Human Factors in Design, Engineering, and Computing, Vol. 238, 2026, 435–444
https://doi.org/10.54941/ahfe1008276

Visual Quality, Perceived Safety, and Spatial Vitality in Shichahai, Beijing: A Day–Night Assessment of a Cultural Waterfront

Jingjing Ma, Yuwei Wu, and Yan Huang

Tsinghua Academy of Fine Arts, China

ABSTRACT

Cultural waterfronts require evaluation after dark, yet photographs, human perceptions and observed activity represent different kinds of evidence. This study develops an image-based audit of Shichahai, Beijing, as the environmental component of a framework linking visual quality, perceived safety and spatial vitality. An archive of 1,556 photographs was screened; the newest available daytime imagery at each location was compared with author-confirmed 2026 nighttime photographs. After removing a duplicated daytime source, the primary analysis contained 71 nominally corresponding sites and 568 images. A fixed ADE20K-pretrained SegFormer-B0 model estimated water-class pixel coverage, while photographic proxies described darkness and the spatial distribution of brightness. Mean site-level dark-area coverage increased from 29.34% to 45.83%, a difference of 16.49 percentage points, and the grid brightness ratio declined from 0.259 to 0.214. Predicted water coverage averaged 1.43% by day and 1.70% at night, with no stable directional difference supported by the descriptive resampling interval. Brightness perturbations exposed sensitivity in water predictions. Historical daytime dates, unequal camera views and absent segmentation ground truth limit interpretation. No perception ratings or activity observations were collected. The contribution is a reproducible environmental audit and an explicit account of the validation needed before inferring human experience or waterfront use.

Keywords: Shichahai, Cultural waterfront, Nighttime imagery, SegFormer, Visual environment, Perceived safety

INTRODUCTION

Waterfront public spaces combine views of water with movement, recreation and social contact. Their suitability for evening use depends on what pedestrians can identify along paths, at the water edge and in surrounding public spaces. Research on blue space has linked the presence of water in visual scenes with preference, affect and restorative evaluations (White et al., 2010). That relationship motivates attention to water visibility, but it does not establish that a larger water-pixel fraction always represents a better environment, particularly at night.

Lighting and human experience also require separate measurements. Fotios et al. (2015) reviewed evidence connecting road lighting with pedestrian

Received November 4, 2026; Revised November 19, 2026; Accepted November 26, 2026; Available online December 1, 2026

© 2026 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

reassurance and highlighted methodological problems in attributing responses to lighting alone. Portnov et al. (2020) examined reported lighting attributes and feelings of safety through an interactive field survey. More recently, Navarrete Hernandez et al. (2025) used image ratings to investigate differences between daytime and nighttime safety perceptions. These studies support the relevance of nighttime conditions while showing why photographs cannot, by themselves, supply public perception scores.

For a cultural waterfront, an additional difficulty is the visual behavior of water. Dark water may blend with its surroundings, while reflections can introduce bright localized features. Vegetation, boats and railings can further interrupt the visible water surface. Semantic segmentation offers a reproducible way to identify candidate water pixels, but a pretrained model may respond to photographic conditions as well as scene content. SegFormer combines a hierarchical Transformer encoder with a lightweight multilayer perceptron decoder, making it a practical baseline for this audit (Xie et al., 2021).

This paper asks two empirical questions: how do photographic environmental indicators differ between the latest available daytime imagery and 2026 nighttime photographs at nominally corresponding sites, and how sensitive are nighttime water predictions to monotonic brightness transformations? A third, conceptual question concerns how the measured indicators can inform a subsequent study of perceived safety and spatial vitality. The present evidence addresses image conditions and model behavior. Human responses, visitation intentions, observed activity and actual safety remain outside the measured outcomes.

STUDY AREA AND DATA

The study concerns the supplied Shichahai waterfront image collection in Beijing. The full archive contains 1,088 daytime JPG files and 468 nighttime JPG files. Daytime folders identify years from 2013 to 2022, with four named viewing directions per site-year. The supplied collection script requests Baidu street-view images at 1,024 × 512 pixels and nominal headings of 0°, 90°, 180° and 270°. Nighttime photographs are 4:3 images at 4,032 × 3,024, 4,096 × 3,072 or 5,712 × 4,284 pixels. The author confirmed that the nighttime photographs were taken in 2026; this year is not independently recoverable from the image EXIF metadata. The location of Shichahai and the mapped lake zone are shown in Figure 1.

Following the author's selection rule, the latest available daytime year was selected separately for each site. This retained 288 daytime images at 72 identifiers. Although all 72 identifiers have nighttime counterparts, SHA-256 screening found identical daytime photographs under ID_2350 and ID_2619 in both 2017 and 2022, accounting for eight repeated files in the full archive. ID_2619 was excluded from the primary comparison so that one daytime source was not counted as an independent location twice. The resulting primary dataset contains 71 sites, each with four daytime and four nighttime images, or 568 images in total. All 756 selected raw images, including supplementary nighttime views, were processed by the model; only the 568 primary images enter the site comparison.

The primary daytime years comprise five sites from 2013, one from 2015, two from 2016, 28 from 2017, 25 from 2019 and ten from 2022. Thus, the newest-image rule reduces avoidable historical repetition but leaves gaps of four to thirteen years relative to the nighttime photographs. Location identifiers and heading names provide nominal correspondence, rather than verified camera registration. Camera height, field of view, season, weather and exposure were not standardized or documented. The analysis therefore compares two photographic records at matched identifiers, without isolating an effect of nightfall.

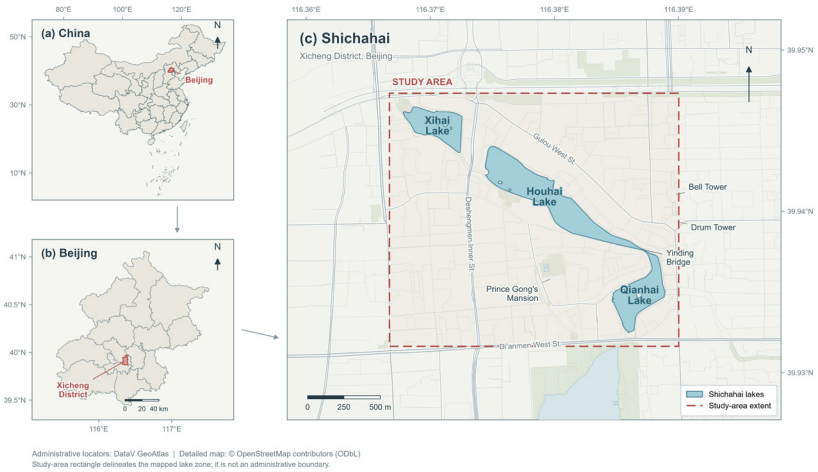


Figure 1: Location of Shichahai, Beijing. (a) Beijing within China; (b) Xicheng District within Beijing; (c) the mapped lake zone, including Xihai, Houhai and Qianhai. The dashed rectangle indicates the mapped extent, not an administrative boundary. Administrative boundaries: DataV GeoAtlas. Detailed water and road data: © OpenStreetMap contributors (ODbL).

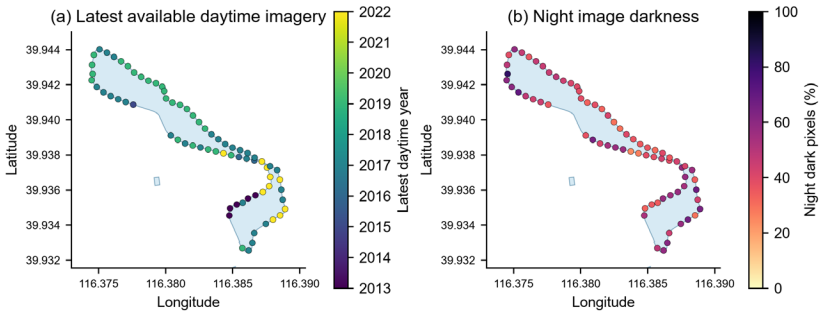


Figure 2: The 71 primary site identifiers mapped using the supplied coordinate table. Panel a shows the newest available daytime year; panel b shows site-average nighttime dark-area coverage. Water polygons are from the supplied OpenStreetMap extract, © OpenStreetMap contributors, ODbL. Points represent sampling identifiers rather than verified nighttime camera coordinates.

METHODS

A fixed SegFormer-B0 checkpoint pretrained for the 150-class ADE20K benchmark was used without local training or fine-tuning (Zhou et al., 2017; Xie et al., 2021). The checkpoint is nvidia/segformer-b0-finetuned-ade-512-512, revision 489d5cd81a0b59fab9b7ea758d3548ebe99677da. Its 15,036,944-byte safetensors file and SHA-256 hash are retained in the reproducibility manifest. Processing used Python 3.14.7, PyTorch 2.14.1+cpu and Transformers 4.57.1 on a Windows CPU with four computation threads. All inference was local; photographs were not uploaded to an inference service.

Images were oriented from available EXIF orientation, converted to RGB and reduced proportionally to a maximum side of 768 pixels using Lanczos resampling. These working images were resized bilinearly to 512 × 512 pixels for the model, with channels scaled to [0,1], normalized by means [0.485,0.456,0.406] and standard deviations [0.229,0.224,0.225]. Class logits were upsampled bilinearly to the working image dimensions with align_corners=False before selecting the maximum-logit class. Binary water masks combine the checkpoint labels water, sea, river and lake, with zero-based IDs 21, 26, 60 and 128. Combining these labels avoids treating distinctions among water types as differences in water presence.

For image i, predicted water coverage is $W_i = N_i(\text{water classes})/N_i(\text{all pixels})$. This quantity describes the model output, not a validated estimate of the water actually visible in the photograph. No Shichahai annotation set was available, so segmentation accuracy, precision, recall and benchmark intersection over union are not reported. This distinction is especially relevant under nighttime domain shift: adverse-condition benchmarks demonstrate the challenge of applying scene-understanding systems across visual conditions (Sakaridis et al., 2021).

Photographic brightness was computed from normalized JPEG RGB code values as $Y = 0.2126R + 0.7152G + 0.0722B$. Dark-area coverage is $D\tau = N(Y < \tau)/N$, with $\tau = 0.20$ for the main analysis and 0.10 and 0.30 as sensitivity thresholds. To describe the spatial distribution of brightness, each working image was partitioned into an 8 × 8 grid. If μg denotes each cell's mean Y, the grid brightness ratio is $U = P10(\mu g)/P90(\mu g)$, with U set to zero for an entirely black image. Values closer to one indicate less variation among grid cells. Because the calculation uses gamma-encoded photographic values and a changing mixture of surfaces and sky, Y is not calibrated luminance and U is not physical illuminance uniformity.

Within each site and period, the four views were averaged with equal weights. Across sites, these averages were again weighted equally. For each metric, the descriptive difference is the site-average nighttime value minus the site-average daytime value. A percentile interval was computed by sampling the 71 complete site records with replacement 5,000 times, using seed 569. This preserves the four-view grouping, but the interval is conditional on the sampled sites and does not account for spatial dependence, camera differences, model error or historical change. It is therefore a descriptive resampling interval, not a causal confidence statement about lighting.

Two sensitivity checks were specified before aggregating the complete results. First, nighttime RGB values were transformed as $RGB' = RGB^{\gamma}$ with $\gamma = 0.7$ and 1.3, respectively, and the unchanged model was rerun on all 284 primary nighttime views. These global transformations brighten or darken the input without changing scene geometry. We compared predicted water fractions and the proportion of pixels whose binary water label changed. Agreement with the raw mask is a stability diagnostic, not accuracy against ground truth. Second, the daytime–nighttime comparison was repeated descriptively for the ten sites whose newest daytime imagery was from 2022. Original source images were preserved throughout. Figure 3 summarizes the completed image-analysis workflow and the proposed perception and activity validation.

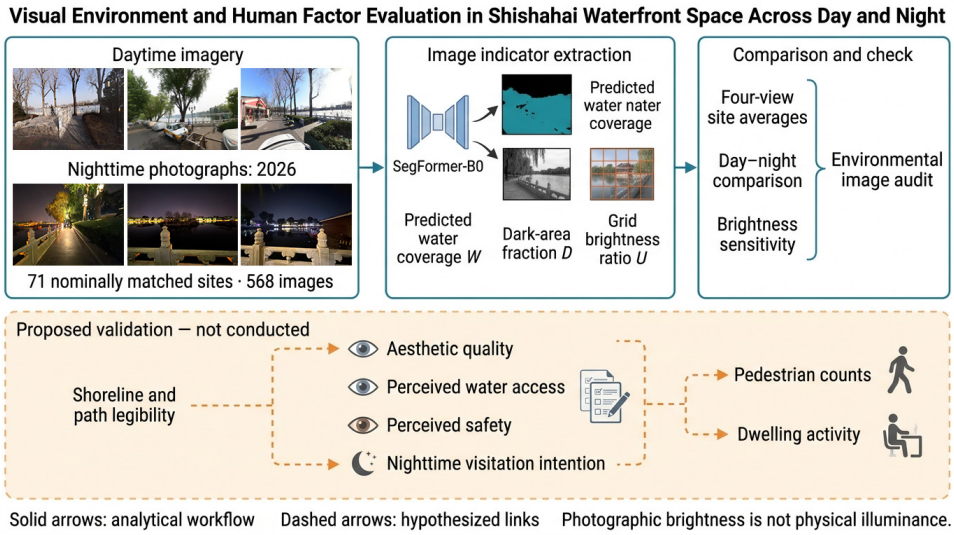


Figure 3: Research framework for the Shichahai day–night assessment. The upper panel shows daytime and nighttime image inputs, SegFormer-B0 water predictions, photographic brightness indicators and site-level comparisons. The dashed lower panel identifies proposed perception ratings and activity observations that were not collected. Photographic and mask examples are illustrative. Solid arrows show the analytical workflow; dashed arrows denote hypotheses requiring validation rather than established causal effects.

RESULTS

Photographic darkness was higher at night for most of the primary sites. The mean site-level fraction of pixels below Y = 0.20 was 29.34% in the daytime record and 45.83% in the nighttime record. The mean difference was 16.49 percentage points, with a descriptive 95% resampling interval of 12.79 to 20.24 points. Darkness increased at 59 of the 71 sites. The mean JPEG brightness proxy decreased from 0.427 to 0.282. These differences describe the photographs; they do not establish a proportional decrease in field illumination.

The brightness distribution also became more uneven on average. The mean grid ratio fell from 0.259 by day to 0.214 at night, with a mean difference of −0.045 and a descriptive interval of −0.070 to −0.022. The ratio decreased at 47 sites and increased at 24, indicating heterogeneity rather than a uniform response throughout the sampled waterfront. Figure 4 displays every primary site, retaining the variability that would be hidden by an overall average.

Table 1: Equally weighted site summaries, n = 71. Percent-based differences are percentage points. Intervals are 5,000-resample site-level percentile intervals; SD is variation among site means.

Indicator	Latest Day Mean (SD)	2026 Night Mean (SD)	Night – Day 95% Interval
Predicted water (%)	1.43 (2.42)	1.70 (3.19)	0.27 [-0.57, 1.19]
Dark pixels Y < 0.20 (%)	29.34 (12.43)	45.83 (11.80)	16.49 [12.79, 20.24]
Grid brightness ratio	0.259 (0.077)	0.214 (0.067)	−0.045 [-0.070, −0.022]
Mean JPEG luma	0.427 (0.080)	0.282 (0.057)	−0.145 [-0.168, −0.122]

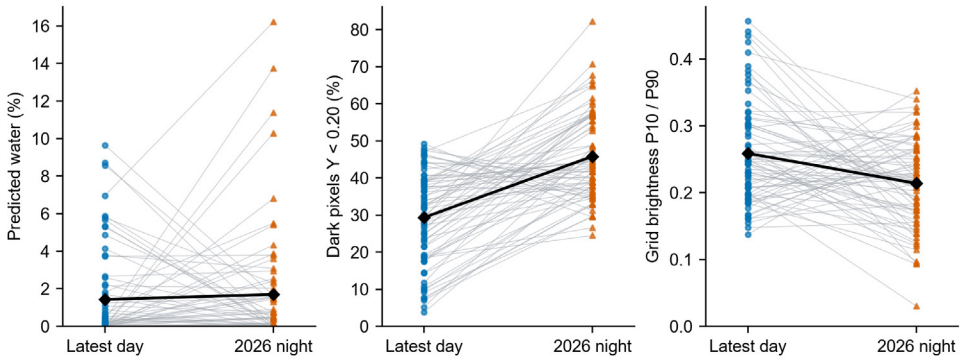


Figure 4: Site-average indicators for the primary sample. Thin lines join nominally corresponding sites; blue circles show daytime imagery and orange triangles show nighttime imagery. Black diamonds show the equally weighted mean across 71 sites. Each point averages four views. View geometry and capture years differ between periods.

Water predictions did not support a clear overall direction of difference. Mean site coverage was 1.43% by day and 1.70% at night; the difference was 0.27 percentage points, with a resampling interval from -0.57 to 1.19. Site-level medians were much smaller, 0.13% and 0.37%, respectively, showing that a limited number of water-positive views influenced the means. Of the 284 primary images per period, 203 daytime images and 209 nighttime images had no water-labelled pixels. A zero prediction cannot distinguish a genuinely water-free view from an occluded, dark or misclassified water surface.

Nighttime water predictions changed under global brightness transformations (Figure 5). With $\gamma = 0.7$, average predicted coverage across the 284 nighttime images changed from 1.70% to 1.86%, and the mean

absolute within-image change was 0.24 percentage points. With $\gamma = 1.3$, the corresponding average was 1.49% and the mean absolute change was 0.26 points. Changes of at least five percentage points occurred in 0 and 3 images, respectively. Mean binary mask disagreement was 0.29% and 0.29% of image pixels. These results demonstrate input sensitivity; they do not identify which version is more accurate.

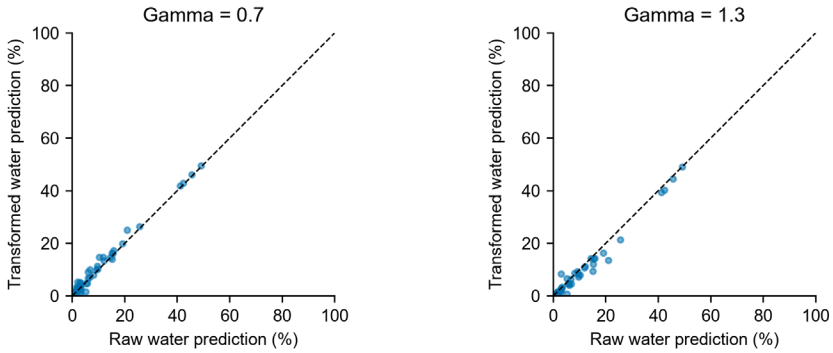


Figure 5: Water predictions before and after RGB power transformations, with 284 primary nighttime images in each panel. Points on the dashed line have unchanged predicted coverage. Movement away from the line reflects model sensitivity to brightness processing, without an observed change to the physical scene. Overlapping zero predictions are retained.

The darkness comparison persisted under the alternative thresholds. The night-minus-day mean difference was 11.58 percentage points for $Y < 0.10$ and 18.25 points for $Y < 0.30$. At the ten sites with 2022 daytime imagery, mean dark-area coverage at $Y < 0.20$ was 29.48% by day and 49.05% at night. Their mean grid ratio was 0.275 and 0.219, respectively. This subset provides a temporally closer descriptive check but remains small and geographically selective. Its water prediction difference was -0.11 percentage points, so the already uncertain aggregate water difference should not be treated as robust.

DISCUSSION

The strongest empirical pattern is a shift in the photographic distribution of dark and bright regions. Higher nighttime darkness and lower grid brightness ratios identify images that merit closer inspection of paths, shoreline edges, vegetation and localized light sources. Research on lighting and pedestrian reassurance provides a rationale for examining how users respond to illumination and its spatial distribution (Fotios et al., 2015; Portnov et al., 2020). The image indicators do not identify unsafe locations automatically. An image can contain a dark sky or dark water while retaining a clearly visible pedestrian route; a bright commercial frontage can coexist with a poorly defined edge. The whole-image indicators used here cannot separate those conditions without further semantic and field validation.

The weak and brightness-sensitive water results are themselves informative for method selection. Water-area estimates from a generic scene model should not be assumed to transfer reliably to night photographs. The ACDC benchmark documents difficulties in semantic scene understanding under adverse conditions, including nighttime (Sakaridis et al., 2021). Although that benchmark concerns driving scenes, it reinforces the need to validate transfer to a different domain such as waterfront imagery. The combination of small medians, many empty masks and changes under RGB power transformations warrants independent annotation before the indicator is used as a measure of actual water visibility. The sensitivity experiment also shows why photographic enhancement cannot be described as improved segmentation performance without labelled reference masks. A useful next validation would sample scenes across darkness levels and water appearances, annotate water boundaries independently and report class-specific errors separately for day and night.

For the human factors framework, three evidence streams should be linked without substituting one for another. The environmental stream contains image-derived candidates such as dark regions and possible water visibility. A perception stream would measure aesthetic quality, perceived opportunities to approach water, perceived safety and nighttime visitation intentions using participant responses. An activity stream would record actual waterfront use through time-stamped counts or other appropriately governed observations. The present study completes an exploratory environmental audit and defines the connections to be tested. It provides no empirical correlation between the three streams, and no basis for equating visitation intentions with observed vitality. Figure 3 illustrates this conceptual connection and distinguishes the completed image audit from the proposed validation.

Navarrete Hernandez et al. (2025) used an image-based randomized experiment to examine day–night safety perceptions, illustrating the value of collecting participant responses directly. A subsequent Shichahai study should use standardized contemporary photographs of the same viewpoints and randomize image order. Aesthetic quality, perceived water access, perceived safety and visitation intention should remain separate outcomes. Participant and site dependence would need explicit modelling when people rate multiple scenes. Activity observations should distinguish movement from lingering and record observation timing and duration. These data could test whether scenes that appear attractive by day are less reassuring at night; that mismatch remains a research question here, rather than an established result.

The principal limitations concern comparability, measurement validity and coverage. The four-to-thirteen-year temporal gap can include changes to vegetation, buildings, shoreline use or commercial activity. Different camera systems, aspect ratios and unknown exposures affect both brightness values and segmentation. The 512×512 model input also changes image aspect ratio before predictions are restored to the working-image frame. Averaging four views reduces dependence on a single named heading but does not standardize the field of view. No camera calibration, independent nighttime coordinate

survey or manual segmentation reference was available. Resampling intervals omit these uncertainties and possible spatial dependence. Supplementary nighttime photographs and duplicated or ambiguous identifiers cannot repair these limitations.

Practical interpretation should therefore remain diagnostic. The results justify checking whether water edges and walking routes remain legible in the scenes with extensive dark regions, and whether local lighting contrasts require a site visit. They do not justify increasing illumination everywhere, identifying crime risk, ranking demographic vulnerability or estimating the economic vitality of shoreline businesses. Design decisions require field measurements and user evidence, including accessibility needs and the consequences of additional artificial light for the waterfront setting.

CONCLUSION

An image audit of Shichahai compared 71 nominally corresponding sites using the newest available daytime imagery and author-confirmed 2026 nighttime photographs. The photographic dark-area fraction increased by an average of 16.49 percentage points, while the grid brightness ratio declined. SegFormer-B0 water predictions showed no stable overall directional difference and were sensitive to brightness transformations. These are findings about photographic records and a fixed segmentation model. The study supplies a reproducible environmental component for investigating visual quality, perceived safety and spatial vitality; the relationships with perceptions and observed use require new empirical data and more comparable imagery.

REPRODUCIBILITY AND EVIDENCE SCOPE

The accompanying reproducibility materials contain image-level metrics, site aggregates, selection records, model revision and weight hashes, processing scripts and binary prediction masks. Original photographs and map-provider images are not redistributed in the experiment package. Source imagery remains subject to the relevant permissions. No questionnaire, behavioral observation or crime dataset was analysed, and no claim of participant ethical approval is made. AI assistance was used in analysis scripting, drafting and language editing; generated numerical claims were checked against the computed output tables. The authors are responsible for source rights, scientific interpretation and approval of the final manuscript.

ACKNOWLEDGMENT

This work was supported by the 2026 Special Fund for Independent Research in the Tsinghua University Humanities under the project “Research on the Digital Construction of a Visual Genealogy of Traditional Chinese Water Culture” (Project No. 2026THZWC021).

REFERENCES

Fotios, S., Unwin, J. and Farrall, S. (2015). Road lighting and pedestrian reassurance after dark: A review. Lighting Research & Technology, 47(4), 449–469. https://doi.org/10.1177/1477153514524587

Navarrete Hernandez, P., Dong, Y. and Moresco, E. (2025). When daylight fades: How nighttime, sociodemographics, and urban zones shape safety perceptions of the built environment. Landscape and Urban Planning, 264, 105460. https://doi.org/10.1016/j.landurbplan.2025.105460

NVIDIA / Hugging Face (n.d.). nvidia/segformer-b0-finetuned-ade-512-512. Model checkpoint, revision 489d5cd81a0b59fab9b7ea758d3548ebe99677da. https://huggingface.co/nvidia/segformer-b0-finetuned-ade-512-512 (accessed 26 October 2026).

Portnov, B. A., Saad, R., Trop, T., Kliger, D. and Svechkina, A. (2020). Linking nighttime outdoor lighting attributes to pedestrians' feeling of safety: An interactive survey approach. PLOS ONE, 15(11), e0242172. https://doi.org/10.1371/journal.pone.0242172

Sakaridis, C., Dai, D. and Van Gool, L. (2021). ACDC: The Adverse Conditions Dataset With Correspondences for Semantic Driving Scene Understanding. Proceedings of the IEEE/CVF International Conference on Computer Vision. https://doi.org/10.1109/ICCV48922.2021.01059

White, M., Smith, A., Humphryes, K., Pahl, S., Snelling, D. and Depledge, M. (2010). Blue space: The importance of water for preference, affect, and restorativeness ratings of natural and built scenes. Journal of Environmental Psychology, 30(4), 482–493. https://doi.org/10.1016/j.jenvp.2010.04.004

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M. and Luo, P. (2021). SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers. Advances in Neural Information Processing Systems, 34. https://proceedings.neurips.cc/paper_files/paper/2021/hash/64f1f27bf1b4ec22924fd0acb550c235-Abstract.html

Zhou, B., Zhao, H., Puig, X., Fidler, S., Barriuso, A. and Torralba, A. (2017). Scene Parsing Through ADE20K Dataset. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. https://doi.org/10.1109/CVPR.2017.544