Risky Jaywalking - Literature Review

Introduction

Pedestrian safety in urban environments remains a persistent challenge, with a majority of severe incidents occurring at midblock, unsignalized locations rather than at designated crosswalks. Traditional safety analysis relies heavily on crash data, which is sparse, reactive, and often fails to identify dangerous locations *before* an accident occurs. Recent advances in deep learning–based pedestrian intention prediction enable proactive detection of crossing behavior, offering an opportunity to map risky crossing zones based not on collisions but on actual human behavior.

This project proposes a system that combines (1) micro-level pedestrian crossing intent detection using state-of-the-art vision models, and (2) macro-level risk zone identification based on infrastructure context (crosswalk presence and road speed). The goal is to identify high-risk pedestrian crossing areas, focusing especially on jaywalking and informal midblock crossings.

Research Question

How can we integrate real-time pedestrian crossing intent prediction with map-based infrastructure analysis to identify and characterize high-risk pedestrian crossing zones—particularly locations where pedestrians frequently cross without nearby crosswalks or where vehicle speeds increase severity of potential conflicts?

Core sub-questions:

1. Micro-level: How accurately can deep learning models detect pedestrian crossing intent in naturalistic driving datasets (PIE, JAAD)?

2. Macro-level: How can we spatially aggregate detected crossing events and correlate them with infrastructure factors like crosswalk availability and road speed?

3. Risk scoring: How can these factors be combined into an interpretable and actionable risk score for each location?

Inspiration and Background

This research is inspired by two major developments:

- Advances in pedestrian intention prediction models- Recent models such as TCL, TAMFormer, TrEP, CAPformer, and Coupling Intent-Action provide highly accurate early detection of a pedestrian's intent to cross. These models leverage Transformers, temporal clustering, multi-modal fusion, and uncertainty estimation to detect subtle behavioral cues.
- Emergence of behavior-based safety mapping - Frameworks like CHAMP demonstrate that repeated detection of pedestrian behavior across city streets can reveal informal crossing hotspots. Such hotspots often correspond to areas where infrastructure is lacking (e.g., no crosswalk for hundreds of meters) and where drivers do not expect pedestrians.

Together, these fields highlight the potential of combining AI-based pedestrian behavior understanding with city-level infrastructure analysis to proactively identify danger zones.

Positioning of This Research

This project is positioned at the intersection of:

- Computer Vision (pedestrian intention modeling)
- Transportation Engineering (risk zone identification)
- Urban Informatics (map-based behavioral analytics)

While previous work typically focuses on either:

- micro-level intent detection (AV/ADAS research), or
- macro-level hotspot detection (traffic engineering),

no existing work integrates both into a unified risk scoring system that:

1. Detects actual pedestrian crossing attempts,
2. Aggregates them spatially,
3. Scores each location by infrastructure-related risk factors.

Thus, our project fills a gap by creating a behavior-driven, infrastructure-aware risk mapping system.

Prior Work Relevant to Our Approach

Pedestrian Intention & Action Prediction

Key models reviewed:

- TCL (Temporal-Contextual Event Learning) – event-level temporal clustering + contextual attention; SOTA on PIE/JAAD.
- TAMFormer – multi-modal Transformer with learned temporal attention masks.
- TrEP – uncertainty-aware Transformer for intent prediction.
- CAPformer – first strong Transformer baseline for crossing prediction.
- Coupling Intent-Action – multi-task model predicting future actions & intent jointly.

These models show that crossing intent can be detected reliably before the pedestrian actually steps onto the road, making them ideal for hotspot detection.

High-Risk Location & Jaywalking Research

Relevant findings:

- Crossing patterns cluster naturally in certain informal midblock locations.
- Absence of crosswalks, long walking detours, and long signal delays drive jaywalking.
- Road speed strongly influences fatality risk; most pedestrian deaths occur on high-speed roads.
- Computer vision systems (CHAMP, CCTV-based analysis) can detect risky behaviors and near-misses.
- OpenStreetMap provides infrastructure data such as crosswalk locations and speed limits.

This prior work directly supports our macro-level risk evaluation.

Our Approach: Key Differences

Our approach differs from prior work in several ways:

1. Joint Micro–Macro Integration

We combine deep learning intent detection (micro) with infrastructure-aware clustering (macro). Prior studies rarely merge these two perspectives.

2. Behavior-Driven Risk Mapping

Instead of relying on crash statistics or manually annotated exposure metrics, we derive hotspots purely from observed pedestrian behavior using modern intention models.

3. Infrastructure and Speed-Based Risk Scoring

Our initial risk score focuses on:

- Presence or absence of crosswalks, and
- Speed limits (as proxy for severity).

This creates a simple but powerful baseline risk map that can be expanded later.

4. Use of PIE/JAAD Annotations for Ground Truth

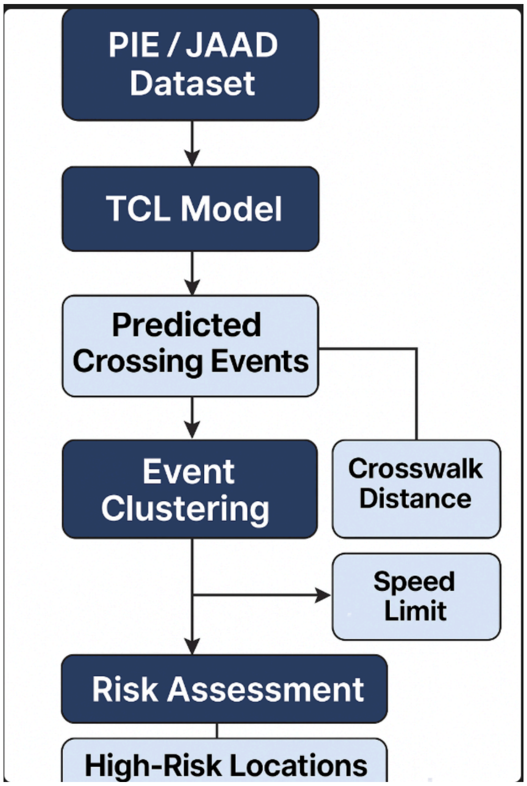
Since PIE already includes intent, crossing, and location annotations, our system can be:

- evaluated,
- validated,
- and trained using these labels directly.

This makes our project much more feasible at early stages.

Methodology: High-Level Architecture and Pipeline Steps

Below is the proposed pipeline, aligned with your constraints (TCL + PIE/JAAD + crosswalk + speed).



Pipeline Overview

Step 1 — Data Acquisition

- Use the PIE and JAAD datasets.
- Extract sequences with:
    - pedestrian bounding boxes,
    - intent labels,
    - crossing annotations,
    - GPS or relative location if available,
    - frame timestamps.

Step 2 — Micro-Level Event Detection (via TCL)

- Train or fine-tune the TCL model on PIE/JAAD.
- For each frame sequence, run TCL to detect:
    - *crossing intent*,
    - *crossing start event*,
    - *confidence*.

This yields a list of predicted crossing events, each with a location.

Step 3 — Event Spatialization

(For datasets that lack GPS, approximate camera location or use known calibration.)

- Convert crossing detections into spatial points:
    - If using dashcam: project detections to global/GPS map.
    - For PIE/JAAD: use scene coordinate mapping.

Output: a dense set of crossing points across all video sequences.

Step 4 — Hotspot Detection (Cluster Crossing Events)

- Use DBSCAN or kernel density estimation to find areas with repeated crossings.
- Each hotspot represents an informal or behaviorally significant crossing location.

Step 5 — Infrastructure Context Extraction

For each hotspot:

- Find the nearest marked crosswalk using OpenStreetMap.
- Compute:
    - distance to nearest crosswalk,
    - road type,
    - road speed limit.

Step 6 — Risk Scoring

Baseline risk score:

> Risk = f(crossing frequency, absence of crosswalk, speed limit)

Example scoring logic:

- If no crosswalk within 50–100 m → +High risk
- If speed limit > 40 km/h → +High risk
- If high crossing density → +High exposure

Step 7 — Output Hotspot Map and Analysis

- Generate a city map showing:
    - hotspot locations,
    - risk score,
    - event counts,

    - crosswalk distances.

- Provide actionable insights:
    - recommend new crosswalks,
    - recommend speed reduction,
    - identify dangerous midblocks.

Final Deliverable

✔ A technical pipeline that uses TCL + PIE/JAAD to detect micro-level crossings.
✔ A data-driven hotspot detection algorithm.
✔ A crosswalk & speed–based risk score for each hotspot.
✔ A map of high-risk pedestrian crossing zones.
✔ A foundation for expanding into full risk modeling (visibility, near-misses, multi-modal context).