Notice of the Final Oral Examination
for the Degree of Master of Science

of

SPENCER ROSE

BSc (University of Alberta, 1999)
BA (University of Victoria, 2003)
MA (University of Western Ontario, 2005)

“An evaluation of deep learning semantic segmentation for land cover classification of oblique ground-based photography”

Department of Computer Science

Tuesday, September 15, 2020
10:30 A.M.
Remote Defence

Supervisory Committee:
Dr. Yvonne Coady, Department of Computer Science, University of Victoria (Co-Supervisor)
Dr. Eric Higgs, School of Environmental Studies, University of Victoria (Co-Supervisor)

External Examiner:
Dr. Chris Bone, Department of Chemistry, UVic

Chair of Oral Examination:
Dr. Ralph Evins, Department of Civil Engineering, UVic

Dr. Stephen Evans, Acting Dean, Faculty of Graduate Studies
Abstract

This thesis presents a case study on the application of deep learning methods for the dense prediction of land cover types in oblique ground-based photography. While deep learning approaches are widely used in land cover classification of remote-sensing data (i.e., aerial and satellite orthoimagery) for change detection analysis, dense classification of oblique landscape imagery used in repeat photography remains undeveloped. A performance evaluation was carried out to test two state-of the-art architectures, U-net[1] and Deeplabv3+[2], as well as a fully-connected conditional random fields model[3] used to boost segmentation accuracy. The evaluation focuses on the use of a novel threshold-based data augmentation technique, and three multi-loss functions selected to mitigate class imbalance and input noise. The dataset used for this study was sampled from the Mountain Legacy Project (MLP) collection, comprised of high-resolution historic (grayscale) photographs of Canada’s Western mountains captured by the Geological Survey of Canada from the 1880s through the 1950s[4] and their corresponding modern (colour) repeat images. Land cover segmentations manually created by MLP researchers were used as ground truth labels. Experimental results showed top overall F1 scores of 0.835 for historic models, and 0.912 for repeat models. Data augmentation showed modest improvements to overall accuracy (+3.0% historic / +1.0% repeat), but much larger gains for under-represented classes.