Metropolitan Street Space Quality Evaluation in Shenzhen

Qianqin Huang, Yuda Qiu, Yicong Wang and Shang Zhen
School of Science and Engineering
The Chinese University of Hong Kong, Shenzhen
218012012, 218012028, 218012038 and 218012058@link.cuhk.edu.cn

Abstract

The quality of street space affects people's behavior, habits, public health and urban culture. In order to deepen the existing research on street space quality and improve the quality of street environment in cities, this paper adopt neural network to evaluate urban streets. On the one hand, this paper objectively analyses the quality of street space by introducing neural network, and uses the segmentation of street image data to measure the quality of street space in Shenzhen, and finds the distribution characteristics of spatial quality in different districts of Shenzhen. On the other hand, through GAN transform existing street space style into German style, we could see the feeling of changing of urban space quality.

1. Introduction

United Nations Human Settlements Programme propose that streets play an important role in social productivity, public life and supporting facilities. It indicates that streets are not only traffic space, but also an important part of urban public space. Improving the quality of street space is closely related to promoting urban prosperity[1].

In the process of rapid urbanization, China lacks of dedicated designs of urban space and there is an inadequate street quality in many aspects, which limited its effect. Because the Urban Development Mode has changed from 'Scale Growth' to 'Quality Growth', it is very important to optimize street quality.

Understanding streets (landscape characteristics, advantages and disadvantages) is the first step in optimizing streets. It can help us quickly acquire the overall landscape image of urban street system and lay a solid foundation for the following administration. However, a practical problem is that with the increasing scale of cities, the number of streets in cities is huge. In this case, the traditional method of investigation by human is inefficiency[2].

In recent years, neural networks (like CNN, GAN) have entered the forefront of urban planning research, because of the increasing street view pictures[3]. Visual understanding of complex urban street scenes from street view pictures is an enabling factor for a wide range of applications. Object detection has benefited enormously from large-scale datasets, especially in the context of deep learning. This leads the central idea of this paper, evaluatint street space quality by neural networks.

2. Related Work

Based on gamming theory GAN (Generative adversarial networks)[4] has become a new research hotspot in the field of deep learning. The significance of this model is that data can be generated through unsupervised learning. Nowadays, GAN model was applied to do image restoration, text to image, image to image, and video generation. In our report we would like to use Cycle GAN model[5] to transfer the style of images, in order to help designer to do blue picture design during the urban planning.

Semantic segmentation is to classify the image at pixel level. It requires a dense pixel-wise prediction. Recently, researches have built kinds of deep learning network to tackle such a task. In 2014, Long et al. [6] proposed Fully Convolutional Networks (FCN), a CNN architecture without any fully connected layers. It enables a faster segmentation and is capable for flexible size of input. After that, most of CNN structures [7][8][9][10] are based on it.

There are also lots of datasets for semantic task. Specially, for semantic urban scene understanding, Cityscapes [11] is widely used.

3. Data and Research methods

This paper chooses street space to evaluate in five districts of southwestern Shenzhen (Nanshan, Futian, Luohu, Longgang, Baoan) . Datasets includes: Map of Shenzhen Road Network, Street view pictures. Cityscape.

Map of Shenzhen Road Network: This map is simplified by myself to do research. It contains five districts main road and secondary road. The red line is main road. And the white line is secondary road.It is shown as Figure 1

Figure 1. Shenzhen Road Network

Street view image: The data used for quality evaluation are street view pictures from Baidu map. We take a point every 200 meters on the road network with equal spacing. Also, we take four street pictures at this point, which is shown in Figure 2.

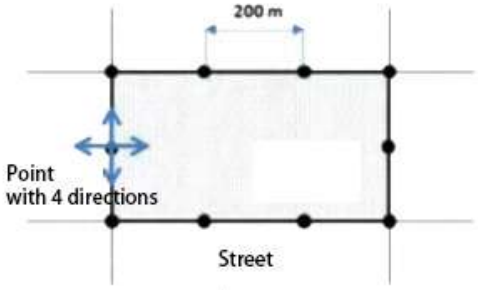

Figure 2. data prepocessing

4. Segmentation

4.1. Algorithm

The critical part of our analysis is accuracy of segmentation. To this end, we compare two kinds of model to extract the semantic information.

One of them is a generation method, CycleGAN. For our task, it aims to learn the mapping from street photo dominant to semantic dominant. Detail of the mechanism will be discussed in part 4.

Another model is a convolutional neural network, DeepLab V3. It's a variant for fully convolution network, the first end-to-end segmentation network, which is shown as Figure 3

DeepLab applies a ResNet-50 to extract features from inputs. The ResNet has been pretrained on ImageNet task. Notice that standard convolution kernel in Block4 of original ResNet is replaced with atrous convolution kernel. Such kernel enables dense feature extraction with the same quantity of parameters. In addition, DeepLab appends an atrous spatial pyramid pooling (ASPP) behind the Block4. The ASPP exploits multi-scale features by employing multiple

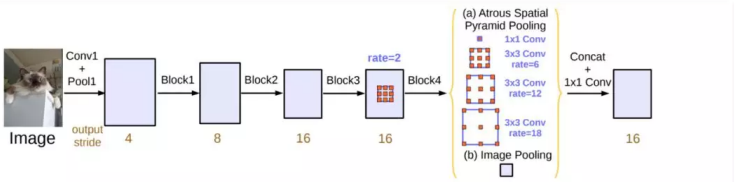

Figure 3. end-to-end segmentation network

parallel filters with different rates. Figure 4 shows the network structure.

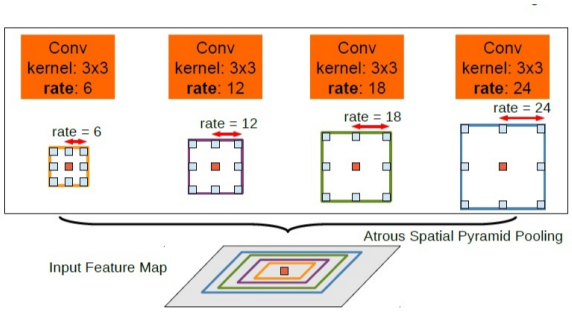

Figure 4. Network Structure

With these improvements, DeepLab V3 is able to segment street photo in fine scale.

4.2. Comparison

In our experiment, we train two models in Cityscapes. It's a street photo dataset of Germany. There are 5000 fine-labelled images with up to 29 different labels, classified in 8 categories. After training, CycleGAN achieves 0.51 on Mean IoU, while DeepLab V3 obtains 0.74, which is shown in Figure 5.

In the rest work, we employ DeepLab V3 to finish the segmentation work. 5210 street photos of Shenzhen, obtained from BAIDU API, are fed into the model. The results will support our statistical analysis on Shenzhen, which is shown in Figure 6.

4.3. Analysis

Through the semantic segmentation model, we get the objects label and its proportion in the street view image. Based on the cityscape's label, we get the seven categories in street view image. Because some labels do not account for a high proportion in our street view dataset, we finally selected three categories of architecture, nature and sky to analyze the street view of main road in Shenzhen.

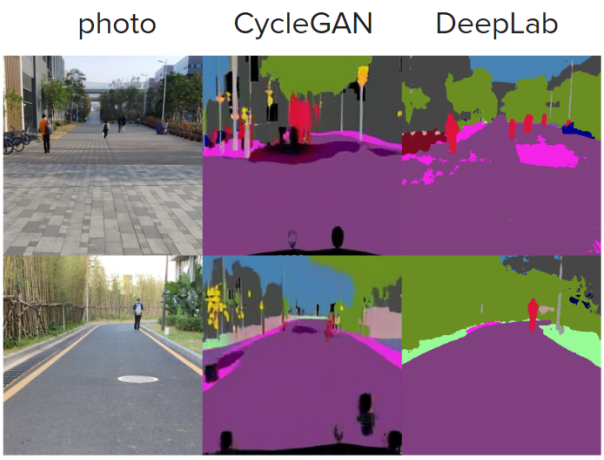

Figure 5. Comparison

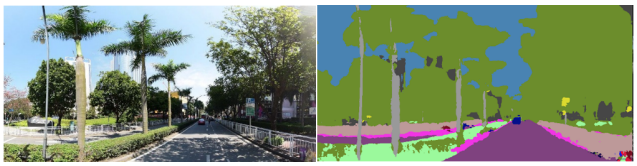

Figure 6. segmentation image

4.3.1 Spatial Image of Construction

Construction label of the street view image is mainly composed of build, wall and bridge label. Figure 7 is the heat map. In the map, red indicates that the building proportion of the location is higher. It can be seen the high proportion of construction is located four place that have been marked by the circles. Besides, Figure 8, it can find the proportion of construction label is higher in the place corresponding to the center of the district.

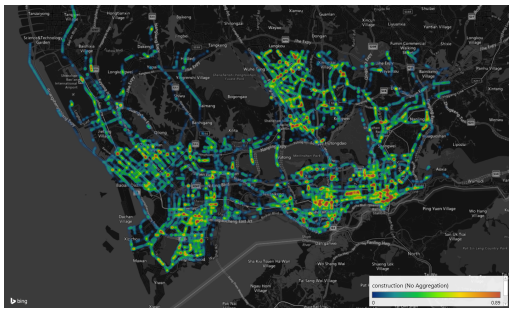

Figure 7. Heat map of Construction

4.3.2 Spatial Image of Sky

About the proportion of the sky label in the street view image, in our dataset, it has some the street view with high sky ratio and the street view with low sky ratio because of the urban structure. Figure 9 is the heat map of the sky label. It can be seen that the main road always has a higher sky proportion than the secondary road. In addition, the proportion of highways along the coast is also the highest.

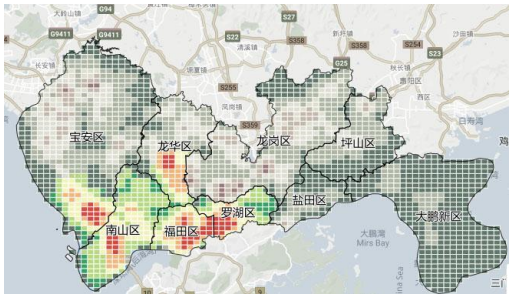

Figure 8. Distribution of Construction

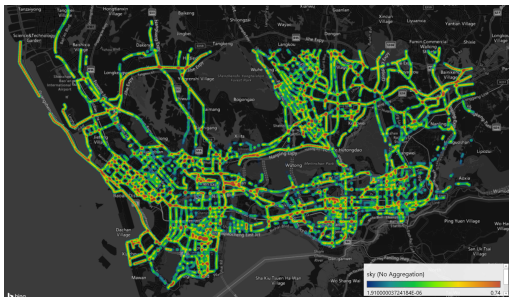

Figure 9. Heat map of Sky

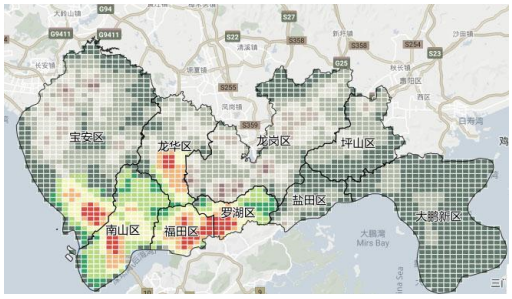

Figure 10. Distribution of Sky

4.3.3 Spatial Image of Nature

About the natural proportion of the street, it consists of soil and plants label. Figure 11 is also a heat map of plants. It can be seen that the natural distribution of the 5 district is relatively uniform, but the green ratio of Longgang District and Longhua District is relatively low. We supposed the reason why is that it is far from the city center and the development is relatively not so well. Figure 12 is the green-

ing plan of Shenzhen from 2010 to 2020, which matches the left figure.

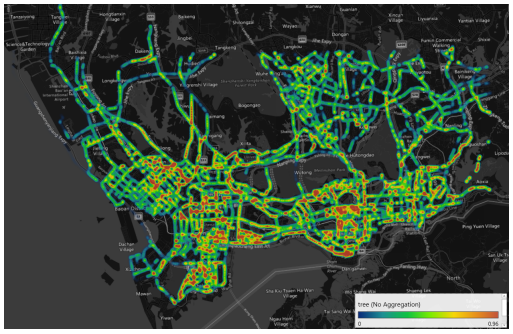

Figure 11. Heat map of Nature

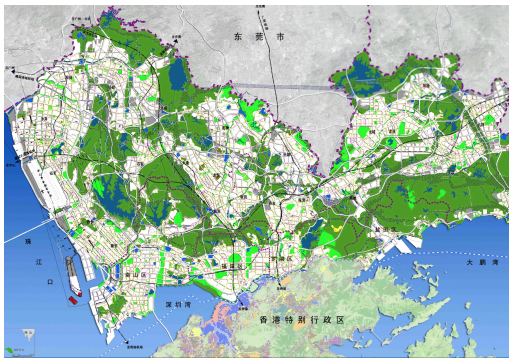

Figure 12. Distribution of Nature

5. Style Transformation

5.1. GAN

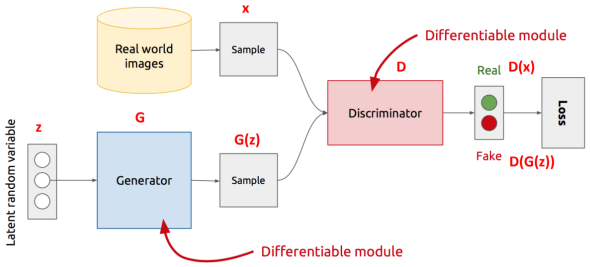

Figure 13. Structure of GAN

Figure 13 shows the Structure of GAN. The principle of common GAN is to generate images of random noise z through generator G, and the generated images are denoted as G(z). Discriminator D is responsible for determining whether an image is real or not, and for binary classification of image x and G(z). In the process of training, generator G aims to generate images to deceive discriminator D. Discriminator D aims to distinguish real images and generated image. The whole process of GAN could be considered as a game theory between generator and discriminator. At the end we would stop the GAN at an Nash equilibrium point, and use generator to generate images.

Suppose the distribution of real picture data is $p_{data}(x)$and the distribution of noise z is $p_z(z)$. Therefore, according to the cross-entropy loss, we can construct the following loss function as the loss of the common GAN.

Based on the common GAN, if we replace the input from random noise to paired image, we can get pix2pix model. In addition, we added the L1 loss in the loss function. Therefore, the loss function of pix2pix is shown below:

$$L(D,G) = E_{x\ p_{data}(x)}[lnD_X] + E_{z\ P_{data}(z)}[ln(1 - D(G(z)))] \quad (1)$$

$$L_1(G,X,Y) = E_{x\ p_{data}(x)}[||G(F(x)) - x||_1] \quad (2)$$

$$L = L(D,G) + L_1(G,X,Y) \quad (3)$$

5.2. CycleGAN

CycleGAN is an improvement of ordinary GAN. CycleGAN can use unpair images to do training and transfer one kind of picture into another kind of picture. The structure of CycleGAN is shown in Figure14

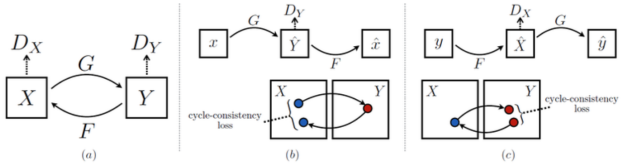

Figure 14. Structure of CycleGAN

CycleGAN aims to learn the mapping from dominant X to dominant Y. Let the mapping from X to Y be G, the first generator in CycleGAN. The Generator G map x in picture X to the picture $\widehat{Y} = G(X)$, and then use the first discriminator D_Y to identify it. Hence, we could get the first loss function:

$$L_{GAN}(G,D_Y,X,Y) = E_{y\ p_{data}(y)}[lnD_Y] + E_{y\ p_{data}(y)}[ln(1 - D_Y(F(x)))] \quad (4)$$

Similarly, we can obtain a mapping F from Y to X as the second generator. The Generator F maps picture y in Y to the picture $\widehat{X} = F(y)$. we also will use the second discriminator Dx to identify it. Therefore, we could get the second loss function:

$$L_{GAN}(G,D_Y,X,Y) = E_{x\ p_{data}(x)}[lnD_X] + E_{x\ p_{data}(x)}[ln(1 - D_X(F(y)))] \quad (5)$$

At the same time, there are also two Generator loss functions in CycleGAN. After the combination, the following loss functions can be obtained:

$$L_{cyc}(G, D_Y, X, Y) = E_{x\ p_{data}(x)}[||F(G(x)) - x||_1] + E_{x\ p_{data}(x)}[||F(G(y)) - y||_1] \quad (6)$$

Finally, three loss functions are merged to form the final loss function,

$$L = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_Y, X, Y) + L_{CYC}(F, G, X, Y) \quad (7)$$

Since CycleGan does not need a one-to-one corresponding sample, we can use loss function to train a network that could transfer one category image to another category image.

5.3. Result

Figure 15 shows changes of pictures from CUHKsz style to German style by using pix2pix and CycleGAN both after 300 epochs trainning:

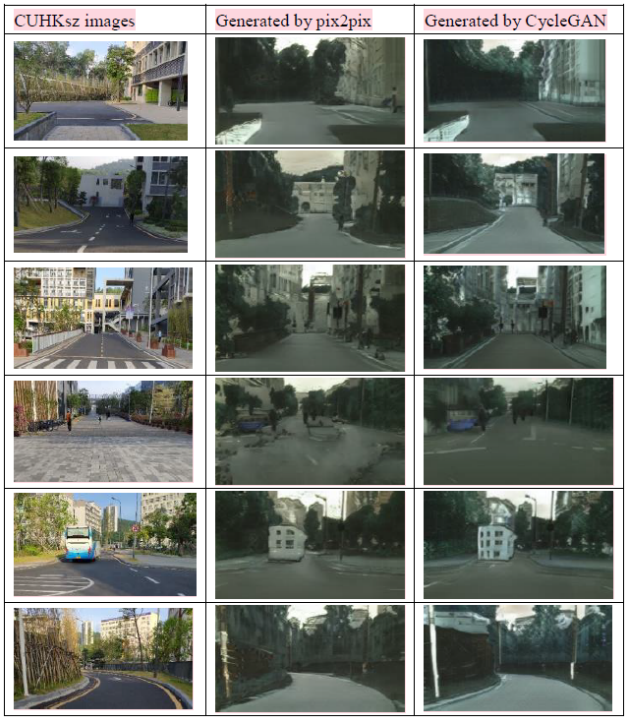

Figure 15. Style transormation

Comparing pix2pix results and CycleGAN results, we can find that pix2pix will loss a lot of detail in the images, and the robustness of pix2pix model is worse than CycleGAN model. In addition, according to generated result, we can find that CycleGAN model can transfer the 2-dimensional stuff in the image well, but result of 3 dimensional and overlapping stuffs are not good enough.

6. Conclusion

In our experiment, DeepLab could output accurate segmentation results in most scenes. Our statistics results accord with some authority data. However, it performs poorly some scenes that are rare in Cityscapes, such as construction site. To improve the performance, a feasible method is to train the model on a Chinese dataset. For example, we could train it on Apolloscape, a dataset published by Alibaba.

Because the limitation of computing power (only one 1080ti GPU), the performance of transfer of the city style is not as good as our expectation: Images are not high definition; generated results of overlapping stuffs are not good enough; different stuffs may be transferred incorrected (e.g. bus was transferred to house). Therefore, if we have greater computing power in the futrue, we might try pix2pix HD or CycleGanHD, and would to use more time to do the training. At that time the result of style transfer will be much better.

Reference

[1] Huang Jianzhong, Hu Gangyu. Comparisons and Reflections on Walkability Measurements of Urban Built Environment [J]. Journal of Human Settlements Environment in West China, 2016, 31 (01): 67-74.

[2] Tang Jingxian, Long Ying. Measurement of Street Space Quality in Central District of Metropolitan City — Taking Beijing Second and Third Rings and Shanghai Inner Ring as Examples [J]. Planner, 2017, 33 (02): 68-73.

[3] Long Ying. New Thoughts on Urban Research and Planning and Design under the New Data Environment of Street Urbanism [J]. Times Architecture, 2016 (02): 128-132.

[4] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[J]. arXiv preprint, 2017.

[5] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[J]. arXiv preprint, 2017.

[6] Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 3431-3440.

[7] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation[J]. arXiv preprint arXiv:1511.00561, 2015.

[8] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[9] Zhao H, Shi J, Qi X, et al. Pyramid scene parsing network[C]//IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). 2017: 2881-2890.

[10] Chen L C, Papandreou G, Schroff F, et al. Rethinking atrous convolution for semantic image segmentation[J]. arXiv preprint arXiv:1706.05587, 2017.

[11] Cordts M, Omran M, Ramos S, et al. The cityscapes dataset for semantic urban scene understanding[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 3213-3223.