Pointpillars:

## **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** | 64.40% | 61.43% | 57.62% |
| **AOS** | 49.35% | 46.73% | 43.84% |
| **BEV AP** | 59.11% | 54.32% | 50.50% |
| **3D AP** | 51.35% | 47.98% | 43.80% |

* **Drop‐off**: Only a ~6.8 pp drop from Easy→Hard in 2D AP, which is fairly graceful.
* **2D vs. BEV vs. 3D**: Performance degrades as you go “deeper” (2D→BEV→3D), reflecting the added localization and height/orientation challenges.
* **AOS lag**: AOS is ~15 pp lower than 2D AP, indicating orientation estimation for pedestrians is substantially harder than simply detecting them.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **View / Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 86.24% | 73.06% | 70.17% |
| **AOS** | 85.02% | 69.08% | 66.28% |
| **BEV AP** | 84.41% | 67.13% | 63.74% |
| **3D AP** | 81.76% | 63.66% | 60.90% |

* **Strong Easy performance**: Cyclists are detected very reliably when large and unoccluded (~86 pp).
* **Larger drop**: There’s a ~16 pp drop Easy→Hard in 2D AP, more pronounced than for pedestrians.
* **Orientation**: AOS tracks AP closely (within ~1 pp), showing that once cyclists are detected, orientation is estimated reasonably well in easy cases.

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

|  |  |  |  |
| --- | --- | --- | --- |
| **View / Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 90.65% | 89.33% | 86.66% |
| **AOS** | 90.48% | 88.68% | 85.73% |
| **BEV AP** | 89.96% | 87.88% | 85.77% |
| **3D AP** | 86.63% | 76.74% | 74.17% |

* **Very high AP**: Even at a strict 0.7 IoU, cars are detected with >89% AP in Easy/Moderate.
* **Narrow drop**: Only ~4 pp drop Easy→Hard in 2D AP—cars remain easy to localize.
* **3D gap**: 3D AP drops by ~13 pp from Easy→Hard, reflecting height/orientation challenges for distant or occluded cars.

## **3. Overall Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| **Metric** | **Easy** | **Moderate** | **Hard** |
| **2D bbox AP** | 80.43% | 74.61% | 71.48% |
| **AOS** | 74.95% | 68.16% | 65.29% |
| **BEV AP** | 77.83% | 69.78% | 66.67% |
| **3D AP** | 73.25% | 62.79% | 59.62% |

* **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**

1. **Cars are easiest** (largest, well‐defined shapes), while pedestrians and cyclists pose greater challenges.
2. **Orientation estimation** lags detection accuracy, especially for small objects—address with finer angular bins or dedicated orientation heads.
3. **3D localization** remains the hardest task, highlighting the importance of improved post‐processing or multi‐modal inputs (RGB+LiDAR).