Evaluating the Performance of Deep Learning in Segmenting Google Street View Imagery for Transportation Infrastructure Condition Assessment

Wei Y., Liu K.
Stevens Institute of Technology, U.S.A.
Kaijian.Liu@stevens.edu

Abstract. Understanding the relationships between the condition of transportation infrastructure and the well-being of citizens in society is of significant importance towards restoring the aging and deteriorating transportation infrastructure in a way that enhances well-being. However, attaining such understanding is challenging because it relies on large-scale transportation infrastructure condition assessment. The broad spatial coverage of Google Street View (GSV) imagery offers a unique opportunity for such large-scale assessment. However, despite the richness of deep learning methods that can segment and recognize transportation infrastructure assets from GSV imagery for subsequent condition assessment, the performance of these methods typically varies. As such, this paper focuses on conducting performance evaluation of representative deep learning-based image segmentation methods to identify the optimal methods for recognizing transportation assets from GSV imagery. The preliminary evaluation results show that the ResNet + UNet and MobileNet + UNet methods achieved the highest intersection over union (IOU) of 0.87.

1. Introduction

The aging and deteriorating transportation infrastructure has disproportionate impacts on the well-being of citizens in society. For example, according to the American Society of Civil Engineers' Infrastructure Report Card, more than 40% of the U.S. transportation infrastructure, such as roadways and highways, is in poor or mediocre condition (American Society of Civil Engineers, 2022a). Such poor conditions result in unreliable and inefficient transportation services, increased travel time and costs for businesses and households, and, in turn, lead to higher prices for goods and lower disposable household income (American Society of Civil Engineers, 2022b). Consequently, the well-being of people gets negatively affected as resources necessary to maintain and improve well-being become less accessible and affordable. Most often, the negative impacts on well-being get passed along disproportionately to people from traditionally underserved groups (American Society of Civil Engineers, 2022b).

There is, therefore, an evident need to quantitatively understand the relationships between the condition of transportation infrastructure and the well-being of citizens. Such an understanding would allow for restoring transportation infrastructure in a way that equitably enhances the well-being of different population groups in society. However, attaining the understanding is challenging because it requires assessing the condition of transportation infrastructure at a large scale (e.g., the city scale). Existing methods for transportation infrastructure condition assessment typically do not scale well. Visual assessment by human inspectors, although it is still common in practice, is too costly and time-consuming to be conducted at scale (Yeum et al., 2021). On the other hand, recent advancement in automated condition assessment technology has largely reduced human involvement in the assessment process. Such technologies include visual imaging, ground penetrating radar, infrared thermography, LiDAR, and hyperspectral imaging. For example, ground penetrating radar has been used to assess the thickness of road pavements

https://doi.org/10.7146/aul.455.c228

(Willett et al., 2006). Visual imaging, infrared thermography, and hyperspectral imaging technologies have been used to assess the conditions of road damages, such as cracks and delamination (Schnebele et al., 2015). However, these technologies have not been widely applied because of technology adoption barriers. Studies (e.g., Li et al., 2017) show that initial investments in assessment equipment and subsequent costs of training professional equipment operators are currently the principal factors hindering the wide adoption and application of these automated condition assessment technologies in practice.

The availability of Google Street View (GSV) imagery offers a great promise for large-scale transportation infrastructure condition assessment. First, GSV imagery is readily and publicly accessible. This advantage liberates transportation agencies from investing in new equipment and training specialized operators and, thus, largely eliminates technology adoption barriers. Second, GSV imagery covers transportation assets in almost any place around the world. Currently, GSV can already provide images that cover more than 10 million miles of roads in the world (Raman, 2017). The broad spatial coverage of GSV imagery lends itself a unique edge in conducting large-scale condition assessment. Third, GSV imagery is of high resolution. GSV images are usually taken using high-resolution camera systems, such as 20MP cameras. The high resolution not only allows for accurately segmenting and recognizing transportation assets from GSV images, but also enables the detection, localization, and quantification of damages on the recognized assets.

However, despite the promise of GSV imagery, there is a lack of studies that evaluate the performance of different deep learning methods in segmenting transportation infrastructure assets from GSV images. Such an evaluation is critical to identify the optimal learning methods that can be used to better conduct the segmentation to support subsequent condition assessment (i.e., damage detection, localization, and quantification). Existing research efforts have focused on images captured using hand-held cameras or cameras mounted unmanned aerial vehicles (UAVs). For a few studies that use GSV images, they are mainly committed to a single learning method and are limited in additionally comparing other alternative methods. To address this limitation, this paper focuses on conducting performance evaluation of representative deep learning-based segmentation methods to identify the optimal ones for segmenting transportation assets from GSV imagery. In the remainder of this paper, the evaluation methodology is introduced in detail, and the evaluation results are discussed.

2. State of the Art and Knowledge Gaps

A body of research efforts have been undertaken in the field of image-based transportation infrastructure condition assessment. According to recent survey studies by Spencer et al., (2019), existing image-based assessment methods mostly rely on images captured using hand-held cameras or cameras mounted on unmanned aerial vehicles (UAVs). For example, recent studies (e.g., Li et al., 2019; Sukanya et al., 2020; and Mahmud et al., 2021) have focused on detecting and segmenting roads from images captured using UAVs. However, a limited number of studies have explored the use of GSV imagery for condition assessment of transportation infrastructure. Recent studies that used GSV imagery for the assessment include Ma et al., (2017) and Alipour and Harris (2020).

However, despite the importance of existing research, there is a lack of understanding of the performances of different deep leaning-based image segmentation methods in segmenting

transportation infrastructure assets from GSV images. Such an understanding is critical to identify the optimal learning methods that can reliably segment transportation assets from GSV images to support subsequent condition assessment. On one hand, existing research efforts have focused on comparing and understanding the performances of different learning methods in segmenting general objects (e.g., humans, trees, and sky in natural scenes) from general images rather than GSV images. For example, Ahmed et al., (2020) evaluated the performance of several representative deep learning methods in segmenting humans from top-view images. However, the comparison results achieved using general images typically do not generalize to GSV images because, compared to general images, GSV images are often associated with severe occlusions (e.g., roads occluded by traveling vehicles) and are captured in dynamically changing environments (e.g., sunny days with an abundant amount of lights vs. cloudy/rainy days without decent lights). On the other hand, existing studies that use GSV images for transportation infrastructure condition assessment are limited in comparing alternative deep learning methods. For example, Ma et al., (2017) leveraged convolutional neural networks to detect pavement damages from GSV images; Campbell et al., (2019) used MobileNet to detect traffic signs from GSV images; and Alipour and Harris (2020) exploited deep residual networks (ResNet) to detect road defects from GSV images. However, these studies are committed to using a single learning method and did not additionally compare the performance of the chosen method to other deep learning-based segmentation methods. They are, thus, limited in identifying optimal methods for GSV-based transportation asset segmentation.

3. Evaluation Methodology

An evaluation methodology was developed and followed to evaluate the performance of representative deep learning methods in segmenting transportation infrastructure assets from GSV images. The methodology included three main steps: (1) data preparation, (2) deep learning method selection, and (3) deep learning method implementation and evaluation.

3.1 Data Preparation

Data preparation aimed to create a dataset with annotations for deep learning-based segmentation algorithm training and evaluation. Data preparation included four steps. First, a dataset, which includes a total of 500 GSV images, was created. The images were purposively sampled from main streets in Manhattan, New York City (NYC), so that each image captures key transportation infrastructure assets such as roads, sidewalks, bike lanes, and vehicles. Images for NYC were used in this study because road/street scenes in NYC are typically more complex than those in other cities, making the use of such images more suitable in comparing different segmentation algorithms. Figure 1 shows a sample of collected GSV images. Second, image preprocessing was conducted to resize raw images into the same size of 256 by 256 to allow for mini-batch-based segmentation model training. Third, the resized images were manually annotated using the Visual Geometry Group Image Annotator, which is a commonly used annotation tool for adding class labels to each pixel in an image. Each image pixel was annotated into one of the five classes that cover key transportation asset categories, including “Road”, “Sidewalk”, “Bike Lane”, “Vehicle”, and “Background”. Fourth, the dataset was split into a training set and a testing set at a ratio of 4:1, which is a commonly used ratio in image segmentation. The annotations in the testing set were used for evaluation purposes only.

Figure 1: Examples of Google Street View (GSV) Images.

3.2 Deep Learning Method Selection

Deep learning method selection aimed to select representative deep learning methods (for image segmentation) for the subsequent performance evaluation. The selection included two steps. First, representative deep learning methods for extracting visual features from images were selected. The selection focused on convolutional neural networks (CNN)-based feature extraction methods, because CNN is one of the most successful and widely used architecture for visual feature extraction. Based on the survey study by Minaee et al., (2020), four well-known CNN-based feature extraction methods were selected, including convolutional neural networks (CNN) (Fukushima, 1980), very deep convolutional networks (VGG) (Simonyan et al., 2015), residual neural networks (ResNet) (He et al., 2015), and MobileNet (Howard et al., 2017). Second, representative deep learning methods for learning from extracted visual features to segment images into pre-defined segmentation classes were selected. Four representative methods were selected, including fully convolutional networks (FCN)-8 (Long et al., 2015), FCN-32 (Long et al., 2015), SegNet (Badrinarayanan et al., 2015), and UNet (Ronneberger et al., 2015). FCNs were selected because they are one of the first deep learning methods for image segmentation and have been commonly used as a benchmark (Minaee et al., 2020). SegNet was selected because, unlike traditional FCN (which leverages a shallow network architecture for up-sampling), it uses a decoder network that includes multiple up-sampling layers and a pixel-wise classification layer for segmentation. Each up-sampling layer uses the pooling indices from its corresponding down-sampling layer to conduct non-linear up-sampling without the need to learn how to up-sample. Such an up-sampling architecture configuration allows for reducing the number of parameters to be learned to improve the performance of segmentation models. UNet was selected because it leverages a contracting path with down-sampling to capture visual contexts and an expanding path with multiple up-sampling layers that also use down-sampled feature maps as input to capture visual patterns to enable precise localization and segmentation. The use of the two paths allows for using a small number of training images to achieve more precise segmentation.

3.3 Deep Learning Method Implementation and Evaluation

The selected deep learning methods were implemented to develop models for segmenting GSV images into the pre-defined segmentation class (i.e., “Road”, “Sidewalk”, “Bike Lane”, “Vehicle”, and “Background”). The implementation included three steps. First, the selected

visual feature extraction methods were implemented. Figure 2 shows the specific deep learning architecture used to implement each extraction method. Second, the selected segmentation methods were implemented. Figure 3 shows the specific learning architecture used to implement each segmentation method. In a segmentation learning architecture, a visual feature extraction architecture was used as an "encoder" to down-sample raw input images to extract feature maps, and the "decoder" architecture of the selected segmentation method was used to up-sample the feature maps for pixel-wise classification/segmentation. For example, in the CNN + FCN-32 architecture, the CNN architecture was used to extract feature maps, and the 32X up-sampling layer in the FCN-32 was used to up-sample the feature maps from the 5th pooling block of the CNN for pixel-wise classification. As a result, a total of 16 segmentation models were implemented, and each model was implemented using a combination of a selected visual feature extraction method and a selected segment method. The Python code implementation from Divam (2019) was used for the implementation. Third, the segmentation models were separately trained using the training dataset. During the model training, the training pixel accuracy was monitored to ensure that the model is fully trained to converge. Pixel accuracy, as per Equation (1), is the ratio of the number of correctly segmented pixels to the total number of pixels. In Equation (1), C is the number of pre-defined segmentation classes, and P_{ij} is number of pixels of class i that are classified as class j. Figure 4 shows the training pixel accuracy against the training epoch for each model. As per Figure 4, at epoch = 10, the training pixel accuracy for each segmentation model was over 90% and became stable, which indicates that the model is trained to converge.

	CNN	VGG	ResNet	MobileNet
Pool 1	Zero Padding 3x3, 64 Conv. BatchNorm. ReLU Activation MaxPooling	3x3, 64 Conv. 3x3, 64 Conv. MaxPooling	Zero Padding 7x7, 64 Conv.	3x3, 32, Conv. Block 64, Depthwise Conv. Block
Pool 2	Zero Padding 3x3, 128 Conv. BatchNorm. ReLU Activation MaxPooling	3x3, 128 Conv. 3x3, 128 Conv. MaxPooling	3x3, 64-64-256, Conv. Block 3x3, 64-64-256, Identity Block 3x3, 64-64-256, Identity Block	128, Depthwise Conv. Block 128, Depthwise Conv. Block
Pool 3	Zero Padding 3x3, 256 Conv. BatchNorm. ReLU Activation MaxPooling	3x3, 256 Conv. 3x3, 256 Conv. 3x3, 256 Conv. MaxPooling	3x3, 256-256-1024, Conv. Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block	256, Depthwise Conv. Block 256, Depthwise Conv. Block
Pool 4	Zero Padding 3x3, 256 Conv. BatchNorm. ReLU Activation MaxPooling	3x3, 512 Conv. 3x3, 512 Conv. 3x3, 512 Conv. MaxPooling	3x3, 256-256-1024, Conv. Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block 3x3, 256-256-1024, Identity Block	512, Depthwise Conv. Block 512, Depthwise Conv. Block 512, Depthwise Conv. Block 512, Depthwise Conv. Block 512, Depthwise Conv. Block
Pool 5	Zero Padding 3x3, 256 Conv. BatchNorm. ReLU Activation MaxPooling	3x3, 512 Conv. 3x3, 512 Conv. 3x3, 512 Conv. MaxPooling	3x3, 512-512-2048, Conv. Block 3x3, 512-512-2048, Identity Block 3x3, 512-512-2048, Identity Block	1024, Depthwise Conv. Block 1024, Depthwise Conv. Block

Figure 2: Learning Architectures for the Selected Deep Learning-based Visual Feature Extraction Methods.

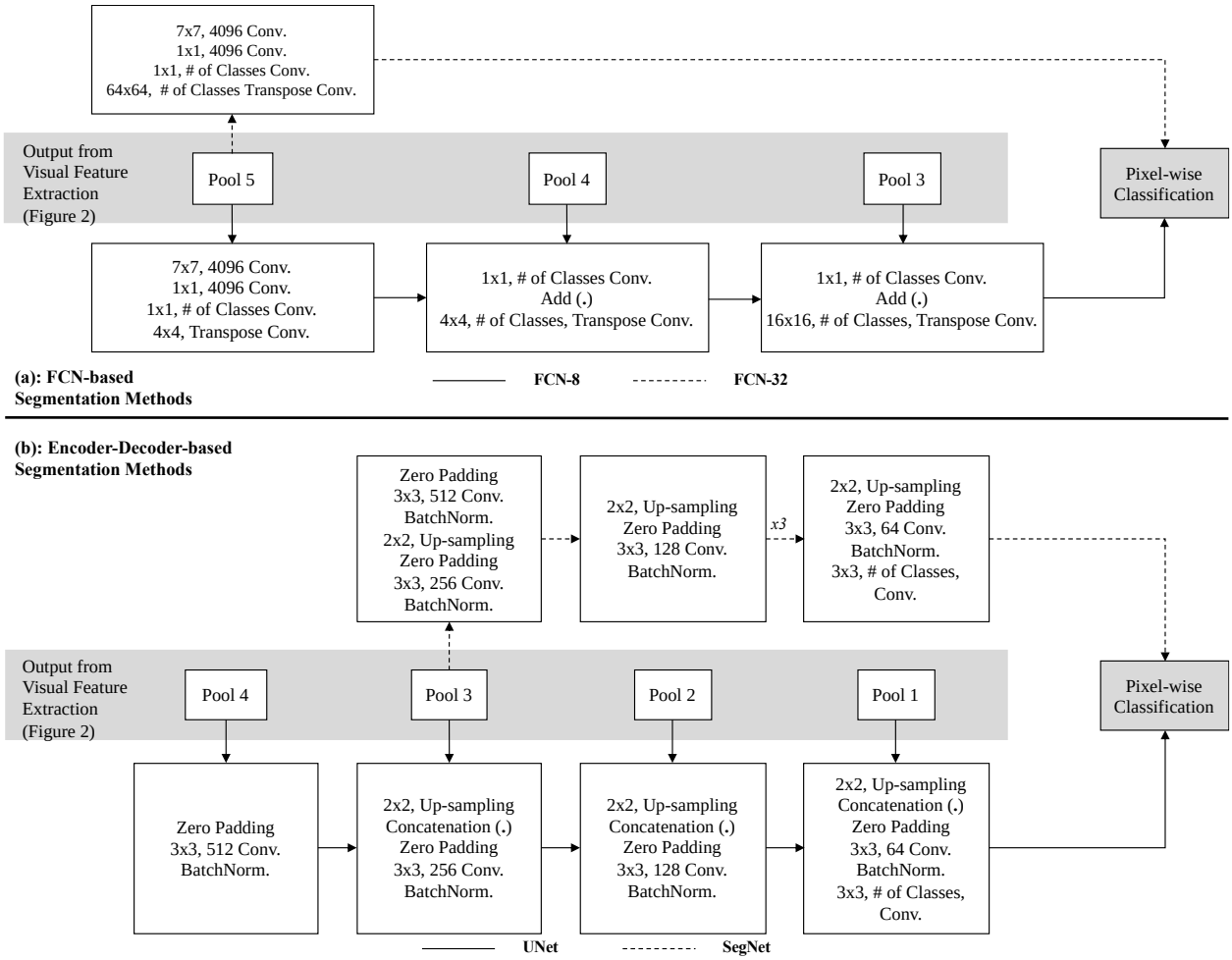

Figure 3: Learning Architectures for Selected Deep Learning-based Segmentation Methods.

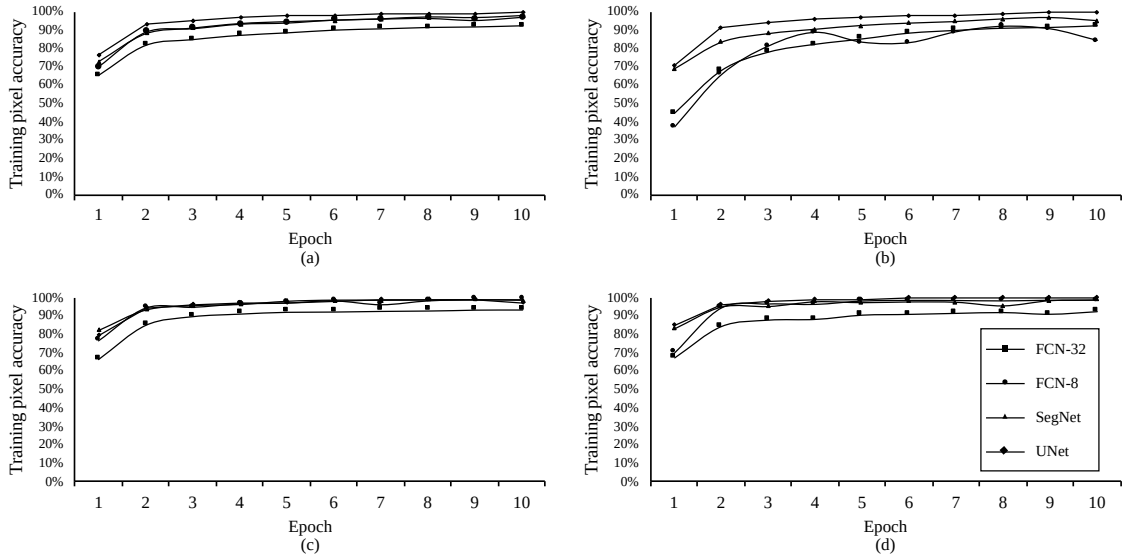

Figure 4: Training Pixel Accuracy of Segmentation Models: (a) CNN-based Models, (b) VGG-based Models, (c) ResNet-based Models, and (d) MobileNet-based Models.

The trained segmentation models were evaluated using two commonly used metrics for image segmentation: class-wise intersection over union (IOU) and mean IOU. For each segmentation class, class-wise IOU, as per Equation (2), is the ratio of the area of intersection between the classified segmentation map for the class and the gold standard segmentation map for the class to the area of union between the two maps. Mean IOU, as per Equation (3), is the average of class-wise IOUs over all the segmentation classes. In Equations (2) and (3), *i* is a segmentation class, *A* is classified segmentation map for class *i*, and *B* is gold standard segmentation map for the class *i*.

$$Accuracy = \frac{\sum_{i=0}^{C} P_{ii}}{\sum_{i=0}^{C} \sum_{j=0}^{C} P_{ij}} \quad (1)$$

$$Class\text{-}wise\ IOU(i) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

$$Mean\ \ IOU = \frac{1}{C} \sum_{i=0}^{C} Class\text{-}wise\ IOU(i) \quad (3)$$

4. Evaluation Results and Discussion

The results for evaluating the performance of the selected deep learning methods in segmenting transportation infrastructure assets from GSV imagery are presented in Table 1. Overall, compared to the other combinations of the feature extraction and segmentation methods, the combinations ResNet + UNet and MobileNet + UNet achieved the highest mean IOU of 0.87.

Two important observations are drawn from the evaluation results. First, using skip connections and separable depthwise convolutions are important strategies to extract representative visual features from GSV images to support subsequent transportation asset segmentation. As seen in Table 1, the ResNet-based and MobileNet-based models achieved an average mean IOU (across all the segmentation methods) of 0.82 and 0.82, respectively. The CNN-based and VGG-based models only achieved an average mean IOU of 0.76 and 0.64, respectively. ResNet uses a set of residual blocks (i.e., convolution and identify blocks, as per Figure 2), which formulate the output layers of each block to learn residual functions with reference to the layer inputs. Such a residual learning configuration allows for substantially increasing the depth of the networks to better capture distinctive visual features, yet without negatively affecting the efficiency and performance of model optimization. On the other hand, MobileNet utilizes depthwise separable convolutional layers, which apply a separate filter only at one input channel (instead of at multiple channels at once). Depthwise separable convolution layers can generate outputs same as those generated by traditional convolution layers, but they require much less parameters to reduce the size of the model and improve the performance of feature extraction. The evaluation results also suggest that incorporating depthwise separable convolutions into residual blocks could allow for benefiting from the advantages of both ResNet and MobileNet to further improve the performance of visual feature extraction from GSV images.

Second, concatenating down-sampled feature maps and up-sampled feature maps to further up-sample the maps for pixel-wise classification shows effectiveness in segmenting transportation infrastructure assets from GSV images. As seen in Table 1, the UNet-based models achieved an average mean IOU (across all the feature extraction methods) of 0.85, which is higher than the IOUs by the FCN-32-based, FCN-8-based, and SegNet-based models (IOU = 0.57, 0.80 and 0.82,

respectively). SegNet, compared to FCN-8, achieved a slightly higher IOU. This could be because SegNet uses more up-sampling and transpose convolution operations than FCN-8. UNet, compared to SegNet, further conducts convolutions on the concatenated feature maps (i.e., up-sampled maps and their corresponding down-sampled maps), which allows UNet to achieve a 3% higher IOU than SegNet. The evaluation results, thus, suggest that encoder-decoder-based segmentation methods, such as SegNet and UNet, are generally more effective than FCN-based methods in segmenting transportation assets from GSV images. In future research efforts, other types of encoder-decoder-based methods, such as VNet and WNet, can be evaluated to assess if they can further improve the performance of GSV image segmentation.

Table 1: Results for Evaluating the Performance of Deep Learning Methods in Segmenting Transportation Infrastructure Assets from GSV Images.

Feature extraction method	Segmentation method	Class-wise IOU					Mean IOU
		Background	Road	Sidewalk	Vehicle	Bike Lane	
CNN	FCN8	0.86	0.89	0.72	0.45	0.82	0.75
	FCN32	0.82	0.86	0.70	0.39	0.71	0.70
	UNet	0.95	0.86	0.87	0.53	0.82	0.81
	SegNet	0.90	0.91	0.79	0.50	0.85	0.79
VGG	FCN8	0.89	0.92	0.79	0.45	0.86	0.78
	FCN32	0.49	0.27	0.05	0.00	0.00	0.15
	UNet	0.96	0.93	0.54	0.55	0.92	0.84
	SegNet	0.91	0.90	0.76	0.51	0.85	0.79
ResNet	FCN8	0.95	0.96	0.86	0.55	0.91	0.84
	FCN32	0.81	0.87	0.71	0.38	0.76	0.71
	UNet	0.98	0.97	0.91	0.58	0.93	0.87
	SegNet	0.95	0.94	0.59	0.56	0.90	0.85
MobileNet	FCN8	0.93	0.94	0.85	0.54	0.88	0.83
	FCN32	0.83	0.88	0.75	0.43	0.75	0.73
	UNet	0.98	0.97	0.90	0.58	0.93	0.87
	SegNet	0.95	0.94	0.59	0.56	0.90	0.85

5. Conclusions, Limitations, and Future Work

In this paper, a set of representative deep learning methods were evaluated in segmenting transportation infrastructure assets from GSV images. The evaluation results show that the ResNet + UNet and MobileNet + UNet methods achieved the highest mean IOU of 0.87 – indicating the suitability of these methods for segmenting transportation assets from GSV images. In addition, two conclusions were also drawn from the results. First, incorporating depthwise separable convolutions into residual blocks could allow for better extracting distinctive visual

features from GSV images to support subsequent segmentation. Second, encoder-decoder-based methods are more suitable than FCN-based methods in segmenting GSV images.

Two main limitations of this study are acknowledged. First, as a pilot study, this paper focused on evaluating CNN-based feature extraction methods and FCN-based and encoder-decoder-based segmentation methods. Other methods, such as multiscale and pyramid network-based and attention-based methods, were not evaluated. In their future work, by following the evaluation methodology presented in this paper, the authors plan to further evaluate the performance of other prominent deep learning methods in segmenting GSV images. Second, the size of the dataset used in the evaluation is limited. For example, the dataset in this study mainly covers transportation assets in Manhattan, NYC, but does not cover other areas in the nation. A larger and more diverse GSV dataset can be curated and used for evaluation in the future.

In their future work, the authors will focus their research efforts on three main directions. First, developing a new deep learning-based image segmentation method, based on the findings of this study, to better segment transportation assets from GSV images. The segmentation is the first step toward large-scale transportation infrastructure condition assessment. Hence, the segmentation method to be developed will need to achieve an IOU of over 95% to reduce the number of errors propagating into the subsequent assessment steps. In addition to GSV images, the method will also be applied to segment transportation assets from images captured by the cameras on autonomous vehicles and/or reported by citizens. Such images would also be valuable for transportation infrastructure condition assessment, because they are more dynamic and can better capture update-to-date infrastructure conditions. Second, developing new deep learning methods for detecting, localizing, and quantifying transportation asset damages in the segmented images that include key transportation assets (instead of assessing damages using unsegmented images to improve performance of damage condition assessment). Third, conducting data-driven investigations to understand the relationship between transportation infrastructure condition and citizen well-being to support well-being informed infrastructure investment decision making.

References

Ahmed, I., Ahmad, M., Khan, F.A., Asif, M. (2020). Comparison of deep-learning-based segmentation models: Using top view person images, IEEE Access, 8, pp. 136361136373.

Alipour, M., Harris, D.K. (2020). A big data analytics strategy for scalable urban infrastructure condition assessment using semi-supervised multi-transform self-training, Journal of Civil Structural Health Monitoring, 10 (2), pp. 313–332.

American Society of Civil Engineers (2022a). Report Card for America's Infrastructure, https://infrastructurereportcard.org/, accessed January 2022.

American Society of Civil Engineers (2022b). Failure to Act Economic Reports, https://infrastructurereportcard.org/resources/failure-to-act-economic-reports/, accessed January 2022.

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), pp. 2481–2495.

Campbell, A., Both, A., Sun, Q.C. (2019). Detecting and mapping traffic signs from Google Street View images using deep learning and GIS, Computers, Environment and Urban Systems, 77, p. 101350.

Divam, G (2019). A Beginner's Guide to Deep Learning Based Semantic Segmentation Using Keras, https://divamgupta.com/image-segmentation/2019/06/06/deep-learning-semantic-segmentation-keras.html, accessed December 2021.

Fukushima, K., Miyake, S. (1980). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, Biological Cybernetics, 36, pp. 193–202.

He, K., Zhang, X., Ren, S. and Sun, J. (2016). Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, 2016, Las Vegas, NV.

Howard, A.G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M. and Adam, H. (2017). Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv: 1704.04861.

Li, M., Faghri, A., Ozden, A., Yue, Y. (2017). Economic feasibility study for pavement monitoring using synthetic aperture radar-based satellite remote sensing: Cost-benefit analysis, Transportation Research Record, 2645 (1), pp. 1–11.

Li, Y., Peng, B., He, L., Fan, K., Tong, L. (2019). Road segmentation of unmanned aerial vehicle remote sensing images using adversarial network with multiscale context aggregation, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 12 (7), pp. 2279–2287.

Long, J., Shelhamer, E. and Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In: IEEE Conference on Computer Vision and Pattern Recognition, 2015, Boston, MA.

Ma, K., Hoai, M. and Samaras, D. (2017). Large-scale continual road inspection: Visual infrastructure assessment in the wild. In: British Machine Vision Conference, 2017, London, United Kingdom.

Mahmud, M.N., Osman, M.K., Ismail, A.P., Ahmad, F., Ahmad, K.A. and Ibrahim, A. (2021). Road image segmentation using unmanned aerial vehicle images and DeepLab V3+ semantic segmentation model. In: 11th IEEE International Conference on Control System, Computing and Engineering, 2021, Penang, Malaysia.

Minaee, S., Boykov, Y.Y., Porikli, F., Plaza, A.J., Kehtarnavaz, N. and Terzopoulos, D. (2021). Image segmentation using deep learning: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence.

Raman A (2017). Cheers to Street View 10th birthday! https://www.blog.google/products/maps/cheers-street-views-10th-birthday/, accessed August 2021.

Ronneberger, O., Fischer, P. and Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, 2015, Munich, Germany.

Schnebele, E., Tanyu, B.F., Cervone, G., Waters, N. (2015). Review of remote sensing methodologies for pavement management and assessment, European Transport Research Review, 7 (2), pp. 1–19.

Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv: 1409.1556.

Spencer Jr, B.F., Hoskere, V., Narazaki, Y. (2019). Advances in computer vision-based civil infrastructure inspection and monitoring, Engineering, 5 (2), pp. 199–222.

Sukanya and G. Dubey. (2020). Segmentation and detection of road region in aerial images using hybrid CNN-random field algorithm. In: 10th International Conference on Cloud Computing, Data Science & Engineering, 2020, Uttar Pradesh, India.

Willett, D.A., Mahboub, K.C., Rister, B. (2006). Accuracy of ground-penetrating radar for pavement-layer thickness analysis, Journal of Transportation Engineering, 132 (1), pp. 96–103.

Yeum, C.M., Choi, J., Dyke, S.J. (2019). Automated region-of-interest localization and classification for vision-based visual assessment of civil infrastructure, Structural Health Monitoring, 18 (3), pp. 675–689.