Pedestrian Detection with Autoregressive Network Phases

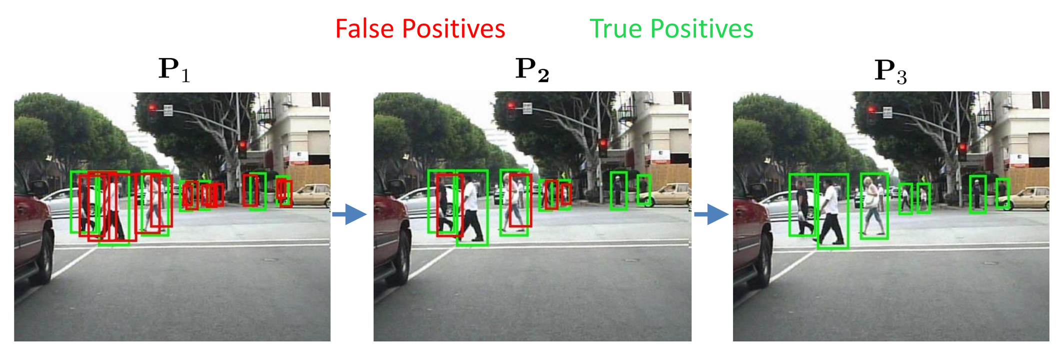
Garrick Brazil, Xiaoming Liu
Computer Vision Lab, Michigan State University
{brazilga, liuxm}@msu.edu

Problem & Contributions

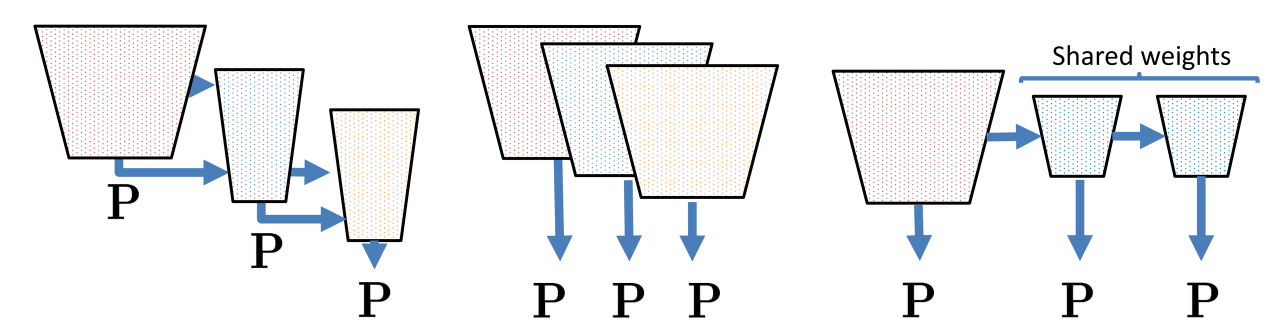
Autoregressive flow of pedestrian detection predictions can boost the precision and robustness of detections without impeding efficiency.

Our main contributions are:

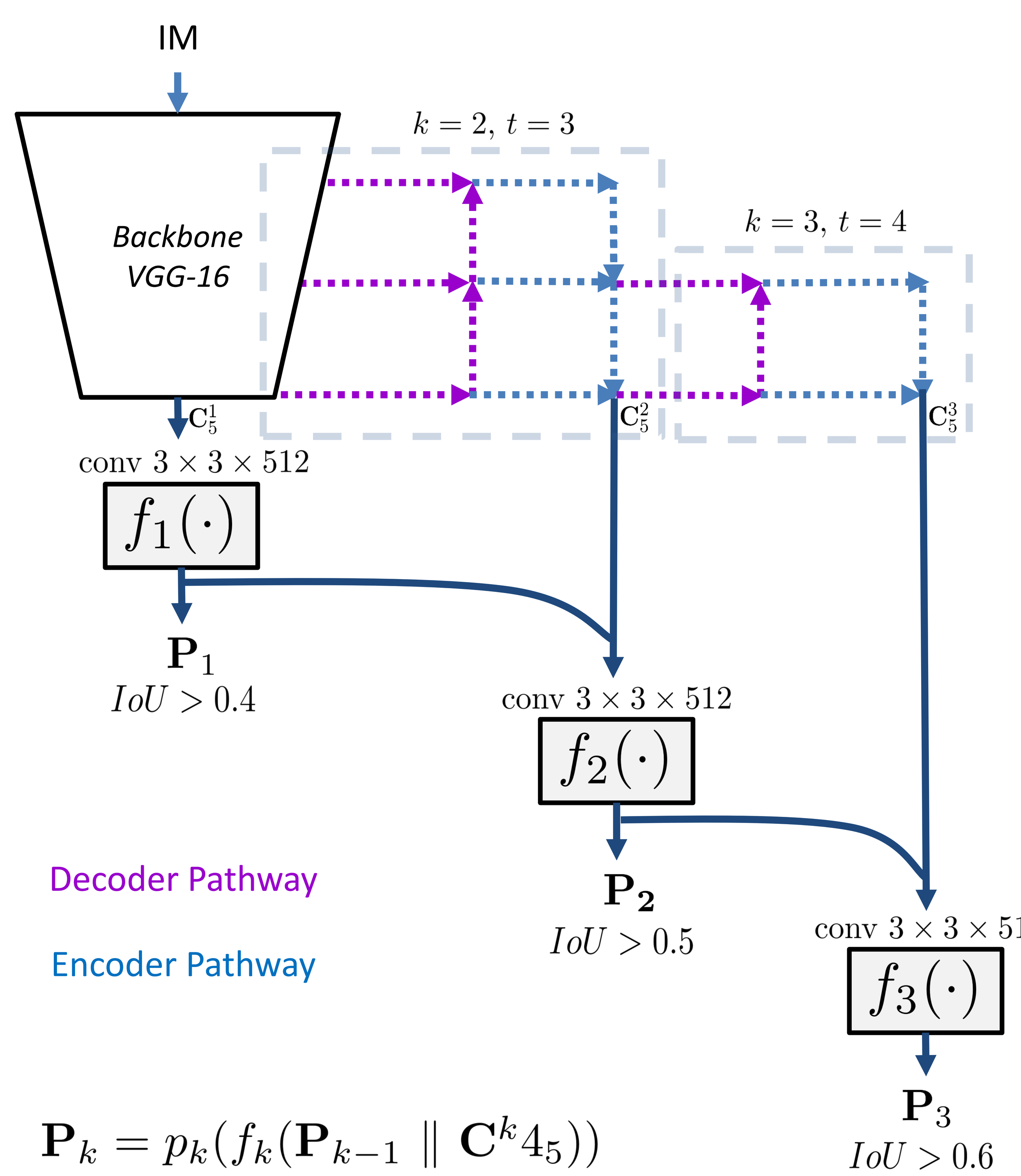
- Autoregressive framework for pedestrian detection using incremental supervision of increasingly *precise* labeling policies.
- Lightweight decoder-encoder module to facilitate feature map refinement and message passing between each autoregressive phase.



Method & Comparison

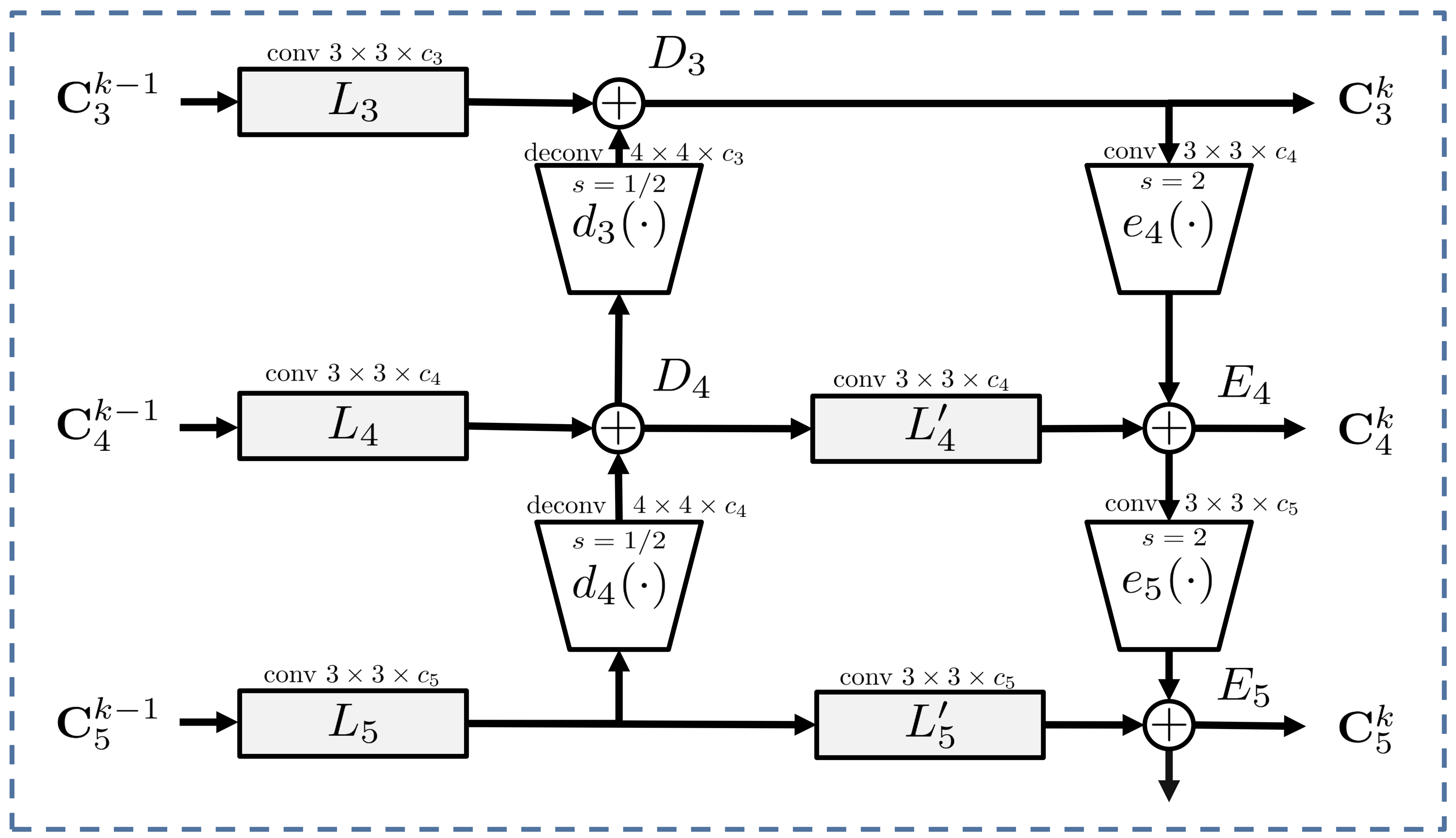


Autoregressive Detection



$$\mathbf{P}_k = p_k(f_k(\mathbf{P}_{k-1} \parallel \mathbf{C}^k 4_5))$$

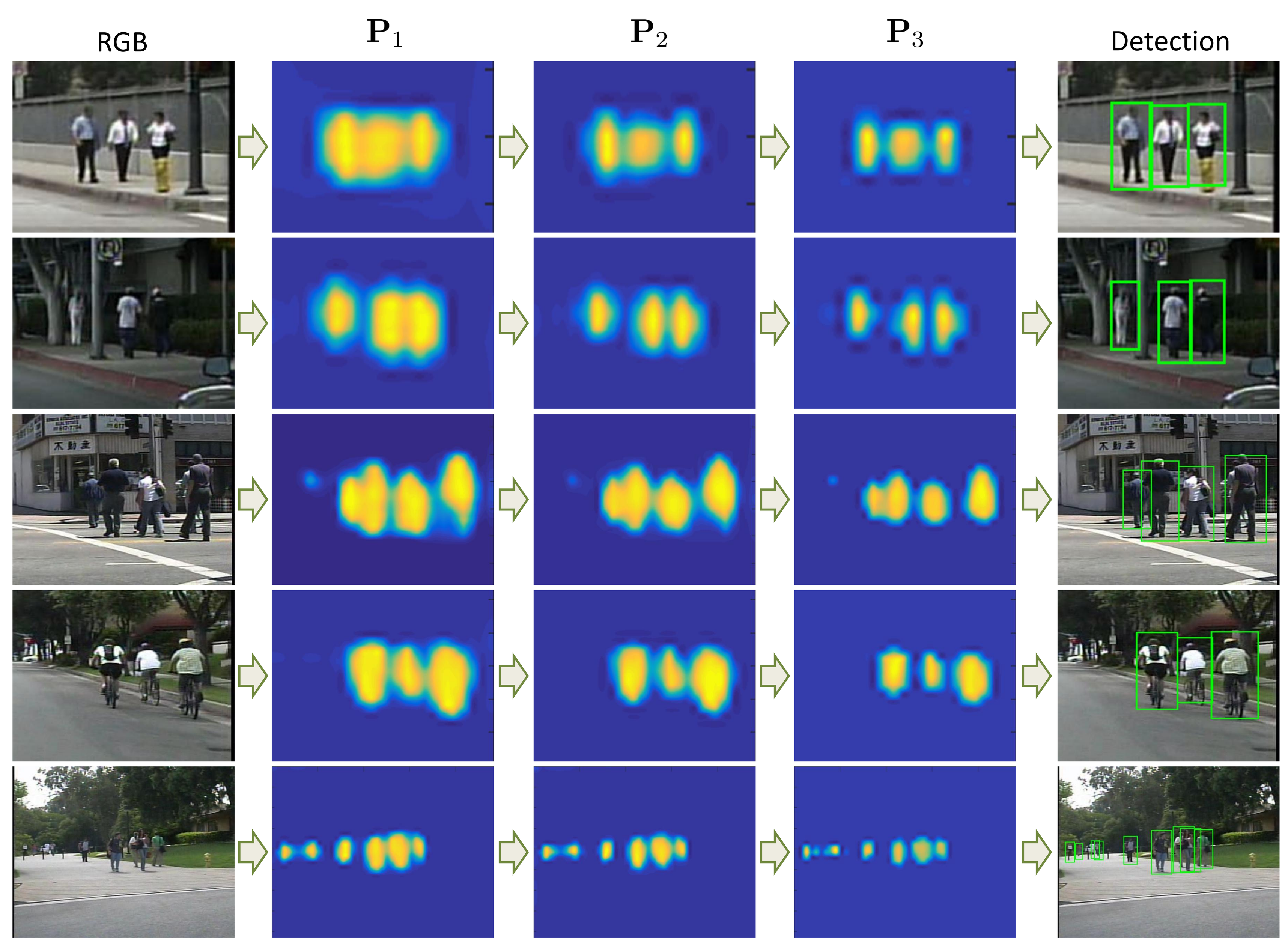
De-Encoder Module



$$\mathbf{D}_i = d_i(\mathbf{D}_{i+1}) + \mathbf{L}_i. \qquad \mathbf{E}_i = e_i(\mathbf{E}_{i-1}) + \mathbf{L}_i'$$

The de-encoder module consist of top-down and bottom-up pathways with inner-lateral convolution between pathways to produce diversified features, as well as convolutional re-sampling layers (s denotes convolutional stride) e_i and d_i for memory-efficient feature generation.

Phase Visualizations



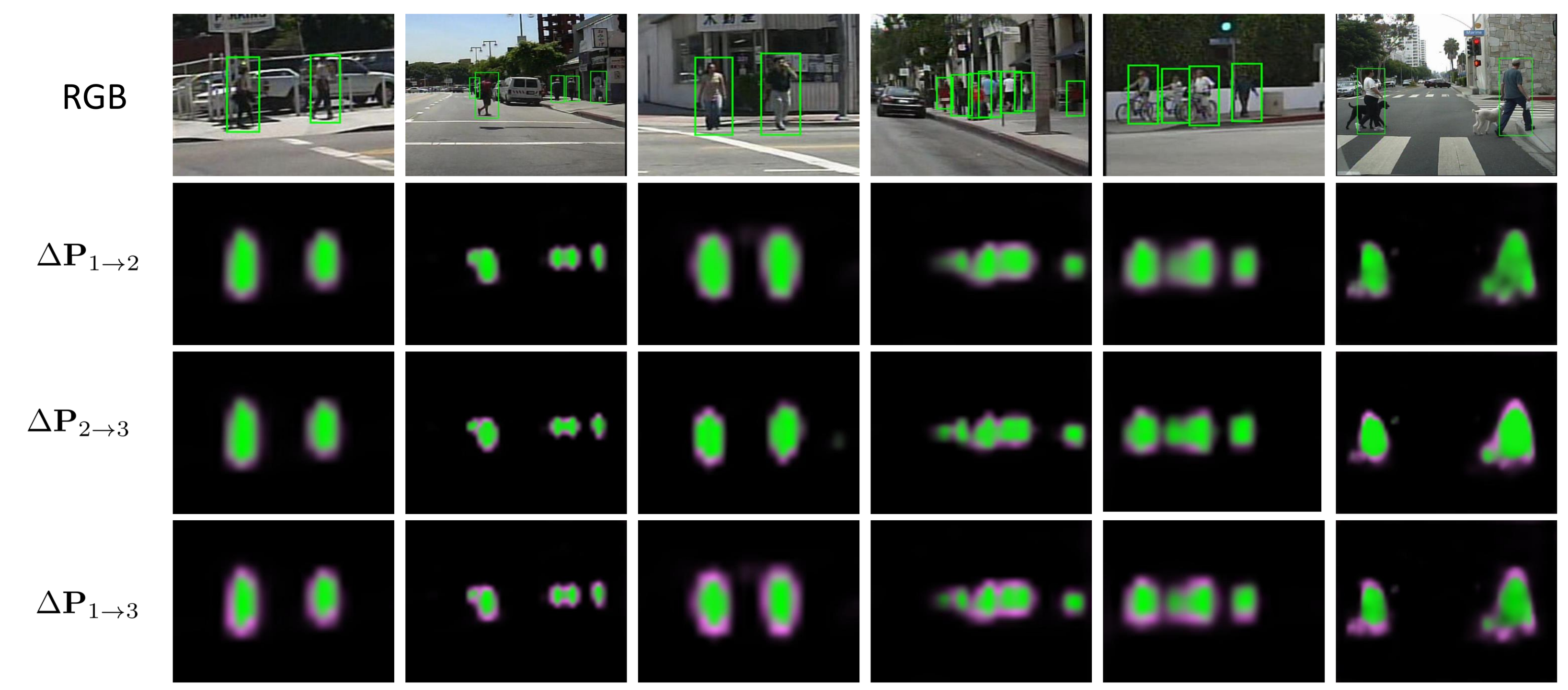
We visualize the prediction maps of each phase using the max foreground scores across all anchors at each location. We use scaled blue to yellow colors, where yellowness indicates high confidence.

Benchmarks

	Caltech Reasonable				Caltech Occlusion			KITTI		
	MR^O_{-2}	MR^O_{-4}	MR^N_{-2}	MR^N_{-4}	Partial^O	Heavy^O	RT (ms)	Easy	Mod.	Hard
MS-CNN	9.95	22.45	8.08	17.42	19.24	59.94	64	*83.92*	*73.70*	**68.31**
RRC	–	–	–	–	–	–	75	–	**75.33**	–
RPN+BF	9.58	18.60	7.28	16.76	24.23	74.36	88	75.58	61.29	56.08
F-DNN	8.65	19.92	6.89	*14.75*	15.41	*55.13*	–	–	–	–
TLL(MRF)+LSTM	7.40	–	–	–	–	–	–	–	–	–
ALFNet	–	–	6.10	–	–	–	–	–	–	–
SDS-RCNN	*7.36*	*17.82*	6.44	15.76	14.86	58.55	95	–	63.05	–
RepulsionLoss	–	–	*5.00*	–	–	–	–	–	–	–
FRCNN+ATT-vbb	10.33	–	–	–	–	45.18	–	–	–	–
PDOE+RPN	7.60	–	–	–	*13.30*	44.40	–	–	–	–
GDFL	7.85	19.86	–	–	16.74	*43.18*	–	**84.61**	68.62	66.86
DSSD+Grid	10.85	18.20	–	–	24.28	**42.42**	–	–	–	–
AR-RPN (ours)	8.01	21.62	5.78	15.86	16.30	58.06	86	–	–	–
AR-Ped (ours)	**6.45**	**15.54**	**4.36**	**11.39**	**11.93**	48.80	91	83.66	73.44	*68.12*

Comprehensive comparison of our frameworks and the state-of-the-art on the Caltech and KITTI benchmarks, in accuracy and runtime (RT). **Boldface**/*italic* indicate the best/second best performance.

Ablation Study



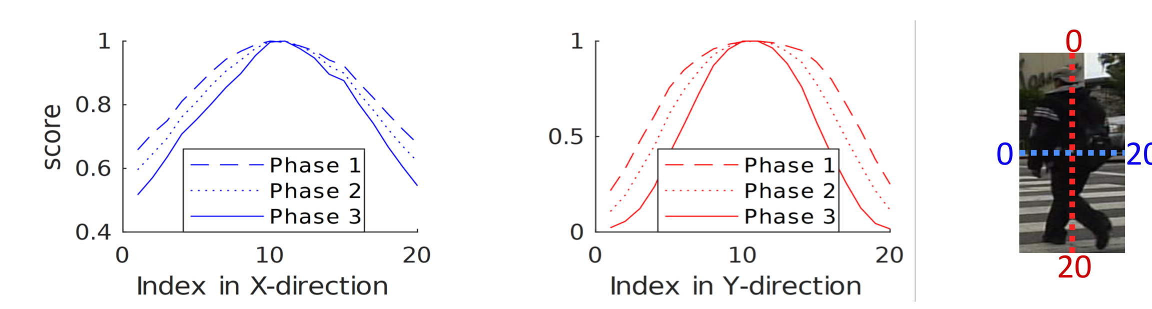
We analyze the prediction disagreements between each phase. We use the green to represent foreground agreement and magenta for **suppression** between each consecutive phase.

N_k	**c** size	MR^O_{-2}	MAC (G)	Runtime (ms)
1	M	10.16	217.9	68
2	M	8.32	429.3	80
3	S	8.62	255.3	74
3	M	8.01	321.3	86
3	L	8.33	429.3	115
4	M	8.68	355.9	97

Experimental settings channel width (c) at small, medium, large, and across N_k number of phases.

Labeling Policy	MR^O_{-2}
no autoregressive	9.06
strict → lenient	9.03
moderate → moderate	8.94
strict → strict	8.43
lenient → strict	8.01

Ablation of different labeling policy techniques of varying strictness.



We analyze the mean prediction score of 20 uniformly sampled points along the center lines of X-direction (left) and Y-direction (right).

project website
+ source code