“Decoding (urban) form and function using spatially explicit deep learning”

Abstract: This paper explores how can geographical dimension be in- corporated into deep learning designed to understand the composition of urban landscapes based on Sentinel 2 satellite imagery. Compared to standard computer vision, satellite imagery is unique as images sampled from the data form a continuous array, rather than being fully independent. We argue that the spatial configuration of the images is as important as the content of each image when attempting to capture a pattern that reflects the structure of the urban environment. We propose a series of approaches explicitly incorporating spatial dimen- sion in the predictive pipeline based on the EfficientNetB4 convolu- tional neural network (CNN) and experimentally test their effect on model performance. The experiments in this study cover the scale of the sampled area, the effect of spatial augmentation, and the role of modelling (logit ensemble and histogram-based gradient-boosted classifiers) with and without the spatial context on the outputs of the neural network-generated vector of probabilities while trying to predict spatial signatures, a classification of primarily urban landscape based on form and function. The results suggest that certain ways of embedding spatial information, especially in the modelling step, consistently significantly improve the prediction accuracy and shall be considered on top of standard CNNs.

Key words: spatial signatures, classification, remote sensing, artificial intelligence, open data