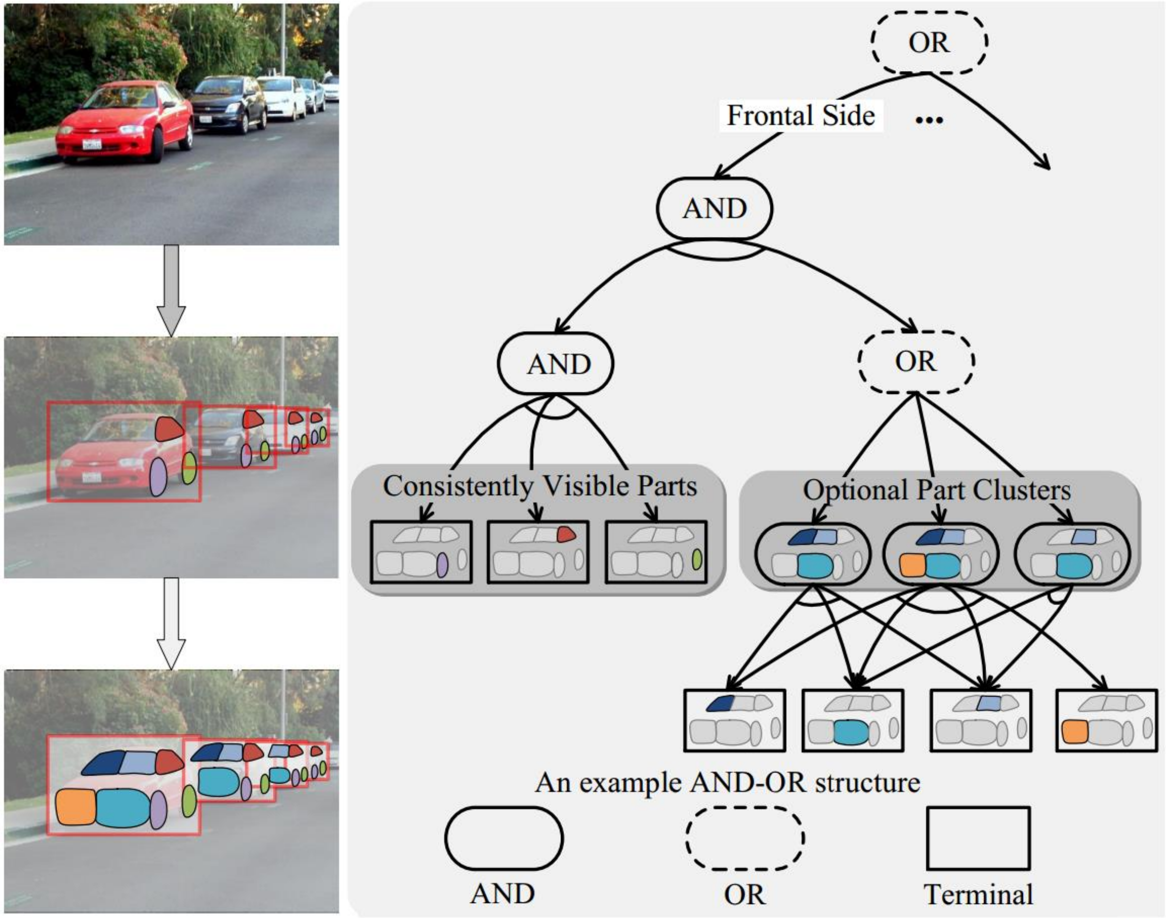
1, Occlusions have different regularities (e.g., consistently visible parts for a given viewpoint).

2, Car-to-car layouts (i.e., N-car together) with reconfigurable single cars and car parts can address the intra-class variations caused by occlusions.

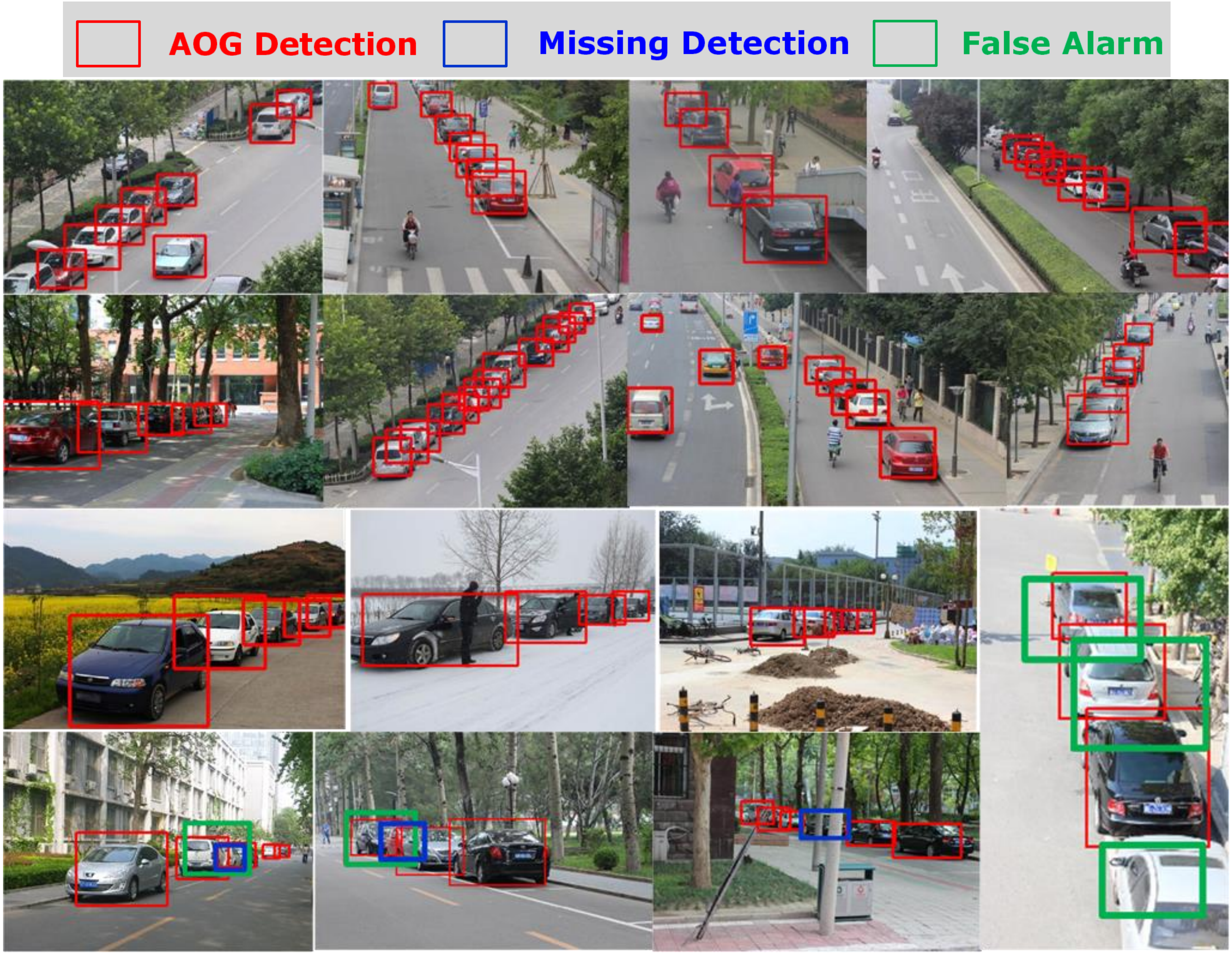
Modeling Occlusion by Discriminative And-Or Structures

B. Li et al., ICCV 2013

S.C. Zhu and D. Mumford, FTCGV 2006

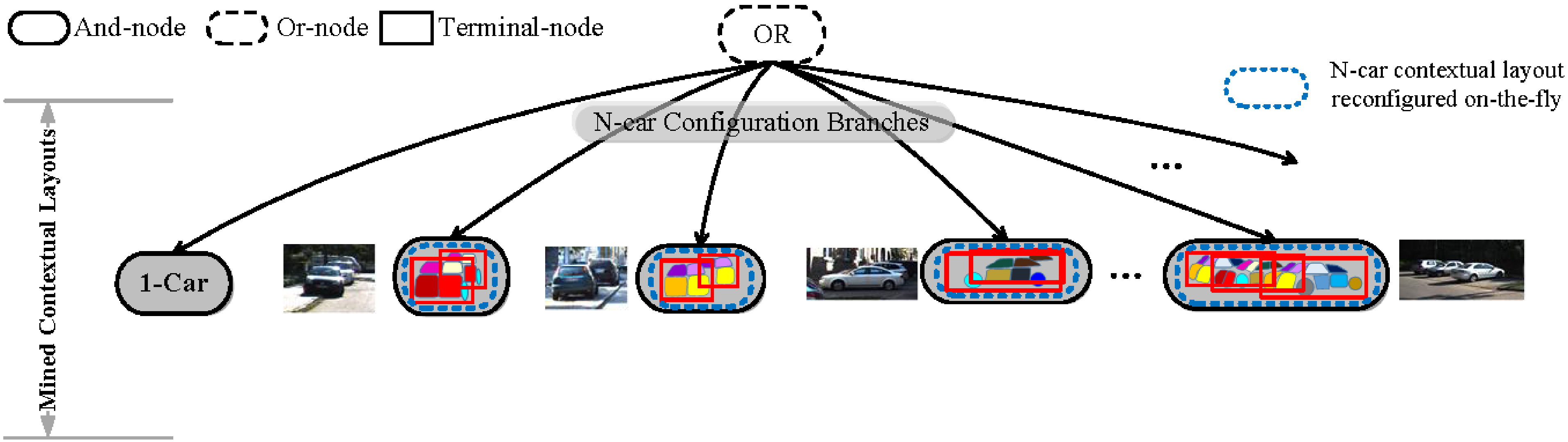


Results on Street-Parking-Car Dataset



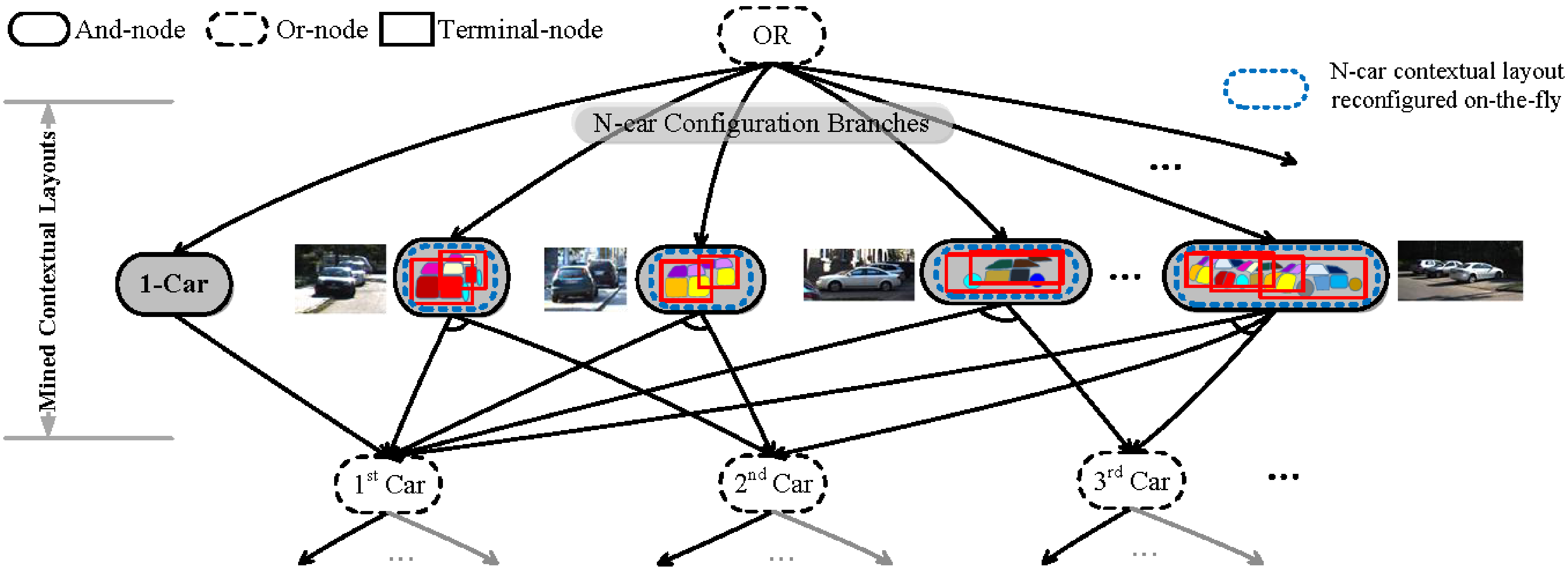
Integrating Occlusion and Context

Representation: Hierarchical and Reconfigurable And-Or Models

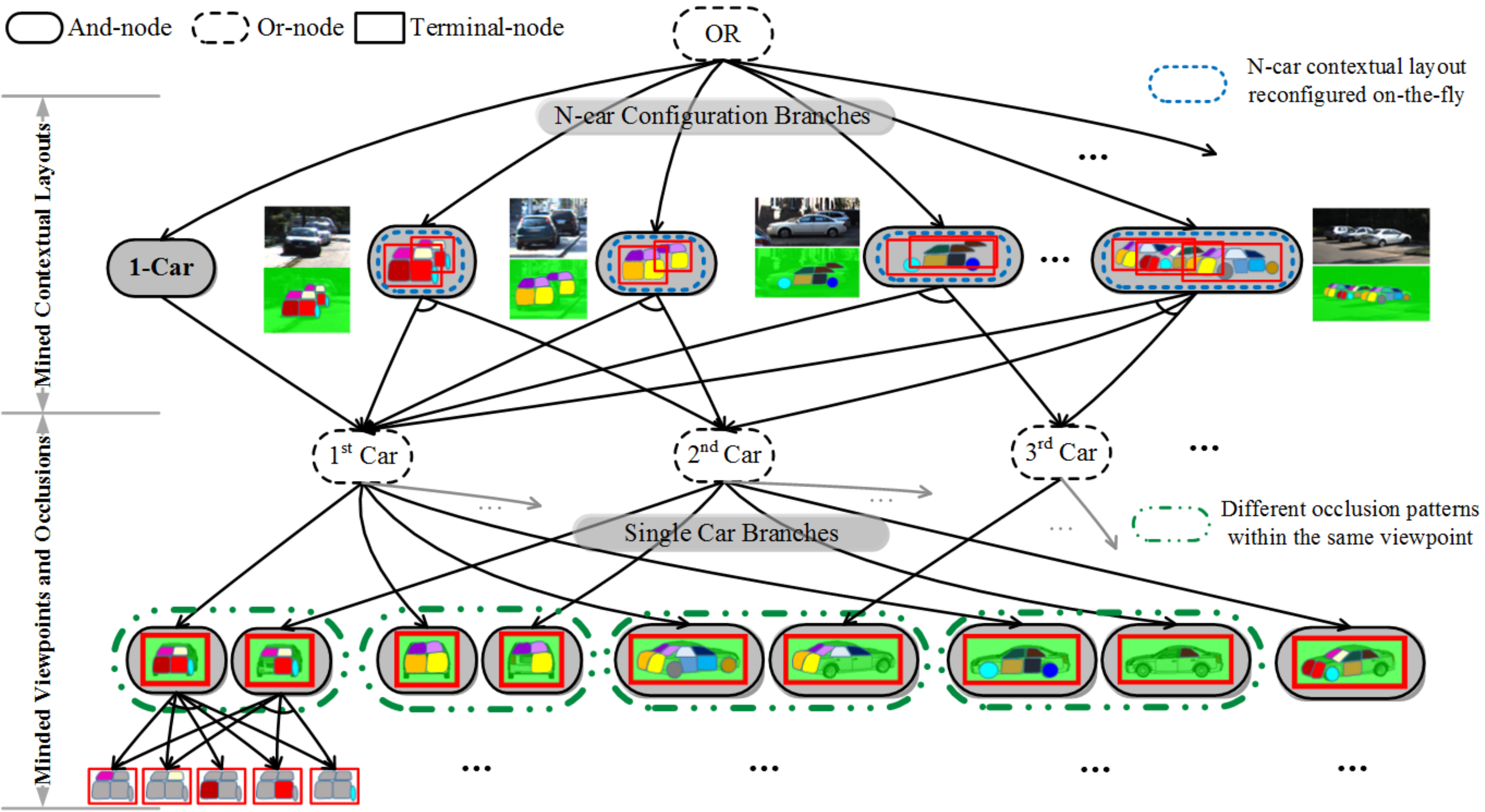


Car-to-Car Context:

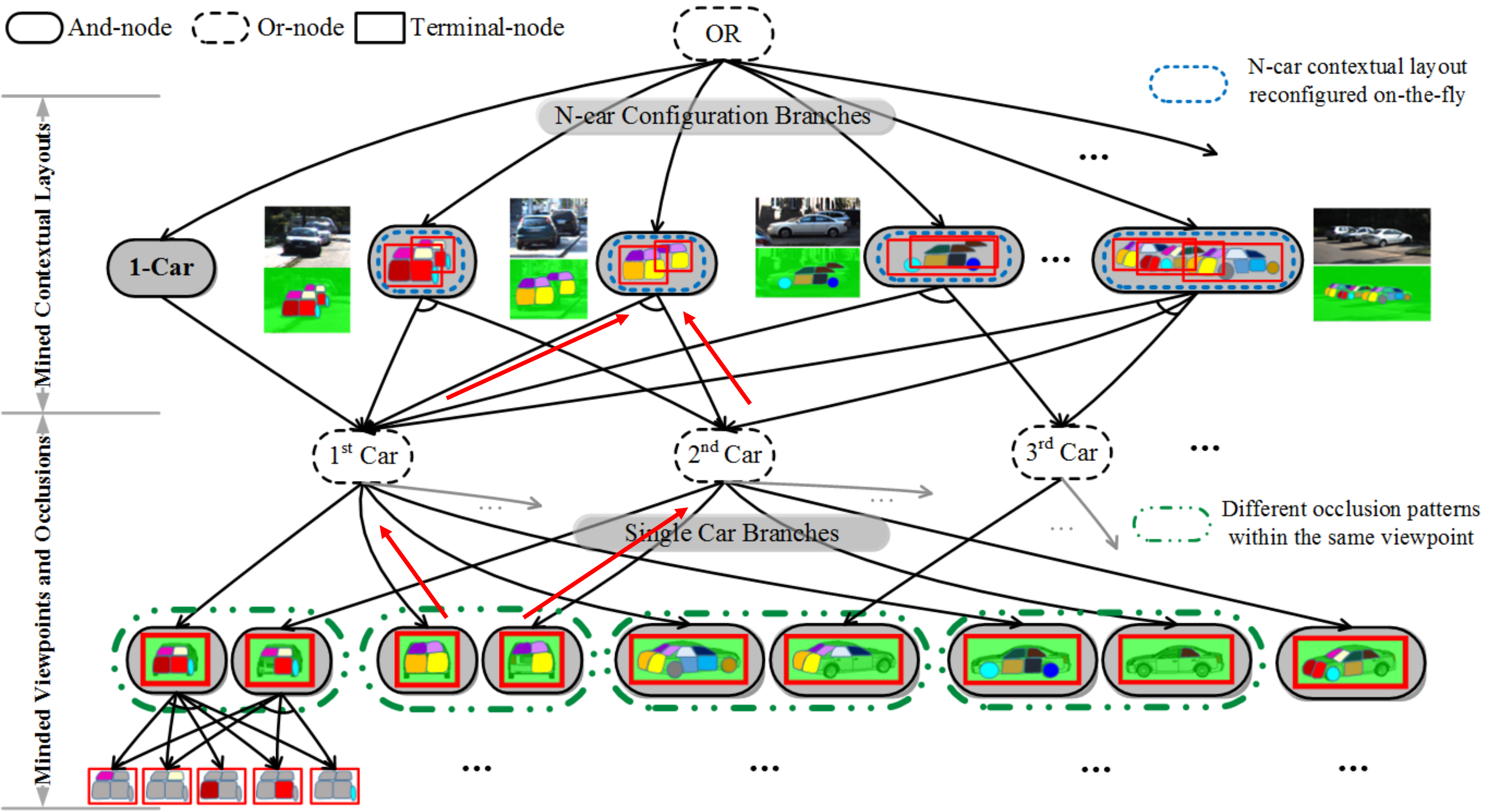
Representation: Hierarchical and Reconfigurable And-Or Models



Representation: Hierarchical and Reconfigurable And-Or Models

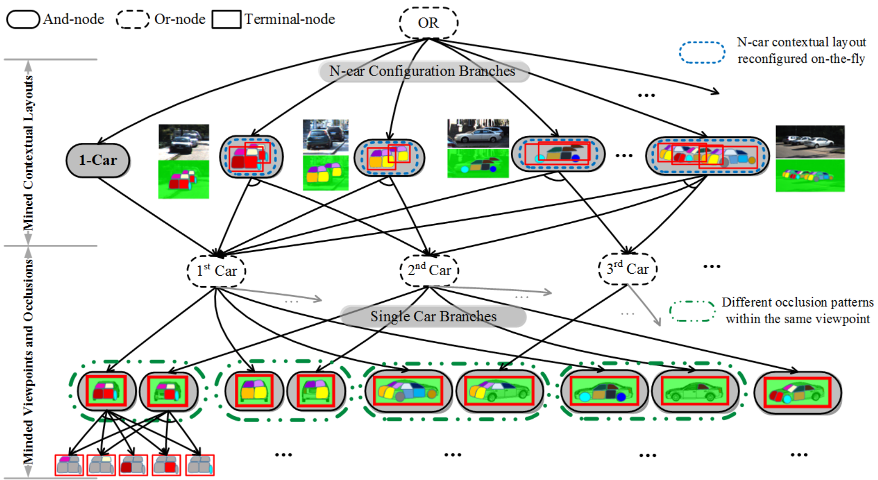


Representation: Hierarchical and Reconfigurable And-Or Models

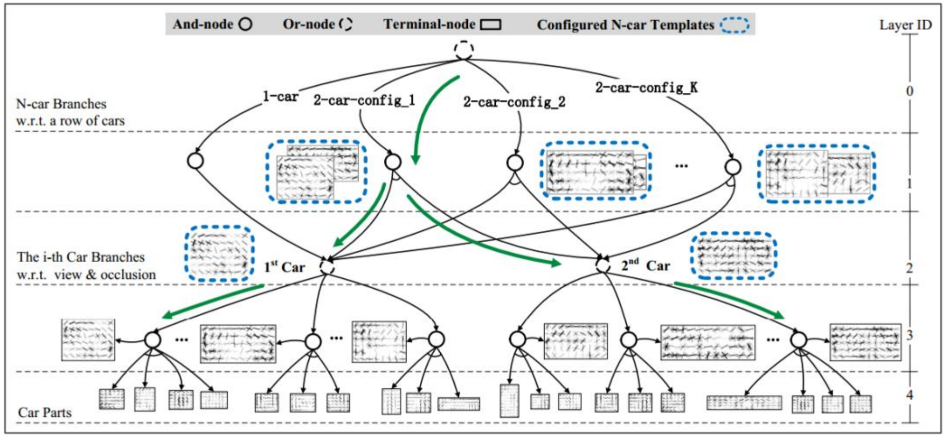


Integrating Occlusion and Context

(1) Learning the And-Or Structure by mining car-to-car context patterns and viewpoint-occlusion patterns



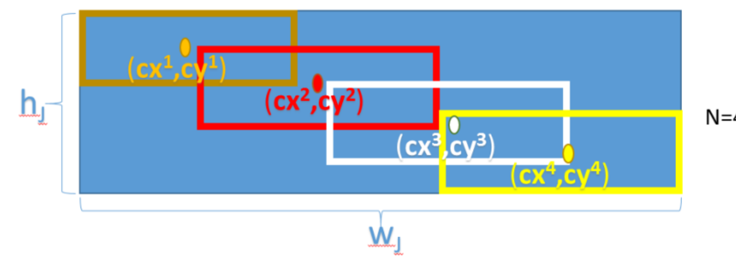
(2) Learning the parameters for appearance, deformation and bias



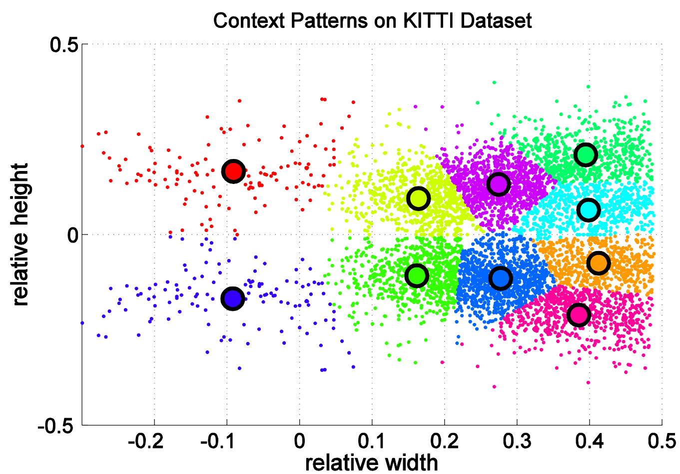
Learning : Mining Car-to-Car Context Patterns

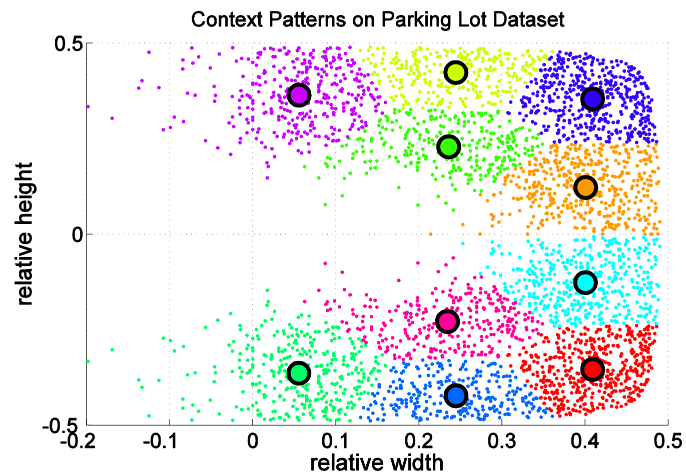
- Context Feature Vector:

$$\left[\frac{cx_i^2 - cx_i^1}{w_J}, \frac{cy_i^2 - cy_i^1}{h_J}, \cdots, \frac{cx_i^N - cx_i^1}{w_J}, \frac{cy_i^N - cy_i^1}{h_J}\right]$$

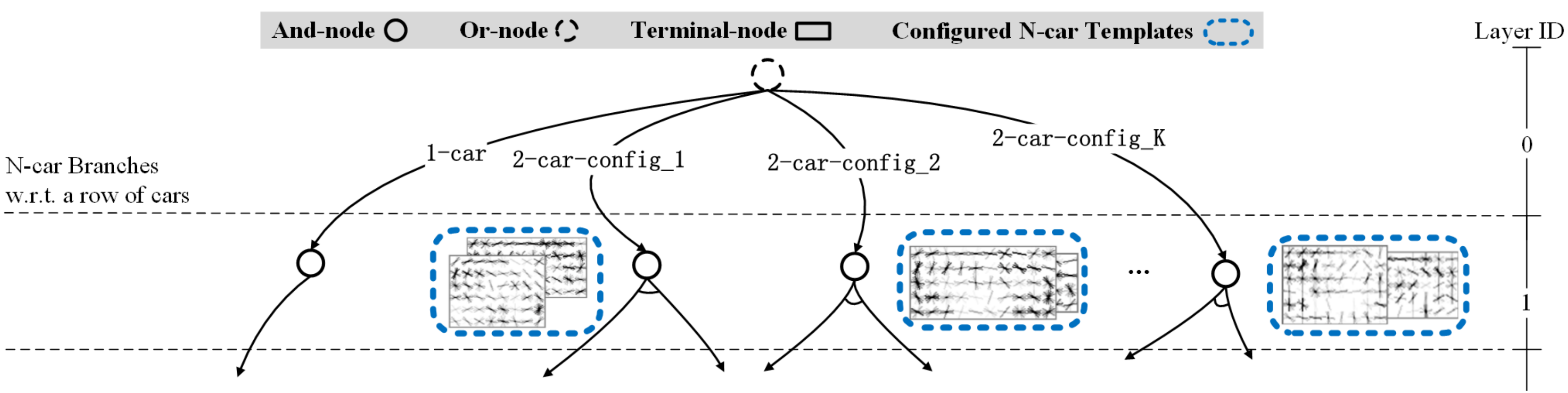


- Clustering



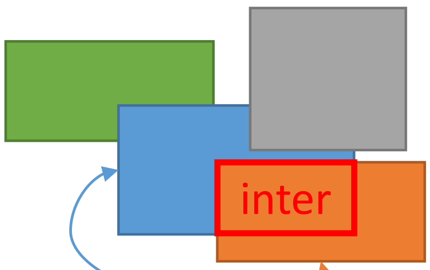


- Each cluster corresponds to an And-node at Layer 1

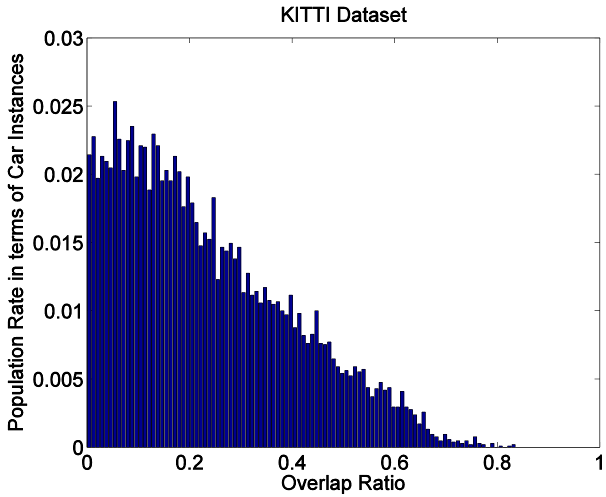


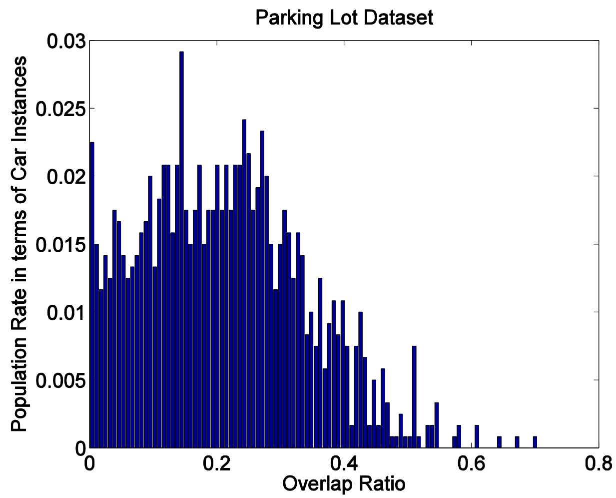
Learning : Mining viewpoint-occlusion Patterns

- For a single car example, we use the maximum overlap ratio between it and its neighbor cars as its occlusion ratio



$$ov = \max_{i \in N(ex)} \frac{A \cap A_i}{A \cup A_i}$$





- Two methods to model occlusion patterns
  - Greedy part selection (Felzsenszwalb, et al., PAMI 2010)
  - And-Or Structure Learning

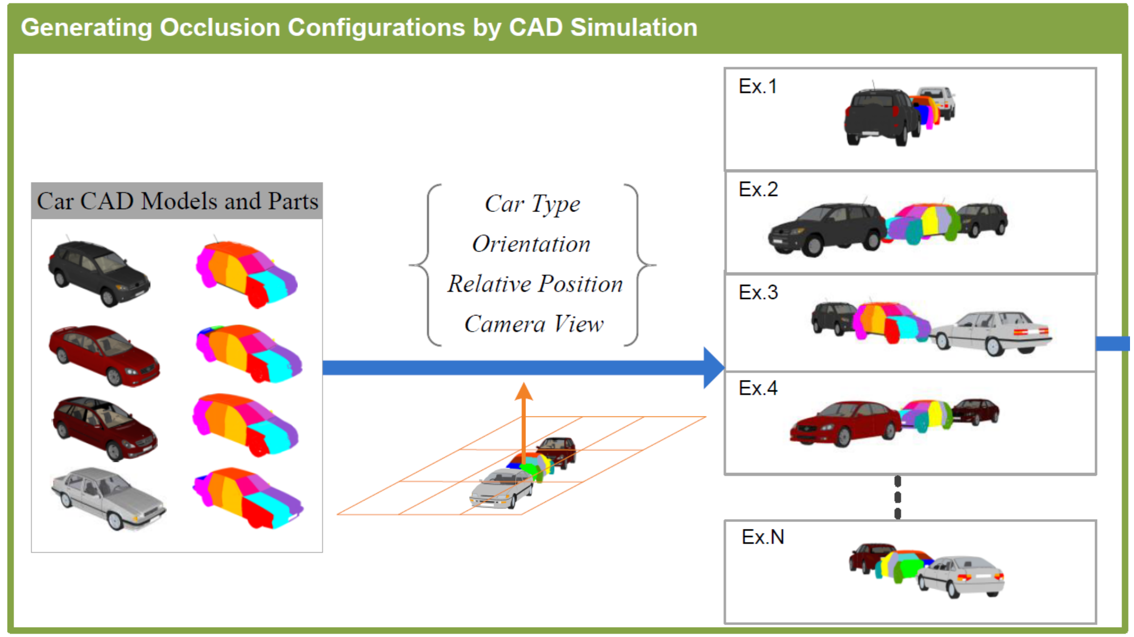
And-Or Structure Learning

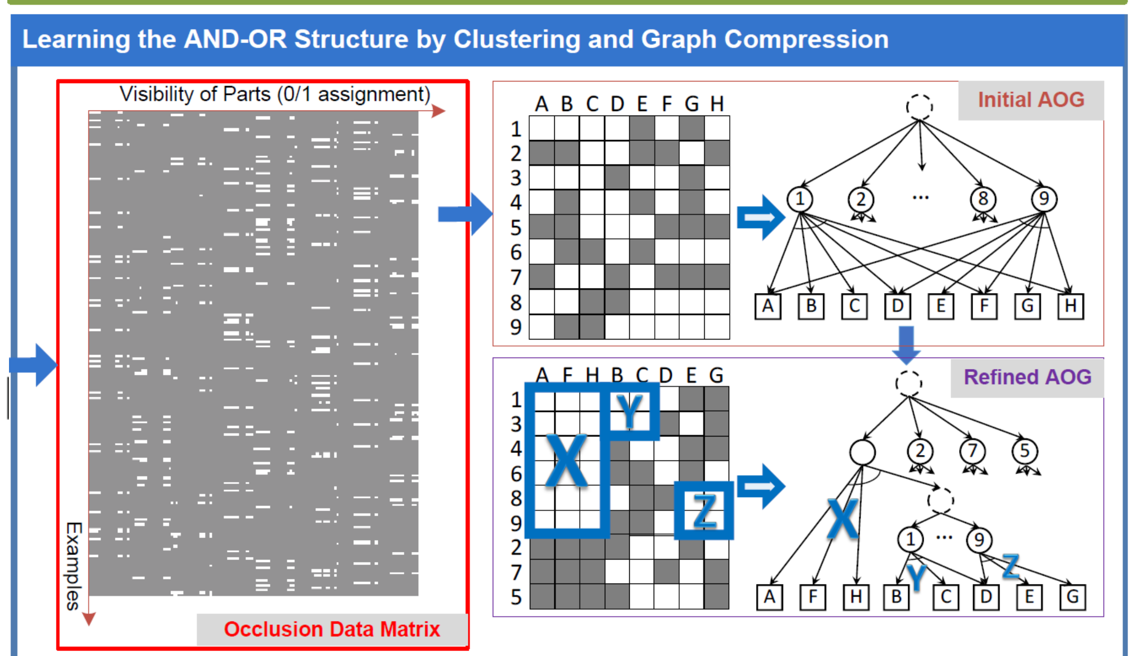
Generating CAD Examples
& Building Data Matrix

Information projection
& Graph Compression

$$\min \sum_{i}^{N} \mid v_i - v_i(AOT) \mid_2^2 + \lambda \mid AOT \mid$$

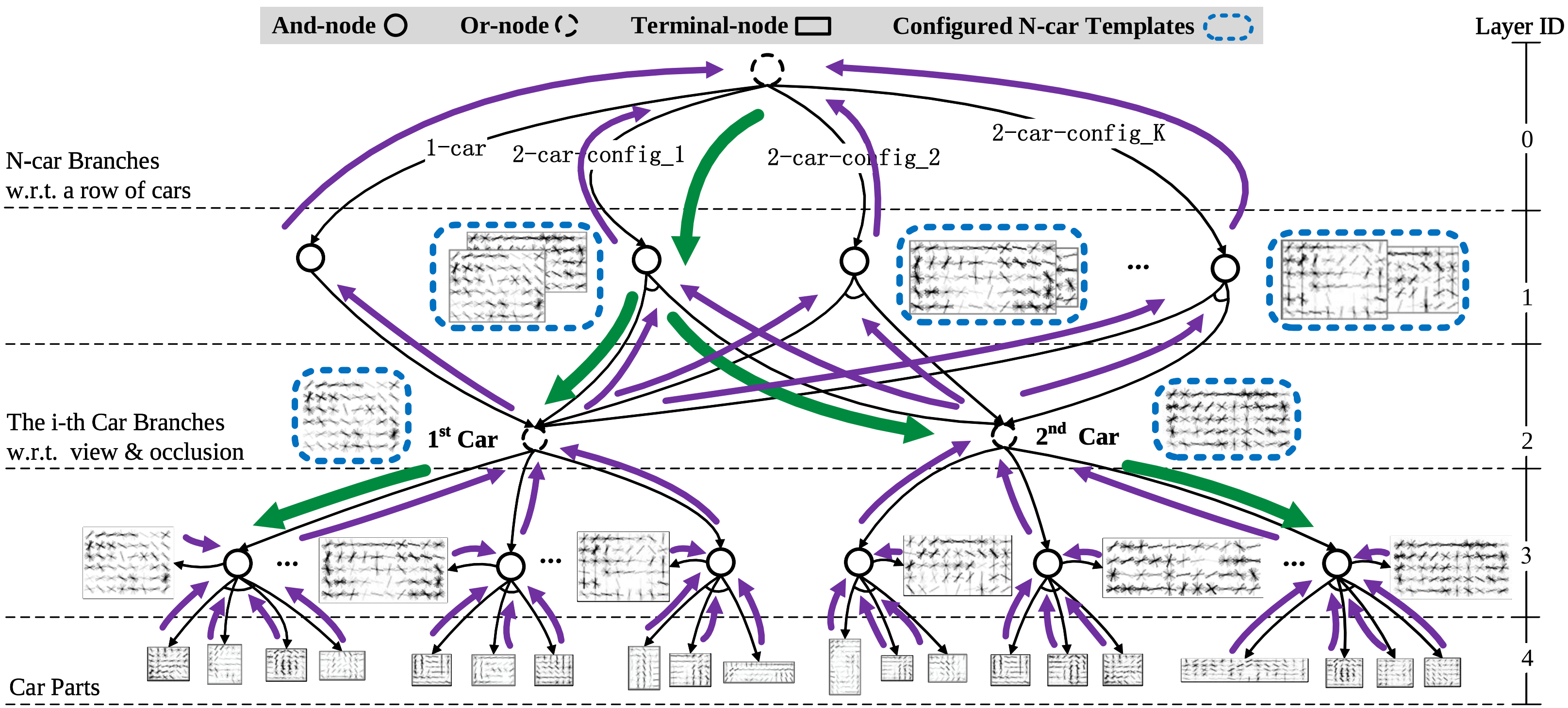
Z. Si & S.C. Zhu., TPAMI (2013)





Inference : the DP algorithm

- Bottom-up Pass – Depth-First-Search
  - Compute appearance & deformation score maps for And-Or nodes
- Top-down Pass – Breadth-First-Search
  - Retrieve parse trees with scores above the threshold



Parameters Learning

- Weak-Label Structural SVM (WLSSVM) Girshick et al., In: NIPS (2011)

$$\mathcal{E}(\Theta) = \frac{1}{2}\|\Theta\|^2 + C\sum_{i=1}^{M} L'(\Theta, x_i, y_i)$$

$$\begin{aligned} L'(\Theta, x, y) = & \max_{pt \in \Omega_{\mathcal{G}}} [score(x, pt; \Theta) + L_{margin}(y, box(pt))] - \\ & \max_{pt \in \Omega_{\mathcal{G}}} [score(x, pt; \Theta) - L_{output}(y, box(pt))] \end{aligned}$$

$$L_{\ell,\tau}(y, box(pt)) = \begin{cases} \ell & \text{if } y = \perp \text{ and pt } \neq \perp \\ 0 & \text{if } y = \perp \text{ and pt } = \perp \\ \ell & \text{if } y \neq \perp \text{ and } \exists\ B \in y \text{ with } ov(B, B') < \tau, \forall B' \in box(pt) \\ 0 & \text{if } y \neq \perp \text{ and } ov(B, B') \geq \tau,\ \forall\ B \in y \text{ and } \exists B' \in box(pt) \end{cases},$$

1) During learning, we run the DP inference to find the optimal parse tree

2) For a N-car example, the loss function is enforced to consider N car bounding boxes together

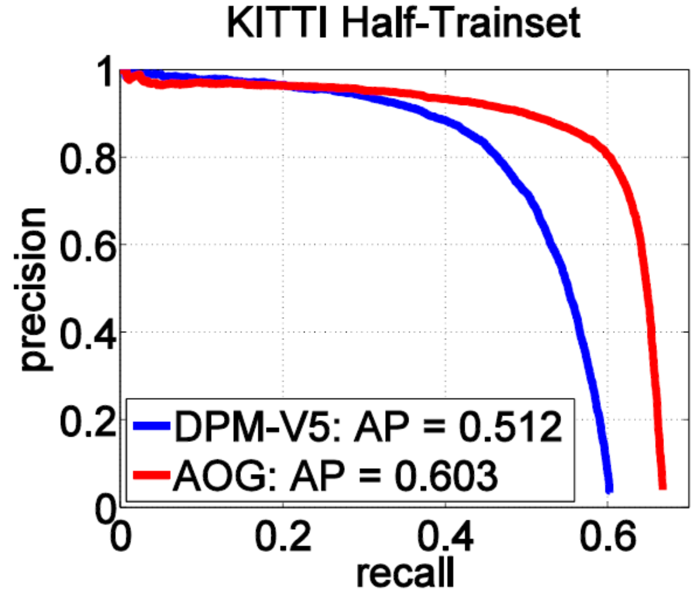
Experimental Results

- Public Datasets
  - KITTI, Street Parking, Pascal VOC 2007
- Self-collected Parking Lot Dataset
  - More Cars per Image
  - More Car-to-Car Context
  - More Occlusion
  - More Diverse Viewpoints

Results: KITTI & Pascal VOC 2007

- Training and Testing by Splitting the Trainset

- Training on Half-Trainset, and Testing on the benchmark



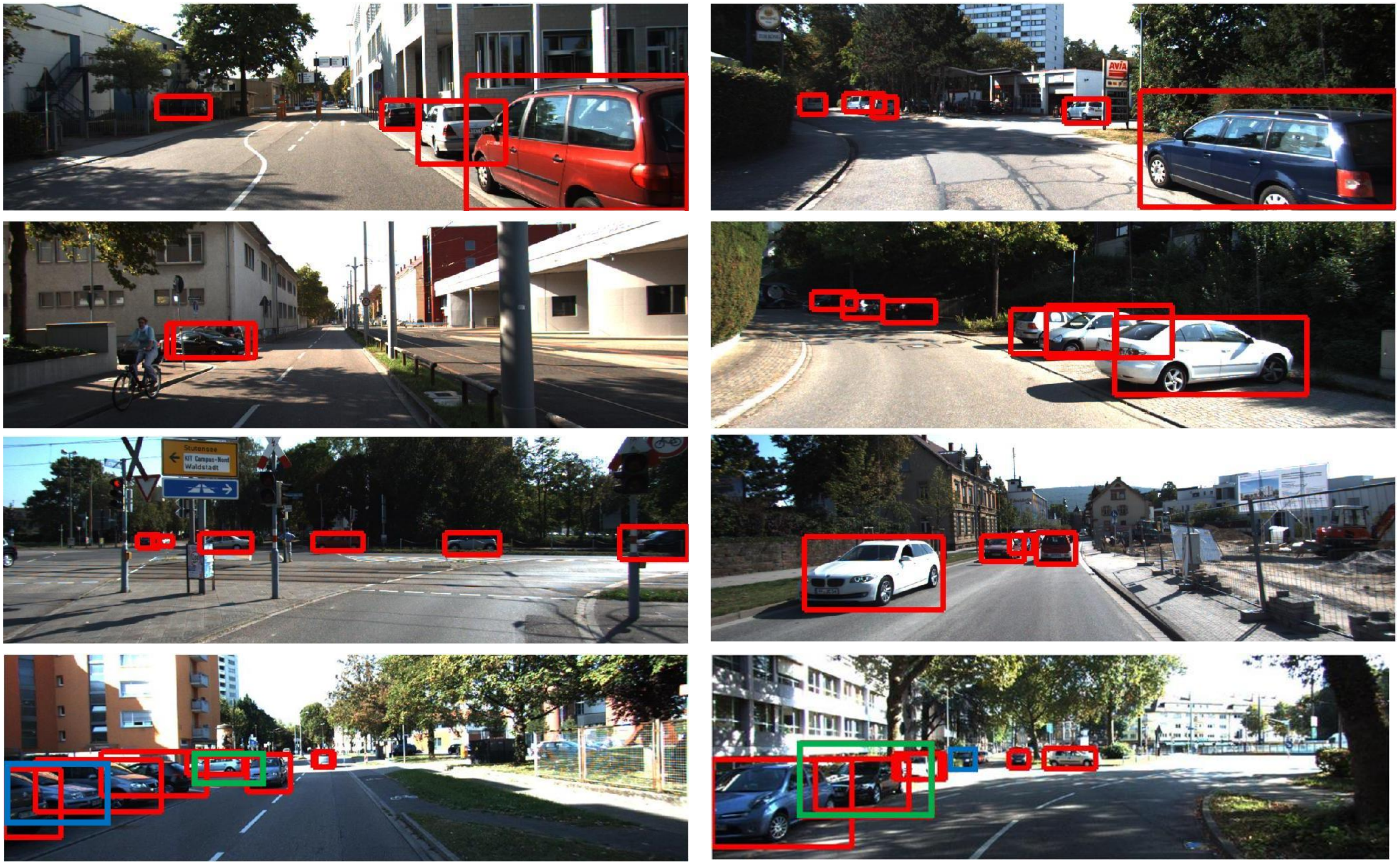
Methods	Easy	Moderate	Hard
mBoW	36.02%	23.76%	18.44%
LSVM-MDPM-us	66.53%	55.42%	41.04%
LSVM-MDPM-sv	68.02%	56.48%	44.18%
MDPM-un-BB	71.19%	62.16%	48.43%
OC-DPM	74.94%	65.95%	53.86%
DPM (trained by us)	77.24%	56.02%	43.14%
AOG	**80.26%**	**67.03%**	**55.60%**

Pascal VOC 2007

	DPM [14]	AOG
AP	58.2%	60.6%

Results: KITTI

Missing Detection　　True Positive　　False Alarm



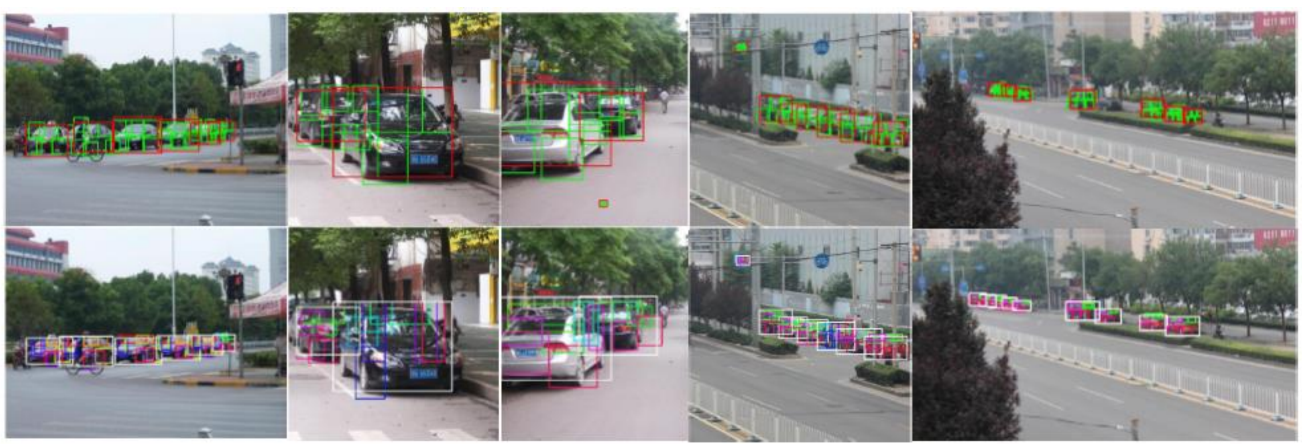
Results: Street Parking

- Quantitative Results

	DPM-V5	And-Or Structure	Our AOG[1]	Our AOG[2]
AP	52.0%	57.8%	62.1%	65.3%

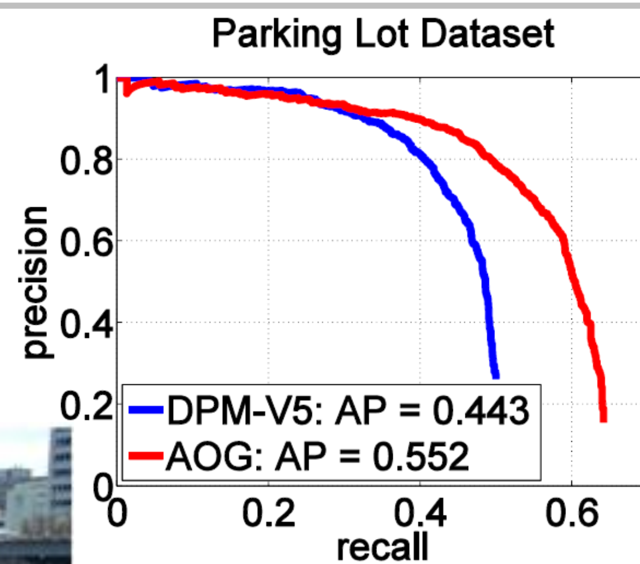
- Qualitative Results

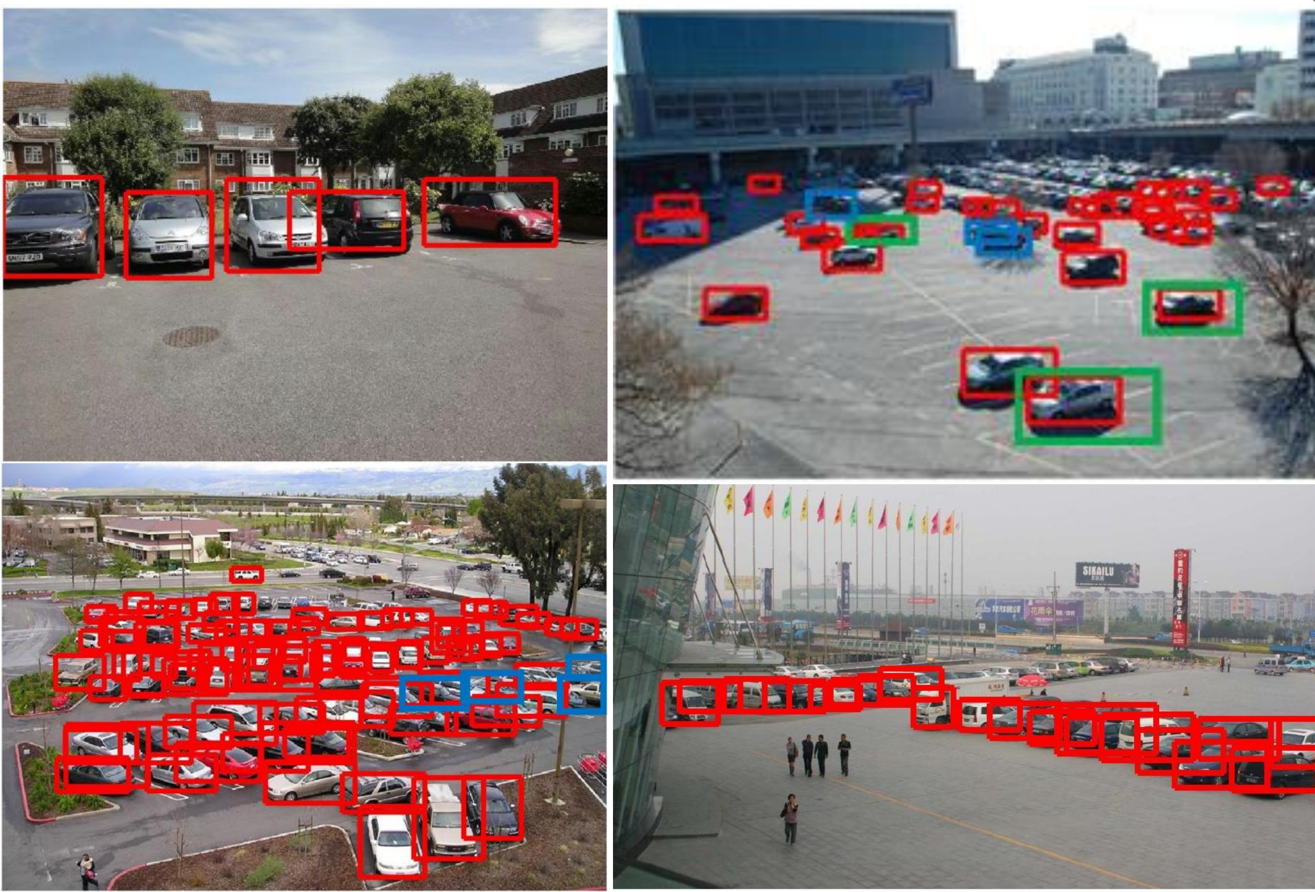
- Part Layouts

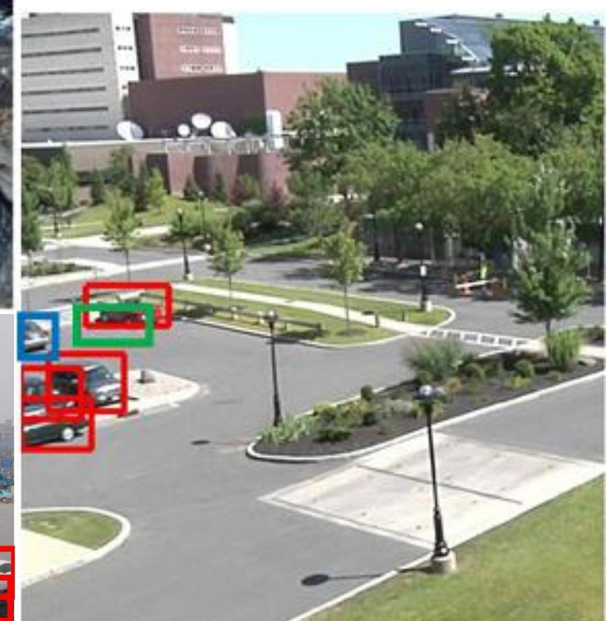


Results: Parking Lot

- Evaluation Settings
  - Overlap threshold: 0.60
  - Minimum area: 1000
  - minimum height: 24
- Qualitative Results







Conclusion

- A flexible & reconfigurable And-Or model
- Mine context and occlusion patterns
- A new parking lot dataset

- Future Work
    - Apply on other categories
    - New ways to mine context & occlusion pattern
    - Integrating other context cues

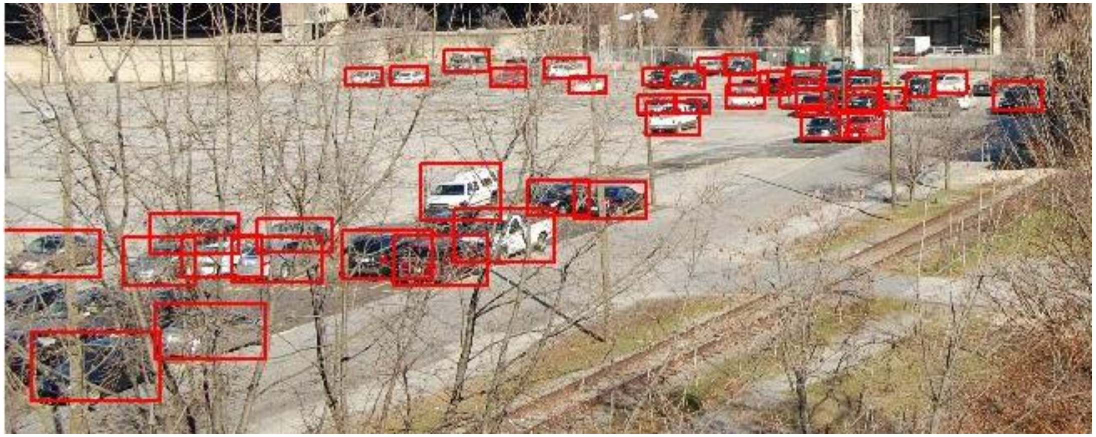
Thank you!

And Questions?

Poster: P4B-46

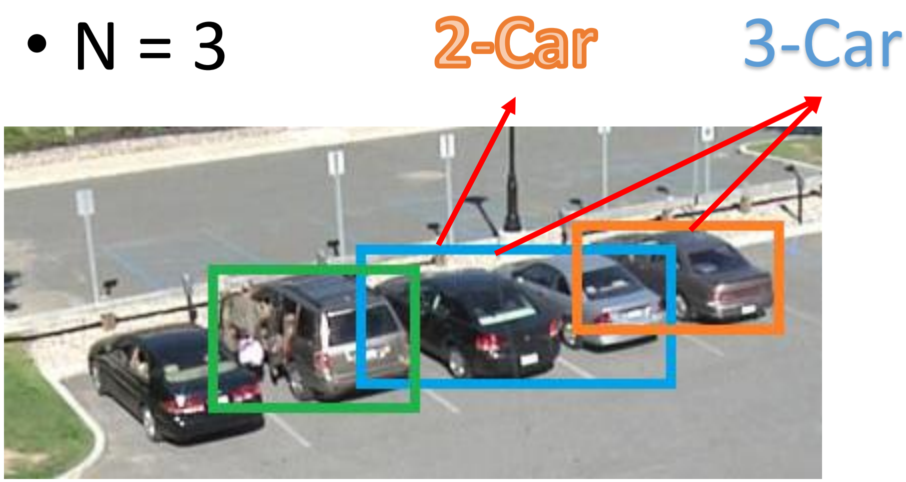
The code and data will be available on the following project page:
http://www.stat.ucla.edu/~tfwu/project/OcclusionModeling.htm

Learning : Generating *N*-car Positive Samples



- N = 2

- N = 3



- N = 4, …

ooo

Generating N-car Positive Sample

- For N > 2, the algorithm is similar

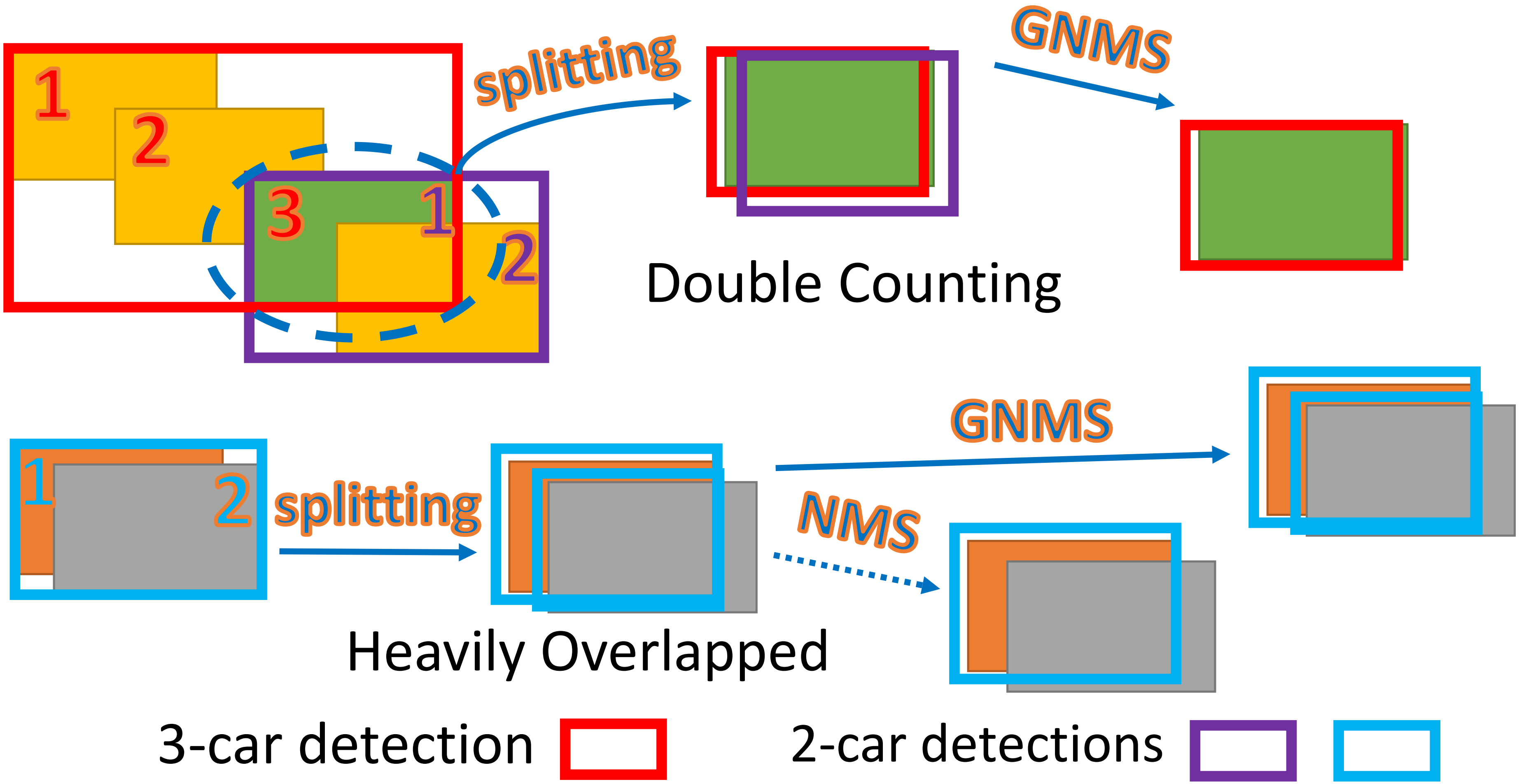
input : $(N-1)$-car positive samples:
$D^{+}_{(N-1)-car} = \{(I_1, \mathbb{B}_{K_1}), \cdots, (I_m, \mathbb{B}_{K_m})\}$,
$|K_l| = N-1, l \in \{1, \cdots, m\}$
output: N-car positive samples: $D^{+}_{N-car} = \{(I_i, B_i^J);_{k_i \geq N, J \subseteq [1,k_i], |J|=N, B_i^J \subseteq \mathbb{B}_i, i \in [1,n]}\}$

for $i \leftarrow 1$ **to** n **do**
  select the i-th training image;
  for $u \leftarrow 1$ **to** m **do**
    $K \leftarrow K_u$;
    set B_i^K *as the seed*;
    set $j = 1$, $maxov = 0$;
    for *each elem* $k \in K$ **do**
      find B_i^v *with*
      $v = \text{argmax}_{s \in \mathcal{N}_{B_i^k}} ov(B_i^s, B_i^k)$;
      if $maxov < ov(B_i^k, B_i^v)$ **then**
      $maxov \leftarrow ov(B_i^k, B_i^v), j = v$;
    end
    add $(I_i, B_i^{K \cup \{j\}})$ *to* D^{+}_{N-car}
  end
end

Algorithm 2: Generating N-car ($N > 2$) positive examples

Post-processing

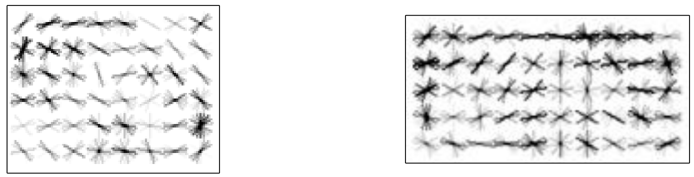
- N-car guided non-maximum suppression (GNMS)




SIKAILU
国际在线
8211 8811

Reconfigurable Context Patterns

Single Car Templates



Configured Car-to-Car Context