Street Parking Presence Inference from Street-Level Imagery

via Multi-Cue Detection and Geo-Aggregation

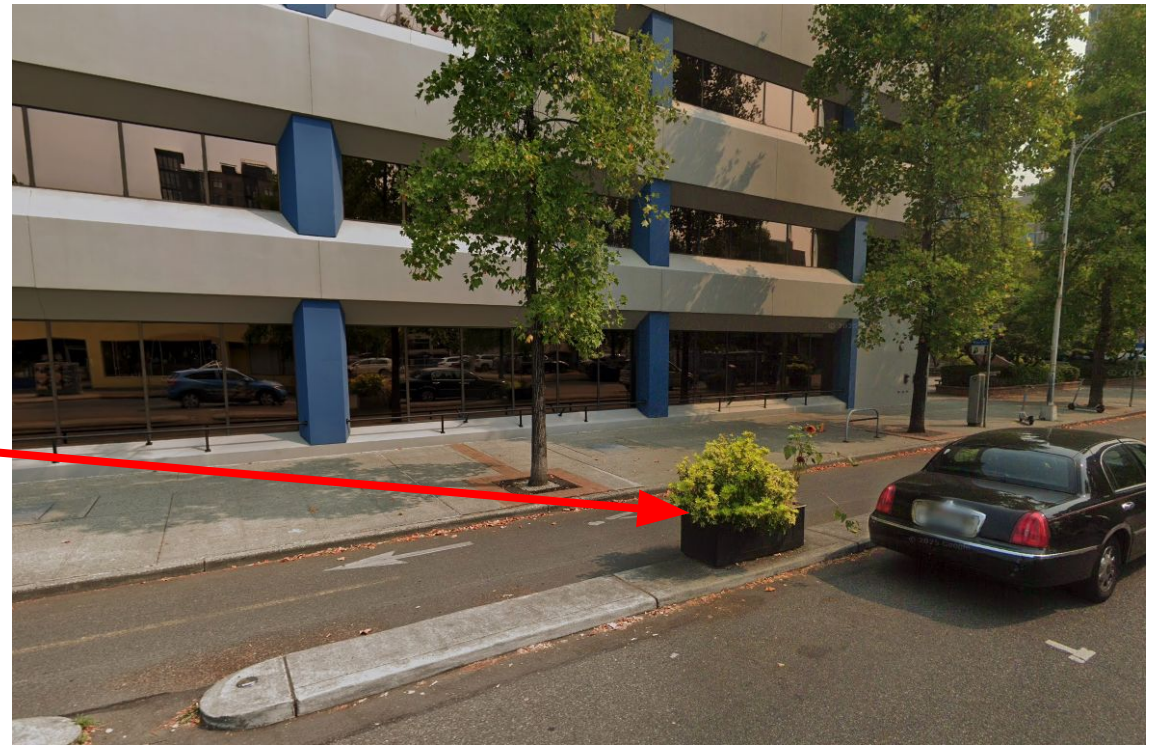
Chirag Jain · **Ritik Singh**

University of Wisconsin–Madison · CS 766 Computer Vision · Spring 2026

quasars06.github.io/cs766-street-parking

Does this street allow parking?

View 1 — parking sign visible.



View 2 — same curb, sign out of frame.

A "no parking visible" prediction from one image often just means the cue isn't in this frame.

⇒ Treat the segment as the unit of inference, not the image.

Why this matters

The practical cost

Drivers cruising for parking are a meaningful share of urban congestion — fuel, emissions, time. The data on where parking is allowed is often incomplete or out of date.

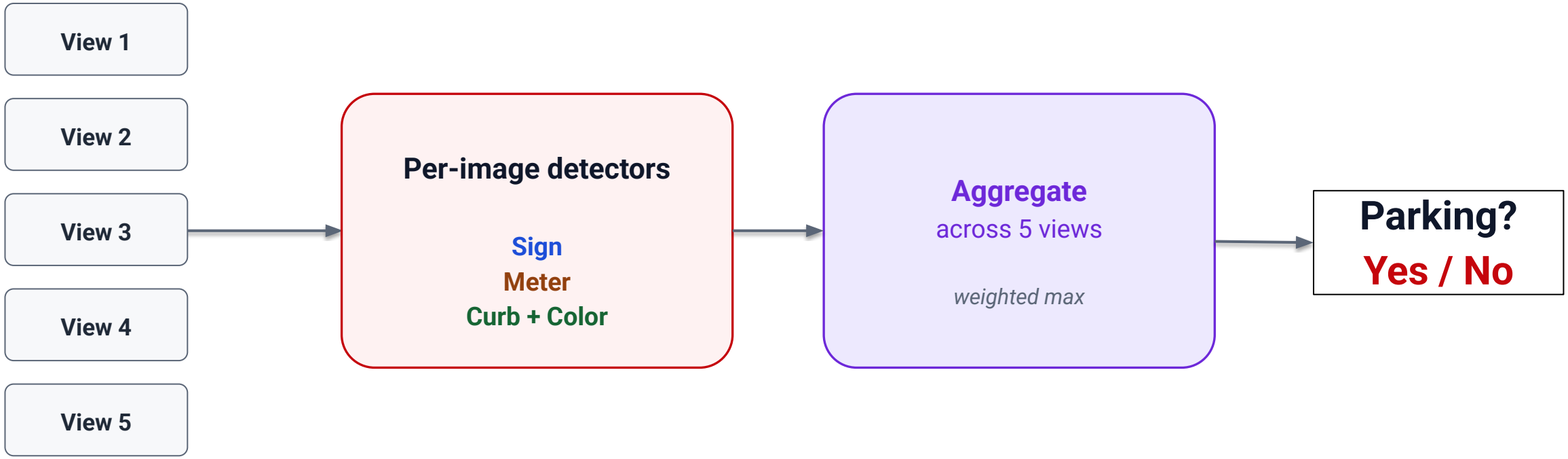
Curb management

Cities are trying to allocate curb space across parking, loading zones, ride-share pickup, and emergency access. Today: manual surveys + static GIS.

The CV angle

Parking cues are small, sparse, and viewpoint-dependent. Evidence is naturally distributed across views — multi-view reasoning is the natural framing.

Our pipeline at a glance



Datasets

- **MTSD** (Mapillary Traffic Sign Dataset) — labeled traffic signs, source for parking-sign training.
- **Mapillary Vistas** — rich curb annotations + parking-meter labels for evaluation.
- **Our manual real-world set** — 6 segments × 5 views, collected by us. No public dataset has the segment-level grouping we needed.

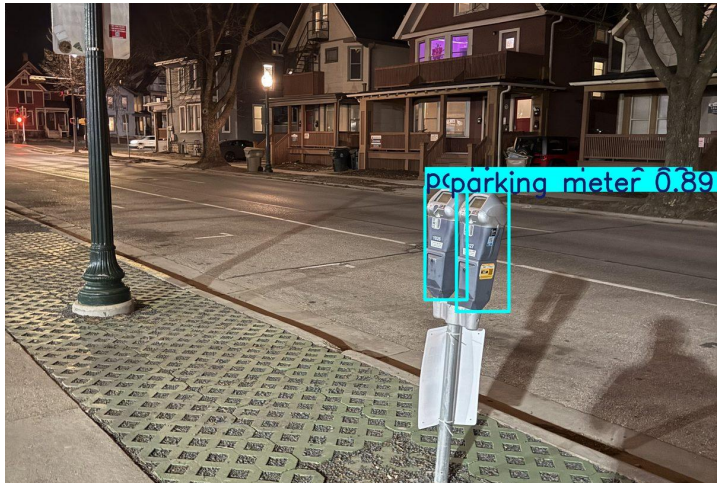
Three image-level cues

Sign

Parking sign

YOLOv8m, **supervised** on a binary parking-sign task derived from MTSD. **Strongest cue.**

Meter



Parking meter

YOLO11x zero-shot (COCO-pretrained), evaluated on Mapillary Vistas. **Useful but noisy** — pole-shaped objects cause false positives.

Curb

Curb structure + color

U-Net segments curb pixels → **HSV** color analysis on the boundary. Most curbs unpainted → conservative **"unknown"** fallback.

Segment-level aggregation

Aggregation

For each cue, take the **strongest evidence** found across all 5 views. Trust **signs > meters > curb**.

$$S\_seg = \max(\max s_i , \; 0.6 \cdot \max m_i , \; 0.4 \cdot \max c_i) \geq 0.15$$

si - Parking Sign *mi - Parking Meter* *ci - Curb*

Weights reflect cue reliability: signs (1.0) most reliable, meters (0.6) noisy, curb color (0.4) most indirect.

How we differ from prior work

Existing parking-detection literature is overwhelmingly **sign-centric and per-image**. Our contribution: combine multiple complementary cues, ***and*** aggregate across views.

Headline result: aggregation improves recall

Evaluation setup: **synthetic 225-segment benchmark** built from detector/evaluation pools.
Each segment has **5 images**; labels include 80 negative and 145 positive segments across sign, meter, curb, and multi-cue cases.

Method	Precision	Recall	F1
Single-image baseline	0.784	0.200	0.319
Segment aggregation	0.753	**0.924**	**0.830**

Main interpretation: recall jumps because parking evidence is sparse — a single view often misses the cue, but at least one of five views usually captures it.
Precision drops only slightly because max-pooling can also promote a few noisy meter/curb detections to segment positives.

Takeaway: **aggregation trades a tiny precision loss for a massive recall gain.**

Real-world segments we collected

seg_001 — meter-only success ✓ aggregation rescues the segment

Sign **0.00** · Meter **0.72** · Curb **0.36** → Combined **0.43** ✓

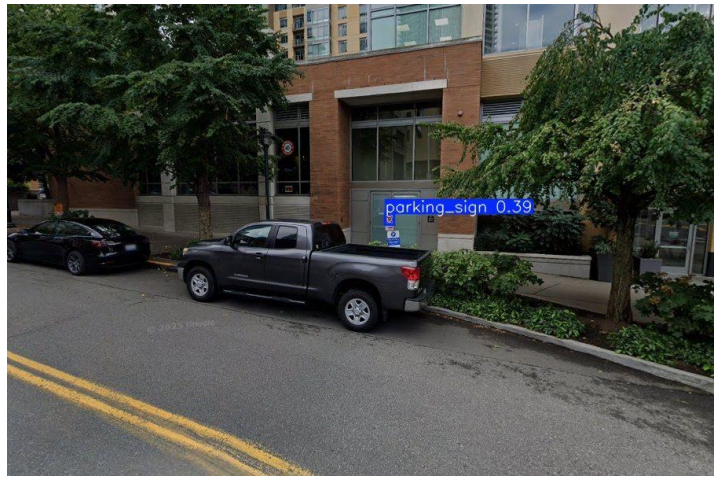
seg_005 — out-of-distribution failure ✗ honest limit of the approach

Sign **0.00** · Meter **0.00** · Curb **0.09** → Combined **0.04** ✗

Aggregation rescues sparse-cue cases — but cannot recover concepts the detector never learned.

Qualitative analysis: what the detectors are actually doing

imgsz=640 · **misses both**



imgsz=960 · **catches one**

imgsz=1280 · **catches both**

Scale sensitivity

The same scene changes as resolution changes. The detector has learned the concept, but small signs disappear in wide views.

Sub-icon / OCR limitation

The sign detector often boxes the salient P/no-parking icon, not the full rule text. For full rule understanding, OCR or VLM-based sign reading is the next step.

In the interest of time, we only show the key qualitative patterns: scale sensitivity, partial sign localization, and the role of multi-view evidence.

What we learned

① **Multi-view rescue is the real win.** *Bigger gain than any single-detector tweak.*

② **Detectors are weirdly scale-sensitive.** *Lowering imgsz from 1280 → 160 sometimes improved distant-sign detection.*

③ **Sign detector boxes the salient sub-icon (the “P”).** *Fine for cue presence, limiting for full rule parsing.*

④ **For curb color, conservative beats aggressive.** *“Unknown” was the right answer about half the time.*

⑤ **Aggregation has a ceiling.** *It can't recover concepts outside the detector's training distribution.*

Honest take: these results aren’t production-ready. But the architectural principle — multi-cue, multi-view — clearly holds.

Future work: VLMs for compositional sign reading · georeferenced segments · learned cue-fusion weights

Thank You

Questions?

quasars06.github.io/cs766-street-parking