DECODING PEDESTRIAN SAFETY IN FLORIDA: INTEGRATING COMPUTER VISION AND NETWORK ANALYTICAL APPROACHES INTO DISCRETE REGRESSIONS

Ryan Lester
Department of Urban and Regional Planning
University of Florida
1480 Inner Rd, Gainesville, FL 32601
Email: ryan.lester@ufl.edu

Yuebing Liang
The Singapore-MIT Alliance for Research and Technology
1 CREATE Way, 10-01 CREATE Tower, Singapore 138602
Email: ybliang@mit.edu

Jaehyun Ha
Sol Price School of Public Policy
University of Southern California
Ralph and Goldy Lewis Hall 311 Los Angeles, California 90089-0626
Email: jaehyunh@usc.edu

Meiqing Li
School of Public Administration
University of Central Florida
Dr. Phillips Academic Commons, 528 W Livingston St, Orlando, FL 32801
Email: meiqing@ucf.edu

Shenhao Wang
Department of Urban and Regional Planning
University of Florida
1480 Inner Rd, Gainesville, FL 32601
Email: shenhaowang@ufl.edu

Word Count: 1993 words

ABSTRACT

Researchers have long explored the relationship between pedestrian safety and the built environment using traditional statistical tools. However, past studies often rely on limited data sources and fail to capture the full range of built environment features. This study integrates computer vision, using Mask2Former with the Mapillary Vistas dataset for semantic segmentation, to analyze how the built environment, sociodemographic factors, and travel behavior influence pedestrian crashes. We leverage a dataset from the Greater Orlando area (2012-2021) and combine these computer vision-derived features with other data in hurdle regression models, accounting for excess zeros and overdispersion in crash data. The results show significant associations, including a lower likelihood of crashes in areas with higher sidewalk-to-street ratios and greener environments captured by the green view index. The analysis also indicates that census tracts with a higher proportion of disabled residents have fewer pedestrian crashes, while commercial land use consistently correlates with increased crash risk. Population density and vehicle miles traveled show positive associations with crashes, likely due to exposure in dense or high-traffic areas. This study demonstrates the value of integrating computer vision with traditional pedestrian safety research to provide a more detailed assessment of sprawled urban contexts.

Keywords: Pedestrian safety, computer vision, Mask2Former, Mapillary Vistas, sociodemographics, urban planning.

1. INTRODUCTION

Pedestrian crashes, defined as incidents where at least one pedestrian was involved (*4*), continue to be a major safety concern in the United States. Despite initiatives like Vision Zero and Complete Streets, crashes and fatalities in Southern states, particularly Florida, remain high. Cities like Orlando have adopted these programs, but factors such as car-dependent urban design, light trucks, higher speed limits, and nighttime activity continue to contribute to high fatality rates (*1, 6, 12, 22*). Florida consistently ranks among the highest for pedestrian crashes and fatalities, particularly in metro areas like Orlando, Tampa, and Miami (*5, 20*).

Traditional studies examining the relationship between sociodemographic factors and the built environment (*11, 16, 21, 24*) often miss critical infrastructure details, such as crosswalks, curb cuts, and greenery. While urban areas may have access to such data, sprawled regions face limitations. Recent advances in computer vision offer more granular analysis of how the built environment impacts safety.

This study uses computer vision to generate variables for non-linear regression analysis, aiming to better understand how sociodemographic and built environment factors influence pedestrian crashes. By integrating computer vision with traditional data, we assess crashes in sprawling areas like Orlando. Using Mask2Former and the Mapillary Vistas dataset, this research seeks to address the question: How can computer vision-derived variables improve our understanding of pedestrian crashes and how can these insights guide targeted interventions to enhance safety for vulnerable populations in sprawled urban contexts?

2. LITERATURE REVIEW

The integration of computer vision techniques in transportation research, such as semantic segmentation, has allowed for per-pixel analysis of images using fully convolutional networks (*15*). Transportation research has been using these models alongside Google Satellite and Street View (GSV) imagery to analyze built environment factors like crosswalks and pedestrian infrastructure (*14*), neighborhood change (*18*), poverty (*9*), and urban green space exposure (*8*).

Computer vision models have long been used to analyze car infrastructure, but their application to pedestrian infrastructure is a more recent development gaining traction. As the field advances, pedestrian analysis is becoming increasingly effective. For example, Li and Rodriguez (13) incorporated the GSV and computer vision generated dataset to a change-on-change regression model that evaluates the impact of station area crosswalk visibility enhancement on station-level ridership change between 2010 and 2018. Hu et al. (7) used street view images (SVIs) to capture pedestrian exposure, showing that higher exposure leads to more frequent crashes but reduces injury severity, supporting the "safety in numbers" effect. Given the effectiveness of using GSV for street level pedestrian safety audit (*17*), and the potential of employing computer vision models to automate the process at a significantly larger spatial and temporal scale, this study demonstrates the potential of combining both approaches.

3. DATA

We examine pedestrian fatality trends in the Greater Orlando Area using data from multiple public sources spanning 2012-2021. Our primary datasets include the American Community Survey (ACS) for sociodemographic variables at the census tract level, Florida Department of Transportation (FDOT) data for pedestrian crashes, panoramic images from Google Street View, and land use classifications from the Florida Department of Environmental Protection.

3.1. Sociodemographics

Sociodemographic variables were retrieved through the Census API and included data at the census tract level for population characteristics such as age distribution, disability status, racial composition, educational attainment, and travel behavior.

3.2. Crashes

Pedestrian crash data was collected via the FDOT API and included crash locations and severity ratings across the Greater Orlando Area. Crashes were classified using FDOT's injury severity ratings, with levels 1-4 aggregated for analysis, and level 5, indicating fatalities, excluded. The primary outcome was whether a crash occurred, regardless of severity level.

3.3. Computer Vision

Detailed features of the built environment were captured using Google Street View imagery. Nearly 24,000 panoramic images were randomly sampled across the study area using the Google Street View API. Each panoramic image was split into four angles (front, back, left, right) using a cube map format to reduce distortion and better focus on relevant surroundings, as illustrated in Figures 2, 3, and 4.

The built environment features extracted from the images included crosswalks, sidewalks, bike lanes, vegetation, and curb cuts. Mask2Former, a segmentation model from Meta (*3*), was applied to the images using the Mapillary Vistas dataset (*19*) to detect these features, which are often absent from traditional datasets.

3.4. Land Use

Land use classifications were obtained from the Florida Department of Environmental Protection, which categorized areas into residential, commercial, and industrial zones.

3.5. Exposure

Exposure measures were also considered in the analysis, including variables such as vehicle miles traveled (VMT), population density, and employment centers. These metrics are meant to account for areas that may represent more pedestrian activity.

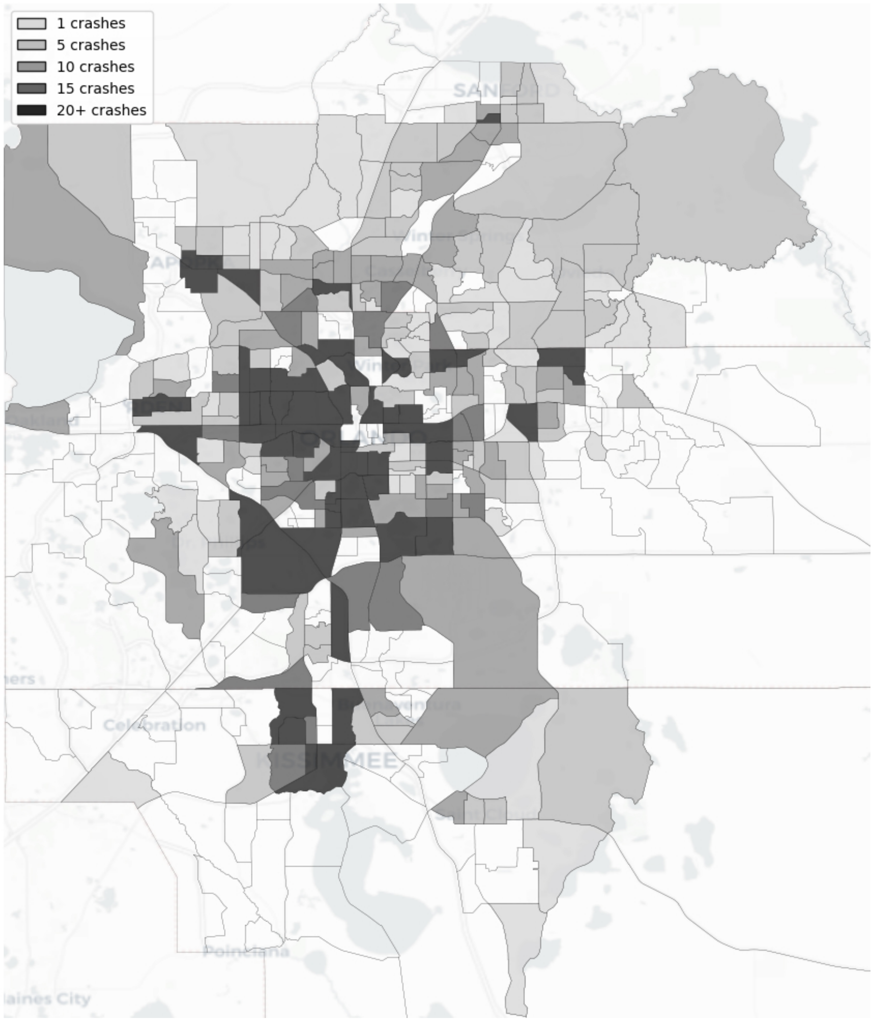

FIGURE 1: Pedestrian crashes in the Greater Orlando Study Area (2012-2021).

FIGURE 2: Panoramic Street View Image.

FIGURE 3: Converted Street View Image into Cube Map Format.

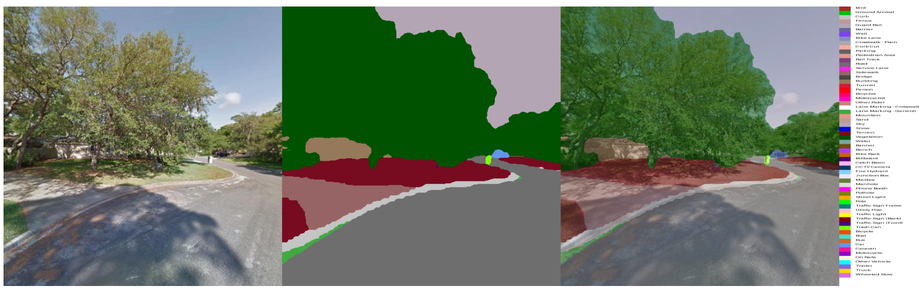

FIGURE 4: Mask2Former and Mapillary Vistas Results.

TABLE 1: Variable Definitions for the Model

Variable	Description
Sidewalk to Street Ratio	Ratio of sidewalk pixels to the total street-related pixels (road, curb, parking). (*10*)
Green View Index	Proportion of green pixels to the total pixels in the image. (*2*)
Crosswalk (Dummy)	Binary variable indicating the presence (1) or absence (0) of a plain crosswalk in the image.
Curb Cut (Dummy)	Binary variable indicating the presence (1) or absence (0) of a curb cut in the image.
Bike Lane (Dummy)	Binary variable indicating the presence (1) or absence (0) of a bike lane in the image.
Pedestrian Area (Dummy)	Binary variable indicating the presence (1) or absence (0) of a pedestrian area in the image.
Traffic Light	Number of traffic lights in the image.
Black Residents (%)	Proportion of Black or African American residents in the population.
Hispanic Residents (%)	Proportion of Hispanic or Latino residents in the population.
Asian Residents (%)	Proportion of Asian residents in the population.
Alaska Native Residents (%)	Proportion of Alaska Native or other Indigenous residents in the population.
Unemployed (%)	Proportion of unemployed residents in the population.
College Educated (%)	Proportion of residents with a college education (bachelor's degree or higher).
Older Adults (65+) (%)	Proportion of residents aged 65 and older in the census tract.
Disabled Residents (%)	Proportion of residents with a disability in the census tract.
Commercial Land Use	Binary variable indicating if the photo location is within a commercial land use area.
Residential Land Use	Binary variable indicating if the photo location is within a residential land use area.
Vehicle Miles Traveled (VMT)	Total vehicle miles traveled in the census tract.
Population Density	Number of persons per square mile in the census tract.
Employment Center	Binary variable indicating if the location is within a major employment center.

TABLE 2: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
Computer Vision					
Sidewalk to Street Ratio	239,246	0.0598	0.0638	0.0000	1.0000
Green Vegetation Index	239,246	0.2144	0.1375	0.0000	0.8929
Crosswalk Plain (Dummy)	239,246	0.0409	0.1980	0.0000	1.0000
Curb Cut (Dummy)	239,246	0.5682	0.4953	0.0000	1.0000
Bike Lane (Dummy)	239,246	0.0183	0.1341	0.0000	1.0000
Pedestrian Area (Dummy)	239,246	0.1286	0.3348	0.0000	1.0000
Traffic Light	239,246	0.0149	0.0849	0.0000	1.6000
Sociodemographics					
Black Residents (%)	239,246	17.28	20.41	0.00	99.57
Hispanic Residents (%)	239,246	25.74	16.33	0.00	86.58
Asian Residents (%)	239,246	3.97	3.42	0.00	25.58
Alaska Native Residents (%)	239,246	0.24	0.60	0.00	11.40
Unemployed Residents (%)	239,246	4.23	2.51	0.11	24.11
College Educated Residents (%)	239,246	20.76	10.39	2.05	55.43
Older Adults (65+) (%)	239,246	324.36	184.15	4.00	1259.00
Disabled Residents (%)	239,246	5507.14	2313.33	626.00	16994.00
Land Use					
Commercial Land Use	239,246	0.1178	0.3223	0.0000	1.0000
Residential Land Use	239,246	0.4888	0.4999	0.0000	1.0000
Exposure					
VMT	239,246	514,505,000	462,705,400	934,290	2,097,029,000
Population Density	239,246	0.0011	0.0007	0.0000	0.0070
Employment Center	239,246	0.0238	0.1524	0.0000	1.0000

4. METHODOLOGY

We use a hurdle regression model to account for the large number of zeros in the pedestrian crash data. This model combines logistic regression to handle the zero counts and negative binomial regression for the positive counts.

For the logistic regression, the dependent variable represents the probability (P_i) of a crash occurrence, where the outcome is coded as 1 if there is at least one crash and 0 otherwise. The logistic regression equation transforms these probabilities using the logit function, as defined in (*23*):

$$\text{logit}(P_i) = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_K X_{K,i} \tag{1}$$

Here, β_0 is the model constant, and $\beta_1, \ldots, \beta_K$ are the coefficients corresponding to the explanatory variables $X_{1,i}, \ldots, X_{K,i}$, which include socio-demographic factors and built environment characteristics. To stabilize the logistic regression and manage multicollinearity, we applied L1 regularization (Lasso), which shrinks the coefficients of less important variables toward zero.

In the second stage of the hurdle model, to address the issue of overdispersion (i.e., the variance being larger than the mean) in the count data, we used an L1-regularized negative binomial regression model. L1 regularization again helped stabilize the parameter estimation, reduce model complexity, and assist in model convergence. The negative binomial model is specified as:

$$\lambda_i = \exp(\beta X_i + \varepsilon_i) \tag{2}$$

where λ_i represents the expected number of pedestrian crashes for census tract i, β represents the coefficients, X_i includes the explanatory variables, and ε_i is a gamma-distributed error term with mean 1 and variance α. The probability distribution function for the negative binomial model is:

$$P(y_i) = \frac{\Gamma\left(\frac{1}{\alpha} + y_i\right)}{\Gamma\left(\frac{1}{\alpha}\right) y_i!} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^{y_i} \tag{3}$$

In this equation, Γ is the gamma function, y_i is the observed count of pedestrian crashes, and α is the overdispersion parameter that allows the variance to differ from the mean.
To further stabilize the estimation of parameters and address multicollinearity, L1 regularization is applied to the negative binomial model as well. The log-likelihood function for the L1-regularized negative binomial model is:

$$\mathscr{L}(\lambda_i) = \prod_i \left[\frac{\Gamma\left(\frac{1}{\alpha} + y_i\right)}{\Gamma\left(\frac{1}{\alpha}\right) y_i!} \left(\frac{\frac{1}{\alpha}}{\frac{1}{\alpha} + \lambda_i}\right)^{\frac{1}{\alpha}} \left(\frac{\lambda_i}{\frac{1}{\alpha} + \lambda_i}\right)^{y_i} \right] - \lambda \sum_{j=1}^{K} |\beta_j| \tag{4}$$

Here, λ is the regularization parameter that controls the strength of the L1 penalty, and $\sum_{j=1}^{K} |\beta_j|$ is the L1 norm of the coefficients. The regularization term reduces the magnitude of the coefficients, helping to prevent overfitting and improve model generalization.

5. RESULTS

5.1. Sociodemographic Factors and Pedestrian Crashes

The hurdle regression analysis in Table 3 shows several associations between sociodemographic factors and pedestrian crashes. The proportion of residents with disabilities stands out as one of the most significant variables, consistently showing that census tracts with higher proportions of disabled residents experience fewer crashes. This finding persisted even after accounting for other sociodemographic and built environment characteristics.

The percentage of college-educated residents was also negatively associated with crashes, showing fewer incidents in areas with higher educational attainment. Race variables showed mixed results. The proportion of Black and Hispanic residents in census tracts displayed positive associations with crashes in certain models, although these relationships were not consistent when controlling for other factors. The percentage of older adults was marginally significant in some models, showing a negative association with crashes.

5.2. Land Use and Exposure

Commercial land use was consistently associated with higher crash frequencies, which is expected, considering that commercial areas have higher traffic and pedestrian activity.

Exposure-related variables such as vehicle miles traveled (VMT) and population density

TABLE 3: Hurdle Regression Results

	Pedestrian Crashes					
Variable	Model 1 (Logit)	Model 1 (NB)	Model 2 (Logit)	Model 2 (NB)	Model 3 (Logit)	Model 3 (NB)
Sociodemographics						
Black Residents (%)	0.1734***	0.1971***			0.3031**	0.0297
	(0.031)	(0.033)			(0.100)	(0.032)
Hispanic Residents (%)	0.1206***	0.1191***			-0.0271	-0.0897**
	(0.032)	(0.034)			(0.036)	(0.034)
Asian Residents (%)	-0.0090	-0.0186			0.0673*	0.0049
	(0.026)	(0.027)			(0.031)	(0.026)
Alaska Native Residents (%)	-0.0119	-0.0122			-0.0291	-0.0423+
	(0.020)	(0.021)			(0.021)	(0.023)
Unemployed Residents (%)	-0.0424+	-0.0583*			0.0411	0.0339
	(0.026)	(0.027)			(0.025)	(0.026)
College Educated Residents (%)	-0.1558***	-0.1374**			-0.2555***	-0.2581***
	(0.039)	(0.040)			(0.041)	(0.042)
Older Adults (65+) (%)	-0.0771+	-0.0645			-0.0743+	-0.0858*
	(0.039)	(0.041)			(0.041)	(0.042)
Disabled Residents (%)	-0.2690***	-0.3004***			-0.2414***	-0.2789***
	(0.039)	(0.042)			(0.041)	(0.043)
Land Use						
Commercial Land Use	0.1103***	0.1364***			0.1374***	0.1339***
	(0.022)	(0.023)			(0.020)	(0.021)
Residential Land Use	0.0048	0.0156			-0.0309	-0.0039
	(0.027)	(0.028)			(0.028)	(0.029)
Exposure						
VMT	0.1734***	0.2115***			0.1868***	0.2130***
	(0.025)	(0.026)			(0.026)	(0.026)
Population Density	0.3013***	0.3438***			0.4001***	0.4217***
	(0.024)	(0.027)			(0.025)	(0.026)
Employment Center	0.0430*	0.0347			0.0772***	0.0887***
	(0.022)	(0.023)			(0.020)	(0.020)
Computer Vision						
Sidewalk to Street Ratio			-0.0391	-0.0527+	-0.3509***	-0.3339***
			(0.026)	(0.027)	(0.036)	(0.036)
Green View Index			-0.1048***	-0.1193***	-1.0437***	-1.0767***
			(0.025)	(0.026)	(0.046)	(0.047)
Crosswalk (Dummy)			-0.0043	-0.0039	0.0978***	0.0918***
			(0.022)	(0.023)	(0.014)	(0.015)
Curb Cut (Dummy)			0.0621*	0.0636*	0.4187***	0.3929***
			(0.025)	(0.026)	(0.031)	(0.031)
Bike Lane (Dummy)			0.0240	0.0128	0.0613***	0.0606***
			(0.020)	(0.022)	(0.013)	(0.015)
Pedestrian Area (Dummy)			-0.0002	-0.0019	-0.1200**	-0.1026**
			(0.024)	(0.025)	(0.036)	(0.036)
Traffic Light			0.0673***	0.0782***	0.2509***	0.2985***
			(0.017)	(0.019)	(0.008)	(0.011)
Log Likelihood	-10907.00	-11672.00	-11224.00	-11991.00	-9203.00	-9921.60
R-squared / Pseudo R-squared	0.03049	0.02902	0.00234	0.00252	0.18260	0.17520
Observations	238802	238802	238802	238802	238802	238802
Standard errors in parentheses						
Note:	+ p<0.1	* p<0.05	** p<0.01	*** p<0.001		

showed positive associations with pedestrian crashes. Again, these results were expected as the increase in exposure will likely see more pedestrian activity.

5.3. Built Environment Characteristics

Several built environment features measured through computer vision found significant associations. Some were expected such as the sidewalk-to-street ratio being negatively associated with crashes in the negative binomial models, showing areas with more sidewalks saw fewer incidents. The green view index, capturing vegetation presence, was another significant factor with negative associations across multiple models.

Infrastructure such as crosswalks, curb cuts, and bike lanes had mixed results. Crosswalks showed a positive association in some models, while curb cuts were linked to higher crash frequencies in the logit models but did not reach significance in the negative binomial models. Traffic lights were consistently associated with higher crash frequencies across the models.

6. CONCLUSION AND DISCUSSION

This study offers a framework for assessing pedestrian crashes in Greater Orlando by integrating Mask2Former for computer vision with traditional statistical data. The segmentation results capture detailed built environment features like sidewalks, crosswalks, and curb cuts more efficiently than manual methods.

These findings give way to several areas for further research. The negative association between disability prevalence and pedestrian crashes warrants closer examination to understand the reasoning behind this relationship. Future research could discuss whether specific interventions in sprawled areas with higher disabled populations contribute to reduced crash risk or if other factors are influencing this association, such as vehicle reliance.

The consistently positive association of commercial land use with pedestrian crashes has been studied extensively, although future work should explore variations within sprawled commercial areas, such as parking minimums, street design, or availability and quality of pedestrian infrastructure.

Green view index results suggest a possible link between the presence of greenery and safety, potentially related to calming traffic or enhancing visibility. Future work should explore whether adding greenery has safety benefits or if these associations reflect other neighborhood characteristics not captured.

Lastly, future research should address the downsides of using current Google Street View imagery, which may not account for changes over time in the built environment, potentially affecting the accuracy of pedestrian safety assessments.

Overall, this study provides an initial step towards understanding the unique factors associated with pedestrian crashes in the Greater Orlando Area, especially within the context of urban sprawl. While the findings found certain associations, they also point to the need for more research into how built environment characteristics, land use, and sociodemographic factors interact to influence crash risk.

AUTHOR CONTRIBUTION STATEMENT

The authors confirm contribution to the paper as follows: study conception and design: R.L.; data collection: R.L.; analysis and interpretation of results: R.L., Y.L., J.H.; modeling and commenting: R.L., Y.L., J.H.; draft manuscript preparation: R.L., Y.L., J.H., M.L.; supervision: S.W.. All authors reviewed the results and approved the final version of the manuscript.

CONFLICT OF INTERESTS

The authors do not have any conflicts of interest to declare.

REFERENCES

1. Anderson, M. (2008). Safety for whom? the effects of light trucks on traffic fatalities, *Journal of Health Economics* **27**(4): 973–989. **URL:** *https://www.sciencedirect.com/science/article/pii/S0167629608000040*
2. Bai, Y., Cao, M., Wang, R., Liu, Y. and Wang, S. (2022). How street greenery facilitates active travel for university students, *Journal of Transport & Health* **26**: 101393.
3. Cheng, B., Schwing, A. G. and Kirillov, A. (2021). Per-pixel classification is not all you need for semantic segmentation.
4. Du, B., Zhang, C., Sarkar, A., Shen, J., Telikani, A. and Hu, H. (2024). Identifying factors related to pedestrian and cyclist crashes in act, australia with an extended crash dataset, *Accident Analysis & Prevention* **207**: 107742.
5. Fatality Analysis Reporting System (FARS) (2021). 2021 ranking of state pedestrian fatality rates - state: Usa, `https://www.nhtsa.gov`. Accessed: 2024-06-12.
6. Ferenchak, N. N. and Abadi, M. G. (2021). Nighttime pedestrian fatalities: A comprehensive examination of infrastructure, user, vehicle, and situational factors, *Journal of Safety Research* **79**: 14–25. **URL:** *https://www.sciencedirect.com/science/article/pii/S002243752100092X*
7. Hu, Y., Chen, L. and Zhao, Z. (2024). How does street environment affect pedestrian crash risks? a link-level analysis using street view image-based pedestrian exposure measurement, *Accident Analysis & Prevention* **205**: 107682.
8. Jang, K. M., Kim, J., Lee, H.-Y., Cho, H. and Kim, Y. (2020). Urban green accessibility index: A measure of pedestrian-centered accessibility to every green point in an urban area, *ISPRS International Journal of Geo-Information* **9**(10). **URL:** *https://www.mdpi.com/2220-9964/9/10/586*
9. Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B. and Ermon, S. (2016). Combining satellite imagery and machine learning to predict poverty, *Science* **353**(6301): 790–794. **URL:** *https://www.science.org/doi/abs/10.1126/science.aaf7894*
10. Koo, B. W., Guhathakurta, S. and Botchwey, N. (2022). How are neighborhood and street-level walkability factors associated with walking behaviors? a big data approach using street view images, *Environment and Behavior* **54**(1): 211–241.
11. Lee, J., Abdel-Aty, M. and Jiang, X. (2015). Multivariate crash modeling for motor vehicle and non-motorized modes at the macroscopic level, *Accident Analysis & Prevention* **78**: 146–154.
12. Lefler, D. E. and Gabler, H. C. (2004). The fatality and injury risk of light truck impacts with pedestrians in the united states, *Accident Analysis and Prevention* **36**(2): 295–304. **URL:** *https://www.sciencedirect.com/science/article/pii/S0001457503000071*

13. Li, M. and Rodriguez, D. A. (2024). Marked crosswalks, station area built environments, and transit ridership: Associations between changes in 877 US TOD stations, 2010–2018, *Transportation Research Part A: Policy and Practice* **179**: 103926.
URL: *https://www.sciencedirect.com/science/article/pii/S0965856423003464*

14. Li, M., Sheng, H., Irvin, J., Chung, H., Ying, A., Sun, T., Ng, A. Y. and Rodriguez, D. A. (2023). Marked crosswalks in us transit-oriented station areas, 2007–2020: A computer vision approach using street view imagery, *Environment and Planning B: Urban Analytics and City Science* **50**(2): 350–369.
URL: *https://doi.org/10.1177/23998083221112157*

15. Long, J., Shelhamer, E. and Darrell, T. (2015). Fully convolutional networks for semantic segmentation, *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3431–3440.

16. Miranda-Moreno, L. F., Morency, P. and El-Geneidy, A. M. (2011). The link between built environment, pedestrian activity and pedestrian–vehicle collision occurrence at signalized intersections, *Accident Analysis & Prevention* **43**(5): 1624–1634.

17. Mooney, S. J., DiMaggio, C. J., Lovasi, G. S., Neckerman, K. M., Bader, M. D. M., Teitler, J. O., Sheehan, D. M., Jack, D. W. and Rundle, A. G. (2016). Use of Google Street View to Assess Environmental Contributions to Pedestrian Injury, *American Journal of Public Health* **106**(3): 462–469.

18. Naik, N., Kominers, S. D., Raskar, R., Glaeser, E. L. and Hidalgo, C. A. (2017). Computer vision uncovers predictors of physical urban change, *Proceedings of the National Academy of Sciences* **114**(29): 7571–7576.

19. Neuhold, G., Ollmann, T., Rota Bulo, S. and Kontschieder, P. (2017). The mapillary vistas dataset for semantic understanding of street scenes, *Proceedings of the IEEE international conference on computer vision*, pp. 4990–4999.

20. Smart Growth America (2024). Dangerous by design 2024, `https://smartgrowthamerica.org/resources/dangerous-by-design-2024/`. Accessed: 2024-06-12.

21. Su, J., Sze, N. and Bai, L. (2021). A joint probability model for pedestrian crashes at macroscopic level: Roles of environment, traffic, and population characteristics, *Accident Analysis & Prevention* **150**: 105898.

22. Tyndall, J. (2021). Pedestrian deaths and large vehicles, *Economics of Transportation* **26-27**: 100219.
URL: *https://www.sciencedirect.com/science/article/pii/S2212012221000241*

23. Washington, S., Karlaftis, M. G., Mannering, F. and Anastasopoulos, P. (2020). *Statistical and econometric methods for transportation data analysis*, Chapman and Hall/CRC.

24. Wier, M., Weintraub, J., Humphreys, E. H., Seto, E. and Bhatia, R. (2009). An area-level model of vehicle-pedestrian injury collisions with implications for land use and transportation planning, *Accident Analysis & Prevention* **41**(1): 137–145.