SECOND:

**1. Terminology & Metrics**

* **AP@X**: Average Precision at IoU threshold X (e.g. 0.5 for pedestrians/cyclists, 0.7 for cars). Higher is better.
* **AOS@X**: Average Orientation Similarity (combines detection + correct orientation) at the same IoU threshold.
* **bbox\_2d**: 2D bounding‐box AP in image space.
* **bbox\_bev**: Bird’s‐Eye View detection AP.
* **bbox\_3d**: Full 3D bounding‐box AP.

**Difficulty levels**

* **Easy**, **Moderate**, **Hard** correspond to increasing occlusion, truncation, distance, and smaller object sizes.

**2. Per‐Category Insights**

**Pedestrians (**[**AP@0.5**](mailto:AP@0.5)**)**

|  |  |  |  |
| --- | --- | --- | --- |
| **View / Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 69.21% | 66.12% | 63.43% |
| **AOS** | 65.40% | 61.91% | 58.93% |
| **BEV AP** | 63.11% | 56.77% | 53.83% |
| **3D AP** | 58.68% | 53.90% | 49.75% |

* **Drop‐off**: The max drop from Easy→Hard was ~9.3 pp om BEV AP.
* **AOS Quality**: AOS for SECOND is ~15 pp better than PointPillar implying it much more reliably can determine pedestrian orientation.
* **AOS lag**: AOS for SECOND is ~15 pp better than PointPillar implying it much more reliably can determine pedestrian orientation.

**Cyclists (**[**AP@0.5**](mailto:AP@0.5)**)**

|  |  |  |  |
| --- | --- | --- | --- |
| **View / Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 86.66% | 76.53% | 72.67% |
| **AOS** | 86.33% | 75.98% | 72.12% |
| **BEV AP** | 83.95% | 69.92% | 66.34% |
| **3D AP** | 80.72% | 66.56% | 62.22% |

* **Strong Easy performance**: Cyclists are detected very reliably when large and unoccluded (~86 pp). This holds
* **Larger drop**: There’s large drops across all the measurements in AP. This suggests cyclicts are much harder to identify as conditions affecting visibility degrade.
* **Comparatively**: AP for detecting cyclists is almost on par with car detection in easy conditions, but is much more difficult to do under harder conditions. It tends to still be easier to identify cyclists than pedestrians.

**Cars (**[**AP@0.7**](mailto:AP@0.7)**)**

|  |  |  |  |
| --- | --- | --- | --- |
| **View / Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 90.81% | 89.98% | 89.18% |
| **AOS** | 90.80% | 89.90% | 89.01% |
| **BEV AP** | 89.88% | 87.83% | 86.47% |
| **3D AP** | 88.52% | 78.61% | 77.33% |

* **Very high AP**: Even at a strict 0.7 IoU, cars are detected with >89% AP in Easy/Moderate.
* **Narrow drop**: Only ~1 pp drop Easy→Hard in 2D AP—cars remain easy to localize.
* **3D gap**: 3D AP drops by ~10 pp from Easy→Hard, reflecting height/orientation challenges for distant or occluded cars.

**3. Overall Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 82.23% | 77.54% | 75.09% |
| **AOS** | 80.84% | 75.93% | 73.35% |
| **BEV AP** | 78.98% | 71.51% | 68.88% |
| **3D AP** | 75.97% | 66.36% | 63.10% |

* **Performance hierarchy**: 2D > AOS > BEV > 3D, as each “dimension” adds complexity (orientation, ground‐plane projection, full 3D estimation).
* **Graceful degradation**: AP drops ~9 pp Easy→Hard in 2D, but ~14 pp in 3D—models retain robustness for detection but struggle more on 3D localization under difficult conditions.
* **Room for improvement**: The larger drop in AOS and 3D (especially for pedestrians/cyclists) suggests future work should focus on better orientation/height estimation, perhaps via multi‐sensor fusion or improved regression losses.

**Key Takeaways**

The high AP for cars even in hard conditions suggests that autonomous driving systems relying on this model might have a relatively robust perception of vehicles.

The larger performance drops for pedestrians and cyclists in challenging conditions underscore the need for further research to improve the safety of vulnerable road users in autonomous systems.

The significant gap between detection accuracy (2D bbox AP) and 3D localization (3D AP), particularly for pedestrians and cyclists, could impact downstream tasks like trajectory prediction and collision avoidance, emphasizing the importance of better 3D understanding.