Roadside Forest Modeling Using Dashcam Videos and Convolutional Neural Nets

Durga Joshi¹, Chandi Witharana¹

Department of Natural Resources and the Environment, University Of Connecticut, Storrs, CT, - (durga.joshi,

Chandi.witharana)@uconn.edu

KEY WORDS: Dashcam videos, Roadside Forest, Vegetation management, Deep learning, U-Net, YOLOV5

ABSTRACT: Tree failure is a primary cause of storm-related power outages throughout the United States. Roadside vegetation management is therefore critical to electric utility companies to prevent power outages during extreme weather conditions. It is difficult to execute roadside vegetation management practices, at the landscape level, without proper monitoring of roadside forests' physical structure and health condition. Remote sensing images and LiDAR are widely used to characterize the forest edge; however, the limitation on the temporal and spatial resolution for most of that dataset is a big challenge. Also, there is a need for a ground-level dataset that provides the vertical profile of the forest trees so that we can more accurately characterize the forest structure and health and recommend the optimal management strategies according to the local forest conditions. For the first time, we introduced Dashcam videos as an alternative to the existing aerial remote sensing data sources to characterize the roadside forest condition using the deep learning (DL) convolutional neural net (CNN) algorithms. In this study, we used dashcam videos taken during the leaf-on and leaf-off conditions and various weather conditions along the roadside. We trained a DLCNN model based on the U-Net and YOLO v5 architectures to classify the multilayer vegetation and detect utility poles and tree trunks alongside the road. Our experiment results suggest that a dashcam can be a viable alternative and complementary way to characterize the roadside vegetation and can be used in the management of roadside forests as a cost-effective data acquisition mechanism for utility companies.

1. INTRODUCTION

Roadside vegetation is defined as the area which consists of all trees and shrubs along with all types of roads on all types of land ownerships that extend up to the distance from the road to which a mature tree could impact the road and utilities (Eric, 2012). Roadside vegetation not only has aesthetic importance, but it functions as home to different wildlife (Cadenasso & Pickett, 2000), as well as provide various ecosystem services (Salmond et al., 2016; Weber et al., 2014), improves air quality (Smith, 2012), blocks pollutants from vehicles (Jin et al., 2014; Tong et al., 2015), works as a noise barrier (Ow & Ghosh, 2017), etc. However, there are major concerns regarding roadside vegetation like highway safety and visibility (Forman & McDonald, 2007). Also, utility companies are badly affected due to the destruction caused by roadside vegetation like a tree falling and ground contact by climbers, etc. Proper management of threatful roadside vegetation that could be a threat to the utility companies is very necessary. For this, we need to have an idea of the status of the vegetation and its proximity to the infrastructures.

Remote sensing images are served as indispensable resources in quantifying the forest structure and health condition. Remote sensing datasets acquired from multiple platforms, such as satellite images (Ingram et al., 2005), aerial images (Hall et al., 2003), or (Unmanned Aircraft System) UAS images (Belmonte et al., 2020) is widely used in the field of forestry. Satellite images usually come with less temporal and spatial resolution. For instance, Landsat at 30m resolution with 15-day frequency Aerial images are mostly high resolution compared to publicly available satellite images, but they lack temporal frequency and are cost-prohibitive. While UASs offer super-resolution imagery, they are hampered by scalability and deploying ability over a large area. Both distribution and transmission lines could occupy tens of thousands of kilometers. For instance, the state of Connecticut has xxx long distribution network. Thus, repeated monitoring of roadside vegetation along powerline corridors inevitably demands data streams acquired at high spatial resolution and temporal frequencies without compromising the geographical extent. Generally, conventional remote sensing data, imagery, and/or LiDAR come with many limitations for the precise assessment of the roadside vegetation. These datasets only provide the aerial view of the objects of interest (e.g., trees), thus, they may overlook actual conditions on the ground. The vertical profile of the objects on the earth's surface may differ from the image captured above. The aerial view of tree crowns from the image cannot provide understory vegetation. While LiDAR could capture the vertical structure of upper canopies effectively, shortcomings in point densities could limit the accuracy of canopy height models, derivation of useful forest metrics (e.g., stem density), and potentially prohibit seeing understory vegetation. It is very important to monitor the status and growth of the vegetation so that utilities can timely manage and prevent any harm from their growth to the surrounding infrastructure. Thus, there is a need for a ground-level dataset that provides the vertical profile of the forest trees to accurately describe the forest characteristics. Ground-level images/videos are the potential datasets that can fulfill this demand. These data have higher spatial resolution than conventional remote sensing data and are providing more information on the entire tree that is comparable to the expert view while assessing the tree manually. A sensor on the ground could gain a distinctive perspective on the spatial composition and characteristics of targets along the roadside. Thus, a roadside scene captured as an image or video would closely provide a novel and highly valuable assessment of forest structure (stem density, trees/branches along electricity lines, cantilevered crown, canopy cover) and condition (unhealthy or dead trees, broken/damaged trees) relative to conventional remote sensing data streams. Such a street viewpoint would also allow for characterization of the surrounding infrastructure and its proximity to hazardous trees/branches. A dashboard camera (commonly known as and hereafter dashcam) is a low-cost vision sensor, usually mounted inside a vehicle to record street-level visual observation from the driver's point of view. Besides the objects on the road, dashcams can provide unique perspectives of roadside objects that are particularly important to vegetation risk modeling. Since the data acquisition can be done on-demand basis and it can be an important source of information to open a new perspective in forest characterization and management.

One of the prominent sources of ground-level information is the dashcam used in public and privately-owned vehicles. A dashcam is widely used in developed countries like Russia, Taiwan, China, etc. (Rea et al., 2018) in vehicles for safety purposes. They serve as an important source of evidence during the disputes and to claim insurance for the big commercial transportation companies like Uber, Lyft, etc. in the United States (Lyft, 2022; UBER, 2022). It is an affordable ad ubiquitous technology. Current day, dashcam videos are getting popular in the field of transportation to characterize and monitor the status of the roads, bridges, traffic signs, cracks, etc. (Dadashova et al., 2021; Hou et al., 2022). Dashcam-based video streams can be used in the assessment of the roadside vegetation and the status of utility infrastructures.

Trees' growth rate depends on various factors such as species type, microclimate, soil types, nutrient availability, etc. The branches and crown size of the trees also depend on such factors. It is found that the crown size of the tree extends more in the open areas (MacFarlane & Kane, 2017), mainly for the deciduous trees. Roadside gives trees more opportunity to extend outwards towards shoulder boundary easily. They can impact the visibility of the road as well as they can easily encounter the utility infrastructure. It can be a serious threat to the powerlines that may lead to the power outage in normal as well as severe weather conditions (thunderstorms, winter storms). Mostly, mature tree trunks that are close and high enough to touch the powerlines are major threats. Assessing and monitoring growth conditions like proximity to the infrastructures, tree height, and size, greenness, etc. is very necessary. Timely management of such vegetation growth is critical to ensure safety from all possible threats. Most of the time, assessment of such factors is done manually via on-foot scouting, which is time-/labor/cost-intense). Given the extensive coverage of powerlines, it is almost impractical to employ a real-time assessment of vegetational growth in repeated intervals. There is a strong need for automated vegetation scouting methods (or digital scouting) to cut down the cost, save time, and inform data-driven vegetation management decision-making. A dashcam can be one of the prudent alternatives to achieve the goal. In this regard, from the dashcam technology, we can automate the monitoring of the roadside vegetation and can be further used to plan the management strategy.

Automated detection of the roadside vegetation (e.g., invasive species, dead trees, threatful branches) electric infrastructure (e.g., poles, lamps, wire), and traffic-related infrastructures (e.g., traffic signs, lights, roads, shoulders, etc. from the roadside images is very important to plan the management strategies. Utility companies need to know vegetation status and pre-plan vegetation management activities (e.g., tree trimming and hazard tree removal) to reduce tree-related power outages during severe weather conditions. Also, near real-time vegetation assessments help in developing necessary interventions such as the management of plant diseases, fire, and invasive species. Even though it is very challenging to automate accurate segmentation of roadside conditions, dashcams can be taken as a possible alternative. In recent studies, dashcam data sources are widely utilized for recognition and localization purposes. Many algorithms and datasets, such as CamVid (Brostow et al., 2008), Leuven (Leibe et al., 2007), and Daimler Urban Segmentation (Scharwächter et al., 2013) are successfully used in the semantic understanding of urban scenes. A benchmark dataset for vision called KITTI (Geiger et al., 2012) is also available that can study vision-based self-driving tasks including object detection, multiple-objects tracking, road/lane detection, semantic segmentation, and visual odometry.

The central goal of this exploratory study is to detect and segment the roadside vegetation with the use of readily available vision data sources i.e., dashcam, and try to see the feasibility of such data sources as citizen data science in the future. We aimed to address two specific objectives 1) to differentiate the canopy structure of the roadside vegetation. 2) to detect tree trunks and electricity poles. To our knowledge, this is the first research effort to explore the utility of dashcam videos in roadside forest risk analysis,

2. METHODS

2.1 Data Collection:

We collected dashcam video using Think ware U1000 Dashboard camera along the roads of Mansfield, Connecticut, US. Thinkware U100 is an off-the-shelf, GPS-enabled, camera that records 4K video at 60fps and has 150° of wide viewing angles. Video dataset was taken in leaf-on (Summer/Fall, 2021) and leaf-off (Spring, 2021) conditions under different daylight conditions (late morning, afternoon, and early evening) and weather conditions (rainy, cloudy, bright sunny day, fog, light snow). This was done to train the DLCNN algorithm in heterogeneous imaging conditions. Acquired videos were further converted to image frames using FFmpeg algorithms (Tomar, 2006). We used the 30th frame and the 15th frame per second for leaf-on and leaf-off conditions for objectives 1 and 2, respectively. An online web tool "VGG Image Annotator" (Dutta et al., 2016) was used to annotate the image frames. For Objective 1, We created a freehand polygon around the target whereas for objective 2 we annotated with a rectangle polygon around the target. The overall methodology used in the study is depicted in Figure 1.



 Data Generation
 Training data
 Algorithms
 Predictions

 Figure 1: Schematic showing the overall methodology used in the study.
 Study.
 Study.

2.2 Model Training

The U-Net algorithm and the YOLO v5 algorithm were used in Objective 1 and Objective 2, respectively. The former involved semantic segmentation of roadside forest canopy while the latter involved the detection of tree trunks and utility poles.

2.2.1 U-Net Model

We applied a transfer learning strategy to re-train the U-Net model. The model has a Resnet101 backbone that has already been trained based on ImageNet data. Transfer learning helps faster adaptation of an existing model using a limited number of training samples. Using the VGG annotator, we delineated roadside vegetation as the high canopy and low canopy from dashcam video frames. The high canopy class captures the trees that have a height taller than the utility wires and the low canopy class engulfs shrubs, saplings, and other vegetation with a smaller height than the utility wires. Since we are in the preliminary phase of our analysis, we only used 64 image frames for the annotation purpose with around 525 of a total number of targets. Annotated video frames into high and low canopy classes are shown in Figure 2.



Figure 2: a) A video frame from the dashcam b) Annotated image into different classes using VGG image annotator

We utilized a Python implementation of the U-Net model. The model was based on the TensorFlow library. Some of the specifications used for the U-Net model are presented in Table 1:

Table 1: Key experimental parameters used in the U-Net model

Parameters	Value
Number of Images	64
Data Split ratio	80:10:10
Backbone	Resnet
Epochs	100
Batch size	8
Classes	3(high, low, road)
Augmentation	Horizontal flip, vertical flip, rotation by 90°

2.2.2 YOLO v5 Model

For Objective 2, the YOLO v5 algorithm was utilized to train our dataset. YOLO v5 is the frontline algorithm of the YOLO series (Redmon et al., 2016). In this detection task, we mainly targeted 1) tree trunks on the roadside and 2) utility poles during leaf-off conditions. For annotation, we used 514 image frames from different videos and were given a bounding box annotation in the VGG image annotator as shown in Figure 3. A total of 743 poles and 1750 trunks were digitized manually.



Figure 3: Tree trunks and Poles annotated from dashcam Image frames

Table 2 depicts key parameters used in the YOLO v5 algorithm during training. No data augmentation was involved at this point to see how well we can detect the tree trunk.

Table 2: Some Experimental parameters used for YOLOv5
algorithm

Parameters	Value
Number of Images	514
Data Split ratio	70:15:15
Epochs	300
Classes	2 (l= tree trunk, p=pole)
Augmentation	none

2.3 Results and Discussion

We were able to successfully stratify the vegetation using the U-Net model from the dashcam video. The overall IOU - score was more than 85% and the F1- score was above 91%. For the stratification of the roadside vegetation canopy, evaluation for training and validation progress is provided in Figure 4. Examples of the predicted mask and annotated mask over the test data are provided in Figure 5.



Figure 4: Training and Validation curve for Loss, IOU, and F1-Score.



Figure 5: a) Annotated video frame b) Prediction mask for high canopy (pink color) and low canopy (light green color).

The class-wise accuracy metric achieved during the analysis is provided in Table 3:

Table 3: Results achieved in the classification of vegetation strata from dashcam videoframes

Classes	Precision	Recall	F1-score
Background	0.909	0.913	0.911
High Vegetation	0.916	0.919	0.917
Low Vegetation	0.925	0.953	0.939
Road	0.969	0.980	0.974
Accuracy	0.923	0.923	0.923
Weighted.	0.912	0.923	0.917
average			

For Objective 2, i.e., detection of the poles and tree trunks along the roadside, our algorithm was able to detect most of the tree trunks and utility poles. Figure 6 shows the loss performance during the training and validation of the algorithm.



Figure 6: Training loss and validation loss for the class in YOLO v5

As seen in the confusion matrix (Figure 7), the detection of the pole was more accurate than the tree trunks. Almost 92 % of the background false positive rate was due to the tree trunk and it was producing background false negatives as well. However, poles had very less contribution to the false positive rate. Data annotation was done by multiple individuals. Even though we had a protocol in place for annotations, still there was a bias among the annotators toward annotating biforked/multiforked trees into one or multiple tree trunks in the stacked forest areas. It could be the possible source of noise in detection. Additionally, the presence of deep background tree trunks within the bounding boxes could be another possible reason behind the false positive and false negative rates for the tree trunks.



Figure 7: Confusion matrix showing the accuracy rate for two classes: tree trunk (l) and poles (p)



Figure 8: Detection of tree trunk and pole on an independent dataset

Lack of sufficient training samples, the difference in the background contrast of frames due to variation of daylight time, glare on the windshield, and sun angle, could have probably contributed to low detection accuracies. Even though the mAP value for each class is low, we still were able to detect a substantial amount of roadside tree trunks Figure 8).

3. CONCLUSIONS

To the best of our knowledge, no other study has yet explored the utility of dashcam video streams in roadside forest monitoring applications. This work introduces the U-Net-based stratification of the roadside vegetation and the YOLOv5-based tree trunk and pole detection method. U-Net model was found to be promising in the classification of the high and low canopy of the vegetation even with a very limited training dataset. Similarly, the YOLOv5 detection method was also found as a feasible method to detect tree trunks and utility poles. Since this is the initial phase of the study,

the level of accuracy we achieved from both tasks is motivating. In future work, we will expand the training data space via the collection of a large volume of video data and hand annotation of targets of interest. We will practice standardized protocols to maintain consistency in the annotation process among analysts. Such measures will help elevate the generalization ability of CNN models. Also, various augmentation techniques will be used to inflate the training sample space while providing robust scenarios to the CNN algorithms. This novel work has opened a costeffective data acquisition avenue to characterize roadside forest's physical structure and health condition. When fused appropriately, Dashcam video streams are complementary to conventional remote sensing data. Further, the ubiquitous nature of this everyday technology places it as an ideal citizen science tool for engaging the public in the process of near real-time roadside forest monitoring.

ACKNOWLEDGEMENT

We would like to acknowledge the Department of Natural resources and the environment, the University of Connecticut for the facilities, Eversource Energy Center, Storrs for the funding, Dr. Rajitha Udawalpola, and graduate assistants Harshana Wedagedara, Shashika Himandi for their enormous support in training data generation.

REFERENCES

- Belmonte, A., Sankey, T., Biederman, J. A., Bradford, J., Goetz, S. J., Kolb, T., & Woolley, T. (2020). UAV-derived estimates of forest structure to inform ponderosa pine forest restoration. *Remote Sensing in Ecology and Conservation*, 6(2), 181-197.
- Brostow, G. J., Shotton, J., Fauqueur, J., & Cipolla, R. (2008). Segmentation and recognition using structure from motion point clouds. Paper presented at the *European Conference on Computer Vision*, 44-57.
- Cadenasso, M. L., & Pickett, S. (2000). Linking forest edge structure to edge function: Mediation of herbivore damage. *Journal of Ecology*, 88(1), 31-44.
- Dadashova, B., Dobrovolny, C. S., Tabesh, M., & through Disruption, S. (2021). No title. Detecting Pavement Distresses using Crowdsourced Dashcam Camera Images,
- Dutta, A., Gupta, A., & Zissermann, A. (2016). VGG image annotator (VIA). URL: *Http://Www.Robots.Ox.Ac.Uk/Vgg/Software/Via, 2*
- Eric, H. (2012). State vegetation management task force final report. (). https://www.ct.gov/deep/cwp/view.asp?a=2697&q=503040 &deepNav_GID=1631

- Forman, R. T., & McDonald, R. I. (2007). A massive increase in roadside woody vegetation: Goals, pros, and cons.
- Geiger, A., Lenz, P., & Urtasun, R. (2012). Are we ready for autonomous driving? the KITTI vision benchmark suite. Paper presented at the 3354-3361. <u>https://doi.org/10.1109/CVPR.2012.6248074</u> <u>https://ieeexplore.ieee.org/document/6248074</u>
- Hall, R. J., Davidson, D. P., & Peddle, D. R. (2003). Ground and remote estimation of leaf area index in rocky mountain forest stands, kananaskis, alberta. *Canadian Journal of Remote Sensing*, 29(3), 411-427.
- Hou, Y., Shi, H., Chen, N., Liu, Z., Wei, H., & Han, Q. (2022). Vision image monitoring on transportation infrastructures: A lightweight transfer learning approach. *IEEE Transactions* on Intelligent Transportation Systems,
- Ingram, J. C., Dawson, T. P., & Whittaker, R. J. (2005). Mapping tropical forest structure in southeastern madagascar using remote sensing and artificial neural networks. *Remote Sensing of Environment*, 94(4), 491-507.
- Jin, S., Guo, J., Wheeler, S., Kan, L., & Che, S. (2014). Evaluation of impacts of trees on PM2. 5 dispersion in urban streets. *Atmospheric Environment*, 99, 277-287.
- Leibe, B., Cornelis, N., Cornelis, K., & Van Gool, L. (2007). Dynamic 3d scene analysis from a moving vehicle. Paper presented at the 2007 IEEE Conference on Computer Vision and Pattern Recognition, 1-8.
- Lyft. (2022). *Recording device policy*. <u>https://help.lyft.com/hc/en-us/articles/115012923127-Safety-policies</u> (2022)).
- MacFarlane, D. W., & Kane, B. (2017). Neighbour effects on tree architecture: Functional trade-offs balancing crown competitiveness with wind resistance. *Functional Ecology*, 31(8), 1624-1636.
- Ow, L. F., & Ghosh, S. (2017). Urban cities and road traffic noise: Reduction through vegetation. *Applied Acoustics*, 120, 15-20.
- Rea, R. V., Johnson, C. J., Aitken, D. A., Child, K. N., & Hesse, G. (2018). Dash cam videos on youtube[™] offer insights into factors related to moose-vehicle collisions. *Accident Analysis & Prevention*, 118, 207-213.
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. Paper presented at the *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 779-788.
- Salmond, J. A., Tadaki, M., Vardoulakis, S., Arbuthnott, K., Coutts, A., Demuzere, M., Dirks, K. N., Heaviside, C., Lim, S., & Macintyre, H. (2016). Health and climate related

ecosystem services provided by street trees in the urban environment. *Environmental Health*, 15(1), 95-111.

- Scharwächter, T., Enzweiler, M., Franke, U., & Roth, S. (2013). Efficient multi-cue scene segmentation. *Pattern recognition* (pp. 435-445). Springer Berlin Heidelberg. <u>https://doi.org/10.1007/978-3-642-40602-7 46</u>
- Smith, W. H. (2012). Air pollution and forests: Interactions between air contaminants and forest ecosystems. Springer Science & Business Media.
- Tomar, S. (2006). Converting video formats with FFmpeg. *Linux Journal*, 2006(146), 10.
- Tong, Z., Whitlow, T. H., MacRae, P. F., Landers, A. J., & Harada, Y. (2015). Quantifying the effect of vegetation on near-road air quality using brief campaigns. *Environmental Pollution*, 201, 141-149.
- UBER. (2022). Can I use a video camera . Can I use a video camera <u>https://help.uber.com/partners/article/can-i-use-a-video-</u> <u>camera-?nodeId=efaad152-cbb6-45fe-9d7d-911842d21c8b</u>
- Weber, F., Kowarik, I., & Säumel, I. (2014). A walk on the wild side: Perceptions of roadside vegetation beyond trees. Urban Forestry & Urban Greening, 13(2), 205-212.