Safe Route Estimation from Bicycle-Mounted Camera in Pedestrian-Active Environments

Ryoichi Kimura[a], Chokiu Leung[b], Tiago Rodrigues[c], Hitesh Pandya[c], Bruno Coelho[c], and Yoshinari Kameda[d]

[a]Master's Program in Intelligent and Mechanical Interaction Systems, University of Tsukuba,1-1-1 Tennodai, Tsukuba, Ibaraki, Japan
[b]Institute of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai Tsukuba, Ibaraki, Japan
[c]Engineering and R&D Services, Capgemini Japan K.K, 1-23-1 Toranomon, Minato City, Tokyo, Japan
[d]Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba,Ibaraki, Japan

ABSTRACT

We propose a method for estimating a safe cycling route in pedestrian-active environments from forward-facing bicycle-mounted camera footage. In environments with pedestrians and obstacles, deriving a highly safe route requires not only object detection but also region evaluation based on risk levels. The proposed method integrates information such as pedestrian positions and movement directions, obstacle locations, and road surface types to generate a "risk potential map" that represents risk levels and uses it to suggest a safe route. A video dataset was constructed using a virtual environment that simulates real-world conditions, and a Pix2Pix-based image generation model was demonstrated to enable real-time image generation.

Keywords: Route Estimation, Bicycle, ITS, Risk Assessment, Pedestrian Detection, Cycling Safety

1. INTRODUCTION

In recent years, bicycle use has increased, and bicycles have gained renewed attention as a mode of transportation. At the same time, the proportion of bicycle-related accidents is rising,[1] making effective safety measures essential. Among these, collisions with pedestrians are particularly critical in sidewalks and shared spaces, where pedestrians constantly change their positions, orientations, and walking directions. Cyclists must perceive these movements in real time, yet they ride in proximity to pedestrians and have limited braking capability, making delayed perception or misjudgment directly lead to collision risk.

Despite these hazards, safe cycling still relies heavily on the rider's perceptual abilities and experience. This motivates the need for technology that objectively assesses the surrounding environment and suggests a safe route. A forward-facing monocular camera provides an objective view aligned with the cyclist's perspective, enabling continuous observation of pedestrians and road conditions and supporting hazard recognition and route guidance.

Although previous studies on bicycle safety have primarily focused on detecting or warning about hazardous situations, they do not provide actionable guidance on how cyclists should adjust their riding trajectories. In contrast, the automotive field has made substantial progress in trajectory estimation. However, bicycles lack lane constraints and operate in proximity to pedestrians, making car-oriented approaches difficult to apply. Furthermore, although there has been research on risk quantification and spatial representation based on environmental perception, these methods are not specifically tailored to bicycles, and there remains a need to establish bicycle-specific approaches.

Further author information: (Send correspondence to Ryoichi Kimura)
Ryoichi Kimura: E-mail: kimura.ryoichi@image.iit.tsukuba.ac.jp
Yoshinari Kameda: E-mail: kameda@ccs.tsukuba.ac.jp

International Workshop on Advanced Imaging Technology (IWAIT) 2026,
edited by Masayuki Nakajima, et al., Proc. of SPIE Vol. 14072
 · 0277-786X · doi: 10.1117/12.3102250

In this study, we propose a method for estimating a safe bicycle route in pedestrian-active environments using only forward-facing monocular images. Based on pedestrian information extracted from the camera, we generate a risk potential map that spatially quantifies collision risk. Using this map as input, we construct a framework that generates a safe route by jointly considering the rider's intended direction and collision avoidance.

The contributions of this study are as follows:

- We propose a risk potential map generation method that dynamically estimates collision risk from pedestrian positions, orientations, and distances, and represents it as a continuous 2D map.
- We develop a learning-based route estimation model that generates a safe bicycle path by integrating the cyclist's intended direction with the estimated risk distribution.
- We design a unified framework that visualizes both the risk distribution and the generated route within the same image-coordinate space.
- We propose a bicycle-specific route estimation framework that produces a single continuous trajectory from the risk distribution, reflecting bicycles' high maneuverability and lane-free characteristics.

2. RELATED WORK

Risk detection for bicycles has been investigated in several studies.[2] Ibrahim et al. proposed CyclingNet, which detects near-miss events using forward-facing bicycle cameras.[3] While effective for alerting cyclists, these methods only detect hazardous situations without suggesting how cyclists should adjust their riding trajectory.

Trajectory estimation has been actively investigated in the automotive domain.[4] Hu et al. proposed UniAD, an autonomous driving framework that integrates perception, prediction, and planning.[5] However, bicycles operate with higher maneuverability and without strict lane constraints, often in proximity to pedestrians. Thus, car-oriented approaches cannot be directly applied to bicycles.

Spatial hazard representation using risk maps has also been explored.[6] Kozuka et al. produced pixel-wise risk maps estimating potential pedestrian intrusion zones.[7] Although these studies focus on spatial risk visualization, they do not provide risk map representations tailored to bicycle navigation.

Therefore, an integrated framework that unifies risk assessment and safe route generation is required.

3. ROUTE ESTIMATION

System Overview: Figure 1 shows the processing pipeline of the proposed method. Based on pedestrian information obtained from the forward-facing camera, the system estimates hazardous regions and generates a risk potential map $M_{i\{u,v\}}$ (G1), where (u, v) denote pixel coordinates in the image plane. The map is then used to generate a continuous safe route (G2) that avoids pedestrian collisions while following the cyclist's intended direction.

Risk Potential Map: We first define the ground-plane risk potential function $R_g(x, y) \in [0, 1]$, which assigns a collision risk value at each ground coordinate (x, y) based on pedestrian position, orientation, walking direction, and distance relative to the cyclist. Here, (x, y) denotes the ground-plane coordinates in the Unity environment. This ground-plane risk distribution is then projected onto the forward-facing camera image to obtain an image-plane risk potential map $M_{i\{u,v\}}$, which provides a risk value at each image coordinate (u, v).

In this study, hazardous regions are computed in the Unity ground coordinate system and subsequently projected onto the camera images to construct ground-truth risk map labels. This projection preserves spatial consistency between the ground and image coordinates, enabling effective learning of risk in the image domain.

The proposed map generation is designed to satisfy four requirements: (i) spatial consistency across ground and image coordinates, (ii) integration of pedestrian position, orientation, walking direction, and distance, (iii) continuous representation as a scalar field, and (iv) dynamic updates with pedestrian motion.

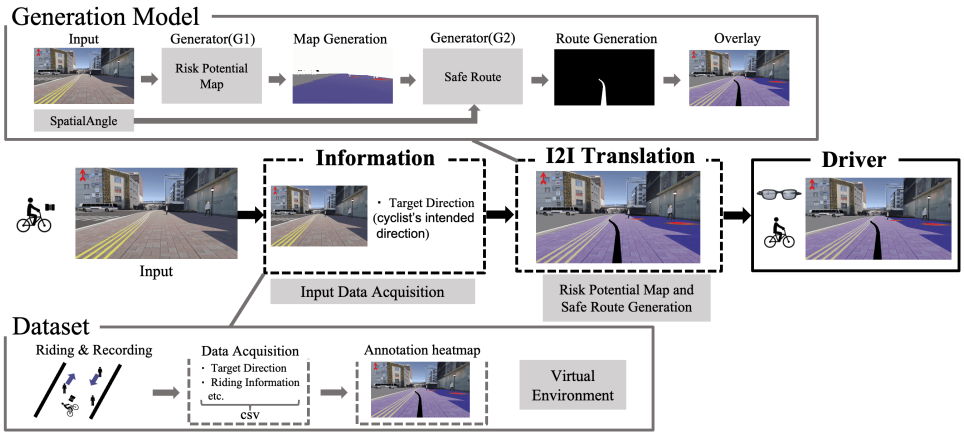

Figure 1. Overview of the safe route estimation framework.

For generating the image-plane risk potential map, we adopt a conditional image-to-image translation model, where the RGB image is the input and the projected risk distribution $M_{i\{u,v\}}$ is the output. The resulting map serves as the spatial foundation for subsequent safe route generation.

Safe Route: The safe route is designed to meet four requirements: (i) consistency with the cyclist's intended direction, (ii) avoidance of high-risk regions, (iii) confinement within feasible walking/cycling areas, and (iv) representation as a single continuous curve.

We employ a conditional image-to-image translation model that takes the risk potential map $M_{i\{u,v\}}$ as input and outputs a continuous route as an image. This enables unified execution of risk estimation and route generation from a single forward-facing camera, providing intuitive and interpretable safe route suggestions even in pedestrian-active environments.

4. IMPLEMENTATION

Cycling Environment: As the cycling environment, we use a virtual model of the Tsukuba Station area constructed in Unity. We reproduce mixed traffic by placing pedestrian and bicycle agents in the scene.

A bicycle agent is driven along a predesigned safe route in the Unity environment, and the resulting trajectory is recorded as ground-truth safe route. At the same time, the ground-plane risk potential function $R_g(x, y)$ is computed in the ground coordinate system based on pedestrian information and road-surface attributes in the virtual scene. In this study, the risk value $R_g(x, y)$ is defined as

$$R_g(x,y) = \min\left(1,\ \sum_{p\in\mathcal{P}}\left[\min\left(1,\ w_s \cdot A_{\varphi}^{(p)}(x,y)\cdot G_{\mathrm{dist}}^{(p)}(x,y)\cdot M_{\mathrm{cone}}^{(p)}(x,y)\cdot M_{\mathrm{vis}}^{(p)}(x,y)\right)\right]^{\beta_r}\right) \tag{1}$$

where $\mathcal{P}$ denotes the set of all pedestrian agents in the scene and $p \in \mathcal{P}$ indexes each pedestrian. A_{φ}, G_{dist}, M_{cone} and M_{vis} are weighting terms based on pedestrian orientation, distance attenuation, collision-relevant directional cone, and visibility, respectively. w_s and β_r are scalar weight parameters that control the overall magnitude and sharpness of the risk contribution.

Ground-truth image-plane risk potential map $M_{i\{u,v\}}$ is obtained by projecting the ground-plane risk potential function $R_g(x, y)$ onto the camera image and applying a traversable-region mask. Training used 2406 risk potential map images and 1496 route images, while 182 paired images were used for testing.

Risk Potential Map and Safe Route Generation: The ground-plane risk potential function $R_g(x, y)$ evaluates pedestrian-induced collision risk within the bicycle's frontal 30 m region in the ground coordinate system. Sidewalks are treated as traversable areas, whereas roads and buildings are treated as non-traversable. Street trees and other obstacles are indirectly handled as regions that are not drawn in the risk map and are therefore non-traversable for the bicycle.

For pedestrians, a 2.5 m region in front of each pedestrian is defined as a high-risk zone to emphasize collision risk when the bicycle approaches. The bicycle is assumed to travel at a constant speed of 3 m/s, and the route endpoint is set 10 m ahead, considering the stopping distance. This formulation allows the system to estimate a route that balances safety and consistency with the rider's intended direction, even in crowded environments.

Generative Model

Risk Potential Map Generation Model: To generate a risk potential map from forward-facing camera images, we adopt a Pix2Pix[8] -based conditional image-to-image translation model with several modifications. Accurate risk estimation requires inferring pedestrian position, orientation, and relative distance from the input image and expressing these cues as a spatial risk distribution on the image plane. Therefore, we employ a ResNet-based generator, which is effective for capturing wide-range contextual information and learning stable relative risk relationships between the bicycle and multiple pedestrians.

In addition, a self-attention mechanism is introduced to enhance risk estimation by modeling interactions among pedestrians. To emphasize hazardous areas during training, we incorporate Red-L1 and Red-Dice losses, encouraging accurate reconstruction of the shape and connectivity of high-risk regions. For the adversarial component, we use LSGAN, updating the generator and discriminator with a ratio of $G : D = 2 : 1$ to achieve stable training.

Safe Route Generation Model: For safe route generation, we use a Pix2Pix-based conditional generative model. The input consists of the image-plane risk potential map and a spatial angle condition map encoding the cyclist's intended heading. Using only the risk map makes the intended direction ambiguous, so the target heading θ is explicitly embedded through a spatial condition map. The spatial angle condition map is defined as

$$C(x_s;\theta) = \mathrm{clip}\left(\alpha_c(a + b\frac{\theta}{\theta_0})(c + dx_s) + \beta_c,\ L,\ U\right), \tag{2}$$

where $x_s \in [0, 1]$ is a normalized spatial horizontal coordinate used solely for generating the condition map. Parameters a, b, c, d control spatial and angle scaling, and α_c, β_c are normalization parameters. We set $a = 10,\ b = 40,\ c = 0.5,\ d = 0.5,\ \alpha_c = \frac{1}{80},\ \beta_c = -0.125,\ \theta_0 = 90$, and clip the output to the range defined by $L = -1$ and $U = 1$. By encoding the target heading as a spatial distribution, the model is given a direction-aware weighting over the feasible region, enabling the generation of a route aligned with the intended direction.

For training, we combine edge loss and route-L1 loss, a pixel-wise L1 loss computed only on the route pixels, to enforce clear route shapes and continuity. This configuration enables unified risk estimation and a safe route generation from a single monocular image.

5. RESULTS

Risk Potential Map Generation: We first evaluated risk potential map generation performance. The baseline model, UNet-Std, uses a UNet generator based on the original Pix2Pix configuration. For comparison, we prepared three variants: UNet-RedLoss, which adds loss terms emphasizing hazardous regions; ResNet-RedLoss, which replaces the UNet generator with a ResNet while using the same extended losses; and Proposed (ResNet-SA-RedLoss), which further introduces a self-attention mechanism.

Table 1 summarizes the quantitative results, showing that the proposed model outperforms UNet-Std on all metrics. In particular, Recall_red and Precision_red, which focus on hazardous regions, are substantially improved, indicating that the proposed model captures risk around pedestrians more accurately in both extent and localization. L1_global and RMSE_global are also reduced, showing that overall reconstruction accuracy is maintained while enhancing detection performance in high-risk areas. These results demonstrate that the proposed architecture effectively improves the reliability of risk distribution estimated from forward-facing camera images.

Figure 2 shows qualitative examples. UNet-Std often produces blurred risk regions and underestimates hazards in front of pedestrians, whereas the proposed model yields sharper and more continuous high-risk regions, especially in locations where collisions are likely to occur.

Safe Route Generation: For safe route generation, we compared three models: Std, the basic Pix2Pix without angle conditioning; ScalarAngle, which adds extended losses and uses a scalar heading; and Proposed, which encodes the heading as a spatial distribution.

Table 2 reports the quantitative comparison. The proposed SpatialAngle model achieves the best performance on all metrics, outperforming both Std and ScalarAngle. Encoding the target heading as a spatial condition enables the model to generate a route that is more consistent with the cyclist's future position while avoiding hazardous regions. Improvements in ADE_2nd_half and FDE_mean, which emphasize the latter part of the trajectory, indicate that the proposed method provides safer path suggestions, particularly near the endpoint, where collision risk is critical.

Figure 3 shows qualitative examples. Std, which lacks explicit heading information, sometimes produces trajectories that deviate from the intended direction. In contrast, the proposed model generates a smoother route that follows the target heading while naturally detouring around high-risk areas, confirming its effectiveness for safe route generation.

Table 1. Quantitative evaluation of risk potential map generation.

Model	L1_global ↓	RMSE_global ↓	L1_red ↓	RMSE_red ↓	Recall_red ↑	Precision_red ↑
UNet-Std	2.61	9.30	52.94	64.68	0.261	0.242
UNet-RedLoss	2.02	8.07	50.07	61.53	0.303	0.346
ResNet-RedLoss	1.61	**6.50**	41.25	51.94	0.469	0.532
Proposed (ResNet-SA-RedLoss)	**1.56**	6.71	**41.19**	**51.68**	**0.482**	**0.569**

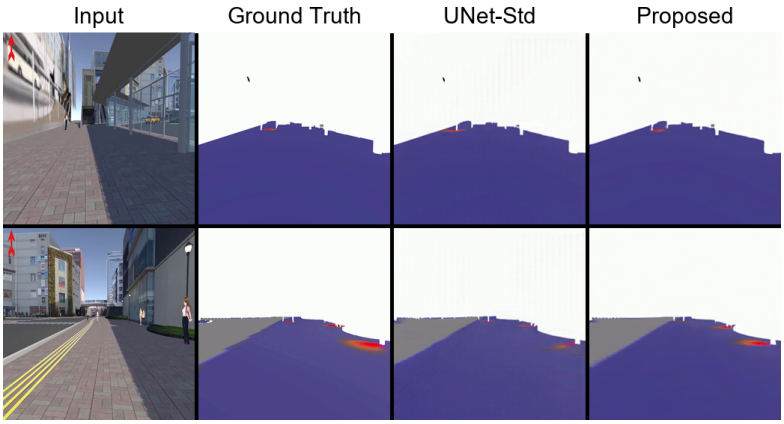

Figure 2. Examples of risk potential maps generated by UNet-Std and the Proposed models.

Table 2. Quantitative evaluation of safe route generation.

Model	ADE_mean ↓	ADE_2nd_half_mean ↓	FDE_mean ↓	Hausdorff_mean ↓
Std	6.77	7.54	9.34	11.72
ScalarAngle	5.15	6.11	8.25	10.07
Proposed (SpatialAngle)	**4.85**	**5.86**	**7.59**	**9.54**

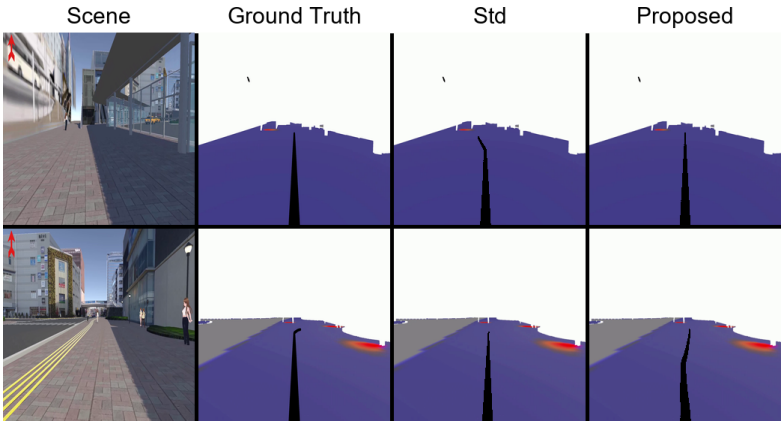

Figure 3. Examples of safe routes generated by the Std and Proposed models.

End-to-End: We also evaluated end-to-end performance by connecting G1 and G2 and processing real forward-facing bicycle video input. On a desktop PC equipped with an Intel Core i7 CPU and an Nvidia RTX 5060 16GB model, the system achieved an average inference time of 0.017 s per frame (about 59 fps). These results confirm real-time operation on a desktop PC and suggest feasibility for practical bicycle-assistance systems, even on lower-power embedded hardware.

Route Generation Under Hazardous Conditions: To evaluate the model's behavior under hazards, we compared route generation with and without pedestrians on the intended path. As shown in Fig. 4, the route remains straight when no pedestrians are present, whereas it detours around the high-risk region when a pedestrian appears. This demonstrates that the proposed model adapts the trajectory according to the current risk distribution.

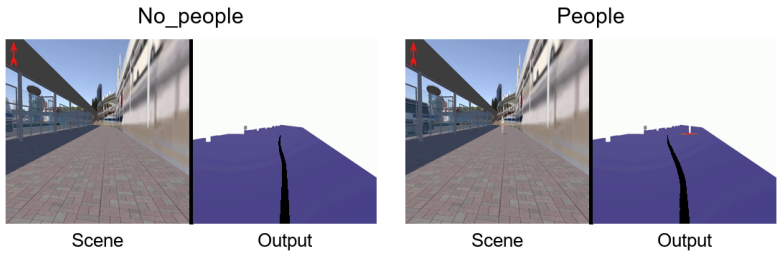

Figure 4. Route generation with and without pedestrians on the intended path.

6. CONCLUSION

In this study, we proposed a method for estimating a safe bicycle route in pedestrian-active environments using only forward-facing monocular images. We defined a risk potential map that represents hazardous regions, and the proposed model improved high-risk estimation accuracy by reducing pixel-wise errors and increasing the Precision and Recall around pedestrians. We also developed a route generation model that uses both the risk map and a spatial angle condition map, achieving lower trajectory errors and enabling a route that balances intended direction and collision avoidance. Finally, the integrated end-to-end pipeline achieved real-time performance at 0.017 s per frame, demonstrating the practical potential of the proposed approach for bicycle safety assistance.

REFERENCES

[1] National Police Agency of Japan, "Bicycle traffic safety information (in japanese)." Available: https://www.npa.go.jp/bureau/traffic/bicycle/info.html (2025). Accessed 20 Nov. 2025.

[2] Smaldone, S., Tonde, C., Ananthanarayanan, V. K., Elgammal, A., and Iftode, L., "The cyber-physical bike: A step towards safer green transportation," in [*Proc. ACM Workshop on Mobile Computing Systems and Applications (HotMobile)*], 56–61 (2011).

[3] Ibrahim, M. R., Haworth, J., Christie, N., and Cheng, T., "Cyclingnet: Detecting cycling near misses from video streams in complex urban scenes with deep learning," *IET Intell. Transp. Syst.* **15**(10), 1331–1344 (2021).

[4] Zeng, W., Luo, W., Suo, S., Sadat, A., Yang, B., Casas, S., and Urtasun, R., "End-to-end interpretable neural motion planner," in [*Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*], 8660–8669 (2019).

[5] Hu, Y., Yang, J., Chen, L., Li, K., Sima, C., Zhu, X., Chai, S., Du, S., Lin, T., Wang, W., et al., "Planning-oriented autonomous driving," in [*Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*], 17853–17862 (2023).

[6] Shim, J., Yu, J., and Lee, K., "Integrated risk grid map for collision avoidance and mitigation maneuvers of autonomous vehicle," *IEEE Access* **13**, 43767–43780 (2025).

[7] Kozuka, K. and Niebles, J. C., "Risky region localization with point supervision," in [*Proc. IEEE Int. Conf. Computer Vision Workshops (ICCVW)*], 246–253 (2017).

[8] Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A., "Image-to-image translation with conditional adversarial networks," in [*Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*], 1125–1134 (2017).