# Analysis of Influencing Factors of Green View Index Based on Street View Segmentation

Xinyuan Xu[1], Lei Niu[1,2,3,4]

[1] Faculty of Geomatics, Lanzhou Jiaotong University, Lanzhou 730070, Gansu, China - liweiyang617@gmail.com; l.niu@lzjtu.edu.cn

[2] Geospatial Research Innovation Development, School of Built and Environment, Faculty of Art and Design, University of New South Wales, Australia, Kensington 2030, lei.niu@unsw.edu.au

[3] National-Local Joint Engineering Research Center of Technologies and Applications for National Geographic State Monitoring, Lanzhou 730070, Gansu, China

[4] Gansu Provincial Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China

**Keywords:** Street View Image; Semantic Segmentation; PSAT-Net Model; Green View Index

**Abstract**

The calculation of green view index in street view image has become an emerging choice for many researchers. This article proposes a PSAT-Net semantic segmentation model and specifically designs a pyramid pooling attention module. This module cleverly integrates channel and spatial information, implicitly implementing a three-dimensional attention mechanism, and significantly improving performance indicators. By performing detailed semantic segmentation on street view image, this article deeply analyzes the impact of different street view parameter configurations on green view index, and studies in detail how these parameters affect the visual perception of urban greening. The experiment found that: (1) The Green view index(GVI) of most sampling points decreases as the FOV increases. The increase in GVI is mainly affected by the crown width and pitch angle, and the high-value points of GVI are more easily affected. (2) Pitch angle usually causes a slight increase in GVI, which is mainly affected by the distribution of tall trees and low vegetation. Tall trees cause an increase in GVI, while low vegetation causes a decrease in GVI. Similarly, high value points of GVI are more likely to produce larger differences. (3) The Heading parameter has no obvious impact on the overall visual impact of GVI. The difference in independent sampling points comes from the angle between the lens acquisition direction and the road forward direction.These findings not only enrich the theoretical connotation of green view index research, but also provide valuable references for practical applications.

## 1. Introduction

Urban street greening has long been considered a design element of urban landscapes(e.g., Fernow, 1910; Schroeder and Cannon, 1983; Wolf, 2005), street greening plays an important role in society, environment and ecology, such as mitigating the urban heat island effect (Guo et al., 2020; He et al., 2019; Lafortezza and Sanesi, 2019; J. Yang et al., 2018; Zardo et al., 2017), reduce flood damage (He et al., 2019; Onishi et al., 2010), increase biodiversity (Aronson et al., 2017; Schneiderman et al., 2015; J. Yang et al., 2018) etc.Understanding how people imagine street greening could help greening programs increase political support (Seymour et al., 2010; Wolch et al., 2014), This can make the difference between the success or failure of a greening program. Methods of measuring people's opinions, attitudes and perceptions about urban green streetscapes include surveys, interviews and audits. Questionnaire survey methods often raise concerns about response bias (Downs and Stea, 1977). Objective assessment methods may be more efficient and accurate. However, there are few objective methods for measuring urban greening. Due to its repeatability and large coverage, remote sensing can calculate indicators such as green space percentage and green space density based on remote sensing images. For example, see (Faryadi and Taheri, 2009), In its early days, it became the most commonly used objective method to measure urban greening,e.g., Gupta et al., 2012. However, remote sensing methods also have limitations. Remote sensing data cannot capture the horizontal cross-section of urban greening. The street view from a human perspective is mostly different from the overhead view of remote sensing methods (Li et al., 2015a).

With the acceleration of urbanization, green view index has become an important indicator for evaluating the quality of urban ecological environment. Traditional green view index calculation methods rely on ground surveys and remote sensing data, but are costly and slow to update.. Yang et al(2009) By developing the Green View Index (GVI), color images were used to assess the visibility of the surrounding urban forest as a proxy for pedestrian perceptions of greenery. This research work is cumbersome and time-consuming, relying on manual collection of street views to extract green areas. The application range of GVI is small, and the viewing angle cannot completely cover the spherical field of view of pedestrians(Asgarzadeh et al., 2012).

With the rise of street view services, street view pictures can provide a perspective similar to that of pedestrians and present the street environment in detail. In particular, Google Street View (GSV) images have been used to map and quantify the amount of street greening (Seiferling et al., 2017; Ye et al., 2019). As SVI now covers half of the world’s population(Goel et al., 2018), It provides a valuable source of large-scale urban data, and SVI is able to examine visual features from a human (horizontal) perspective that other common data sources such as aerial or satellite imagery cannot provide. Since the early days of large-scale SVI service delivery, researchers have recognized that it is well suited for assessing characteristics of the built environment (Kelly et al., 2013). Therefore, it has been accepted by many fields. In addition to GSV, Baidu Maps occupies an important position among SVI providers in China. Baidu Maps is an online map service provided by Baidu and can be regarded as China's Google Maps. Since 2013, it has provided a street view service - Baidu Panorama. While satellite imagery and maps in Baidu Maps have coverage beyond China, SVI is only available in China (Biljecki and Ito, 2021).

The calculation of GVI requires the extraction of vegetation elements, and various algorithms have been developed to depict green plants from RGB images based on spectral and geometric rules. (Li et al., 2015a, 2015b; Woebbecke et al., 1993). However, traditional methods are often unstable for images captured under different lighting conditions, which is a very common problem for street images captured at different times. With the development of computer vision and deep learning semantic segmentation technology, street images can be accurately segmented into different categories. Based on a large amount of training data (street view images collected under different conditions from around the world), the trained network model can quickly distinguish different types of features of street view images. Scholars have used SegNet, PSPNet, FCN and other models to calculate green view index(Lu et al., 2019; Yang et al., 2019; Ye et al., 2019). However, current semantic segmentation technology still has problems such as rough segmentation results and discontinuous segmentation boundaries for street view images. With the emergence of complex semantic scene factors, it poses higher challenges to urban environment perception, so more semantic segmentation models are proposed. High requirements.

Currently, numerous studies utilize semantic segmentation technology for street view segmentation and GVI calculation. However, research on the impact of street view parameters such as heading, FOV (Field of View), and pitch on GVI calculation remains insufficient. The heading represents the camera's orientation angle, FOV denotes the camera's field of view, and pitch refers to the camera's tilt angle. These parameters influence the visibility of green vegetation in street view images, thus affecting the GVI calculation. This paper takes Anning District in Lanzhou City as an example and proposes the PSAT-Net semantic segmentation model. By employing the pyramid pooling attention mechanism to enhance segmentation accuracy, the study explores the impact of street view parameters on the GVI. This provides scientific evidence for urban planning and environmental management.

## 2. Research Area and Data Source

### 2.1 Study Area

Anning District, as one of the core areas of Lanzhou City, is located on the north bank of the Yellow River, adjacent to Chengguan District, Qilihe District, Xigu District, and Gaolan County. The geographical advantages, diverse urban structure, and natural environment of Anning District, along with its representation in the greening development of Lanzhou City and Gansu Province, make it an ideal choice for studying the green view index.

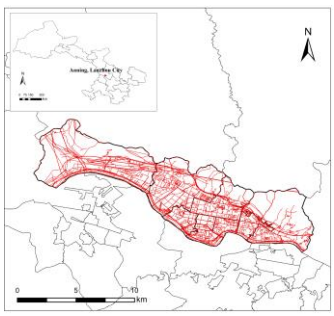


Figure 1. The location and the street map of Anning District (obtained from the OpenStreetMap website)

### 2.2 Data Source

**2.2.1 Street view data**: Street view data, as a rich source of urban scene imagery, is typically collected through various means such as vehicle-mounted cameras, pedestrian-carried devices, or drones. These images provide a comprehensive and three-dimensional perspective of the urban environment, greatly enriching our understanding of urban spaces. Street view data plays an indispensable role in map creation, navigation system design, urban planning, and research in the field of computer vision. In this study, street view imagery data was effectively obtained using the Baidu Maps API interface and Python web scraping techniques.

This study used Baidu Map API and Python crawler technology to obtain street view images, combined with the road network data of OpenStreetMap, preprocessed the road network data to merge and topology, remove tunnels, railways, bicycle lanes and other level roads, and create sampling points at intervals of 100 meters. Accurate street view image data was obtained.

**2.2.2 Sample Data**: This study utilized the open-source Cityscapes dataset for semantic segmentation tasks. The dataset comprises a total of 3,475 finely annotated images, with 2,975 designated for training and 500 for validation. Each image has a resolution of 1024x2048 pixels. The Cityscapes dataset originally has 34 categories, but this paper has set the number of categories to 19 to optimize memory usage and reduce resource consumption during model training. The image sizes were adjusted to three resolutions: 256x256, 512x512, and 1024x1024, used at different training stages. Data augmentation was also applied to the dataset to enhance the model’s generalization capabilities. Finally, the model’s accuracy was evaluated using the validation set to ensure the accuracy and reliability of the research results.

## 3. Street scene segmentation and model accuracy evaluation

### 3.1 Research Framework

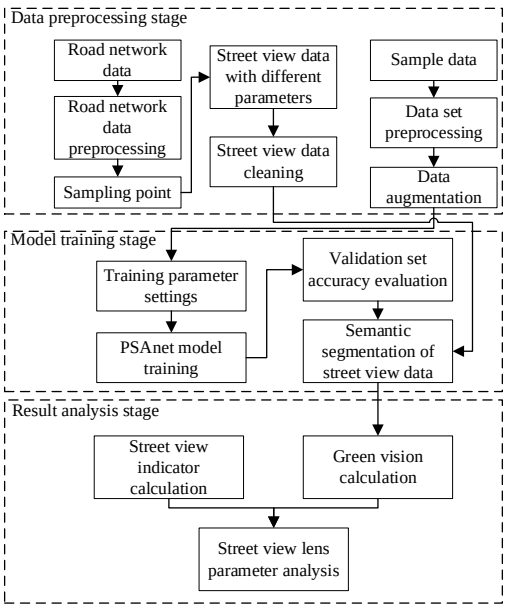


Figure 2. Technological Roadmap

This paper first introduces the PSAT-Net model, which performs semantic segmentation on street view data with different camera parameters, and proposes a method for calculating the green view index, analyzing the impact of different camera parameters on the green view index in street view data. The architecture is shown in the figure below.

### 3.2 Establishment of the PSAT-Net Model

The paper establishes the PSAT-Net model, a semantic segmentation model that integrates a hybrid attention mechanism with a convolutional neural network. It utilizes ResNet50 as the backbone network for feature extraction, with an encoder-decoder as the main architecture. A meticulously designed Feature Refinement Module (FRM) combines spatial and channel attention. The paper employs a minimalist decoder architecture to merge the pyramid pooling feature maps and attention feature maps outputted by the FRM, convolving them to produce the final segmentation map.
The network structure is as follows:

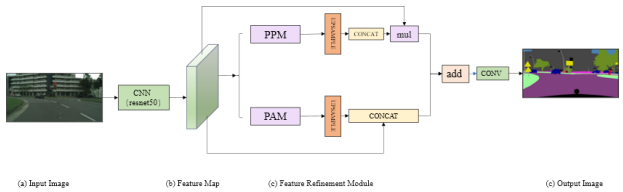


Figure 3. PSAT-Net Network Architecture

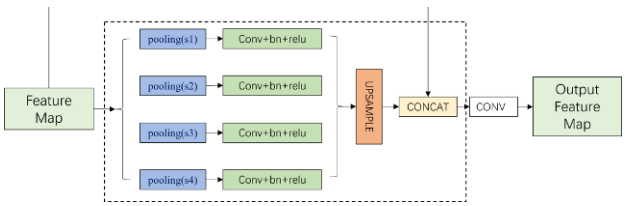


Figure 4.Pyramid Pooling Module

**3.2.1 Pyramid Pooling Module:** The Pyramid Pooling Module (PPM) is one of many multi-scale feature extraction modules, such as ASPP (Atrous Spatial Pyramid Pooling), PPM, FPA, etc. ASPP is a pyramid of atrous spatial convolutions that combines convolutions with different dilation rates and pooling results; PPM, on the other hand, is a special pooling model. By pooling from more to less, it can effectively increase the receptive field and enhance the efficiency of global information utilization.

**3.2.2 Feature Refinement Module:** In convolutional neural networks, attention mechanisms play a crucial role in enhancing meaningful features and suppressing irrelevant information. Various attention mechanisms, such as spatial, channel, self, and hybrid attention, have been extensively explored. For instance, SENet generates attention scores for each channel, focusing mainly on channel attention but overlooking the interaction between element positions. Non-Local calculates the self-correlation of each element in the feature map to establish long-range dependencies, but the attention distribution is the same for each channel. SCSE, an improvement over SENet, introduced cSE channel attention and sSE spatial attention, which have been proven to enhance meaningful features and suppress useless ones.

This paper proposes the Pyramid Pooling Attention Module (PAM), which combines the SE module with PPM (Spatial Pyramid Pooling). PAM extends single-scale channel attention (SE) to multi-scale channel attention, generating attention maps of different scales for each channel dimension. Moreover, the distribution of multi-scale channel attention is no longer limited to a 1x1 size, implicitly achieving spatial attention without compressing the channels, thus significantly enriching the feature information. The feature calibration module merges the feature maps output by the PPM and PAM modules, resulting in features with a larger receptive field and more refined purity.

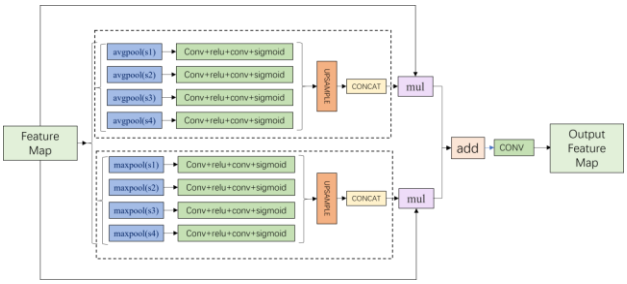


Figure 5. Pyramid Pooling Attention Module

**3.2.3 Decoder Feature Fusion Module**: In the decoder stage, the feature maps outputted by the encoder are upsampled to align spatial information with the decoder feature maps. To further integrate these two feature maps and enhance their information exchange, a 1x1 convolution layer is used for feature fusion. This convolution operation effectively reduces the feature dimensions while maintaining the validity of the features, which helps to improve the expressiveness and efficiency of the model.

### 3.3 Model Training

**3.3.1 Configuration and Environment:** The PyTorch deep learning framework is based on the Python programming language. Code writing and experimental testing are performed using the VSCode software. The computer environment is configured with the Windows 11 operating system, and the virtual environment is configured with 8GB RAM and an NVIDIA GeForce GTX 1050 GPU.

**3.3.2 Training parameter settings:** This article uses transfer learning to fine tune the ResNet50 model. Due to device performance limitations, the parameters of the semantic segmentation model are set as follows: the initial learning rate is 0.01, the defined number of iterations is 500 epochs, and an "early stop" is used to monitor MIOU. The optimization is performed using the SGD optimizer. The remaining parameters are shown in Table 1.

| Training stage | LR decay | Batch size | Image size | Loss function |
|---|---|---|---|---|
| Stage1 | Poly | 8 | 256 | FacalLoss |
| Stage2 | Poly | 4 | 512 | LovaszSoftmax |
| Stage3 | Poly | 2 | 1024 | LovaszSoftmax |

Tabel 1. Training Parameter Configuration

Firstly, use images of 256 * 256 size, train them to stability using Facalloss, load the model parameters for the first time, continue training using images of 512 * 512 size combined with

LovaszSoftmax loss function, and finally use small batches of 1024 * 1024 large size images for fine training.

**3.3.3 loss function:** The loss function is a key measure of model performance, reflecting the difference between predicted and true values. This article selects two loss functions, FocalLoss and LovaszSoftmax, to address different training challenges.

In the phased training strategy, FocalLoss is first used to solve the class imbalance problem in the initial training, providing a solid foundation for the model. As the training deepens, we shift towards the LovaszSoftmax loss function to further improve the performance of the model in semantic segmentation tasks, especially in accurately dividing different class boundaries. This ordered loss function arrangement aims to enable the model to solve specific problems targeted at different training stages, thereby effectively improving training efficiency and the performance of the final model.

### 3.4 Calculation method for green viewing index

Green view index refers to the proportion of plants seen from a human's perspective among all visible objects, which can reflect the level of greenery in a city

$$GVI = \frac{S_{vegetation}}{S_{total}} \quad (1)$$

In the formula, the urban green visibility is the ratio of the pixels of vegetation in the street view image to the total number of pixels.

### 3.5 Street View Data Parameter Settings

To investigate the influence of street view lens parameters on the calculation of green view index, six different lens parameters were set when collecting street view images, as shown in Table 2. The size of all street view images was 512x512 pixels.

| Param | Headings | FOV | Pitch |
|---|---|---|---|
| Param 1 | 0°90°180°270° | 90° | 0° |
| Param 2 | 0°90°180°270° | 60° | 0° |
| Param 3 | 0°90°180°270° | 120° | 0° |
| Param 4 | 0°60°120°180°240°300° | 90° | 0° |
| Param 5 | 0°120°240° | 90° | 0° |
| Param 6 | 0°90°180°270° | 90° | 45° |

Table 2. Street View Data Parameter Configuration

In street scene analysis, the Heading parameter ensures that pedestrians can achieve a 360° horizontal omnidirectional observation, which is crucial for comprehensive sampling.The Pitch parameter describes the pitch angle of the camera relative to the street view vehicle, which is set at or above the horizontal line to ensure that it can capture roadside trees of different heights(Meng et al., 2020). The FOV (Field of View) parameter defines the vertical and horizontal field of view of a street view image. The larger the FOV value, the richer the information contained in the image. The careful setting of these parameters is to ensure that street view data can comprehensively and accurately reflect the natural and architectural characteristics of the city. To ensure that the green view index calculation results on photos taken at different sampling points are comparable, the lens settings (Heading, Pitch, and FOV) under the same parameter should be fixed values(Meng et al., 2020).

## 4. Street scene segmentation and model accuracy evaluation

### 4.1 Street Scene Semantic Segmentation

Using the PSAT-Net model for semantic segmentation of street view features can obtain pixels of categories such as roads, sky, and vegetation in each image. Python is used to calculate the proportion of different features in the street view. For the original image and semantic segmentation results of a street view image, we can extract the street view elements as shown in Figure 6 and use the calculation formula in 3.4 to calculate the green view index.

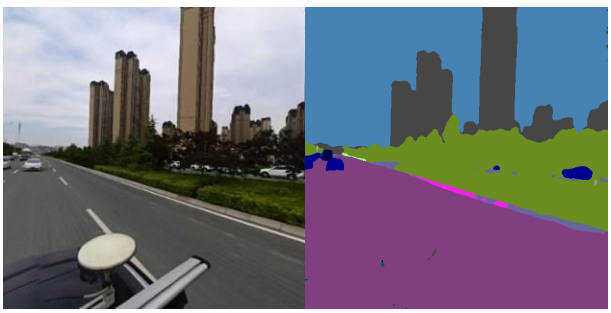

Figure 6. Street Scene Segmentation Results

### 4.2 Model accuracy evaluation

In image segmentation, many standards are commonly used to measure the accuracy of models. Based on the semantic segmentation review by Garcia-Garcia et al. (2017), A. et al. and combined with the research background, the evaluation methods used in this paper are intersection to union ratio, accuracy, recall, and F1 score.

From Tables 3 and 4, it can be seen that the PSAT-Net model proposed in this paper achieves higher segmentation accuracy compared to mainstream models like Unet and DeepLab-V3. It can more accurately segment small objects such as utility poles and traffic lights, while also reducing misclassifications and omissions. The mean Intersection over Union (MIOU) is improved by 5%-7%, and for major street view elements such as roads, sky, buildings, and vegetation, the MIOU can reach over 90%, meeting the requirements of this study.

| Model | Backbone | MIOU | M-Prec | M-Rec | M-F1 |
|---|---|---|---|---|---|
| UNet | ResNet50 | 0.660 | 0.806 | 0.760 | 0.779 |
| PSPNet | ResNet50 | 0.675 | 0.777 | 0.810 | 0.791 |
| DeepLab-V3 | ResNet50 | 0.688 | 0.800 | 0.806 | 0.801 |
| PSAT-Net | ResNet50 | 0.737 | 0.849 | 0.833 | 0.840 |

Table 3. Performance comparison of different models on the Cityscapes validation set

| Classes | MIOU | M-Prec | M-Rec | M-F1 |
|---|---|---|---|---|
| road | 0.979 | 0.991 | 0.989 | 0.990 |
| sky | 0.947 | 0.972 | 0.973 | 0.973 |
| vegetation | 0.930 | 0.960 | 0.968 | 0.964 |
| building | 0.907 | 0.943 | 0.959 | 0.951 |
| sidewalk | 0.751 | 0.871 | 0.846 | 0.858 |
| traffic-sign | 0.770 | 0.898 | 0.845 | 0.870 |
| pole | 0.573 | 0.755 | 0.704 | 0.729 |

Table 4. Accuracy evaluation of each category

| Label / Model | Slice 1 | Slice 2 | Slice 3 | Slice 4 |
|---|---|---|---|---|
| Label |  |  |  |  |
| PSPNet |  |  |  |  |
| UNet |  |  |  |  |
| DeepLab-V3 |  |  |  |  |
| PSAT-Net |  |  |  |  |

Table 5. Segmentation results comparison of different models on the Cityscapes validation set

## 5. Analysis of Green view index

### 5.1 Calculation of Green view index Based on Street View Semantic Segmentation

Based on the semantic segmentation results of the street view, the calculation formula for the street view indicators proposed in 3.4 is used to calculate the green view index, sky width, road width, and building density for each sampling point. Using FOV of 90°, pitch of 0°, and heading of 0°-90°-180°-270° as examples, the calculation results are shown in Table 6.

| Id | Lon | Lat | GVI |
|---|---|---|---|
| 1 | 103.6088447 | 36.15032763 | 0.079 |
| 2 | 103.6096686 | 36.15036756 | 0.095 |
| 3 | 103.6104911 | 36.15043079 | 0.140 |
| … | … | … | … |
| 2458 | 103.6852566 | 36.09535118 | 0.021 |
| 2459 | 103.6845155 | 36.09498863 | 0.006 |
| 2460 | 103.6837744 | 36.09462608 | 0.009 |

Table 6. Calculation Results of GVI

### 5.2 Analysis of Green view index Results under Different Street View Parameter Configuration

Under the fov parameter, g60 represents the green view index with a horizontal viewing angle range of 60°, and the same applies to g90 and g120; Under the pitch parameter, g45 represents the green view index with a pitch angle of 45°, and g0 is the same; Under the heading parameter, g4 represents the green view index in the configuration with a heading of 0°-90°-180°-270°, g3 represents the green view index in the configuration with a heading of 0°-120°-240°, and g6 represents the green view index in the configuration with a heading of 0°-60°-120°-180°-240°-300°.

| Para | Group | Avg-Trend | Max-Inc | Max-Dec | Inc-Ratio | r |
|---|---|---|---|---|---|---|
| Fov | g90-g60 | -0.052 | 0.131 | 0.246 | 0.125 | 0.950 |
| | g120-g90 | -0.042 | 0.133 | 0.183 | 0.105 | 0.957 |
| | g120-g60 | -0.095 | 0.197 | 0.354 | 0.101 | 0.865 |
| Pitch | g45-g0 | 0.022 | 0.452 | 0.303 | 0.548 | 0.898 |
| Head-ing | g4-g3 | -0.002 | 0.09 | 0.156 | 0.458 | 0.980 |
| | g6-g4 | -0.003 | 0.104 | 0.114 | 0.404 | 0.991 |

Table 7. Trend Analysis of GVI with Different Parameters

**5.2.1 The influence of Fov parameters on green view index**: The FOV parameter determines the field of view angle of the image. The study controlled for other lens parameters to remain fixed. As shown in Table 8, an increase in field of view angle leads to a general decrease in green view index, while exhibiting a strong correlation. However, as the difference in field of view angle increases, the correlation also weakens. This phenomenon is due to a broader perspective allowing observers to simultaneously see more non green elements, such as buildings and roads, thereby visually diluting the density of green space.

(a) fov=60°

(b) fov=90°

(c) fov=120°

Figure 7. Comparison of Green view index for High-Value Points with three fov parameters

Although the green view index of most sampling points decreases, there are still a small number of sampling points with an increase in green view index. For large crown trees, this is because most of the trunk is captured when viewed horizontally,

and more of the crown is captured when the field of view angle is expanded. For small crown trees, when the elevation angle is 45 degrees, a large amount of sky is captured. As the field of view angle increases, more tree crowns are captured, and the green visibility rate increases accordingly.

At the same sampling points, in order to explore clear patterns, there are a total of 44 points with g120>g90>g60 and differences greater than 0.01. Among them, there are 24 points with a green view index greater than 0.2 and 7 points with a green view index less than 0.05, all of which are located within the tunnel and will not be considered. When the green view index is at a high value, it is more likely to increase due to an increase in the field of view angle

| g120-g90 & g90-g60 | total | GVI<0.05 | 0.05<GVI <0.2 | GVI>0.2 |
|---|---|---|---|---|
| >0.00 | 87 | 32 | 23 | 32 |
| >0.01 | 44 | 7 | 13 | 24 |
| >0.03 | 8 | 0 | 3 | 5 |

Table 8. Distribution of GVI Difference Generated by FOV Parameters Across Three Levels

**5.2.2 The Influence of Pitch Angle on Green view index**:
The increase in pitch angle usually leads to an increase in green view index. This is because as the observation angle increases, the observer sees more of the tree crown rather than the trunk or ground, while the internal area has more tall trees, enhancing the visual proportion of the green space. According to Table 7, when the Pitch angle increases from 0 ° to 45 ° (g45-g0), the average green visual acuity increases by 0.022, with a maximum increase of 0.452 and a maximum decrease of -0.303. The proportion of increase is 0.548, indicating that more than half of the green visual acuity values are increasing. The sampling points with a decrease in green visual acuity are mainly due to the neglect of low shrubs, grasslands, and green belts after the perspective is raised, resulting in a decrease in green visual acuity. The Spearman correlation coefficient is 0.898, indicating a strong positive correlation between green view indexs. The average upward trend and increasing proportion of green view index reflect the transition from a ground perspective to a higher perspective, making green spaces more prominent visually.

Further explore the sampling point patterns that cause the increase and decrease of green visual rate. By studying the different differences in pitch between 45 degrees and 0 degrees, as shown in Table 9, based on 0 degrees, sampling points with larger differences in rise or fall have a larger proportion at high value points and a lower proportion at low value points.

| g45-g0 | total | GVI<0.05 | 0.05<GVI<0.2 | GVI>0.2 |
|---|---|---|---|---|
| >0.01 | 790 | 116 | 288 | 386 |
| <-0.01 | 664 | 17 | 376 | 271 |
| >0.03 | 536 | 46 | 152 | 338 |
| <-0.03 | 457 | 3 | 230 | 224 |
| >0.05 | 389 | 20 | 73 | 296 |
| <-0.05 | 226 | 0 | 80 | 146 |

Table 9. Distribution of GVI difference generated by Pitch parameters across three Levels

**5.2.3 Impact of Heading Parameters on Green view index**:
According to Table 8, the impact of changes in the horizontal perspective range on green view index is not significant, possibly because the overall visual proportion of green space remains relatively stable in different directions of observation.
The reason for the slight difference in green view index is the disassembly of the lens. For east-west roads, the heading parameter of 90 ° represents that the street view collection vehicle lens and the direction of the road are the same, while the 60 ° and 120 ° lenses will decompose the roadside trees on both sides of the road into two scenes, so the green view index will be slightly increased, as shown in Figure 8; For north-south roads, a heading parameter of 90 ° indicates that the street view collection vehicle lens is perpendicular to one side of the street, while a 60 ° and 120 ° tilt lens will scatter the green vegetation on one side of the street, reducing the green visibility, as shown in Figure 8.


(a) east-west road


(b) north-south road

Figure 8. Example of disassembly of east-west and north-south road lenses

## 6. Conclusion

In this study, we conducted an in-depth analysis of the street view data in Anning District, Lanzhou City, mainly focusing on the green view index and its influencing factors. The following are the main findings and conclusions:

(1) Model validity: We have established a PSAT-Net model and innovatively proposed a pyramid pooling attention module. This module integrates channel and spatial information, implements a three-dimensional attention mechanism, and significantly improves performance indicators. This result demonstrates the effectiveness of our model in street scene semantic segmentation tasks.

(2) Factors influencing Green View Index (GVI): We conducted an in-depth analysis of the effects of different street view parameter configurations on GVI and explored various potential influencing factors. We found that parameters such as field of view angle, pitch angle, and horizontal viewing range significantly affect GVI. The heading parameter has no significant overall visual impact on GVI, and the differences in independent sampling points stem from the angle between the lens acquisition direction and the direction of road progression. The pitch angle typically results in a slight increase in GVI, mainly influenced by the distribution of tall trees and low vegetation. Tall trees cause an increase in GVI, while low vegetation leads to a decrease. Similarly, high-value points of GVI are more likely to produce larger differences. For most sampling points, GVI decreases with increasing FOV, with GVI

increases primarily influenced by canopy width and pitch angle. High-value points of GVI are more susceptible to these influences. These findings provide important references for urban planning and green space design.

The PSAT-Net model proposed in this article can efficiently segment street view images and automatically calculate the green view index. And explored the impact of street view parameters on green visibility, in order to provide more comprehensive information support for urban planning and environmental management.

## References

Aronson, M.F.J., Lepczyk, C.A., Evans, K.L., Goddard, M.A., Lerman, S.B., MacIvor, J.S., Nilon, C.H., Vargo, T., 2017. Biodiversity in the city: key challenges for urban green space management. *Frontiers in Ecology and the Environment*, 15(4), 189–196. doi.org/10.1002/fee.1480.

Asgarzadeh, M., Lusk, A., Koga, T., Hirate, K., 2012. Measuring oppressiveness of streetscapes. *Landscape and Urban Planning*, 107(1), 1–11. doi.org/10.1016/j.landurbplan.2012.04.001.

Biljecki, F., Ito, K., 2021: Street view imagery in urban analytics and GIS: A review. *Landscape and Urban Planning*, 215, 104217. doi.org/10.1016/j.landurbplan.2021.104217.

Downs, R.M., Stea, D., 1977: *Maps in Minds: Reflections on Cognitive Mapping*. Harper & Row,New York.

Faryadi, S., Taheri, S., 2009: Interconnections of Urban Green Spaces and Environmental Quality of Tehran. *International Journal of Environmental Research*, 3(2), 199–208.

Fernow, B.E., 1910. *The care of trees in lawn, street and park, with a list of trees and shrubs for decorative use*. H. Holt and company, New York. doi.org/10.5962/bhl.title.66913.

Garcia-Garcia, A., Orts, S., Oprea, S., Villena Martinez, V., Rodríguez, J., 2017. A Review on Deep Learning Techniques Applied to Semantic Segmentation.

Goel, R., Garcia, L.M.T., Goodman, A., Johnson, R., Aldred, R., Murugesan, M., Brage, S., Bhalla, K., Woodcock, J., 2018. Estimating city-level travel patterns using street imagery: A case study of using Google Street View in Britain. PLOS ONE, 13(5), e0196521. doi.org/10.1371/journal.pone.0196521.

Guo, A., Yang, J., Xiao, X., Xia (Cecilia), J., Jin, C., Li, X., 2020. Influences of urban spatial form on urban heat island effects at the community level in China. *Sustainable Cities and Society*, 53, 101972. doi.org/10.1016/j.scs.2019.101972.

He, B.-J., Zhu, J., Zhao, D.-X., Gou, Z.-H., Qi, J.-D., Wang, J., 2019. Co-benefits approach: Opportunities for implementing sponge city and urban heat island mitigation. *Land Use Policy*, 86, 147–157. doi.org/10.1016/j.landusepol.2019.05.003.

Kelly, C.M., Wilson, J.S., Baker, E.A., Miller, D.K., Schootman, M., 2013. Using Google Street View to Audit the Built Environment: Inter-rater Reliability Results. *Annals of Behavioral Medicine : A Publication of the Society of Behavioral Medicine*, 45(Suppl 1), 108–112. doi.org/10.1007/s12160-012-9419-9.

Lafortezza, R., Sanesi, G., 2019: Nature-based solutions: Settling the issue of sustainable urbanization. *Environmental Research*, 172, 394–398. doi.org/10.1016/j.envres.2018.12.063.

Li, X., Q, C., Zhang, C., Li, W., Kuzovkina, Y., Weiner, D., 2015. Who lives in greener neighborhoods? The distribution of street greenery and its association with residents’ socioeconomic conditions in Hartford, Connecticut, USA. *Urban Forestry & Urban Greening*, 14(4). doi.org/10.1016/j.ufug.2015.07.006.

Li, X., Zhang, C., Li, W., Ricard, R., Meng, Q., Zhang, W., 2015. Assessing street-level urban greenery using Google Street View and a modified green view index. *Urban Forestry & Urban Greening*, 14(3), 675–685. doi.org/10.1016/j.ufug.2015.06.006.

Lu, Y., Yang, Y., Sun, G., Gou, Z., 2019. Associations between overhead-view and eye-level urban greenness and cycling behaviors. *Cities*, 88, 10–18. doi.org/10.1016/j.cities.2019.01.003.

Meng, Q., Wang, X., Sun, Y., Zhang, J., Chen, X., 2020. Construction of green view index model based on street view data and research on its influence factors. *Ecological Science*, 39(1), 146–155.

Onishi, A., Cao, X., Ito, T., Shi, F., Imura, H., 2010. Evaluating the potential for urban heat-island mitigation by greening parking lots. *Urban Forestry & Urban Greening*, 9(4), 323–332. doi.org/10.1016/j.ufug.2010.06.002.

Schneiderman, J.E., He, H.S., Thompson, F.R., Dijak, W.D., Fraser, J.S., 2015. Comparison of a species distribution model and a process model from a hierarchical perspective to quantify effects of projected climate change on tree species. *Landscape Ecology*, 30(10), 1879–1892. doi.org/10.1007/s10980-015-0217-1.

Schroeder, H., Cannon, W., 1983: The Esthetic Contribution of Trees to Residential Streets in Ohio Towns. *Arboriculture & Urban Forestry*, 9(9), 237–243. doi.org/10.48044/jauf.1983.058.

Seiferling, I., Naik, N., Ratti, C., Proulx, R., 2017. Green streets—Quantifying and mapping urban trees with street-level imagery and computer vision. *Landscape and Urban Planning*, 165, 93–101. doi.org/10.1016/j.landurbplan.2017.05.010.

Seymour, M., Wolch, J., Reynolds, K., Bradbury, H., 2010. Resident Perceptions of Urban Alleys and Alley Greening. *Applied Geography*, 30(3), 380–393. doi.org/10.1016/j.apgeog.2009.11.002.

Woebbecke, D.M., Meyer, G.E., Von Bargen, K., Mortensen, D.A., 1993. Plant species identification, size, and enumeration using machine vision techniques on near-binary images: Optics in Agriculture and Forestry. *Proceedings of SPIE - The International Society for Optical Engineering*, 1836, 208–219.

Wolch, J.R., Byrne, J., Newell, J.P., 2014. Urban green space, public health, and environmental justice: The challenge of making cities ‘just green enough.’ *Landscape and Urban Planning*, 125, 234–244. doi.org/10.1016/j.landurbplan.2014.01.017.

Wolf, K.L., 2005. Business District Streetscapes, Trees, and Consumer Response. *Journal of Forestry*, 103(8), 396–400. doi.org/10.1093/jof/103.8.396.

Yang, J., Guan, Y., Xia, J. (Cecilia), Jin, C., Li, X., 2018. Spatiotemporal variation characteristics of green space ecosystem service value at urban fringes: A case study on Ganjingzi District in Dalian, China. *Science of The Total Environment*, 639, 1453–1461. doi.org/10.1016/j.scitotenv.2018.05.253.

Yang, J., Zhao, L., Mcbride, J., Gong, P., 2009. Can you see green? Assessing the visibility of urban forests in cities. *Landscape and Urban Planning*, 91(2), 97–104. doi.org/10.1016/j.landurbplan.2008.12.004.

Yang, Y., He, D., Gou, Z., Wang, R., Liu, Y., Lu, Y., 2019. Association between street greenery and walking behavior in older adults in Hong Kong. *Sustainable Cities and Society*, 51, 101747. doi.org/10.1016/j.scs.2019.101747.

Ye, Y., Richards, D., Lu, Y., Song, X., Zhuang, Y., Zeng, W., Zhong, T., 2019. Measuring daily accessed street greenery: A human-scale approach for informing better urban planning practices. *Landscape and Urban Planning*, 191, 103434. doi.org/10.1016/j.landurbplan.2018.08.028.

Zardo, L., Geneletti, D., Pérez-Soba, M., Eupen, M. van., 2017. Estimating the cooling capacity of green infrastructures to support urban planning. *Ecosystem Services*, 26, 225–235. doi.org/10.1016/j.ecoser.2017.06.016.