Supplementary Material

S1. Additional Experiments

S1.1. Sensitivity Study

Table S1. Sensitivity study on camera rotation speeds.

	200°/s	400°/s	700°/s
RR	41.43%	43.01%	43.66%
Dynamic-oracle	44.02%	45.38%	45.75%
WideEye	50.14%	52.12%	52.74%

Camera rotation speed. As mentioned in subsection 4.1, our evaluation results use a camera with a rotation speed of 200°/s, commonly achievable for commercial PTZ camera offerings. Higher-speed PTZ models are also available, with rotation speeds up to 400°/s [35] or even up to 700°/s [36]. To understand the impact of rotation speed, we compare the performance of WideEye and the dynamic baselines under varying speeds, as shown in Table S1. Both the baselines and WideEye benefit from faster camera rotation – as the speed increases from 200°/s to 700°/s, the accuracy of RR, Dynamic-oracle and WideEye improves by 2.23%, 1.73%, and 2.60% respectively. Despite these gains, WideEye consistently outperforms the baselines across all rotation speeds.

Number of candidate FoVs. By default, the evaluation of RR and WideEye uses 4 candidate FoVs – the minimum needed to cover the WFoV. (Note that Dynamic-oracle was not restricted to this 4-FoV configuration.) We next expand the number of candidate FoVs to 6, arranged as 3 per row, resulting in overlapping FoVs within each row. As the number of candidate FoVs increases from 4 to 6, RR's accuracy drops from 41.43% to 40.91%, and WideEye's accuracy drops from 50.14% to 48.69%. For RR, the reduced accuracy is due to the longer time required to cycle through all FoVs. For WideEye, the drop is due to inconsistencies introduced by overlapping FoVs: WideEye's per-FoV accuracy model assumes all objects in future frames in the FoV are predicted by the trajectory model. With overlapping FoVs, however, a subset of objects in each frame within an FoV are directly observed rather than predicted, violating this assumption. These results suggest that WideEye performs best when configured with a minimal set of non-overlapping FoVs that cover the WFoV.

S1.2. Component-wise Analysis

S1.2.1. Trajectory Model

We evaluate WideEye's trajectory model compared with the two baselines: (1) **Static**, which assumes objects remain at their last-seen location, and (2) **Kalman-Filter**, which uses a Kalman filter [54] to estimate each object's velocity and the rate of change in bounding box size, then extrapolates

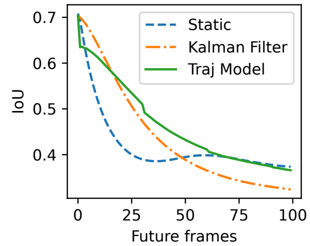

Figure S1. The IoU degradation of various algorithms for predicting bounding boxes into the future.

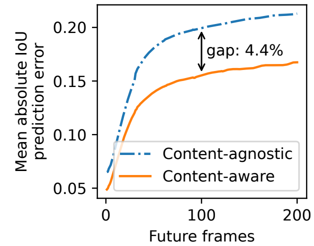

Figure S2. Accuracy of WideEye's content-aware accuracy predictor, vs. a content-agnostic predictor that outputs the mean IoU degradations profiled on each video separately.

these into the future. We take moving windows of $M+N$ frame sequences, providing each algorithm with the bounding boxes for the first M frames and making each predict future bounding boxes of the subsequent N frames, and calculate the per-frame IoU across the objects. In our experiment, we set $M = 30$ and $N = 100$.

Analysis. Figure S1 shows the average IoU at each future frame. WideEye's trajectory model outperforms Static in the short term, achieving 0.12 higher IoU at 30 frames ahead. Interestingly, at the 60th frame (*i.e.* 2 seconds into the future), Static briefly improves due to typical 2-second gaps between vehicles [8], where trailing vehicles move into the bounding boxes of leading ones. Compared to Kalman filter, WideEye is more accurate over longer horizons, with 0.06 higher IoU at 100 frames. Kalman filter's simple motion assumptions (*e.g.* constant velocity) break down as trajectories grow more complex. In summary, WideEye's trajectory model outperforms the baselines in mitigating accuracy degradation across both short- and long-term scenarios, whereas the baselines are effective only in either the short or long term.

S1.2.2. Accuracy Predictor for the Trajectory Model

We next evaluate WideEye's accuracy predictor using windows of $M+N$ frames on the test set videos. The trajectory model predicts bounding boxes for the last N frames based on the first M, and we compute the IoU degradation against a high-accuracy DNN as ground truth. The accuracy predictor is then given the first M frames' features to predict IoU degradation for each of the future N frames. For each frame i, we report the mean absolute error (MAE) between predicted and actual degradation. As a baseline, we compare against a content-agnostic predictor that outputs mean degradation values for each FoV, based on offline profiling a separate video of the same scene.

Analysis. Figure S2 shows that prediction error increases over time for both methods, due to higher uncertainty further into the future. However, WideEye's content-aware predictor consistently outperforms the baseline, with the gap widening over time: at 1 frame into the future, MAEs are 4.9% vs. 6.5%, and at 100 frames, 15.5% vs. 19.9%. This demonstrates the effectiveness of WideEye's accuracy predictor in modeling long-term degradation.