AI-Empowered Research on Healthy Streets: An Iterative Path of Streetscape Perception, Evaluation, and Optimization

Yu Yinqi[1], Liu Liu[1]*
Affiliation: 1. Department of Urban Planning, College of Architecture and Urban Planning, Tongji University, Shanghai 200092, China
Email:yinqi.yu@campus.tu-berlin.de; l.liu@tongji.edu.cn
* Correspondence: l.liu@tongji.edu.cn

Abstract
This study presents a closed-loop AI framework for assessing and enhancing "healthy streets," applied to downtown Haining City, China. First, semantic segmentation of street-view images yields eight quantitative indicators (e.g., greenery, enclosure, openness). Second, an Elo-based pairwise survey combined with Random Forest and CNN models reveals how these indicators predict perceived street healthiness. Third, we use a Low-Rank Adaptation model(Lora) in Stable Diffusion generator to optimise low-scoring scenes by adding health-supportive elements. Results indicate that (1) the eight indicators capture key environmental qualities; (2) Elo-scoring aligns with machine-learning feature importances—especially greenery, human-scale enclosure, and visual diversity; and (3) the generative model improves predicted health scores by enhancing greenery and openness. This iterative approach offers planners a rapid tool for healthy-street design.

Keywords: Healthy streets; Street view imagery; Semantic segmentation; Machine learning; Generative AI

1 Introduction

The idea of the Healthy City has put citizens' physical and mental well-being at the centre of urban planning since the late twentieth century. Among all public spaces, streets—the city's "capillaries"—shape daily activity, social contact, and psychological comfort more than any other setting. When they are walkable, human-scale, green, and orderly, streets foster exercise, lower stress, and lift overall happiness. Empirical evidence underscores this link. Ulrich's controlled experiments showed that simply viewing greenery or water restores positive mood, whereas scenes dominated by bare concrete increase sadness and anger (Ulrich, 1979). In urban design and public-health scholarship, such environments are now labelled healthy streets: places that invite walking and conversation, provide safety and comfort, and deliver a pleasing sensory experience (Chen et al., 2020; Xu & Hu, 2020). Creating and promoting healthy streets has therefore become a core objective of livable-city strategies.

Rapid urbanisation and motorisation have left many streets marred by poor pedestrian conditions, sparse greenery, excessive enclosure, visual clutter, and traffic hazards—factors that undermine public health and well-being(Chen, Zhang & Long, 2020). To address these deficits, researchers increasingly rely on quantitative evaluations that can steer evidence-based upgrades. The explosion of street-view imagery and advances in computer vision now make city-wide, objective assessments possible (Long & Tang, 2019; Ye et al., 2019). A seminal example is MIT Media Lab's Place Pulse, which crowdsourced judgements by asking online participants which of two street scenes appeared safer, livelier, and so forth. The resulting perception dataset underpinned StreetScore, a model that predicts perceived safety and vibrancy directly from images, enabling block-by-block comparisons across entire cities(Quintana, Gu & Biljecki, 2024). Subsequent deep-learning work has generalised these mappings worldwide, showing that AI can "read" visual cues—building massing, tree cover, human presence—and infer how people are likely to feel in those spaces. Such image-based analytics now offer planners a scalable lens on the everyday quality of urban life.

However, most current research is confined to measurement and prediction, with limited attention to how AI could improve street design. A notable exception is Wijnands et al. (2019), who trained a generative adversarial network (GAN) to rewrite street-view images and then estimated the health-related perception shifts caused by those edits. Their experiments showed that GAN-based style transfer can faithfully render "greener" versus "grey" streetscapes, yielding a virtual test bed for design interventions. This pivot from assessment to intervention points to a new frontier: harnessing generative AI to prototype and appraise street-design options before a single curb is moved.

Recently, researchers have made significant progress in two key areas. First, researchers have refined the measurement of objective streetscape features. Long & Tang (2019) showed that computer vision now enables city-wide appraisals with unprecedented scope and precision. Based on this, Ye et al. (2019) extracted fine-grained indicators—greenery cover, spatial openness, façade articulation—from street-view imagery to produce block-level quality maps. Second, studies have demonstrated how these features are translated into health-related outcomes. From a public-health stance, Chen, Zhang & Long (2020) argued that elevating a street's "spatial order"—reducing clutter, improving pedestrian continuity—directly enhances its health performance by removing discomfort factors such as noise, pollution, or disorganised furniture. Complementing this point, Xu & Hu (2020) proposed a "healing streets" model grounded in therapeutic-landscape theory, calling for generous greenery, water features, public art, and accessible open space to help residents relax and recover from stress.

Although existing work supplies robust theories and metrics for healthy-street assessment, three gaps remain. First, the link between objective spatial indicators and subjective well-being is only weakly understood; we still lack models that clarify how quantified image features translate into lived health benefits. Second, most intervention studies are theoretical—design prescriptions are seldom tested on real streetscapes. Third, this field lacks a closed loop that cycles from diagnosis to intervention and back to evaluation. We address these gaps by building an AI-enabled workflow that (i) diagnoses street health, (ii) generates targeted design alternatives, and (iii) iteratively verifies their predicted influence. By fusing multi-sourced imagery with complementary AI models, our approach shifts the agenda from data-driven perception to AI-assisted design, giving planners a scalable toolkit for healthy-street regeneration.

2 Study Area, Data and Methods

2.1 Study area

Haining, a county-level city in Zhejiang Province's Yangtze River Delta, serves as our case study. Nestled between Hangzhou and Shanghai, it hosts about 800 000 residents and an economically active urban core. We focus on four central sub-districts—Xiashi, Haizhou, Haichang, and Maqiao—which together span the historic downtown and its surrounding neighbourhoods. Within this zone, narrow heritage alleys, protected preservation blocks, modern commercial corridors, residential streets, and major arterials coexist. The resulting spectrum—from centuries-old lanes to newly built boulevards—mirrors the diversity found in many medium-sized Chinese cities, making Haining a fitting proxy for comparable contexts.

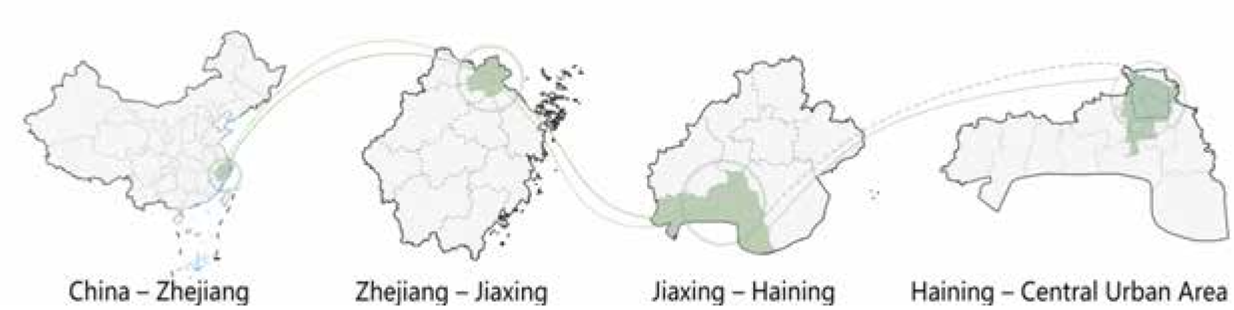


Figure 1. Study area: location and urban layout of Haining City

2.2 Data sources

Our analysis draws on four complementary datasets: geospatial vectors, street-view imagery, human-perception scores, and open-source generative-AI models.

·Geospatial vectors. Street-centreline geometries for the four sub-districts were downloaded from OpenStreetMap and cross-checked against Haining planning documents and statistical yearbooks to ensure coverage of all major land-use zones.

·Street-view imagery. Using the Baidu Street View API, we placed sampling points every 200 m along the network. At each of the 1077 points we captured the panorama and clipped two opposing perspectives aligned with the road axis, yielding 2154 images (1024 × 768 pixels). Shots were taken at pedestrian eye-level in fair daytime conditions during 2021–2022, spanning commercial cores, residential streets, park edges, and a few peril-urban roads (locations without imagery were excluded).

·Perception survey. To obtain subjective "street-health" scores, 50 participants (planners, public-health experts, and volunteers) performed Elo-style pairwise comparisons of randomly drawn image pairs (see section 5.1). The resulting crowd-sourced ratings provided training labels for our predictive models.

·Generative-AI toolkit. For the optimisation stage we employed Stable Diffusion 1.5 enhanced with Low-Rank Adaptation (LoRA) and ControlNet guidance. Pre-trained weights, open-source scripts, and a curated prompt library supplied the basis for domain-specific fine-tuning.

Figure 2 Road network and streetscape observation points in the study area

2.3 Methodological framework

Our study follows a three-stage, closed-loop workflow—Perception, Evaluation, and Optimisation.First perception, we run semantic segmentation on every street-view image to delineate key urban elements and compute objective visual indicators.Second evaluation, an Elo-style pairwise survey first yields subjective "street-health" scores. These labels feed two predictive models: a Random Forest regressor trained on the indicator set and a convolutional neural network (CNN) trained on raw pixels. Together they explain how specific features shape perceived health and allow city-wide scoring. Finally optimisation, low-scoring scenes are redrawn with Stable Diffusion 1.5, fine-tuned via Low-Rank Adaptation (LoRA) and guided by ControlNet. The generator injects health-supportive cues—extra greenery, broader sky views, livelier façades—and the edited images are re-scored by the RF/CNN models (and inspected by eye). If gains fall short, the edit-evaluate cycle repeats until the target threshold is met.

By chaining computer vision, predictive modelling, and generative design, the framework moves seamlessly from diagnosis to intervention and back, offering planners a virtual test bench for healthy-street retrofits. The next sections unpack each stage in detail.

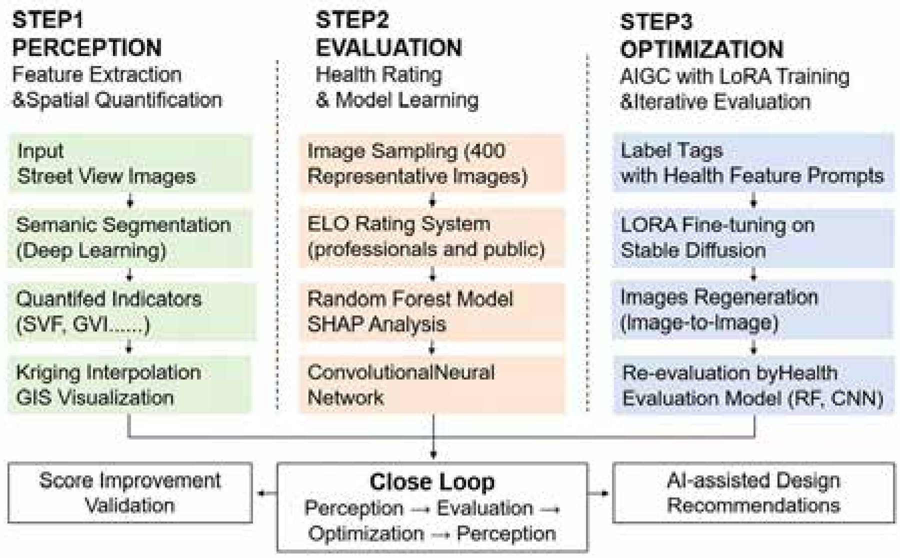


Figure 3 Flow chart of the overall methodological framework of the study

3 Perception: Visual-Indicator Extraction and Analysis

3.1 Semantic segmentation of street-view imagery

We applied semantic segmentation to all 2154 street-view photographs to convert raw pixels into meaningful urban elements. Using DeepLab-v3+ with a ResNet backbone—pre-trained on the Cityscapes corpus and lightly fine-tuned on 200 hand-labelled Haining scenes. Class masks were then aggregated into eight quantitative indicators aligned with four perceptual dimensions:

The eight visual indicators fall into four perceptual dimensions: Vibrancy, represented by Facade Diversity (FD) and Colour Richness (CR), captures the heterogeneity and chromatic variety of street façades; Comfort, measured by Building Enclosure (BE) and Height–Width Ratio (H/W), reflects spatial enclosure and human-scale proportion; Naturalness, via Sky View Factor (SVF) and Green View Index (GVI), assesses visual access to open sky and vegetation; and Safety, defined by Relative Walking Width (RWW) and Traffic-Facility Proportion (TFP), quantifies pedestrian space and the presence of traffic-management elements

To sum up, these metrics furnish a compact, interpretable synopsis of each streetscape's physical affordances for health. Table 1 details indicator formulas and class mappings, while Figure 4 illustrates a typical segmentation output.

Table 1 Definitions and calculation methods for the eight spatial indicators

Category	Indicator Name	Calculation Formula
Vitality	Facade Diversity	$H = -\sum(p_i * \ln p_i)$,
Vitality	Color Richness	Entropy of HSV hue channel / $\log_2 36$, normalized to [0,1]
Comfort	Building Enclosure	Proportion of building elements
Comfort	Height-Width Ratio	(Building + Wall) / Road Elements
Nature	Sky View Factor	Proportion of sky pixels
Nature	Green View Index	Proportion of vegetation pixels
Safety	Relative Walking Width	Sidewalk / (Driveway + Trail)
Safety	Traffic Facilities	Proportion of traffic-related elements

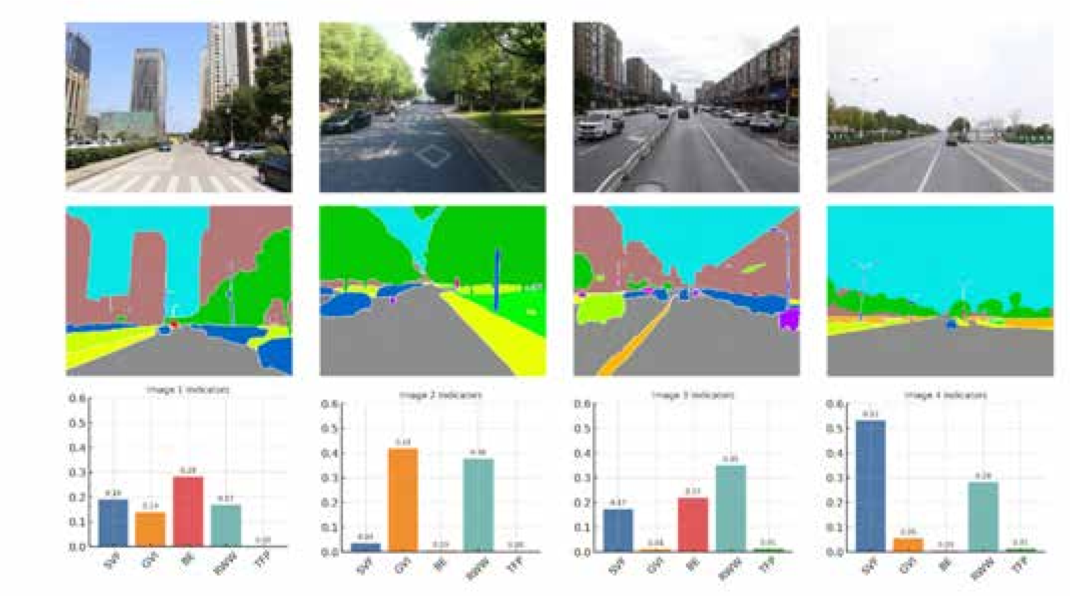


Figure 4 Street view image with semantic segmentation results

3.2 Indicator distribution and inter-relationships

After computing the eight indicators for all 2154 images, we examined their statistical profiles and pairwise links. Figure 5 plots the histograms, and Figure 6 shows the correlation heat-map.

The eight semantic indicators exhibit considerable variation across the dataset, capturing differences in vegetation, enclosure, openness, visual diversity, and pedestrian space. Some indicators, such as green view index and sky view factor, span a wide numerical range, while others, like traffic facility proportion, remain low but meaningful. These metrics provide a comprehensive quantitative basis for analyzing spatial patterns of perceived street healthiness.

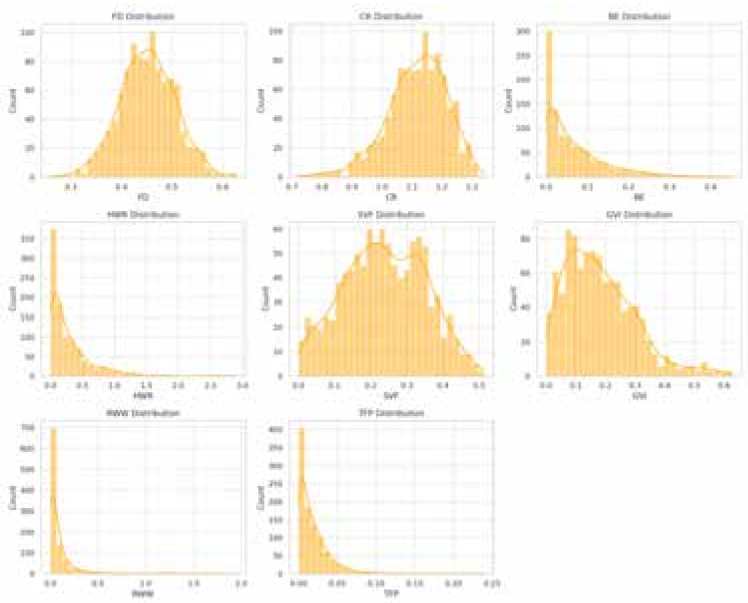


Figure 5 Histogram data distribution by indicator

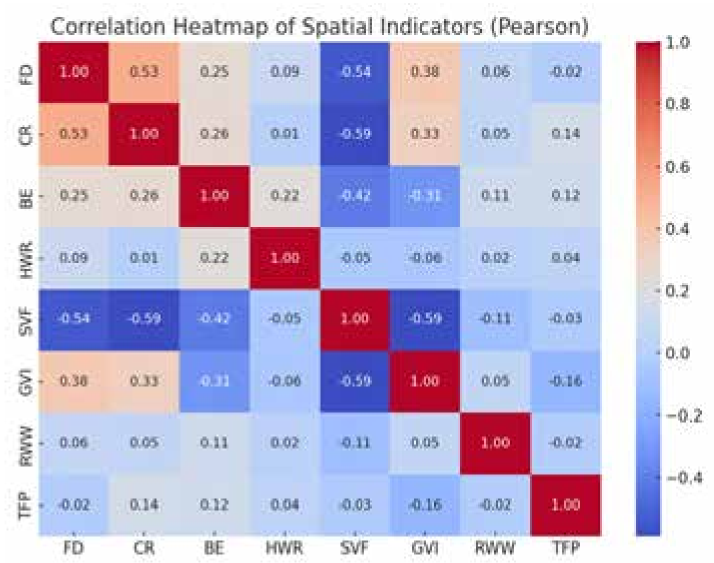


Figure 6 Heatmap of indicator correlations($p < 0.05$)

To read the eight indicators as coherent street profiles, we ran k-means clustering with k = 4. The feature set comprised the five most salient indicators—selected by variance and preliminary importance tests—which retained > 92 % of total information. Table 2 lists the centroid scores, Table 3 sketches the corresponding spatial traits, and Figure 7 plots the four centroids on a radar graph. The clusters form a clear typology. This four-way classification distils Haining's diverse streetscape morphologies into interpretable groups that anchor the subsequent evaluation stage.

Table 2 Clustering means

Cluster	Sky Openness	Green View Index	Building Enclosure	Road Width	Traffic Facilities
0	0.1343	0.1026	0.238	0.3027	0.0166
1	0.1394	0.3405	0.0385	0.2718	0.011
2	0.2283	0.1564	0.0726	0.2676	0.0766
3	0.3275	0.1187	0.0353	0.3296	0.0116

Table 3 Introduction to clustering types

Cluster	Street Type	Feature Keywords	Spatial Interpretation
0	Medium Road, High Enclosure	High enclosure (0.238), Medium road width (0.30)	Dense buildings on both sides, limited view, low greenery and sky, typical urban main road.
1	Open Green Type	Highest GVI (0.3405), Low enclosure & traffic facility	Rich vegetation, open and comfortable, few buildings, suitable for residential or park-side streets.
2	Infrastructure Enhanced	Most facilities (0.0766), Moderate enclosure, Mid-level openness	Strong infrastructure presence, visually active, functional streets or commercial corridors.
3	Wide Open Arterial Road	Highest openness (0.3275), Widest road (0.33), Few buildings	Extremely open, low architectural presence, likely highways or suburban roads.

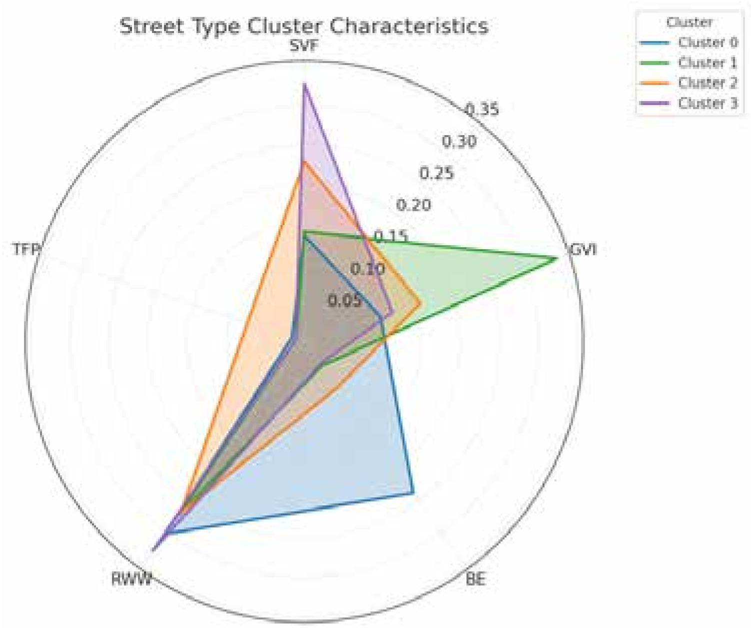


Figure 7 Radar diagram of clustering features

3 .3 Spatial distribution of street indicators

To locate where healthy-street attributes wax and wane across Haining, we interpolated each indicator from the 1 077 sampling points onto a continuous surface and rendered heatmaps (Figure 8).

·GVI ranges from less than 5% to over 60%, with a mean around 25%. Vegetation is heavily concentrated along the northern and northeastern river corridors, where greenways and linear parks flank the water. In contrast, the dense downtown and the industrial southwest show severe vegetation deficits, making them visually "grey" and in need of greening.

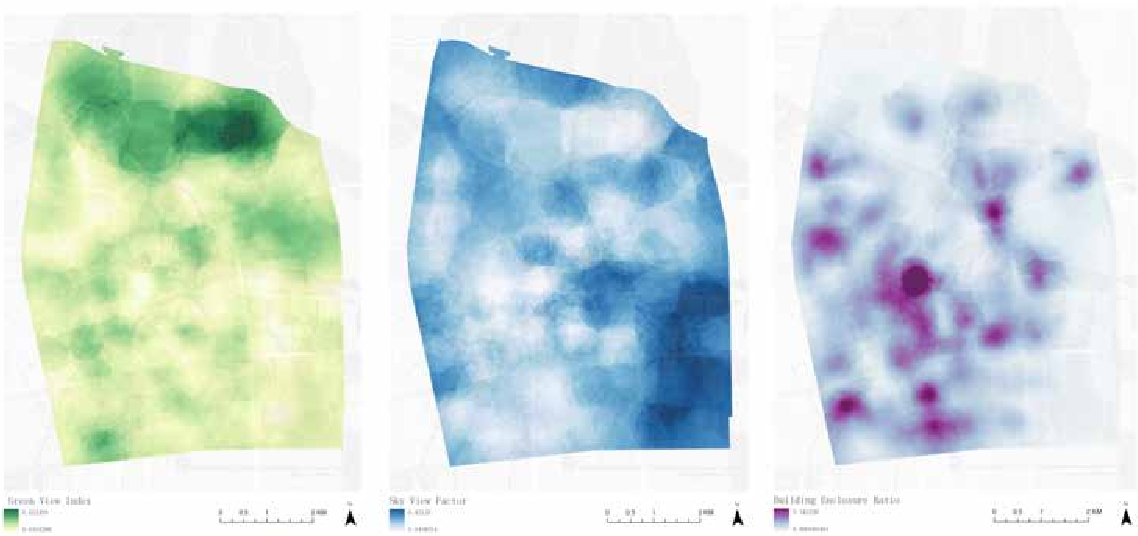
·SVF averages approximately 30%, but varies widely. High SVF values are found in southeast and fringe areas with fewer vertical obstructions, while low values occur in historic quarters where mid-rise blocks obstruct sky views and create enclosed street canyons.

·BE inverts the GVI trend. Traditional urban cores such as Haizhou and Haichang have high enclosure due to tightly built, narrow street networks. Newer districts like Maqiao, with wider setbacks and open cross-sections, register low enclosure values, contributing to a more open spatial feel.

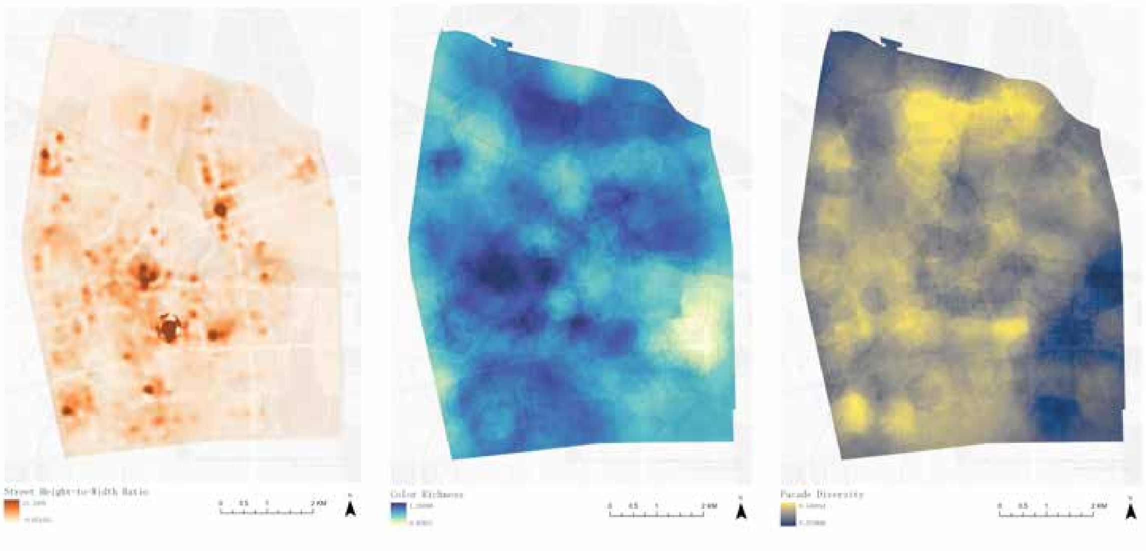
·H/W is strongly correlated with BE. Peaks occur along older corridors like Renmin and Yucai Roads, forming deep urban canyons with narrow view corridors. These areas may feel intimate, but risk reduced comfort under extreme conditions.

·CR and FD rise in commercial and mixed-use zones featuring varied signage, shopfronts, and building types. Uniform residential blocks and industrial areas score lower. These two indicators show a moderate positive correlation, as both are enhanced by land-use diversity and design articulation.

Taken together, the heatmaps reveal a stark spatial contrast in street environments—green, pedestrian-friendly corridors stand in sharp contrast to vegetation-poor, car-oriented zones in the historic core and industrial fringe. These disparities highlight priority areas for intervention, such as adding greenery, widening sidewalks, or improving traffic safety, and set the stage for the perceptual evaluation presented in the next section.



Top row (left to right): Green View Index, Sky View Factor, Building Enclosure



Bottom row (left to right): Height–Width Ratio, Colour Richness, Facade Diversity
Figure 8 Results of spatial interpolation of indicators

4 Evaluation: Street-Health Evaluation via Machine Learning

4.1 Elo-based collection of perceived-health scores

Because "healthy streets" are ultimately judged by people, we first needed a robust measure of perceived healthiness for each scene. We therefore adopted an Elo-style pairwise comparison, a method popular in online gaming and mirrored in the Place Pulse project, to translate thousands of binary choices into a continuous rating scale.
A stratified sample of 400 images—drawn evenly from the four clusters described in § 3.3—served as the test set. An online survey then showed planners, public-health scholars, and lay volunteers random image pairs and asked, Which street looks healthier? Respondents judged greenery, openness, comfort, safety, and overall aesthetics; after each "match" the Elo algorithm updated both images' scores. Ratings stabilised after 10000 comparisons, yielding a scale from about 800 to 1 300. Inter-rater agreement was high (Kendall's W > 0.80), confirming that the scores capture a shared perception of street health. These Elo values can be used as basic truth labels for the predictive models to be presented below.

Figure 9 ELO evaluation system interface

4.2 Random-Forest regression and feature importance

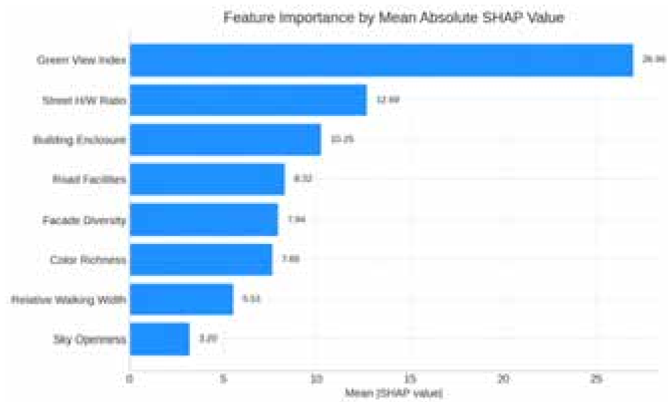
With Elo scores established for 400 street-view scenes, we next linked those subjective ratings to the eight objective indicators. A Random Forest (RF) regressor—well suited to modest sample sizes and non-linear feature interactions—served as the workhorse.

Model set-up. The images were split 80 / 20 into training (n = 320) and test (n = 80) folds. Input features were the eight indicators introduced in § 3.1; the target was the Elo score. A grid search with five-fold cross-validation tuned hyper-parameters. The best configuration—300 trees, maximum depth = 6, minimum samples per leaf = 2, delivered an out-of-sample $R^2 \approx 0.38$. Given the inherently subjective target, this indicates that the indicator set explains a substantial share of perceived street health, while leaving room for latent factors such as noise, smell, or social cues.

To unpack the model's decisions we applied Shapley Additive Explanations (SHAP). Figure 10 ranks features by their mean absolute SHAP value, and Figure 11 plots the full SHAP summary. Three findings stand out:

·GVI dominates the ranking, confirming greenery as the strongest visual cue of street health.
·BE and H/W follow, showing that human-scale enclosure matters, but only up to a point—very high values depress scores.
·CR and FD contribute positively, whereas TFP exerts only a minor influence, perhaps because safety cues are subtle in static imagery.

The RF model quantifies key design principles—more greenery, balanced enclosure, pedestrian-friendly features, and lively façades all contribute to higher perceived street health. These insights validate the indicator set and guide the optimisation stage by highlighting which attributes most efficiently lift perceived health.



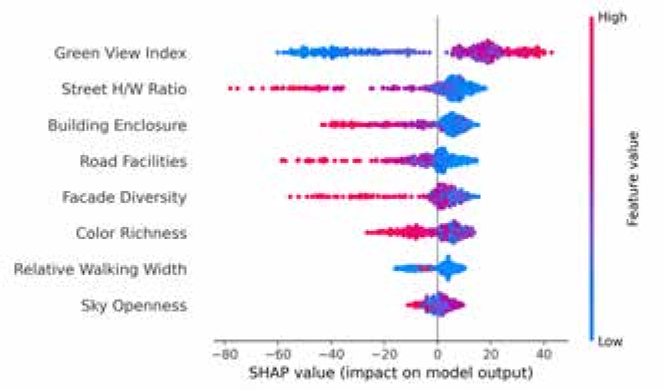


Figure 10 SHAP Importance Score (left)
Figure 11 SHAP Scatter Plot Analysis(right)

4.3 Convolutional Neural Network Prediction

To assess whether a deep model can learn health-related cues without predefined indicators, we trained a ResNet-50 convolutional neural network (with ImageNet weights) to directly regress Elo scores from street-view imagery, using the same 400 labelled images as in the Random Forest model.

The dataset was stratified 80/20 into training (n = 320) and validation (n = 80) sets. Each image was resized to 256 × 256 pixels, with any empty space padded with black pixels to preserve full scene content. The model was trained to minimise mean squared error, with early stopping applied once validation loss plateaued.er the small sample size.The model achieved an R^2 of approximately 0.48 and a mean absolute error of 72 Elo points, slightly outperforming the Random Forest, despite receiving no structured input features. High-scoring predictions consistently aligned with green, bright, and tidy scenes, while grey, empty, or car-dominated streets scored lower—paralleling SHAP findings from § 4.2. This confirms the CNN's capacity to autonomously extract relevant perceptual cues.

Table 4 Demonstration of RF model parameters

Parameter / Metric	**Value**
Model Type	Random Forest Regressor
Best n_estimators	300
Best max_depth	15
Best min_samples_leaf	2
Best max_features	sqrt
R^2 on Test Set	0.381
MSE on Test Set	≈ 1350

Table 5 Demonstration of CNN model parameters

Model	**EfficientNet-B0**
Input Size	224x224
Batch Size	32
Epochs	25
Learning Rate	1e-4
Optimizer	Adam

Loss Function	MSE (weighted)
Train Loss (final)	0.0023
Val Loss (final)	0.0172
MSE	7336.57
MAE	71.81
R^2	0.4792

Across the full 1,754-image dataset, the CNN's predictions correlated moderately with RF results (r ≈ 0.64), with both models identifying consistent high and low performers. Divergences likely reflect finer-grained visual textures perceived by the CNN but not captured by the indicator set, such as ground paving and street tidiness. Together, the two models offer complementary insights: strong agreement reinforces key design priorities (like greenery and human-scale proportion), while their differences point to additional visual nuances—some of which will be explored through generative enhancement in section 5.

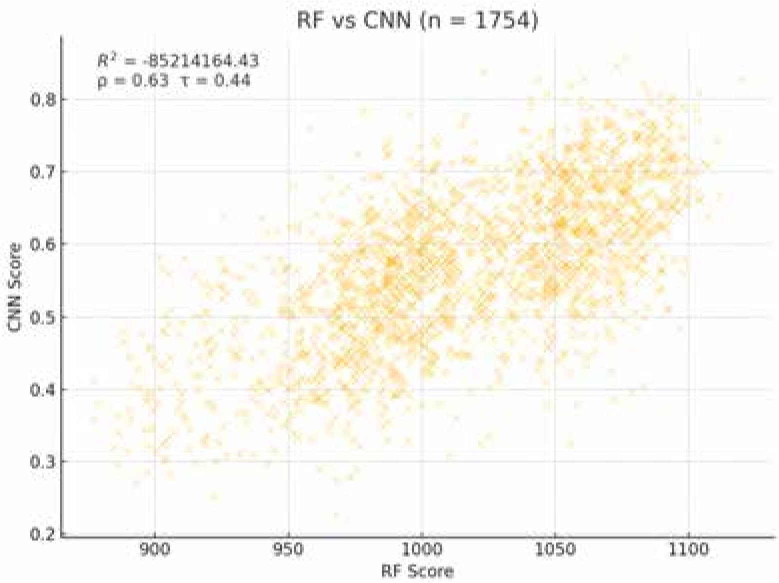


Figure 12 RF and CNN model prediction score correlation

5 Optimization: AI-Based Streetscape Optimisation and Iterative Validation

5.1 LoRA-fine-tuned image generation for healthy streets

With the key health features defined, we shifted from evaluation to visual intervention. To simulate healthier versions of low-scoring streets, we adopted an image-to-image generation approach using Stable Diffusion, and fine-tuned it via LoRA. This allowed the model to learn and apply health-promoting visual styles—while preserving the core structure of the original image.

(1)From Indicator Scores to Labels

The fine-tuning pipeline began with the 400 Elo-scored scenes. Each image was placed into one of three perceptual tiers—high, medium, or low health—defined by the upper, middle, and lower quartiles of the score distribution. Rather than feed the model cryptic numerical features, we distilled the eight indicators into short, human-readable phrases that echoed how survey participants had talked about the scenes. A frame with GVI ≈ 0.05 became "no trees"; GVI ≈ 0.25, "some street trees"; GVI > 0.40, "lush trees on both sides." Streets with BE and H/W at historic-core levels were tagged "narrow, tall buildings," whereas wide suburban corridors read "broad and open." Colour palettes and façade variety were rendered as "dull grey" or "bright and colourful." These textual summaries were paired with the source images so that the generator learned not only which elements depressed a scene's health rating, but also how a healthier variant should look.Table 6 documents how to route from metric scores to specific labels, and Figure 13 illustrates the labels for a typical street map.

Table 6 Metrics and Labeling Transformation Logic

Indicator	Value Range	Level	Tags
facade diversity	≤ 0.3301	Low	plain building walls
facade diversity	0.3301 – 0.4905	Medium	simple facades with some variation
facade diversity	> 0.4905	High	street-level shops and balconies
color richness	≤ 0.9007	Low	gray and monotonous facades
color richness	0.9007 – 1.1294	Medium	some colored elements
color richness	> 1.1294	High	colorful signs and painted walls
building enclosure	≤ 0.2336	Low	buildings far apart
building enclosure	0.2336 – 0.4673	Medium	moderate space between buildings
building enclosure	> 0.4673	High	buildings tightly packed
street height width ratio	≤ 0.0198	Low	wide street with low-rise buildings
street height width ratio	0.0198 – 0.4810	Medium	balanced street scale
street height width ratio	> 0.4810	High	narrow road with tall buildings
sky openness	≤ 0.1904	Low	sky mostly blocked by buildings
sky openness	0.1904 – 0.3808	Medium	partial sky between buildings
sky openness	> 0.3808	High	open sky over wide street
green view index	≤ 0.2357	Low	no visible trees
green view index	0.2357 – 0.4714	Medium	scattered trees
green view index	> 0.4714	High	trees along both sides
relative walking width	≤ 0.0140	Low	narrow pedestrian strip
relative walking width	0.0140 – 0.0670	Medium	standard pedestrian space
relative walking width	> 0.0670	High	wide sidewalk with benches
road facilities	≤ 0.1395	Low	no visible traffic equipment
road facilities	0.1395 – 0.2791	Medium	some road markings
road facilities	> 0.2791	High	traffic lights and street signs
street health	≤ 850	Low	unhealthy street
street health	850 – 1075	Medium	average street
street health	> 1075	High	healthy street

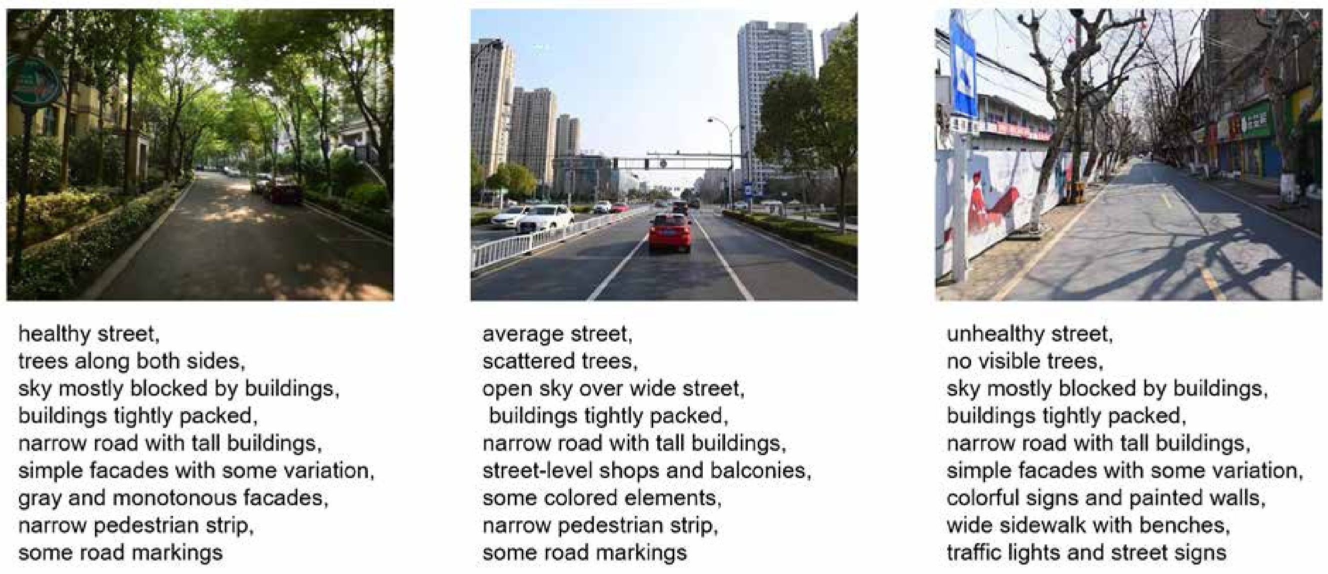


Figure 13 Representing the labeling results for Street View

(2)Lora Model Training

To enable health-aware generation, we fine-tuned the latent diffusion model AWPortrait_v1.4 using Low-Rank Adaptation. This lightweight approach inserts a small set of trainable matrices into the frozen UNet backbone, allowing the model to adapt to the new task—enhancing streetscapes for visual health—without updating the full network.

We used the same 400 Elo-scored images as in earlier models, resizing each to 512 × 512 pixels. Each image was paired with a prompt from the previously constructed label dataset, which provided explicit guidance on how healthy street features are visually expressed.

Training ran for approximately 16000 steps, with a learning rate of 1×10^{-4} and batch size of 4. We adopted an L1 loss function to guide the model towards stylistic enhancement without distorting structural geometry. As shown in Figure 14, the average training loss decreased from 0.30 to below 0.14, with key inflection points around steps 5000, 11000, and 14500. This indicates that the model successfully internalized the stylistic cues of healthier streets—such as greener vegetation, warmer colors, and cleaner appearances.

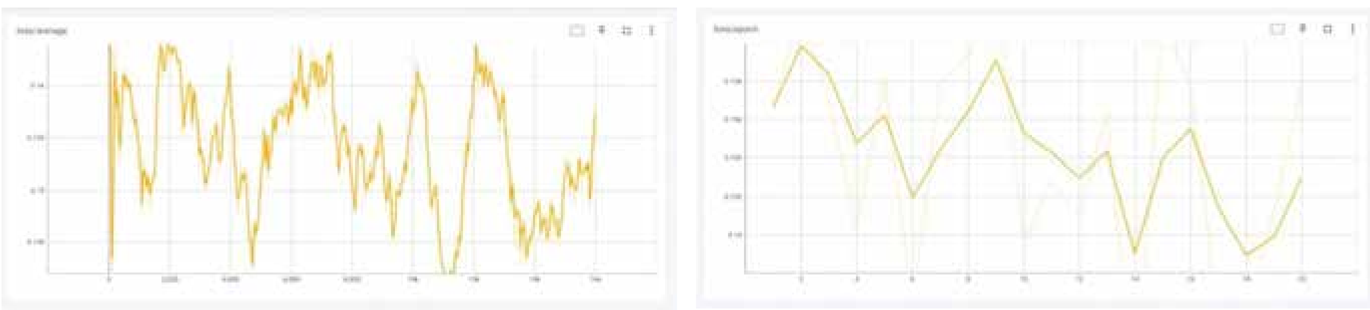


Figure 14 Variation of loss values for Lora model training

(3)Image-to-Image Regeneration

After fine-tuning, the LoRA-enhanced model was used to upgrade low-health street views via an image-to-image approach. Each original image was paired with a descriptive prompt that conveyed its intended transformation (e.g., "add lush trees," "widen sidewalks," "brighten facades"), and mild Gaussian noise was introduced to trigger generative variation.

To preserve spatial consistency, we incorporated ControlNet during inference. The model was switched to a large-scale base model ("Urban Streetview"), and semantic segmentation masks—obtained from previous analysis—were used as ControlNet input. This anchored the spatial structure of key elements such as roads, buildings, sky, and trees, ensuring layout fidelity while allowing stylistic transformation through the LoRA weights.

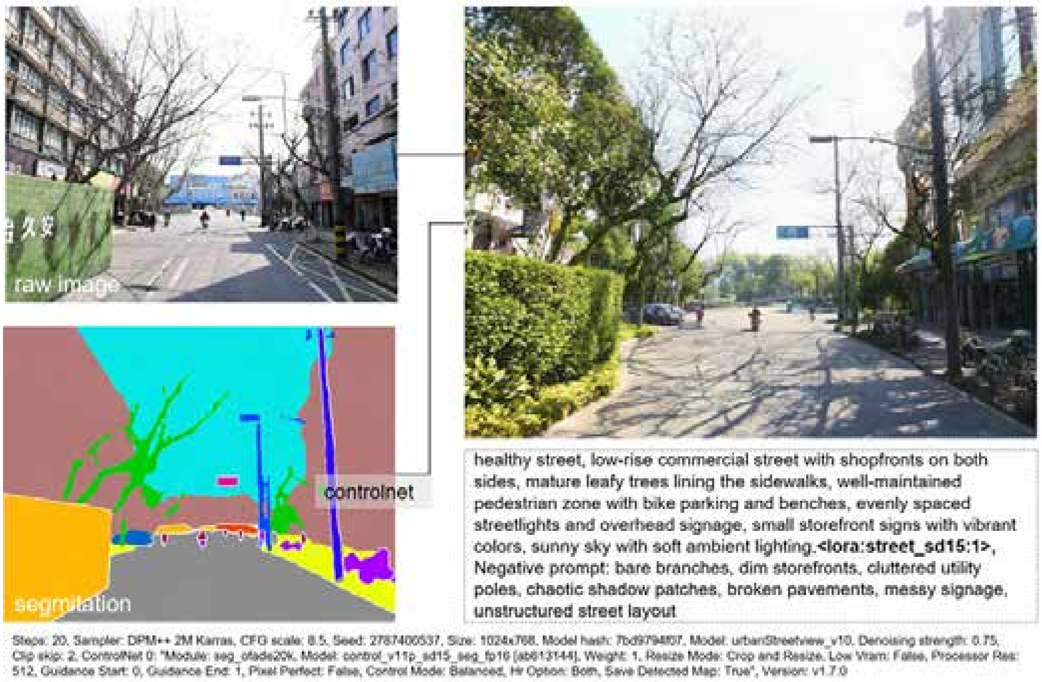


Figure 15 Typical street streetscape optimization process, sample 1

Figure 15 (Commercial Block Transformation) illustrates a typical result. The original scene shows a dull commercial street with gray façades and sparse vegetation. The generated version retains the layout but adds mature street trees, warm color tones, and new signage—creating a livelier and more inviting atmosphere.

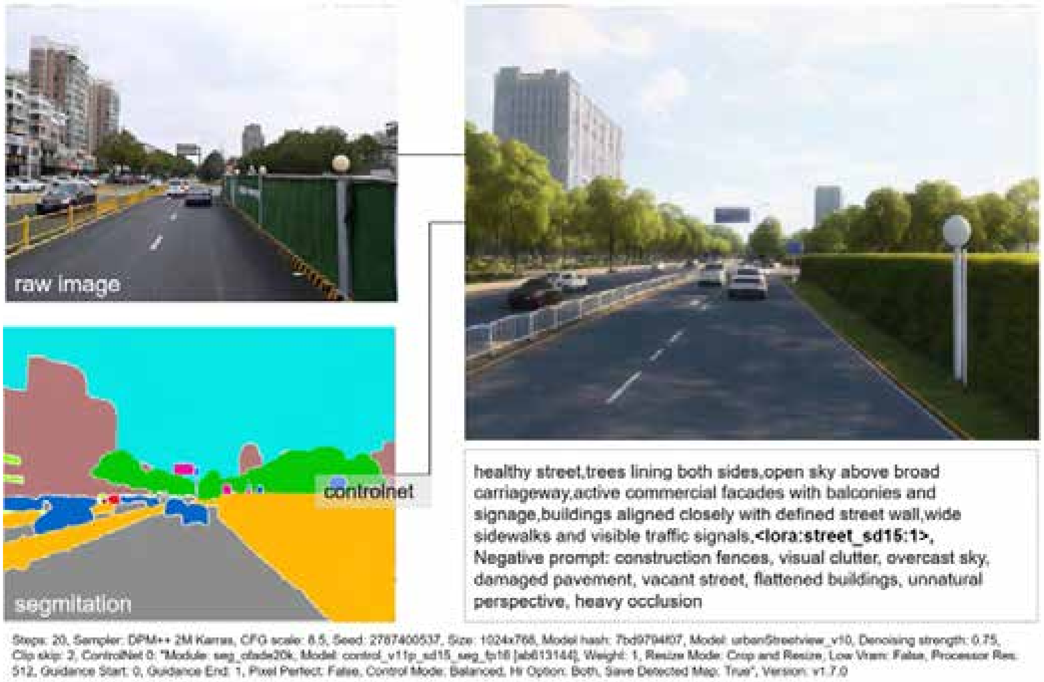


Figure 16 Typical street streetscape optimization process, sample 2

Figure 16 (Arterial Road Renewal) presents a congested urban road flanked by mid-rise buildings. The optimized output introduces planting strips, a brighter sky, and softened tones, improving the spatial feel while keeping the building envelope intact.

These examples confirm that the model has learned a coherent "visual grammar" of healthy streets. With only a semantic layout and a prompt, it can inject greenery, vibrancy, and visual order—offering a rapid, automated method for visual upgrading of urban streetscapes.

5 .2 Batch enhancement: results and evaluation

Encouraged by the single-scene trials, we ran the LoRA-adapted Stable Diffusion model on the remaining 1754 street views—every image outside the Elo survey. Using an NVIDIA RTX 4060, the image-to-image pipeline rendered the entire set in roughly one hour, underscoring the practicality of city-scale, AI-assisted retrofitting.

Figure 17 Batch optimization of Street View images

Figure 17 samples the outcomes. Across commercial corridors, residential lanes, and arterial roads, the AI generator consistently inserted or intensified health-supportive cues—canopy trees, brighter façades, tidier shopfronts, clearer skies—while retaining each street's geometry and landmarks. The resulting portfolio reads as recognisably Haining, yet palpably more verdant, colourful, and inviting.

To judge whether these cosmetic upgrades translate into a higher perception of street health, we re-scored every "before" and "after" image with RF and CNN models introduced in valuation section.

·RF Model: Predicted Elo scores rose by an average of +18.7 points, with a median gain of +26.5. Three-quarters of the images improved by more than +20 points. Only a handful of streets near the original ceiling showed little change, indicating that the model's headroom, not a failure of the generator, limited further gains.

·CNN Model: On the CNN's 0–1 scale, scores increased by ≈ 0.18 on average and ≈ 0.20 at the median. Many streets jumped by 0.4–0.5—an improvement large enough to shift them from the lower to the upper half of the distribution. A few scenes remained flat, typically where the generator's edits fell short of adding sufficient greenery or light.

Table 7 Score improvement under RF and CNN model scoring

Metric	RF Prediction Gain	CNN Prediction Gain
Sample Size	1754	1754
Mean Score Gain	41.89	0.065
Standard Deviation	40.3	0.086
Median Gain	37.59	0.059
Minimum Gain	-104.31	-0.205
Maximum Gain	202.88	0.43
Interquartile Range (25%–75%)	14.07 – 66.12	0.007 – 0.119

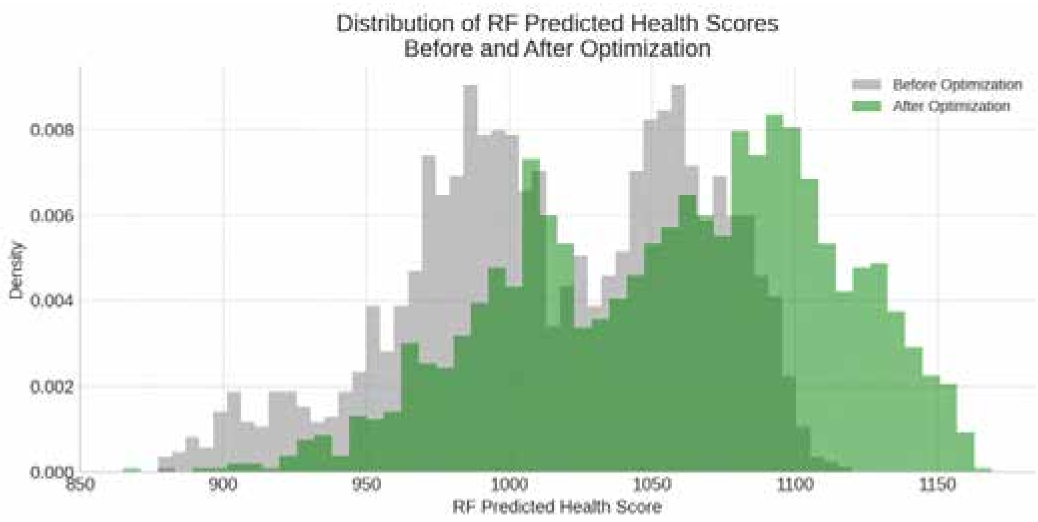


Figure 18 Histogram of RF and CNN model score changes

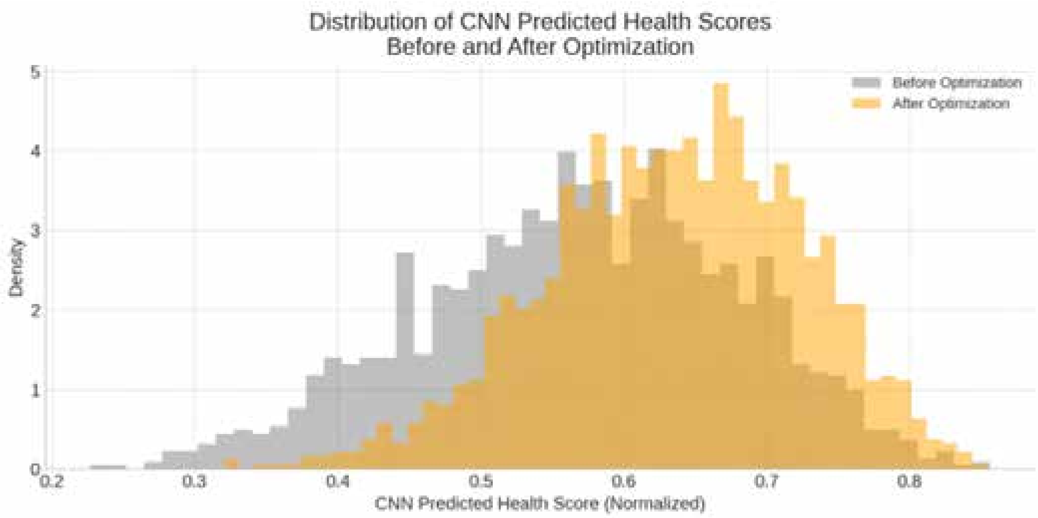


Figure 19 Histogram of RF and CNN model score changes

Taken together, the dual-model evaluation confirms that the AI intervention not only beautifies images but also measurably lifts the predicted healthiness of streets at scale. This quantitative uplift, combined with the rapid run-time, positions the workflow as a practical tool for planners seeking quick, evidence-based previews of healthy-street retrofits.

6 Conclusion

This study introduces an AI-driven workflow that links perception, evaluation and optimisation in a single, self-updating loop, and tests it on the streets of Haining. Deep-learning segmentation translated raw street-view imagery into eight visual indicators—greenness, enclosure, sky exposure, façade colour diversity and so on—whose city-wide maps expose where healthy-street qualities thrive and where they are scarce. Machine-learning models, trained on crowd-sourced Elo judgements, then bridged those objective metrics with subjective well-being: the Random Forest revealed that canopy trees, human-scale proportions, pedestrian infrastructure and visual vibrancy are the strongest predictors of health, while a convolutional network, working directly from pixels, echoed the same hierarchy. Finally, a Stable Diffusion generator, fine-tuned with Lora and anchored by ControlNet masks, recast low-scoring scenes as greener, brighter and tidier streetscapes; re-scoring with the predictive models showed a clear, quantitative uplift. In short, the pipeline moves seamlessly from diagnosis to design suggestion and back to diagnosis, delivering a rapid, evidence-based aid for street retrofits.

This study introduces two major innovations that set it apart from prior work on street environment evaluation and design:
(1)A closed-loop methodology bridging perception, valuation, and intervention.Where many previous studies focus only on visual diagnosis or abstract design ideas, our research establishes a full-cycle framework that integrates objective spatial analysis, subjective well-being evaluation, and AI-based visual enhancement. By building a consistent pipeline—from semantic indicators to perception models to targeted visual regeneration—we enable not just problem identification, but also evidence-based solution generation and feedback verification.

(2) Deep integration of AI across the workflow, including pioneering use of generative models.We leverage state-of-the-art AI technologies throughout the process: deep-learning segmentation for semantic features, machine learning and CNNs for health score prediction, and—most notably—a fine-tuned generative diffusion model (via LoRA) for street-level transformation. Unlike typical AIGC applications focused on artistic or entertainment content, we demonstrate its use for purposeful, rule-informed urban design. By teaching the model to understand and apply the visual logic of healthy streets, we show that AI can not only see and evaluate cities but also reshape them—offering planners a novel and scalable tool for visual experimentation and public engagement.

But this study remains an initial probe. Haining is a medium-sized city with a mid-rise fabric; megacity canyons or small-town main streets may challenge the models in unforeseen ways, so cross-city replication is essential. Our indicator set is intentionally visual, yet everyday street health also hinges on sound, air quality and micro-climate; coupling image analysis with sensor feeds would yield a fuller diagnosis. In addition, the perceptual ground truth comes from a one-off survey. Embedding the system in an online platform where residents can iteratively rate—or even tweak—virtual street views would create a living feedback loop that refines both the evaluators and the generator while deepening civic participation.

Artificial-intelligence tools now let us see, evaluate and redesign the public realm in a single, accelerated workflow. By uniting quantitative perception analysis with generative design, the framework sketched here moves urban practice towards an interactive, evidence-based craft—a step, we hope, toward streets and cities that are not only smarter, but tangibly healthier and more delightful for those who use them.

Acknowledgements
The research was sponsored by Natural Science Foundation of Shanghai (Grant No. 24ZR1469300).

References
Ulrich, R.S. (1979) 'Visual landscapes and psychological well-being', Landscape Research, 4(1), pp. 17–23.

Wijnands, J.S., Nice, K.A., Thompson, J., Zhao, H. and Stevenson, M. (2019) 'Streetscape augmentation using generative adversarial networks: insights related to health and wellbeing', Sustainable Cities and Society, 49, 101602.

Long, Y. and Tang, Y. (2019) 'Research progress on large-scale quantitative measurement of urban street spatial quality based on street view images', Urban Planning, (6), pp. 107–114. (Chinese Journal)

Ye, Y.Y., Zhang, Z.X., Zhang, X.H. and Zeng, W. (2019) 'Measurement of street space quality at human scale: a high-precision evaluation framework combining street view data and new technologies', International Urban Planning, 34(1), pp. 18–27. (Chinese Journal)

Chen, J.J., Zhang, Z.X. and Long, Y. (2020) 'Strategies for improving the quality of urban street space oriented to promoting public health: perspective from spatial disorder', Urban Planning, 44(9), pp. 35–47. (Chinese Journal)

Xu, L.Q. and Hu, Y.Z. (2020) 'Healing streets: a new model of healthy streets', Time + Architecture, (5), pp. 33–41. (Chinese Journal)

Quintana, M., Gu, Y. and Biljecki, F. (2024) 'My street is better than your street: Towards data-driven urban planning with visual perception', in Proceedings of the 11th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys).