Landmark-Based Cyclist Swerving Analysis via Class-Aware One-Shot Segmentation

No Author Given

No Institute Given

Abstract. The growth of urban cycling brings safety challenges, particularly in detecting unsafe behaviors like excessive swerving, which can lead to collisions or running off the road. Identifying these patterns in video data can improve road safety and aid infrastructure planning. This study introduces a framework for automatically measuring swerving patterns by tracking the distance between landmarks defining the road and the bicycle. By leveraging the Segment Anything Model (SAM) fine-tuned with PerSAM, the method involves segmenting relevant landmarks in video frames to calculate the pixel distance between these landmarks and the bicycle, thus illustrating the swerving patterns over time. The methodology is applied to a dataset comprising 54 videos recorded in diverse cycling environments in Groningen, the Netherlands, categorized into three subdatasets based on the type of bicycle path. This research evaluates three segmentation variants: baseline PerSAM for single-class segmentation, Multi-Class PerSAM (MCP) for 2-class segmentation, and MCP for 3-class segmentation, alongside outlier removal techniques and signal smoothing to enhance measurement accuracy. Quantitative and qualitative results reveal that the optimal configuration involves two landmarks, Isolation Forest for outlier removal, and a rolling average for signal smoothing. This configuration demonstrates a strong correlation with ground truth measurements, offering valuable insights for efficient analysis of cyclists' swerving behavior in real-world environments. The code will be available upon the publication of the manuscript.

Keywords: Multi Class PerSAM, Landmark Segmentation, Swerving analysis, Egocentric video, Bicycle Safety

1 Introduction

The large number of cyclists in the Netherlands has introduced significant challenges in ensuring cyclist safety [1–5]. One of the issues is the detection of erratic or unsafe cycling behaviors, such as excessive swerving, which has become a critical focus of research [6–8]. Swerving, defined by lateral deviations from a cyclist's intended path, can signal various underlying factors, including distraction, intoxication, fatigue, or attempts to avoid obstacles. This behavior often serves as an early indicator of more complex events, such as near-misses or collisions [9–11]. The detection of such behaviors from video data holds immense potential to effectively assess cycling behavior, guide infrastructure design, and support

the development of advanced driver-assistance systems. This, in turn, can help better accommodate cyclists [12, 13].

The increase of video data from traffic cameras, dashcams, and wearable devices has opened new avenues for leveraging computer vision and machine learning to analyze cyclist swerving behavior. However, automatic analysis of swerving patterns in video data presents unique challenges. Swerving motions are frequently subtle and may lack clear visual distinctiveness, necessitating the creation of highly sensitive and robust detection algorithms [14–16].

Automatic swerving measurement relies on accurately converting the spatial data of nearby objects, such as segmentation masks of a curb, into coordinates relative to the bicycle in a video frame. This transformation enables direct analysis of the cyclist's trajectory dynamics [17, 18]. The approach involves tracking positional shifts of the landmark over time and computing their deviation relative to the bicycle's position. Through this spatial-to-metric conversion, even minor swerving patterns, otherwise obscured in raw visual data, can be reliably extracted and quantified.

However, tracking landmarks within a single reference (e.g., only a curb) can be challenging when visual obstructions occur, preventing segmentation and subsequent swerving distance calculations [19]. This research aims to automate swerving detection by leveraging multiple landmarks in video data, using spatial pixel information across several reference points to compensate for obstructions. The combined signals will be converted into lateral displacement data to identify swerving patterns accurately. To demonstrate the feasibility of this approach, the study utilizes bicycle video data captured from a first-person perspective. This leads to the following research questions:

- How can the swerving pattern be automatically measured in the egocentric videos to accommodate efficiently the constant changes in the scenery?
- How can multiple swerving signals, each derived from different landmarks, be effectively combined to generate a single, cohesive swerving pattern that compensates for potential obstructions and occlusions in video frames?

1.1 Related work

Automated analysis of cyclists' behavior Measuring swerving cycling behavior is critical for assessing cyclists' behavior. Early work in this domain has primarily focused on robustly detecting bicycles in low-resolution traffic videos. For instance, utilizing methods based on multi-feature and multi-frame fusion, to accurately identify bicycles even in challenging video conditions [20], thereby laying the groundwork for subsequent behaviour analysis in traffic surveillance.

Beyond mere detection, continuous research has increasingly turned to modeling the dynamics of cycling behavior to measure important indicators such as swerving. Excessive swerving can be seen as a form of trajectory anomaly, an excessive deviation from the expected path of travel. In this context, recent frameworks like CyclingNet [21] have been introduced to detect cycling near misses from dashcam or bicycle-mounted camera videos. CyclingNet leverages

spatio-temporal deep learning architectures [22] to capture abrupt trajectory anomalies, which often correspond to unsafe maneuvers.

Complementing these approaches, other methods were explored for the broader challenge of outlier detection in video data using advanced deep learning architectures, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs). These methods have been applied to differentiate between normal and abnormal motion patterns in traffic scenarios, where abnormal cycling behaviors, including swerving, often indicate potential near-miss incidents [23]. While these techniques are typically designed for general outlier detection tasks, they offer valuable insights into modeling subtle variations in landmark trajectories, which are essential for accurately identifying cyclist swerving behavior.

Overall, while significant progress has been made in detecting bicycles and general anomalies in video data, specialized automated methods for recognizing the nuanced behavior of bicycle swerving remain an open research area. Robust swerving detection methods that are efficiently generalizable to new scenes have yet to be developed. This gap highlights the need for more advanced and automated solutions to address the challenges of measuring and interpreting cyclist swerving behavior.

Curb detection An underexplored aspect of measuring swerving behavior is linking vehicle movement to a continuous landmark within a video frame, such as the curb alongside the road. While significant research has been conducted on road detection [24, 25] and lane marking identification [26, 27], the specific relationship between swerving behavior and curb detection has received limited attention.

Traditional approaches to road and curb detection often rely on handcrafted features and geometric models. Early methods utilized edge detection, Hough transforms [28], and intensity thresholding [29, 30] to identify curbs and road boundaries. However, these techniques struggle with generalizing varying lighting conditions, occlusions, and complex urban environments. More recent approaches have shifted towards data-driven methods, particularly deep learning-based segmentation models. For instance, CNNs have been employed to segment road surfaces and detect curbs by learning hierarchical features from large-scale datasets [31]. Despite these advancements, curb detection remains challenging due to the subtle visual cues that differentiate curbs from surrounding surfaces [32].

Recent progress in segmentation models has opened new avenues for curb detection. The Segment Anything Model (SAM) [33], a state-of-the-art foundation model for image segmentation, has demonstrated remarkable generalization capabilities across diverse domains. SAM's ability to segment objects based on sparse user inputs or prompts makes it a promising candidate for curb detection tasks. Building on SAM, the Promptable and Efficient Segment Anything Model (PersAM) [34] further enhances segmentation efficiency and accuracy, particularly in scenarios requiring real-time processing. These models could be

adapted to detect curbs by leveraging their ability to generalize across varying road conditions and perspectives.

Signal outlier filtering Outlier detection and removal in signal processing are critical for ensuring data reliability. Traditional approaches often rely on statistical methods (e.g., Z-score, moving average filters) [35] or machine learning models like One-Class SVM [36]. However, these methods struggle with non-stationary signals or high-frequency noise.

Wavelet transforms [37, 38] are particularly effective at capturing localized variations in both time and frequency domains, rendering them highly suitable for analyzing signals characterized by non-stationary behavior. This includes scenarios where transient phenomena (e.g., impulsive spikes or discontinuities) are present. Another method is the Isolation Forest [39, 40], which detects anomalies by exploiting their susceptibility to isolation in feature space. The algorithm constructs an ensemble of random decision trees, recursively partitioning the data by randomly selecting features and split values. Since anomalies require fewer splits to be isolated, their path lengths in the trees are shorter, allowing the algorithm to flag them as outliers.

2 Methodology

In this research, we use PerSAM and extend it to accommodate multiple classes, and fine-tune the extended version using reference images and corresponding landmark class masks. The fine-tuned models then segment landmarks across video frames. In the next step, the pixel distance between each landmark and the bicycle position is calculated. This process enables the visualization of the relative distance of the bicycle to landmarks over the duration of the video. The resulting distance pattern, presented in the form of a signal, can be interpreted as a swerving pattern, revealing insights into the movement dynamics of the bicycle. The methodology incorporates the necessary techniques to acquire these patterns from the egocentric view of the cyclist.

2.1 Dataset

The dataset used in this research consists of videos recorded from the bicycle's point of view, with a camera mounted on the front of the bicycle facing the direction of movement. The dataset includes a total of 54 videos across three subdatasets: A, B, and C (18 videos per subset). All videos were recorded in the city of Groningen, the Netherlands, with each subdataset representing a different cycling environment. Subdataset A contains video data from a bicycle lane integrated within a car road. Subdataset B features recordings from a dedicated two-way bicycle paths, while Subdataset C consists of videos from a dedicated one-way bicycle path. Examples are displayed in Figure 1.

All videos contain various visual obstructions, ranging from light shadows to very heavy shadows, bus stops, and T-split intersections. Additionally, cars



Fig. 1. Sample frames from the A, B, and C subdatasets; subset A features bicycle lanes integrated with car roads, subset B captures a dedicated two-way bicycle path, and subset C includes footage from a one-way bicycle path.

are sometimes parked on bicycle lanes, obstructing landmarks such as curbs and the bicycle lane itself. From time to time, the cyclist may be overtaken by other cyclists or might overtake other cyclists, adding complexity to the overall visibility of landmarks in the scene.

Each video in the dataset, across all the scenarios, is partially annotated by traffic safety domain experts. The annotations, which were carefully labeled by hand across 30,000 frames in total, include the lateral real-world distance between the bicycle and the curb within each video frame.

For one frame per video, we measured the pixel distance of the bicycle to the curb to calculate a scaling factor for converting between the lateral distances in real-world (as the ground truth distances were manually measured) to the pixel values. We refer to this scaling factor as the 'calibration value', calculated per subdataset.

2.2 Landmark segmentation

To segment curbs and determine the swerving pattern, this research employs *Segment Anything* (SAM) [33], a foundation model for image segmentation that generalizes to new objects and scenes without task-specific training. SAM comprises three key components: an image encoder E, a prompt encoder P, and a mask decoder D. These components work together to generate segmentation masks based on an input image and a prompt.

Given an input image I, the image encoder extracts high-dimensional features, represented as:

$$F = E(I), \tag{1}$$

where F is a feature map with dimensions h, w, and feature depth d. The prompt encoder P processes user-provided prompts p, such as points, boxes, or text, to produce an embedding representation:

$$\mathbf{z}_p = P(p), \tag{2}$$

where $\mathbf{z}_p$ is the prompt embedding. The mask decoder D then generates a segmentation mask M using the feature map and prompt embedding:

$$M = D(F, \mathbf{z}_p). \tag{3}$$

PerSAM (Personalized Segment Anything Model) [34] extends SAM by enabling one-shot personalization, allowing automatic segmentation of a user specified landmark in new images or videos without repeated prompting. Given just a single reference image and a mask (either precise or rough), PerSAM adapts SAM to recognize the target object by leveraging feature matching and similarity learning [34].

To further enhance segmentation accuracy across varying object scales, by default, SAM selects the most appropriate scale for the given prompt, where scale refers to the relative size or resolution at which objects are detected in an image, influencing how large or small features are perceived. However, the multi-scale outputs can also be leveraged for refined analysis, such as fusing predictions to enhance segmentation accuracy across varying object sizes. PerSAM introduces PerSAM-Finetune [34], a scale-aware fine-tuning variant that refines mask predictions through a weighted fusion of three-scale masks (M_1, M_2, M_3) generated by SAM. The final mask M is computed as in Equation 4

$$M = w_1 \cdot M_1 + w_2 \cdot M_2 + (1 - w_1 - w_2) \cdot M_3 \tag{4}$$

Where w_1 and w_2 are learnable weights initialized at 1/3 and fine-tuned on the reference mask. While keeping SAM's pretrained weights frozen to preserve its generalization, this lightweight adaptation allows PerSAM-Finetune to dynamically select the optimal segmentation scale for different objects.

We extended perSAM to a multi-class generalization of PerSAM (MCP) through class-specific parameterization of all adaptive components, including both feature similarity computation and multi-scale mask fusion. Equation 5 presents a class-conditioned reformulation of the mask aggregation process, where distinct scale-weighting parameters are learned for each target class $c \in \{1, \ldots, C\}$.

$$M^{(c)} = w_1^{(c)} \cdot M_1^{(c)} + w_2^{(c)} \cdot M_2^{(c)} + (1 - w_1^{(c)} - w_2^{(c)}) \cdot M_3^{(c)} \tag{5}$$

2.3 Swerving analysis pipeline

Figure 2 outlines the automatic measurement process of swerving. Each stage is described in more detail in the following.

(Multi-Class) PerSAM training The training procedure followed the official PerSAM implementation [34], with both single-class and multi-class variants trained for 1000 epochs [34]. The standard PerSAM model ($P1$) used the right curb (and its pixel-level annotation) in the training image for each scene as the reference landmark. This decision was made following the suggestion of the cycling traffic experts. MCP2 uses the indicated bicycle lane as the extra landmark, and MCP3 uses both the indicated bicycle lane and the left dividing marks on the road as the additional landmarks. An example of a segmentation attempt can be seen in Figure 3, where three landmark masks were predicted.

All model variants were trained for each scene with their weights preserved for subsequent experimentation.

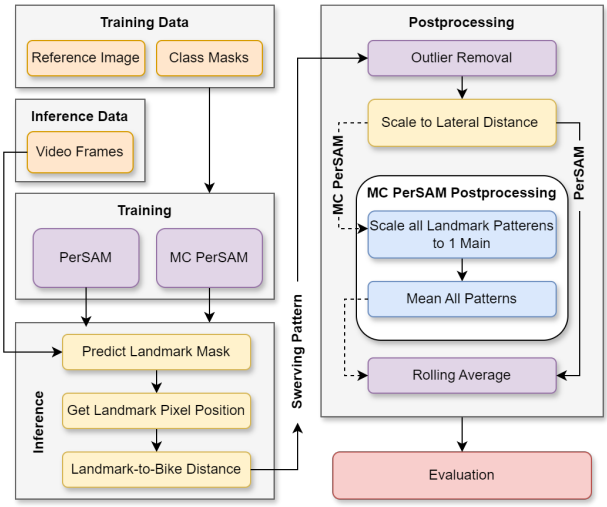

Fig. 2. Schematic representation of the swerving measurement process.

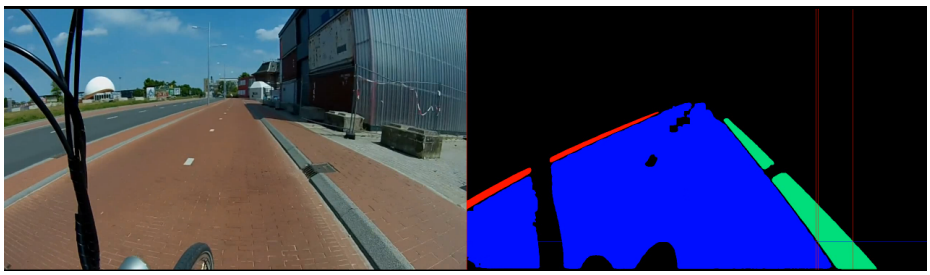

Fig. 3. Example image of the segmentation of multiple landmarks within a frame using Multi Class PerSAM. Left: input frame, Right: predicted landmark masks for the curb, bicycle lane, and the divider.

This research focuses on the application rather than the architectural improvement of PerSAM. Consequently, no segmentation optimization attempts were made, nor were segmentation results evaluated. Referring to the original PerSAM paper, the mean Intersection over Union (mIoU), when evaluated on the PerSEG dataset [34], showed that fine-tuning PerSAM for one class resulted in an mIoU of 95.3 [34], indicating the suitability of this approach. Potential performance enhancements would require either PerSAM architectural modifications or SAM backbone fine-tuning, both beyond the scope of this study.

Landmark position extraction Following the training of segmentation models, all of the frames from the dataset videos were processed using all three model variants - $P1, MCP2$, and $MCP3$ - to predict landmark locations by generating corresponding segmentation masks. To extract the relative position of the bicy-

cle to each landmark, the leftmost side pixel of each landmark at a fixed skyline height is identified. This measured pixel position is then compared against the center of the image (where the camera is located on the bicycle) to convert the pixel values into the pixel distance D between each landmark and the bicycle's position. By calculating the distance D for each frame of the video, a lateral distance signal can be extracted for the entire video, indicating the position of the bicycle with regard to each landmark over time. The distance signal is also referred to as the swerving signal.

Outlier removal In calculating the distance signal, any error in the accurate prediction of the landmarks leads to miscalculating distance, casuing outliers in the swerving signal. To mitigate this effect, the calculated raw swerving signals from different landmarks individually undergo an outlier removal process.

In this work, we used two different approaches for outlier removal. The first approach is the isolation forest algorithm [39, 40] that identifies outliers through recursive partitioning of the feature space, where the anomaly score is derived from a predefined path length, shown in Equation 6.

$$s(x, n) = 2^{-\frac{E(h(x))}{\mu(n)}} \tag{6}$$

The function $s(x, n)$ represents the anomaly score for a data point x given a sample size of n. The term $E(h(x))$ denotes the expected path length for point x across all isolation trees in the forest. The function $\mu(n)$ refers to the average path length of an unsuccessful search in a Binary Search Tree (BST) containing n nodes.

The second approach employs Daubechies wavelets [41], a widely used family of wavelets known for their compact support, orthogonality, and vanishing moments. These properties enable precise time-frequency localization, making Daubechies wavelets particularly effective for tasks such as noise reduction, outlier removal, and multi-scale signal analysis.

Both outlier removal methods employed a 5% threshold criterion. This choice was motivated by the following key considerations: First, under the assumption of Gaussian-distributed noise, approximately 5% of the wavelet coefficients are expected to exceed 2 standard deviations from the mean. Second, this threshold empirically maintains $\geq 95\%$ of legitimate signal components while effectively eliminating significant artifacts.

Signal scaling and alignment After the previous step, for each video, there are three distance signals calculated considering each of the landmarks. Since only the distance of the bicycle to the right curb is manually measured and available as the groundtruth, for evaluation purposes, all these signals need to be aligned with the primary signal identifying the distance of the bicycle to the right curb.

To combine these signals effectively, we propose to first calculate the mean lateral offset between each secondary signal and the primary curb reference.

Each secondary signal is then adjusted by subtracting its characteristic mean offset, aligning all trajectories with the curb-based coordinate system. Finally, the algorithm averages all aligned signals to produce one unified swerving signal that incorporates the information from all detectable landmarks.

Signal smoothing In the last step, a final smoothing is applied to the swerving signal using a rolling average with a window size of 5 frames. This value is selected empirically as a smaller window size might not sufficiently smooth the signal, while a larger one could overly flatten it, eliminating subtle swerving deviations. This effect can be observed in Figure 4. This temporal filtering reduces high-frequency noise while preserving the underlying swerving patterns, resulting in smoother trajectory measurements that better reflect the cyclist's actual lateral movements.

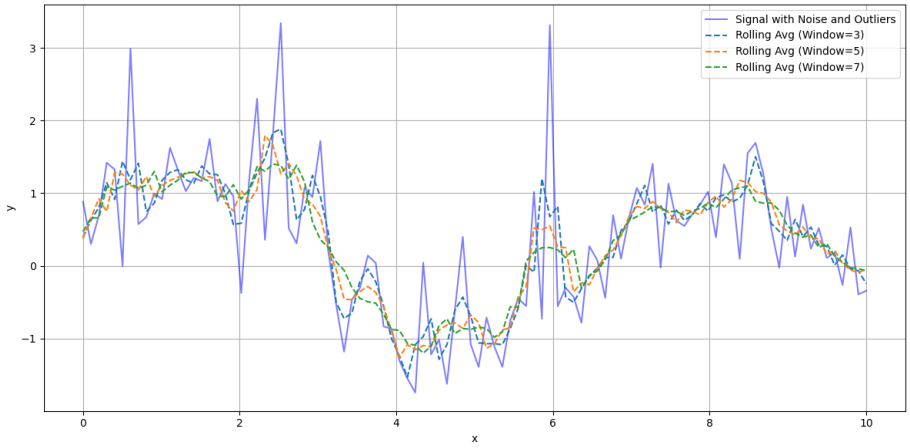

Fig. 4. A signal (solid blue), which includes outliers and noise, is smoothed using rolling averages with varying window sizes: 3 (dotted blue), 5 (dotted orange), and 7 (dotted green).

2.4 Evaluation

To evaluate the performance of the predicted bicycle lateral position trajectory against the annotated ground truth, the following metrics were employed, each providing unique insights into different aspects of the prediction accuracy.

The Root Mean Square Error (RMSE), Equation 7 [42], measures the standard deviation of prediction errors, giving more weight to large errors due to the squaring operation. This is particularly valuable for swerving measurements as it

penalizes significant deviations more heavily. Here, y_i is the annotated swerving distance, $\hat{y}_i$ is the predicted swerving distance, and n is the number of samples.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{7}$$

The Mean Absolute Error (MAE), Equation 8 [43], provides a linear score of average prediction error magnitude, being more robust to outliers than RMSE. This metric gives a straightforward understanding of the typical prediction error in physical units (e.g., lateral distance), which helps assess the practical significance of errors in swerving distance prediction.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{8}$$

The Maximum Absolute Error, Equation 9, identifies the single worst prediction error in a given prediction signal, highlighting local performance extremes.

$$\text{MaxAE} = \max(|y_i - \hat{y}_i|) \tag{9}$$

Spearman's ρ, Equation 10 [44], assesses the relationship between the predicted and the actual swerving movements rank in the same order, where d_i is the difference between ranks of corresponding values. Unlike other metrics, it does not require the relationship to be perfectly linear; it checks if bigger swerves in the annotation generally correspond to bigger swerves in the prediction. Without assuming a linear relationship, Spearman's ρ evaluates whether swerving events in the predictions consistently align with the annotated ground truth, assessing rank-order agreement rather than strict numerical precision.

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{10}$$

3 Experiments, results, and discussion

To systematically evaluate the performance of the proposed swerving detection framework, in this work, we employed a full factorial experimental design examining three critical factors. First, the landmark reference segmentation method was varied between MCP (Multi-Class PerSAM) with 2 and 3 classes, and P (PerSAM) with only one class, with each configuration influencing the bicycle-to-landmark distance estimation in distinct ways. Second, the effect of signal smoothing through application of a rolling average with a 5-frame window size is tested to assess its impact on noise reduction versus real-time responsiveness. Third, three outlier detection approaches were compared: Isolation Forest with a contamination parameter of 0.05 for model-based outlier detection, a Wavelet Transform with a threshold of 0.05 for frequency-domain filtering, and a baseline configuration without any outlier filtering.

Each configuration was evaluated across three distinct subdatasets (A, B, and C) representing diverse urban cycling scenarios that varied in obstacle density, lighting conditions, and traffic patterns. Performance was assessed using four quantitative metrics: MAE, measuring the average deviation in centimeters from ground truth annotations. RMSE, emphasizing larger deviations. Max AE, identifying worst-case detection failures, and Spearman's rank correlation coefficient, evaluating trend agreement between detected and actual swerving patterns.

3.1 Quantitative results

The results of the experiments for each subset of the dataset is outline in Table 1, Table 2, and Table 3. The average results over all three subsets is is presented in Table 4, providing an overall assessment of the performance of swerving measurement under different parameterizations.

Table 1. Performance metrics across method configurations (Dataset A)

Method	Rolling Avg.	Outlier Rem.	MAE ↓	RMSE ↓	Max AE ↓	Spearman's ρ ↑
MCP2	No	Isolation Forest	3.60	5.27	21.53	0.91
MCP2	No	None	4.69	8.81	64.38	0.90
MCP2	No	Wavelet	4.33	7.22	51.18	0.91
MCP2	Yes	Isolation Forest	3.38	4.78	18.41	0.92
MCP2	Yes	None	4.44	6.95	37.92	0.89
MCP2	Yes	Wavelet	4.25	6.80	36.25	0.91
MCP3	No	Isolation Forest	5.52	7.72	42.26	0.84
MCP3	No	None	6.19	9.86	64.89	0.82
MCP3	No	Wavelet	5.98	9.01	55.61	0.83
MCP3	Yes	Isolation Forest	4.43	6.24	27.04	0.84
MCP3	Yes	None	4.81	7.00	34.16	0.82
MCP3	Yes	Wavelet	4.72	6.93	35.55	0.83
P1	No	Isolation Forest	3.29	4.97	26.27	0.91
P1	No	None	4.26	9.29	90.07	0.89
P1	No	Wavelet	4.18	8.87	78.62	0.89
P1	Yes	Isolation Forest	2.90	4.17	19.10	0.93
P1	Yes	None	3.65	6.63	45.18	0.89
P1	Yes	Wavelet	3.64	6.55	43.79	0.90

Across all datasets, the MCP2 method with rolling average and Isolation Forest outlier removal consistently achieved the lowest errors, with MAE of 2.68 cm, RMSE of 3.92 cm, and Max AE of 19.43 cm. This configuration also exhibited the highest Spearman's rank correlation coefficient ($\rho = 0.93$), indicating a strong correlation between predicted and observed values. The inclusion of rolling averaging generally improved performance, reducing MAE compared to

Table 2. Performance metrics across method configurations (Dataset B)

Method	Rolling Avg.	Outlier Rem.	MAE ↓	RMSE ↓	Max AE ↓	Spearman's ρ ↑
MCP2	No	Isolation Forest	2.19	3.54	20.06	0.94
MCP2	No	None	2.89	6.97	78.22	0.93
MCP2	No	Wavelet	2.48	5.61	56.07	0.94
MCP2	Yes	Isolation Forest	2.16	3.27	16.59	0.94
MCP2	Yes	None	2.65	4.55	32.40	0.92
MCP2	Yes	Wavelet	2.41	4.19	26.05	0.93
MCP3	No	Isolation Forest	2.43	4.71	32.46	0.93
MCP3	No	None	3.35	8.18	76.76	0.91
MCP3	No	Wavelet	2.86	6.82	58.69	0.92
MCP3	Yes	Isolation Forest	2.74	5.02	26.32	0.92
MCP3	Yes	None	3.63	7.04	43.41	0.88
MCP3	Yes	Wavelet	3.23	6.43	40.10	0.90
P1	No	Isolation Forest	3.34	6.54	35.15	0.94
P1	No	None	6.07	15.95	133.72	0.94
P1	No	Wavelet	6.07	15.95	133.72	0.94
P1	Yes	Isolation Forest	3.35	6.09	30.18	0.93
P1	Yes	None	5.49	10.39	64.02	0.91
P1	Yes	Wavelet	5.49	10.39	64.02	0.91

Table 3. Performance metrics across method configurations (Dataset C)

Method	Rolling Avg.	Outlier Rem.	MAE ↓	RMSE ↓	Max AE ↓	Spearman's ρ ↑
MCP2	No	Isolation Forest	2.92	4.55	29.09	0.92
MCP2	No	None	5.25	12.00	92.30	0.90
MCP2	No	Wavelet	3.49	6.72	55.80	0.92
MCP2	Yes	Isolation Forest	2.51	3.71	23.29	0.94
MCP2	Yes	None	4.50	7.39	42.74	0.88
MCP2	Yes	Wavelet	3.30	6.40	43.69	0.92
MCP3	No	Isolation Forest	4.47	8.98	71.05	0.88
MCP3	No	None	6.24	14.09	108.85	0.86
MCP3	No	Wavelet	3.97	8.87	87.16	0.90
MCP3	Yes	Isolation Forest	4.01	7.40	50.79	0.85
MCP3	Yes	None	5.50	11.49	86.19	0.81
MCP3	Yes	Wavelet	4.28	9.62	71.09	0.87
P1	No	Isolation Forest	4.25	7.65	39.65	0.88
P1	No	None	9.67	24.89	176.68	0.86
P1	No	Wavelet	7.34	20.20	159.67	0.87
P1	Yes	Isolation Forest	3.77	6.00	34.59	0.90
P1	Yes	None	8.53	15.68	106.47	0.83
P1	Yes	Wavelet	6.89	14.78	107.61	0.86

Table 4. Performance metrics across method configurations across all datasets

Method	Rolling Avg.	Outlier Rem.	MAE ↓	RMSE ↓	Max AE ↓	Spearman's ρ ↑
MCP2	No	Isolation Forest	2.90	4.45	23.56	0.92
MCP2	No	None	4.28	9.26	78.30	0.91
MCP2	No	Wavelet	3.43	6.51	54.35	0.92
MCP2	**Yes**	**Isolation Forest**	**2.68**	**3.92**	**19.43**	**0.93**
MCP2	Yes	None	3.86	6.30	37.69	0.90
MCP2	Yes	Wavelet	3.32	5.80	35.33	0.92
MCP3	No	Isolation Forest	4.14	7.14	48.59	0.88
MCP3	No	None	5.26	10.71	83.50	0.86
MCP3	No	Wavelet	4.27	8.23	67.15	0.88
MCP3	Yes	Isolation Forest	3.72	6.22	34.72	0.87
MCP3	Yes	None	4.65	8.51	54.59	0.84
MCP3	Yes	Wavelet	4.08	7.66	48.91	0.87
P1	No	Isolation Forest	3.63	6.38	33.69	0.91
P1	No	None	6.67	16.71	133.49	0.89
P1	No	Wavelet	5.86	15.01	124.00	0.90
P1	Yes	Isolation Forest	3.34	5.42	27.95	0.92
P1	Yes	None	5.89	10.90	71.89	0.88
P1	Yes	Wavelet	5.34	10.57	71.80	0.89

non-rolling configurations. Isolation Forest outlier detection outperformed both wavelet-based and no outlier detection approaches.

Subdataset-specific analyses revealed notable variations in performance. Dataset B exhibited the lowest errors overall, with MCP2 achieving an MAE of 2.16 cm when using rolling averages and Isolation Forest. In contrast, Dataset C proved more challenging, with higher errors across all configurations, particularly for the P1 method without outlier removal (MAE = 9.67 cm, RMSE = 24.89 cm). The MCP3 method demonstrated intermediate performance, with MAE values ranging from 3.72 to 6.24 cm, suggesting that increasing the model complexity beyond 2 classes did not yield consistent improvements.

The wavelet-based outlier removal method showed mixed results, occasionally improving upon no outlier removal but generally underperforming compared to Isolation Forest. For instance, in Dataset A, wavelet denoising reduced MAE by 7.6% compared to no outlier removal but remained 15.1% higher than Isolation Forest. The maximum error metrics further highlighted the robustness of Isolation Forest, which consistently mitigated extreme prediction errors.

These findings suggest that the combination of MCP2 with rolling averaging and Isolation Forest outlier removal provides the most reliable and accurate predictions across diverse datasets. The consistent performance gains from rolling averaging indicate its utility in smoothing temporal fluctuations, while the superior outlier removal capability of Isolation Forest underscores its importance in handling noisy real-world data.

3.2 Qualitative analysis

The enhanced performance of MCP2 over P1 is exemplified in a video from scene C, as illustrated in Figure 5. Utilizing the optimal configuration of signal post-processing, which includes a rolling average and an isolation forest, it is evident that the right graph depicting the swerving pattern of MCP2 exhibits reduced noise overall compared to when only one landmark is used. This highlights the potential of using more than one reference landmark for automatic swerving measurements.

Additionally, in the same figure, between frames 400 and 500, the reference landmark, the curb, is occasionally obstructed when using PerSAM, resulting in gaps in the prediction. However, with MCP2, other signals are available for the prediction to rely on, leading to a complete swerving prediction without gaps.

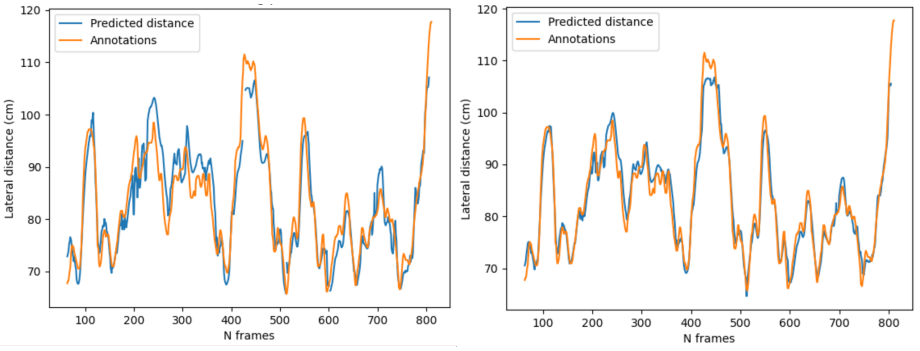

Fig. 5. Blue shows the calculated lateral distance values after postprocessing, and orange is the manually calculated distance measurements. Left: PerSAM trained on only one class. Right: Multi-Class PerSAM with 2 classes (right curb and the bicycle lane).

Another example is shown in Figure 6, where MCP2 (right), although not being closely aligned with the annotations, is still outperforming P1 (left). For this scenario, upon examining the raw segmentation masks of P1, as shown in Figure 7, it is evident that there are heavy shadows in the curb segmentation area due to the unfortunate placement of the cyclist's shadow. This results in mask predictions with a high error rate.

Similarly, the same issue is observed in Figure 8 with MCP3, where the heavy shadow of the cyclist overlaps both the segmentation areas of the curb and the bicycle lane. Even the addition of an extra class, is of no help in this example, as the signal and image are blocked by a bicycle lamp at the location where the third landmark would be.

These examples indicate the importance of choosing the right reference landmarks when conducting segmentation and measurement tasks. Selecting landmarks that are less prone to obstructions and shadows can significantly improve

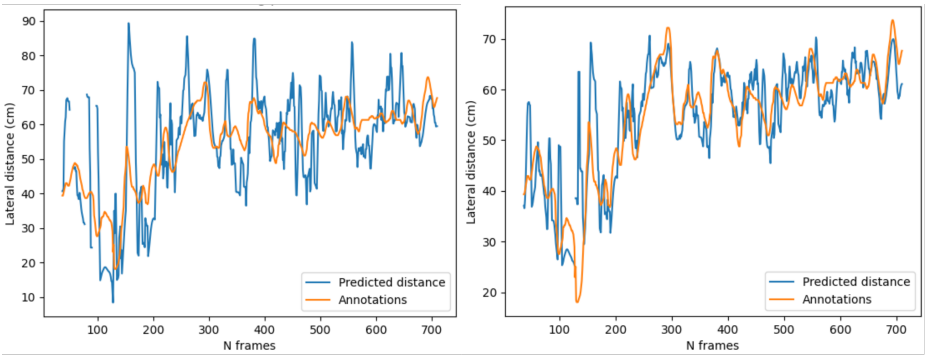

Fig. 6. Blue shows the calculated lateral distance values after postprocessing, and orange is the manually calculated distance measurements. Left: PerSAM trained on only one class. Right: Multi-Class PerSAM with 2 classes (right curb and the bicycle lane).

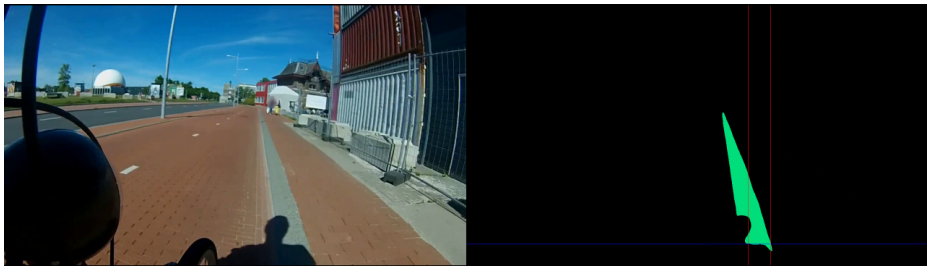

Fig. 7. PerSAM trained on only one class, showing heavy shadows affecting the curb segmentation area, and therefore the lateral distance calculation.

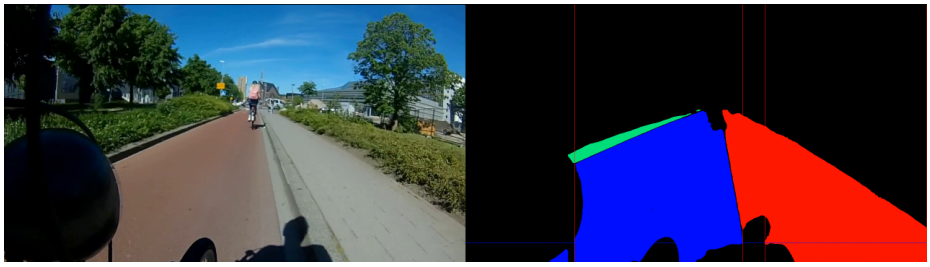

Fig. 8. Segmentation masks of Multi-Class PerSAM with 3 classes showing heavy shadows affecting the curb, bicycle lane, and the left divider segmentation areas.

the accuracy of the measurements. While MCP2 offers improvements over P1, attention to setup details and the choice of reference landmarks remain key factors in maximizing the effectiveness of the proposed model.

4 Conclusion

In this study, the SAM model was fine-tuned using PerSAM with a reference image per scenario and the landmark class masks to segment landmarks in video frames. By measuring the distance between landmarks and the bicycle, the bicycle's swerving pattern was visualized. Experiments used three segmentation variants for bicycle swerving analysis: the baseline PerSAM model, Multi-Class PerSAM (MCP) for 2-class segmentation, and MCP for 3-class segmentation. In addition, outlier removal techniques and signal smoothing were applied to reduce errors throughout the videos.

By segmenting landmarks per frame and calculating the pixel distance from the curb to the bicycle, the bicycle's lateral distance can be estimated. Converting pixel distance to lateral distance using calibration parameters and aggregating all landmark distances yields a swerving pattern of the cyclists. Overall results suggest that two landmarks are optimal for measuring swerving motion. Additional landmarks can introduce noise, and adding a third landmark degrades signal performance.

To optimally combine multiple swerving signals, we first used Isolation Forest with a 5% outlier removal rate, then scaled all signals to a chosen reference signal (right curb). Next, we averaged all the scaled signals at each timestamp, and lastly, applied a rolling average with a window of 5 to smooth the signals. Using Isolation Forest reduces the MAE by over 3 cm, while the wavelet-based method improves it by about 1 cm. The rolling average does only have a noticeable affect on Max AE.

Overall, while absolute performance varies across datasets, the relative benefits of specific configurations remain consistent. Combining multi-class perSAM using two landmarks, signal smoothing, and Isolation Forest-based outlier removal proves to be the most reliable approach.

5 Future work and broader impact

This study offers insights into automatic cyclists' swerving pattern analysis through one-shot landmark segmentation and signal processing techniques. Several promising directions remain for future research. First, conducting an additional analysis over landmarks could help identify specific conditions or environments where each landmark enhances swerving measurement accuracy without adding noise. Primary landmarks such as lane markings or curbs can sometimes be obstructed by parked vehicles, pedestrians, or debris. In such cases, dynamically selecting landmarks based on real-time visibility, environmental context, or cyclist behavior could improve the robustness.

Second, broader application and validation of the proposed methodologies across a wider range of scenarios, such as urban bicycle lanes, high-traffic roads, and shared pedestrian-cyclist spaces, would enhance their generalizability and practical relevance. The current analysis is constrained to three scenes within a single urban setting, which may not reflect the full variability of real-world conditions, including differences in infrastructure, traffic density, and cyclist behavior across cities.

In regulated monitoring environments, safeguarding individual privacy must be a top priority. Those monitored cyclists should be informed about what data is being collected and how it will be used. At the same time, the system may inadvertently capture information about third parties, such as bystanders, vehicle license plates, or other identifiable details. To address these concerns, strong anonymization methods should be implemented to protect all individuals' identities. Additionally, data must be securely stored to prevent unauthorized access. Ensuring transparency in the monitoring process and obtaining informed consent from primary participants are essential components of ethical data collection. To conduct this study, all cyclists first provided informed consent. Video data is anonymized and kept securely in internal protected servers.

References

1. Paul Schepers, Divera Twisk, Elliot Fishman, Aslak Fyhri, and Anne Jensen. The dutch road to a high level of cycling safety. *Safety science*, 92:264–273, 2017.
2. Paul Schepers, Marco Helbich, Marjan Hagenzieker, Bas de Geus, Marco Dozza, Niels Agerholm, Anna Niska, Noora Airaksinen, Francis Papon, Regine Gerike, et al. The development of cycling in european countries since 1990. *European Journal of Transport and Infrastructure Research*, 21(2):41–70, 2021.
3. Sonja Haustein and Mette Møller. E-bike safety: Individual-level factors and incident characteristics. *Journal of Transport & Health*, 3(3):386–394, 2016.
4. Michael Branion-Calles, Trisalyn Nelson, and Meghan Winters. Comparing crowdsourced near-miss and collision cycling data and official bike safety reporting. *Transportation research record*, 2662(1):1–11, 2017.
5. Mohamed Bayoumi Kamel and Tarek Sayed. The impact of bike network indicators on bike kilometers traveled and bike safety: A network theory approach. *Environment and Planning B: Urban Analytics and City Science*, 48(7):2055–2072, 2021.
6. Masayuki Miyadai, Teruo Uetake, and Masahiro Shimoda. How does a cyclist avoid obstacles? *Journal of human ergology*, 41(1_2):95–100, 2012.
7. ZHAO Yuqing. Analysis of driver and cyclist responses in car-to-cyclist conflicts.
8. Frank Westerhuis. Advancing the age of cycling. 2021.
9. Zaili Yang, Zhisen Yang, John Smith, and Bostock Adam Peter Robert. Risk analysis of bicycle accidents: A bayesian approach. *Reliability Engineering & System Safety*, 209:107460, 2021.
10. Paul M Salmon, Mitch Naughton, Adam Hulme, and Scott McLean. Bicycle crash contributory factors: A systematic review. *Safety science*, 145:105511, 2022.
11. Antonella Scarano, Massimo Aria, Filomena Mauriello, Maria Rella Riccardi, and Alfonso Montella. Systematic literature review of 10 years of cyclist safety research. *Accident Analysis & Prevention*, 184:106996, 2023.

12. Adnan Shaout, Dominic Colella, and Selim Awad. Advanced driver assistance systems-past, present and future. In *2011 Seventh International Computer Engineering Conference (ICENCO'2011)*, pages 72–82. IEEE, 2011.
13. Vipin Kumar Kukkala, Jordan Tunnell, Sudeep Pasricha, and Thomas Bradley. Advanced driver-assistance systems: A path toward autonomous vehicles. *IEEE Consumer Electronics Magazine*, 7(5):18–25, 2018.
14. Frank Westerhuis and Dick De Waard. Using commercial gps action cameras for gathering naturalistic cycling data. *Journal of the Society of Instrument and Control Engineers*, 55(5):422–430, 2016.
15. Ryan Toxopeus, Shady Attalla, Sam Kodsi, and Michele Oliver. Driver response time to cyclist path intrusions. Technical report, SAE Technical Paper, 2018.
16. Yan Ke. *A Cyclist Detection and Tracking System for Heavy Goods Vehicles*. PhD thesis, 2022.
17. Ryan De Iaco, Stephen L Smith, and Krzysztof Czarnecki. Safe swerve maneuvers for autonomous driving. In *2020 IEEE Intelligent Vehicles Symposium (IV)*, pages 1941–1948. IEEE, 2020.
18. Ryan De Iaco, Stephen L Smith, and Krzysztof Czarnecki. Universally safe swerve maneuvers for autonomous driving. *IEEE Open Journal of Intelligent Transportation Systems*, 2:482–494, 2021.
19. Yixin Zhang, Nicholas Konz, Kevin Kramer, and Maciej A. Mazurowski. Quantifying the limits of segmentation foundation models: Modeling challenges in segmenting tree-like and low-contrast objects, 2025.
20. Yicheng Zhang and Qiang Ling. Bicycle detection based on multi-feature and multi-frame fusion in low-resolution traffic videos, 2017.
21. Mohamed R Ibrahim, James Haworth, Nicola Christie, and Tao Cheng. Cyclingnet: Detecting cycling near misses from video streams in complex urban scenes with deep learning. *IET Intelligent Transport Systems*, 15(10):1331–1344, 2021.
22. Matthew J Williams and Mirco Musolesi. Spatio-temporal networks: reachability, centrality and robustness. *Royal Society open science*, 3(6):160196, 2016.
23. Shumayla Yaqoob, Salvatore Cafiso, Giacomo Morabito, and Giuseppina Pappalardo. Detection of anomalies in cycling behavior with convolutional neural network and deep learning. *European transport research review*, 15(1):9, 2023.
24. Hui Kong, Jean-Yves Audibert, and Jean Ponce. General road detection from a single image. *IEEE Transactions on Image Processing*, 19(8):2211–2220, 2010.
25. Jannik Fritsch, Tobias Kuehnl, and Andreas Geiger. A new performance measure and evaluation benchmark for road detection algorithms. In *16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013)*, pages 1693–1700. IEEE, 2013.
26. Youcheng Zhang, Zongqing Lu, Xuechen Zhang, Jing-Hao Xue, and Qingmin Liao. Deep learning in lane marking detection: A survey. *IEEE Transactions on Intelligent Transportation Systems*, 23(7):5976–5992, 2021.
27. Mauricio Braga De Paula and Claudio Rosito Jung. Real-time detection and classification of road lane markings. In *2013 XXVI conference on graphics, patterns and images*, pages 83–90. IEEE, 2013.
28. Weifeng Liu, Zhenqing Zhang, Shuying Li, and Dapeng Tao. Road detection by using a generalized hough transform. *Remote Sensing*, 9(6):590, 2017.
29. José M Álvarez Alvarez and Antonio M Ĺopez. Road detection based on illuminant invariance. *IEEE transactions on intelligent transportation systems*, 12(1):184–193, 2010.

30. Alberto Hata and Denis Wolf. Road marking detection using lidar reflective intensity data and its application to vehicle localization. In *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, pages 584–589. IEEE, 2014.
31. Min Zou and Yoichi Kageyama. Road curb detection based on a deep learning framework. In *2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC)*, pages 0259–0262. IEEE, 2023.
32. Lucero M Romero, Jose A Guerrero, and Gerardo Romero. Road curb detection: a historical survey. *Sensors*, 21(21):6952, 2021.
33. Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026, 2023.
34. Renrui Zhang, Zhengkai Jiang, Ziyu Guo, Shilin Yan, Junting Pan, Xianzheng Ma, Hao Dong, Peng Gao, and Hongsheng Li. Personalize segment anything model with one shot. *arXiv preprint arXiv:2305.03048*, 2023.
35. Joseph Ndong and Kavé Salamatian. Signal processing-based anomaly detection techniques: a comparative analysis. In *Proc. 2011 3rd International Conference on Evolving Internet*, pages 32–39, 2011.
36. Yuting Chen, Jing Qian, and Venkatesh Saligrama. A new one-class svm for anomaly detection. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3567–3571. IEEE, 2013.
37. Amirreza Moradi and Sajjad Asiaei Mojarad. A modified wavelet-based method for detection of outliers in time series. *Earth Observation and Geomatics Engineering*, 3(1):77–83, 2019.
38. Stéphane Mallat. *A wavelet tour of signal processing*. Elsevier, 1999.
39. Dong Xu, Yanjun Wang, Yulong Meng, and Ziying Zhang. An improved data anomaly detection method based on isolation forest. In *2017 10th international symposium on computational intelligence and design (ISCID)*, volume 2, pages 287–291. IEEE, 2017.
40. Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. Isolation forest. In *2008 eighth ieee international conference on data mining*, pages 413–422. IEEE, 2008.
41. Ahmad MH Al-Khazaleh, S Al Wadi, and Faisal Ababneh. Wavelet transform asymmetric winsorized mean in detecting outlier values. *Far East Journal of Mathematical Sciences*, 96(3):339–351, 2015.
42. Cort J Willmott and Kenji Matsuura. Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. *Climate research*, 30(1):79–82, 2005.
43. Cort J Willmott. On the evaluation of model performance in physical geography. *Spatial statistics and models*, pages 443–460, 1984.
44. Charles Spearman. The proof and measurement of association between two things. *The American journal of psychology*, 15(1):72–101, 1904.