Visual Data-Based Spatial Analysis of Built Environments

WONG, Daniel KH[1], HABLANI, Chirag[1], SCHROEPFER, Thomas[1]

[1]Singapore University of Technology and Design, Singapore

ABSTRACT: The use of data-driven spatial analysis and design tools and methods to support traditional planning and design heuristics has become increasingly important over the last years. Incorporating data derived from the point of view of the user is vital for the successful planning and designing of urban and architectural spaces. These include data that considers the spatial perspective of users in built environments which can be empirically sensed. Precedents include photography, 2D and 3D isovists, Computer Vision-based approaches such as the mapping of urban green, using Google Street View images for public health indicators, social media photos, and architectural design classification. A multi-disciplinary approach could thus be applied to the analysis of the built environment, by systematically quantifying the users' visual experiences while navigating the built environment and then translating these through the use of data-driven tools and methods into relationships and interactions of a spatial network.

Our research sets up a framework for the empirical mapping and analysis of the built environment using a combination of Computer Vision, 360-degree panoramic and point cloud-based techniques. These can quantify key features such as built structures, greenery, water bodies and sky, and their respective extents. Next, the framework embeds those features' values in a network of spatial nodes. We introduce a *Green-Sky-Building* (GSB) diagram that visualises these features, translated in conjunction with network connectivity measures. The contribution of this research is the combination of existing and newly developed technique across inter-disciplinary fields – scene semantic segmentation recognition with spatial network analysis – as a basis to aid spatial design analysis and decision-making. In our conclusion, we discuss how our novel approach to the analysis of the built environment can provide important insights and support the traditional heuristic for the planning and design of urban and architectural spaces.

KEYWORDS: Spatial Design and Analysis, Visual features, Computer Vision, Spatial network graphs
PAPER SESSION TRACK: **Digital Design and Planning**

1 INTRODUCTION

The design and planning of urban and architectural spaces affects our daily activities in complex ways. Important small-scale components of the city have a greater influence on their surroundings than is commonly thought, and urban ecosystems deliver important benefits for the quality of the urban environment and for the health and psychological well-being of residents (Schroepfer, 2019). A scientific basis for the design of the interaction of these components of cities is thus crucial. The plethora and nature of information used to support decisions in this process is evolving rapidly in the current information technological renaissance. Since its early days, the goal of better analysis of data, deriving procedures of diagnosis, extracting indications, as well as graphical techniques was of a primary concern (Tukey, 1962). The use of data-informed spatial design and analysis methods to support traditional planning heuristics has become increasingly important over the last few years. The rule-of-thumbs of planners and designers are being augmented with empirical data from multiple sources and sensors, replacing or adding on to existing frameworks and techniques of understanding how built environments affect interactions of people, societies, and cities at the fine to larger scale.

Urban analytics now incorporates the structural as well as temporal science of cities, dealing with issues of complexity, scaling, allometry, as well as daily variations in transportation, social networks (Batty, 2019). It has become a mainstream approach in understanding the structure and circulation analysis of a spatial network (Boeing et al., 2021), using existing tools in graph-based spatial analysis, in conjunction with new and evolving computationally aided techniques from different disciplines. A ground-up approach from the spatial user point of view can be invaluable when understanding how individual users interact with their surroundings in a larger complex dynamic system that comprises other people, the built environment, and nature at scale. The research on *Imageability* studied the physical elements of the environment, including paths, edges, districts, nodes, and landmarks, which help people understand, navigate, and orientate themselves in it (Lynch, 1960). By incorporating multiple types of data derived from the diversity of the experience of cities, including that of the user, a more inclusive and successful planning and designing of urban and architectural spaces is facilitated. A multi-disciplinary approach could thus be applied to the design and analysis of the built environment, by systematically quantifying the users' visual experiences at street level, then embedding these analyses into the relationships and interactions of its spatial network.

Precedents have evolved from laborious traditional photography and survey-based research in the case of Lynch; to 2D and embodied 3D isovist simulations (Krukar et al., 2021); to recent Google-Street View (GSV) based approaches. The use of street-view images in documenting urban scenes and architectural frontage has contributed to unique studies that harnessed existing visual datasets more often than generating new images for analysis. They used a combination of Computer Vision (CV)-based, spatial mapping, crowd-sourcing and/or geo-embedded information for various purposes: the urban mapping of green (Seiferling et al. 2017); Green index and Canyon Sky View (CSV) defined by building edges, and meta-data information of the GSV images used in their geolocation and analyses (Li et al, 2016); correlation of GSV building features with geolocated area public health indicators (Nguyen et al. 2019); architectural design classification (Yoshimura et al. 2018); safety, beauty, and historicity of places through crowd-sourced opinions on city streetscape photos which are then used to classify and rate more city images in MIT's Senseable Lab project *Placepulse* (2016); City Bikeability (Ito et al., 2021); social media imagery and text (Liu et al. 2020); In *FaceLift: A Transparent Deep Learning Framework to Beautify Urban Scenes* (Joglekar et al., 2020), a deep learning framework used GSV images and was trained to identify beautiful urban scenes, and subsequently alter and beautify existing views with specific urban elements. These views were then presented to 20 architectural experts who agreed that the results benefited three areas: *decision-making, participatory urbanism, and the promotion of restorative spaces.* In their paper (Liu et al., 2017), wide coverage GSV images were analysed through machine learning models, resulting in a "medium-to-good estimation of people's real experience", with the results applying to different stakeholders such as researchers, planners, and residents. These studies translated empirically sensed data in built environments including the spatial perspective of users into frameworks which can be further quantitatively analysed with some even directly generating resultant visualizations that further gave insight to stakeholders such as planners, designers, policymakers, and residents.

To this end, we study the use of visual data to answer the research questions: How can we use visual data to scientifically aid the design of circulation flow and spatial elements at the urban-architectural scale? How can we harness nascent technology in the field of architecture and urban analytics using visual data for potential real-time analysis of the built environment?

2.0 METHODS AND MATERIALS

2.1. Mapping Visual Data to Spatial Networks

Our research adds on to the framework of spatial network analysis at urban-architectural scale by the empirical mapping of the built environment and relating that to its spatial network for analysis, using a combination of methods. First, a computer vision analysis of 360-degree imaging panorama and point cloud-based techniques, which quantify key features such as urban built structures, greenery, water bodies and sky, and their respective extents. Next, the framework embeds the values of those features into a network of spatial nodes and analyses these nodes in conjunction with their network measures. These features, translated in conjunction with spatial network connectivity measures, are then visualized to identify emergent patterns of related spatial clusters and potential correlation with areas of higher connectivity and hence, potential space use. By quantifying and embedding key features as edge weights in a spatial network graph, we can map and differentiate regions with spatial and visual features that can potentially be used for the analyses of spatial experience and performance. Conversely, when used during the design process to intentionally include features of interest, this process can inform future planning and design strategy.

Our study was carried out in three phases. The first onsite phase included image capturing on the Singapore University of Technology and Design (SUTD) Campus. We mapped out the spatial network of the Campus. The image captures were analysed using a Deep Convolutional Neural Network (DCNN)-based computer vision algorithm with a model implementation pre-trained on the Cityscapes dataset. The results were visualized with Rhinoceros Grasshopper in conjunction with network centrality measures.

2.2. Visual Data Capture: 360-degree Panoramic Imaging

In this context, the 360-degree image capture of a space provided a snapshot in time and immersive experience of the built environment at a single point, simulating the visual experience of a person at the position from a 360-degree point of view. The commercial availability and end-user browser integration of this technology and platform allows quick documentation, representation, analysis, and information representation, for further integration with other platforms such as Virtual, Augmented and Mixed Reality.

Our study captured 360-degree panorama images from the node points of the SUTD Campus using a portable, commercially available camera, the Insta360 One X2. It was placed on a floor stand at about 1.5m height, approximating a person's visual perspective. This 360-degree camera uses dual fisheye lens with 5.7K resolution, H.265 video encoding, image stabilization, and purpose-built algorithms in the commercial app to stitch image captures into both spherical and equirectangular panoramic formats. The spherical output images are user-rotatable in the app simulating an immersive visual environment. We used the equirectangular images in our subsequent analysis (e.g., Figure 1)

Figure 1: 1. Output Images of 360-camera captures. Spherical (first image), equirectangular panorama (second image), and 2 perspectives captured within the scrollable immersive app environment (third and fourth images).

2.3. Computer Vision Segmentation of Images

A subset of machine learning, the field of Computer Vision (CV) studies the automated extraction, analysis and understanding of useful information from images, achieving automatic visual understanding through development of theoretical and algorithmic basis (BMVA, 2017). In LeNet-5, the classic letter recognition Convolutional Neural Network (CNN), input images are subsampled several times into feature map layers, with each 'neuron' unit in a layer receiving inputs from a set of adjacent units in a previous layer (Lecun et al., 1999). Subsequently, a CV pipeline consists of training a CV model on a labelled training image set, validating with a validation set, then testing it on a test dataset to ensure the model shows minimal overfitting on the training set. Recent methods of object detection and recognition use pixel-level labelled semantic segmentation, for the separation of semantically related 'blobs or objects, making possible applications such as augmented reality wayfinding in GSV, virtual backgrounds and mixed reality in online conferences, and even street-embedded game applications. In general, CV semantic segmentation models today use a unique Deep Convolutional Neural Networks (DCNN) CV architecture tailored for it. Segnet, a widely referenced model uses

> an encoder network, a corresponding decoder network followed by a pixel-wise classification layer. The architecture of the encoder network is topologically identical to the 13 convolutional layers in the VGG16 network. The role of the decoder network is to map the low-resolution encoder feature maps to full input resolution feature maps for pixel-wise classification. The novelty of SegNet lies in the manner in which the decoder upsamples its lower resolution input feature map(s). Specifically, the decoder uses pooling indices computed in the max-pooling step of the corresponding encoder to perform non-linear upsampling. This eliminates the need for learning to upsample. The upsampled maps are sparse and are then convolved with trainable filters to produce dense feature maps. *(Badrinarayan et al., 2017)*

As the objective of this study was discerning the extent of relevant object classes, we chose an available existing CV model already pre-trained in city datasets, the Enet Deep Neural Network semantic segmentation model. Based on Segnet, it has the advantage of speed as it was developed for application to mobile video analysis of vehicular scenes in mind. This gives it the future potential of applying to real-time urban analytics. Enet was scripted with Torch7 machine-learning library and uses several CV design principles to speed up the CV process. Enet follows the Segnet approach by downsampling thus reducing memory usage, with filters operating on downsampled images gathering more field context. Early downsampling reduces the input size while using only a small set of feature maps. These and other characteristics (Paszke et al. 2016), contributes to its speed and relative accuracy.

While we used the model that was pre-trained on the Cityscapes dataset (Cordts et al., 2016), the Enet CV model could be trained for specific datasets depending on the required application such as a variety of indoor or outdoor spatial contexts. The Cityscapes dataset consists of 20 classification type labels as follows: *unlabelled, road, sidewalk, building, wall, fence, pole, traffic light, traffic sign, vegetation, terrain, sky, person, rider, car, truck, bus, train, motorcycle, bicycle*.

As the image captures for this study were 360-degree equirectangular panoramas, consideration was given to research addressing specifically this format. We found that some studies used converted 360-degree panoramas such as perspectives (Lai et al., 2018), stereographical projections (Yang et al., 2018), or cube maps from which a saliency map is derived (Monroy et al., 2018), while others used scenes in spherical or equirectangular settings with rotational alignment (Davidson et al., 2020) or omnidirectional approaches (Sekkat et al., 2020). The recognition of indoor objects was also a consideration as the CV model chosen for this study was based on outdoor cityscapes. Hence, we noted for future exploration to pre-train CV model on indoor images, such as the SUN-RGB or 360-Indoor datasets (Chou. et al., 2020).

2.4. Visual Data Capture: Point Cloud Imagery

A 3-dimensional (3D) point cloud is a set of data points in a 3D coordinate system. These data points are usually defined by X, Y, and Z coordinates, and colour values for external surfaces of an object based on laser scanning or photogrammetry. Over the last decade, reality capture techniques such as laser imaging that can generate 3D point cloud data have enabled increasingly accurate and economical point cloud data acquisition. Major applications of point cloud imaging today are 3D model reconstruction and quality inspection of construction works, as well as dimensional quality, surface quality, and displacement inspections. In addition to these major applications, 3D point cloud data has also been used for construction progress tracking, building performance analysis, construction safety management, building renovation, construction automation, heritage applications, and robot navigation. Infra-red and LiDAR scanners have been used for these purposes. There is great potential for point cloud captures in future image classification that can be incorporated into spatial analysis and contextual analysis techniques. Our study captured point clouds of several key nodes in the SUTD Campus spatial network using Faro Scene LiDAR for future exploration (e.g., Figure 2).

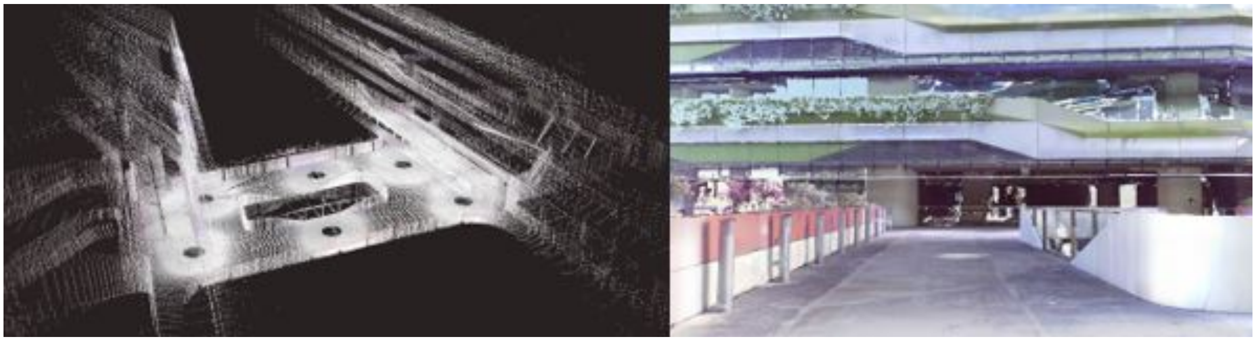

Figure 2: Output Images of LiDAR captures. Point Cloud (left) and combined with colour information (right).

While we applied the Enet CV script to our images, we also did a review of point cloud CV techniques for future exploration. Recent developments in LiDAR-based object inferences are in the field of robotics and autonomous driving and the potential for spatial analysis are due for further exploration. The advantage of using Point Cloud imagery is its accurate dimensional data for the spatial position of objects. In a survey on 3D point cloud deep learning, three broad categories were introduced to correspond to 3D shape classification, 3D object detection and tracking, and 3D point cloud segmentation (Guo et al., 2021). Some examples of CV models for Point Cloud image detection and segmentation follow. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation, PointPainting (Vora et al., 2019) works by projecting lidar points into the output of an image-only semantic segmentation network and appending the class scores to each point, and the resulting point cloud can then be fed to any lidar-only method. PolarNet, a lightweight neural network, quantizes points into grids using their polar coordinates, learns a fixed-length representation for each grid and uses a 2D neural network for point segmentation results. It achieved 57.2% mIoU performance in the SemanticKITTI dataset and addresses the objectives of "(1) the need for near-real-time latency with limited hardware; (2) uneven or even long-tailed distribution of LiDAR points across space; and (3) an increasing number of extremely fine-grained semantic classes" (Zhang et al., 2020).

2.5. **Embedding and Visualizing Data in Spatial Network**

Building on the author's previous research on constructing a graph-based 3D spatial network model from the SUTD campus (Gopalakrishnan et al., 2021), the same spatial network model is used. It characterizes important functional spaces and circulation paths of the Campus as nodes and edges. The network centrality measures were then calculated with the NetworkX algorithm and visualized with the data in a Rhino Grasshopper workflow. We used the following three centralities (degree, closeness, and betweenness centrality) for measuring the connectivity of nodes (Freeman, 1978; Barrat et al., 2004). Degree centrality shows the number of connected links for each node. Closeness centrality measures the distance (in steps) from a node to the rest of the network. Betweenness centrality captures the critical level of a node in terms of being the connection between communities. The three centralities are simple yet useful measurements in complex network analysis for the measurement of important levels of nodes within the network.

3.0 RESULTS AND DISCUSSION

3.1. Image Captures on SUTD Campus

We Captured 73 images of unique locations on various levels in the SUTD Campus with the 360-degree camera. Based on visual assessment, the images were categorized as indoor (34), semi-outdoor (15) and outdoor images (24). The CV classification script was applied to these images (e.g., Figure 3), resulting in pixel counts of each of the 20 object classes for the images. The input images were 1024 x 512 pixels with a total of 524288 pixels for each image. All labelled pixels were tallied to ensure there were no unaccounted pixels in the analysis.

We noted various instances where the classified output images had variances with ground truth. As this was primarily an outdoor cityscape dataset, certain elements in the outdoor scenes such as ground, sidewalk, vegetation, terrain, building, and walls appeared to be more accurately classified. For the interior images however, the results were mixed. Next, we recorded the pixel extents of each class in a table.

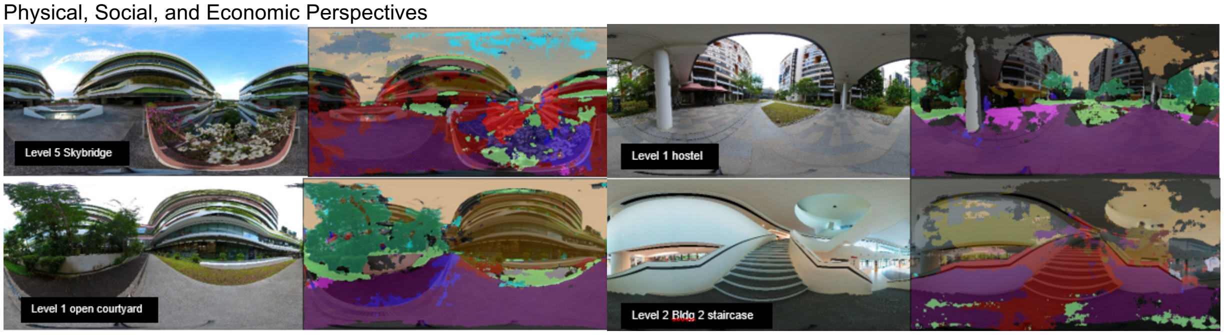

Figure 3: Equirectangular panorama images and detected object classes – outdoor, semi-outdoor and indoor scenes are shown.

We also tested converting some images from equirectangular panorama to cube map formats (e.g., figure 4), and using the same CV segmentation model, noted that the percentage of object class *unlabelled* decreased, thus indicating more accuracy in detecting labelled segments. However, there was variance in the classes, *road*, *vegetation*, *terrain* and *sky* which form most pixels in the images, likely from the mathematical contortion of the image. This would require a more detailed future study in this direction, using models trained on spherical training sets. However, for the purposes of our study, we found the use of equirectangular panoramas sufficient in capturing the proportions of the object class extents.

Figure 4: Comparison of Equirectangular Panorama (left), Cube map (centre), and their variance in Object Class pixels (right).

3.2. Verification of Indoor, Semi-Indoor and Outdoor Scenes

Next, we verified the CV classification of the image set. From the pixel segmentation results, 13 object classes were removed as they were irrelevant to the spatial analysis being carried out in the context of this study, leaving seven classes: *road*, *sidewalk*, *building*, *wall*, *vegetation*, *terrain*, *sky*. These labels correspond broadly to four urban-architectural spatial elements of ground plane (*road* + *sidewalk*), vertical built surfaces (*building* + *wall*), greenery (*vegetation* + *terrain*), and *sky*. For this study, we further removed the ground plane elements that correspond to the *road* and *sidewalk* labels to focus on the other three spatial elements. Figure 5 shows the amounts of greenery, sky, and vertical built surfaces in the indoor, semi-outdoor, and outdoor images. It can be observed that the CV classification corresponds to our expected amount of sky, greenery and built surface views in the three scene types. Thus overall, the CV segmentation classification result is representative of the types of actual views.

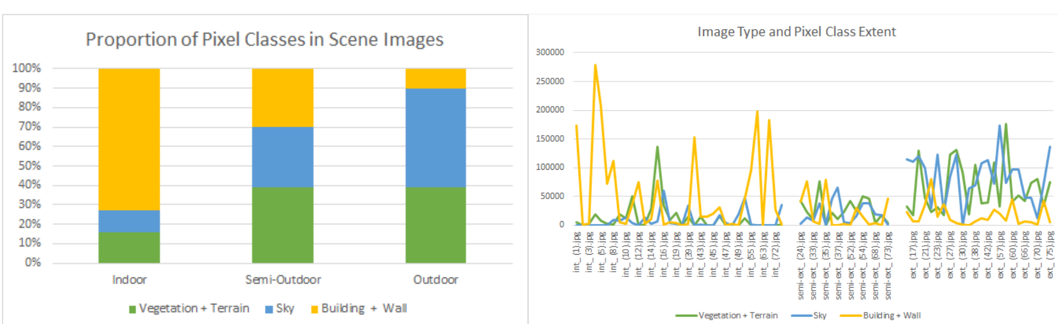

Figure 5: Proportion of 3 types of pixel classes in indoor, semi-outdoor and outdoor scene images (left), and the same comparison, in number of pixels, of individual images (right).

3.3. Urban-Architectural Spatial Elements: Ground, Building, Greenery and Sky

Subsequently we extracted the pixel extents of our defined classes corresponding to *Building*, *Greenery* and *Sky* and applied them to the spatial nodes of SUTD Campus. This spatial network of SUTD was constructed by assigning important functional spaces as nodes and connecting them to adjacent nodes with links, with the lifts and staircase cores serving as vertical all-to-all links in the network. The distribution of the proportion of visible *Building*, *Green*, and

Sky spaces were visualized by overlaying line strokes with the colours red, green, and blue respectively onto the 3D spatial map. The number of coloured line strokes in that space reflects the proportion of object class pixels across the whole collection of images. The visual density of the line strokes in a space is related to both the number of object class pixel numbers and the area of that space. We called this spatial overlay the *Green-Sky-Building* (GSB) diagram (e.g., Figure 6).

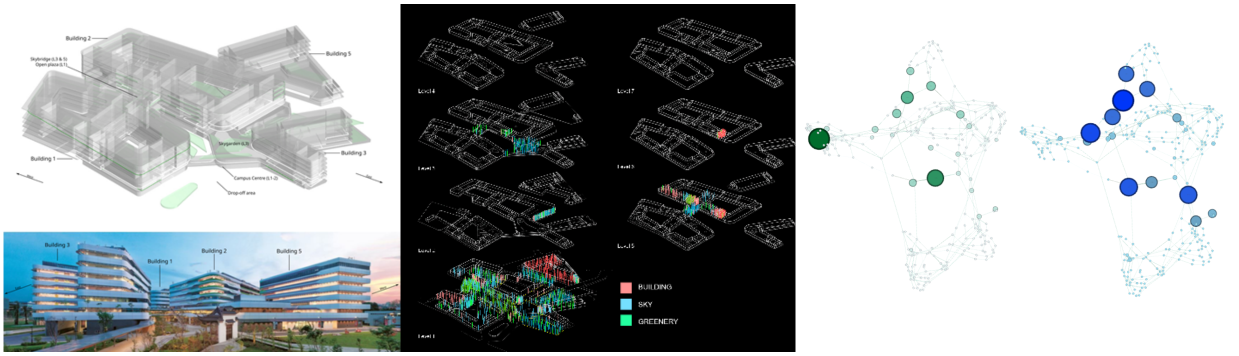

Figure 6: Image and conceptual diagram of SUTD Campus (left; Photograph by Daniel Swee), Image of GSB spatial overlay with selected nodes on all levels (centre), and comparison of graph map of campus nodes overlaid with Green and Sky values (right).

This overlay of GSB information is an indicator of the visual access to the identified categories at a specific location. These data informed our analysis of the experience of the Campus. The analysis is useful for understanding the experience of openness, green and building in each space and in relation to other spaces. When combined with further information such as complexity of pathways, number of fenestrations, materiality of surfaces, signages, presence of seating, and other architectural qualities, an even fuller picture of imageability of spaces can be derived from the visual data. This information can potentially be included using additional visual training datasets with the needed elements. In the context of digital twins, virtual perspective views that incorporate this information are a useful tool that can augment remote sensing of other data in the nodes, such as human activity, and environmental measurements.

3.4. Visualizing Spatial Element Data in Conjunction with Spatial Network Centralities

In this phase, we embedded network centrality information with the visual information as they are indicators of connectivity of the overall spatial system. This provides insights into the potential spatial experience at important nodes and links. The data obtained from the visual analysis can be embedded into the spatial network and visualized as a spatially embedded bar-graph map. In this process, we modified the value line strokes of the GSB diagram with the network centrality measures of the corresponding node to understand the interaction between areas of different connectivity and their visual data. Figure 7 shows the original values of the GSB values, and how they are modified by the relative betweenness and closeness centrality measure of each node, shown by the extended height in red. By quantifying and embedding visual data features in a spatial network graph, we can map and differentiate regions according to their relative importance in the entire spatial network. Conversely, when used as part of the design process, we can analyse a proposed plan for a range of desired spatial and visual features, filtering limits for required design function and spatial experience at key nodes in the urban or architectural plan. This enhances the process of planning and design as well as post-built analyses.

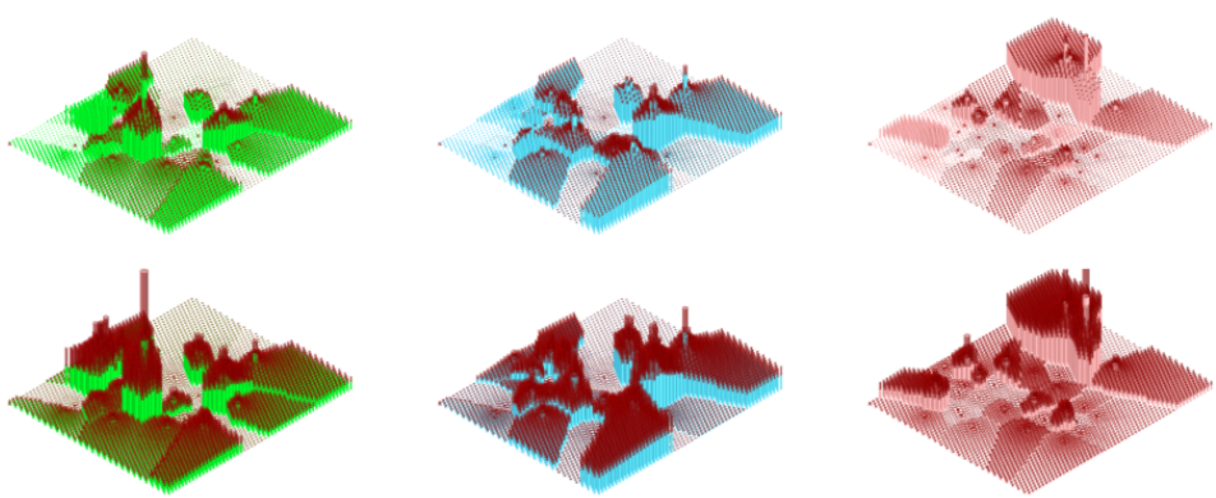

Figure 7: Green(L), Sky(C), Building(R) diagrams of Level 1, modified by Betweenness(top) and Closeness Centralities(bottom)

4.0 CONCLUSION, LIMITATIONS AND FUTURE DIRECTIONS

We studied the use of visual data to answer the following research questions: (1) How can visual data be quantified and used scientifically to inform the planning and design of circulation flow and spatial elements at the urban and architectural scale? (2) How can we harness nascent technology for timely visual analysis of the built environment? We demonstrated that by quantifying and embedding visual data features in conjunction with a spatial network graph, we can map and differentiate regions with spatial and visual features, adding filters for limits in GSB measures. This revealed spaces that correspond to design programme as needed. This can be used for analysing e.g., space use which informs future design urban and architectural design. This framework could be adapted for timely use in various formats, such as table, graphical and visual format, in augmenting design decision-making.

We noted the limitation of available datasets that span the ‘in-between' urban and architectural scale containing both outdoor and indoor spatial elements. We observed examples that corresponded in varying degrees with ground truth. Further study could be conducted on models with relevant datasets that have greater mIoU benchmarking accuracy. For the purpose of this study, we used the CV model pre-trained on the Cityscapes dataset as is. In particular, we observed that the Cityscapes feature class *Bus* was represented disproportionately, due to the building facades of the SUTD Campus having a similar visual morphology to linear bands of colour (i.e., roof, glass, sidewall). This relates to facades that bear resemblance to objects that are prevalent in the class features of a particular CV model and visual training set. However, this also engages the intuition of the similarity of design morphology and proportions across differing design categories (i.e., buildings and vehicles) that designers may not identify as grouped similarly – both the façade design of building and buses could very well result in similar floor to ceiling heights and visual banding as they are both based on the human proportion. We acknowledge the limitations of the perspective view - however, our study also aims to consider what is seen from the point of view of a person. Hence, we plan to further explore the use of combined photo and LiDAR based point-cloud segmentation in visual data extraction augmented with depth data for the analysis of visual connectivity of nodes for more comprehensive analysis.

In conclusion, we showed that our framework for a bottom-up approach that uses visual assessment of the built environment can provide important insights and support the traditional heuristic for the planning and design of urban and architectural spaces. The objective of this study was to create a framework for augmenting nodes in a spatial network with quantified visual feature data within a built environment. By being able to analyse features that affect the performance of spaces, we can inform future spatial planning and design. Ultimately, the development of planning and design strategies that include the classification of visual features at an early stage, is the goal of our research.

ACKNOWLEDGEMENTS

This research is supported by a SUTD grant (Cities: PIE-SGP-CTRS-1803).

REFERENCES

Badrinarayanan, V., Kendall, A., & Cipolla, R. (2017). Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(12), 2481-2495.

Barrat, A., Barthelemy, M., Pastor-Satorras, R., & Vespignani, A. (2004). The architecture of complex weighted networks. *Proceedings of the National Academy of Sciences*, 101(11), 3747-3752.

Batty, M. (2019). Urban analytics defined. *Environment and Planning B: Urban Analytics and City Science*, 43(3), 403-405.

Boeing, G., Batty, M., Jiang, S., & Schweitzer, L. (2021). Urban Analytics: History, Trajectory, and Critique. Trajectory, and Critique. *SSRN,* 3846508.

Chou, S. H., Sun, C., Chang, W. Y., Hsu, W. T., Sun, M., & Fu, J. (2020). 360-indoor: Towards learning real-world objects in 360° indoor equirectangular images. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision* (pp. 845-842).

Cordts, M., Omran, M., Ramos, S., Rehfeld, T., Enzweiler, M., Benenson, R., Franke, U., Roth, S., & Schiele, B. (2016). The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 3213-3223).

Davidson, B., Alvi, M. S., & Henriques, J. F. (2020). 360° Camera Alignment via Segmentation. In *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVIII 16 (pp. 579-595). Springer International Publishing.

Freeman, L. C. (1978). Centrality in social networks conceptual clarification. *Social Networks*, 1(3), 215-239.

Gopalakrishnan S., Wong, D., Manivannan, A., Bouffanais, R. and Schroepfer, T. 2021. “User-driven evaluation of emergent patterns of space use in vertically integrated urban environments.” *Architectural Research Centers Consortium 2021 International Conference*, April 7-10, 2021, Virtual/Tucson.

Guo, Y., Wang, H., Hu, Q., Liu, H., Liu, L., & Bennamoun, M. (2020). Deep learning for 3d point clouds: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(12), 4338-4364.

Joglekar, S., Quercia, D., Redi, M., Aiello, L. M., Kauer, T., & Sastry, N. (2020). FaceLift: A transparent deep learning framework to beautify urban scenes. *Royal Society Open Science*, 7(1), 190987.

Ito, K., & Biljecki, F. (2021). Assessing bikeability with street view imagery and computer vision. *Transportation Research Part C: Emerging Technologies*, 132, 103371.

Krukar, J., Manivannan, C., Bhatt, M., & Schultz, C. (2021). Embodied 3D isovists: A method to model the visual perception of space. *Environment and Planning B: Urban Analytics and City Science*, 48(8), 2307-2325.

Lai, W. S., Huang, Y., Joshi, N., Buehler, C., Yang, M. H., & Kang, S. B. (2018). Semantic-driven generation of hyperlapse from 360-degree video. *IEEE Transactions on Visualization and Computer Graphics*, 24(9), 2610-2621.

LeCun, Y., Haffner, P., Bottou, L., & Bengio, Y. (2000). Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision* (pp. 319-345). Springer, Berlin, Heidelberg.

Liu, L., Silva, E. A., Wu, C., & Wang, H. (2017). A machine learning-based method for the large-scale evaluation of the qualities of the urban environment. *Computers, Environment and Urban Systems*, 65, 113-125.

Liu, Y., Yuan, Y., & Zhang, F. (2020). Mining urban perceptions from social media data. *Journal of Spatial Information Science*, 20, 51-55.

Lynch K. (1960), *The Image of the City*, The MIT Press: Cambridge, Massachusetts.

Monroy, R., Lutz, S., Chalasani, T., & Smolic, A. (2018). Salnet360: Saliency maps for omni-directional images with CNN. *Signal Processing: Image Communication*, 69, 26-34.

Nguyen, Q. C., Khanna, S., Dwivedi, P., Huang, D., Huang, Y., Tasdizen, T., Brunisholz, K. D., Li, F., Gorman, W., Nguyen, T. T., & Jiang, C. (2019). Using Google Street View to examine associations between built environment characteristics and US health outcomes. *Preventive Medicine Reports*, 14, 100859.

Paszke, A., Chaurasia, A., Kim, S., & Culurciello, E. (2016). Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv preprint*, 1606.02147.

Schröpfer, T., Menz, S., Jiang, M., Belcher, R., Erdolu, E., Kaushal, M., Pilsudski, T., Raju P., Suen, E., & Tan, J. K. N. (2019). Dense and Green Building Typologies: Architecture as Urban Ecosystem. In *Indicia 02: Future Cities Laboratory* (pp. 32-42). Lars Müller Publishers.

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.

Sekkat, A. R., Dupuis, Y., Honeine, P., & Vasseur, P. (2020, June). A comparative study of semantic segmentation using omnidirectional images. In *Congrès Reconnaissance des Formes, Image, Apprentissage et Perception (RFIAP)* (ffhal-03088368f). Vannes, France.

The British Machine Vision Association and Society for Pattern Recognition (BMVA), 2017-02-16. What is computer vision? *The British Machine Vision Association and Society for Pattern Recognition*. https://web.archive.org/web/20170216180225/http://www.bmva.org/visionoverview (Accessed on November 11, 2021, via Wayback Machine).

Tukey, J. W. (1962). The future of data analysis. *The Annals of Mathematical Statistics*, 33(1), 1-67.

Vora, S., Lang, A. H., Helou, B., & Beijbom, O. (2020). Pointpainting: Sequential fusion for 3D object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 4604-4612).

Yang, W., Qian, Y., Kämäräinen, J. K., Cricri, F., & Fan, L. (2018, August). Object detection in equirectangular panorama. In *2018 24th International Conference on Pattern Recognition (ICPR)* (pp. 2190-2195). IEEE.

Yoshimura, Y., Cai, B., Wang, Z., & Ratti, C. (2019, July). Deep learning architect: Classification for architectural design through the eye of artificial intelligence. In *International Conference on Computers in Urban Planning and Urban Management* (pp. 249-265). Springer, Cham.

Zhang, Y., Zhou, Z., David, P., Yue, X., Xi, Z., Gong, B., & Foroosh, H. (2020). Polarnet: An improved grid representation for online lidar point clouds semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 9601-9610).