Per-class AP vs. Average Recall (Test)