A Deep Neural Network For Multiclass Bridge Element Parsing In Inspection Image Analysis

Chenyu ZHANG, Muhammad Monjurul KARIM, Ruwen QIN

Monitoring the condition of civil infrastructure is a critical step in assuring its safety and serviceability. Traditional manual bridge inspection requires a crew of inspectors, heavy equipment with lifting capacity, access to dangerous heights, and closure of the road during the inspection. Besides, manual bridge inspection results are subjective, varying from one inspector to another even when they follow the best inspection practice. Limitations of the traditional approach motivate research into automating bridge monitoring and evaluation with advanced technologies such as robotics and image analysis. Researchers use segmentation Deep Neural Networks (DNNs), such as Mask Region-based Convolutional Neural Network (Mask-RCNN), to detect bridge elements in the inspection images and recognize their types. This is an important task of computer vision-based bridge inspection because structural elements and their defects must be associated to produce an overall rating for a whole bridge, according to structure inspection manuals, such as the AASHTO's Bridge Element Inspection Manual. Some datasets focus on structure-level images wherein multiple structural elements are salient objects. Others are defect-level images that present pixel-level details of surface defects. Mobile sensors also take inspection images of bridges from a certain distance so that the images have both recognizable structural elements and apparent defects on the elements, which provide valuable information for the condition assessment. Our study found instance segmentation DNNs, such as Mask-RCNN, have limitations in segmenting multiclass structural elements from such images. In those images, bridge elements of the same type have widely different appearances due to the imperfect and changing view of the camera(s). Multiple types of structural elements may have similar textures and close contact. The bridge also mixes with the cluttered, dynamic background in inspection images. Therefore, multiclass bridge element parsing in inspection images or videos remains challenging. This paper aims to determine if DeepLabv3+, a representative semantic segmentation encoder-decoder network with instance prediction labels, is better than the prestigious instance segmentation network, Mask-RCNN, in analyzing images with the above-discussed challenges. First, the study created an annotated dataset by adding the pixel-level labels for six common elements of steel girder bridges, namely deck, girder, bracing, floor beam, bearing, and substructure, to a portion of data in the COCO-Bridge 2021+ dataset. Then, it designed numerical experiments that varied the amount of training data in transferring DeepLabv3+ and Mask-RCNN, respectively, to the bridge dataset of study. After that, the study compared the performance of these two networks in segmenting multiclass structural elements using accuracy metrics, including the mean values of IOU and pixel accuracy. Results from the comparative experiments indicate DeepLabv3+ has better transferability and segmentation performance in this specific study. The study further developed a human-in-the-loop method to recommend additional training data to boost the transferred DeepLabv3+ to a satisfying level. Qualitative results verify why DeepLabv3+ outperforms Mask-RCNN in segmenting bridge elements with a closer look, for instance, the depthwise separable convolution in both Atrous Spatial Pyramid Pooling and decoder modules.